

INVESTIGAÇÃO OPERACIONAL EM AÇÃO

CASOS DE APLICAÇÃO

RUI CARVALHO OLIVEIRA
JOSÉ SOEIRO FERREIRA
(EDITORES)

IMPRESA DA UNIVERSIDADE DE COIMBRA
COIMBRA UNIVERSITY PRESS

(Página deixada propositadamente em branco)



E N S I N O



EDIÇÃO

Imprensa da Universidade de Coimbra
Email: imprensa@uc.pt
URL: http://www.uc.pt/imprensa_uc
Vendas online: <http://livrariadaimprensa.uc.pt>

COORDENAÇÃO EDITORIAL

Imprensa da Universidade de Coimbra

CONCEÇÃO GRÁFICA

António Barros

INFOGRAFIA

Mickael Silva

IMAGEM DA CAPA

By developer team (Gephi)
[GPL (<http://www.gnu.org/licenses/gpl.html>) or CC-BY-SA-3.0
(<http://creativecommons.org/licenses/by-sa/3.0/>), via Wikimedia Commons

EXECUÇÃO GRÁFICA

NSG – Novas Soluções Gráficas, S.A.

ISBN

978-989-26-0737-5

ISBN DIGITAL

978-989-26-0738-2

DOI

<http://dx.doi.org/10.14195/978-989-26-0738-2>

DEPÓSITO LEGAL

384210/14

OBRA PUBLICADA COM O APOIO DE:



INVESTIGAÇÃO OPERACIONAL EM AÇÃO

CASOS DE APLICAÇÃO

RUI CARVALHO OLIVEIRA
JOSÉ SOEIRO FERREIRA
(EDITORES)

(Página deixada propositadamente em branco)

SUMÁRIO

Preâmbulo.....	7
Parte I – Setor da Agricultura e Pescas.....	15
Caso 1 – Otimização de planos de exploração florestal em baldios do norte de Portugal.....	17
Caso 2 – Planeamento de explorações agropecuárias sustentáveis: uma abordagem multicritério.....	57
Caso 3 – Um modelo para a produção de leite nos Açores	105
Caso 4 – Modelo de apoio à gestão da pescaria de bivalves com ganchorra no Algarve baseado em dinâmica de sistemas	133
Parte II – Setor da Indústria	163
Caso 5 – Modelação e simulação de uma linha de produção de painéis solares	165
Caso 6 – Otimização do escalonamento da produção na indústria de moldes: aplicação ao caso da GECO.....	199
Caso 7 – Sistema de apoio à decisão do orçamento anual de produção na indústria de bebidas.....	247
Caso 8 – Otimização da sequência de veículos na montagem final de uma empresa da indústria automóvel: estudo de um caso	287
Caso 9 – Integração e otimização de processos na indústria de prensados de madeira.....	315
Caso 10 – Otimização de processos na produção de capas para estofos de automóveis: o caso da Coindu.....	349

Parte III – Setor da Logística e Transportes	399
Caso 11 – Projeto de uma rede logística para a recolha de equipamentos elétricos e eletrônicos	401
Caso 12 – Reestruturação de áreas de influência e de rotas de veículos num sistema de recolha de resíduos recicláveis	437
Caso 13 – Transporte multiproduto por oleoduto e gestão de parque de tanques: o caso da companhia logística de combustíveis	475
Caso 14 – Afetação multicritério de recursos a portefólios de atividades da REFER.	515
Parte IV – Setor dos Serviços.....	563
Caso 15 – Desenho de promoções diferenciadas em empresas de retalho recorrendo à segmentação de clientes	565
Caso 16 – Otimização da operação de redes de distribuição: consideração simultânea de variáveis topológicas, tomadas de transformadores e baterias de condensadores utilizando EPSO e heurísticas.....	589
Caso 17 – Otimização do planeamento inverso de tratamentos em radioterapia	609
Caso 18 – Avaliação de desempenho das unidades de saúde para a melhoria dos cuidados de saúde primários.....	637
Caso 19 – Otimização de redes hospitalares: um modelo de localização hierárquico e multi-serviço aplicado ao caso Português.....	673
Caso 20 – Benchmarking dos serviços dos hospitais portugueses: uma aplicação de data envelopment analysis.....	703
Caso 21 – Avaliação e acompanhamento de projetos de I&D numa empresa de telecomunicações	741

PREÂMBULO

Pretende-se com este livro promover e divulgar a Investigação Operacional (IO) enquanto ciência aplicada, contribuindo para demonstrar inequivocamente o seu potencial, através da publicação de um conjunto diversificado de casos de aplicação que relatam atuações em situações de decisão e de otimização bem reais concretizadas em Portugal.

Esta publicação tem como destinatários privilegiados os estudantes em cursos superiores (Licenciaturas, Mestrados, Doutoramentos) envolvendo Investigação Operacional, Métodos de Decisão, Otimização ou Ciências de Gestão, esperando-se que contribua para uma motivação acrescida para o estudo da IO, bem como os seus professores ou orientadores. Mas um outro público-alvo é constituído pelos gestores/decisores que se veem confrontados com problemas decisórios complexos e que, através dos casos publicados, ganhem consciência do potencial desta ciência aplicada e motivação para a explorar.

Esta obra tem os seguintes objetivos:

- i) Incluir uma variedade de casos nas mais diversificadas organizações e empresas que ilustre a riqueza das aplicações efetivamente realizadas;
- ii) Mostrar como se pode lidar com situações reais complexas através de adequadas metodologias de estruturação do problema e modelação para apoio à decisão;
- iii) Desenvolver capacidades para estruturar problemas, encarar e avaliar alternativas de decisão e selecionar métodos de resolução (algoritmos);

- iv) Permitir trabalhar com os casos expostos ou réplicas, discutindo abordagens, propondo outros métodos de solução e resolvendo exemplos concretos;
- v) Alertar os estudantes para as significativas diferenças entre exercícios de aulas (*canned problems*) e problemas reais.

O presente livro surge na sequência de outra obra de natureza idêntica (“Casos de aplicação da Investigação Operacional”), publicado pela McGraw Hill Portugal em 2000 sob coordenação de Carlos Henggeler Antunes e Luís Valadares Tavares, e vem de algum modo atualizar o estado da arte da atividade de IO ao serviço da sociedade em Portugal. O número de autores comuns às duas obras, publicadas com um intervalo de 14 anos, é praticamente residual, o que revela bem a renovação geracional e dinamismo da comunidade de IO em Portugal.

Esta obra nasceu da iniciativa do então presidente da Associação Portuguesa de Investigação Operacional (APDIO), Prof. Joaquim João Júdice, que lançou o desafio aos coordenadores desta edição. Como resultado do convite à submissão dos trabalhos anunciado pela APDIO, os coordenadores receberam mais de duas dezenas de propostas que, após aturado processo de análise e revisão em que estiveram envolvidos mais de 50 membros da comunidade técnico-científica de IO, conduziram à seleção dos 21 casos de aplicação que integram esta obra.

Os casos selecionados representam um largo espectro e diversidade, tanto em termos de objeto de aplicação como de métodos e técnicas resolutivas empregues. Na organização deste livro privilegiou-se uma classificação segundo o setor de atividade em que o caso estudado se integra.

Assim, na Parte I são apresentados quatro casos de aplicação no setor da agricultura e pescas e abordando as seguintes temáticas:

- Caso 1 – Planeamento da exploração florestal;
- Caso 2 – Planeamento de explorações agropecuárias;
- Caso 3 – Planeamento da produção de leite;
- Caso 4 – Apoio à gestão da pescaria de bivalves;

A Parte II é dedicada ao setor industrial, integrando 6 casos de aplicação que abordam as seguintes temáticas:

- Caso 5 – Planeamento da produção de painéis solares;
- Caso 6 – Escalonamento da produção na indústria dos moldes;
- Caso 7 – Orçamentação na indústria de bebidas;
- Caso 8 – Sequenciamento da produção na indústria automóvel;
- Caso 9 – Integração de processos na indústria da madeira;
- Caso 10 – Otimização de processos na produção de estofos para a indústria automóvel.

A Parte III integra 4 casos de aplicação no setor da logística e transportes, com tratamento das seguintes temáticas:

- Caso 11 – Desenho de uma rede de recolha de equipamentos elétricos e eletrónicos;
- Caso 12 – Definição de áreas de influência e rotas de veículos num sistema de recolha de resíduos;
- Caso 13 – Planeamento do transporte por oleoduto e gestão de parque de tanques na logística de combustíveis;
- Caso 14 – Apoio à decisão na afetação de recursos e portefólio de atividades para a gestão de infraestruturas de transporte.

Finalmente, na Parte IV são apresentados 7 casos de aplicação relativos à prestação de serviços tratando as seguintes temáticas:

- Caso 15 – Desenho de promoções e segmentação de clientes – setor do comércio;
- Caso 16 – Otimização da operação de redes de distribuição – setor energético / utilities;
- Caso 17 – Planeamento de tratamentos de radioterapia – setor da saúde;
- Caso 18 – Avaliação do desempenho de unidades de saúde;
- Caso 19 – Otimização de redes hospitalares;

- Caso 20 – Benchmarking dos serviços hospitalares portugueses;
- Caso 21 – Avaliação de projetos de I&D numa empresa de telecomunicações.

Para além da diversidade temática que com esta síntese de objetos de aplicação fica bem ilustrada, os casos selecionados exploram uma grande variedade de métodos e técnicas próprios da IO, nomeadamente:

- programação matemática nas suas diversas vertentes (linear e não linear, inteira mista, multiobjectivo);
- métodos heurísticos e meta-heurísticos;
- modelos de simulação e de dinâmica de sistemas;
- métodos de exploração de dados (“*data mining*”) e de avaliação comparativa do desempenho (“*data envelopment analysis*”);
- metodologias multicritério de apoio à decisão.

Assim, os casos de aplicação da IO em Portugal que constituem este livro revelam bem o carácter transversal e a multidisciplinaridade que, desde os seus primórdios, caracterizam a IO e constituem um seu elemento distintivo: a busca, por vezes obsessiva, das melhores soluções para os problemas que o mundo real, nos mais diversos contextos, nos coloca, desenvolvendo formulações ajustadas ao problema em apreço e explorando métodos e técnicas analíticas que permitem, de forma objetiva e cientificamente fundamentada, comparar os méritos das soluções alternativas e otimizar decisões.

Espera-se assim que esta obra sirva os propósitos que presidiram à sua edição. Desde logo, contribuir para a divulgação e promoção da IO junto de gestores e decisores, demonstrando o potencial desta ciência aplicada para abordar problemas desafiantes e complexos. Mas também para professores e, sobretudo, estudantes destas matérias, alargando os seus horizontes para além dos exercícios tipicamente utilizados nas aulas, motivando-os e dando-lhes oportunidade para explorar abordagens alternativas aos casos reais apresentados. Nesse sentido, são disponibilizados no site da APDIO (www.apdio.pt) os dados relativos aos casos incluídos nesta obra e para os quais foi obtida autorização da sua divulgação.

Infelizmente, em outras situações tal não foi possível, por razões de confidencialidade invocadas pelas entidades para os quais os trabalhos relatados foram desenvolvidos.

Os coordenadores desta edição não podem deixar de agradecer os diversos contributos que tornaram possível esta publicação. À Direção da APDIO, nas figuras do seu anterior presidente, Prof. Joaquim João Júdice que lançou esta iniciativa, e do atual presidente, Prof. Domingos Cardoso, por todo o incentivo e apoio recebidos. Ao Prof. Carlos Henggeler Antunes, que estabeleceu as pontes com a Editora da Universidade de Coimbra que, em boa hora, acolheu a publicação desta obra. À direção da APDIO e ao Fundo Isabel Themido, pelo apoio financeiro a esta edição.

Ao conjunto, numeroso e ilustre, de revisores que apoiaram os coordenadores no processo de seleção e revisão (anónimas) dos trabalhos submetidos para publicação e que muito contribuíram para assegurar a elevação da qualidade e rigor das versões finais dos textos publicados. À Patrícia Nunes e Maria da Glória Valle, pelo inestimável apoio de secretariado.

E, *last, but not the least*, aos autores dos casos publicados, sem cujo contributo, bem como paciência e perseverança ao longo deste demorado processo de seleção, revisão e edição final, esta obra não poderia obviamente ver a luz do dia. A todos o nosso agradecimento, na expectativa de que, em conjunto, tenhamos dado um contributo significativo para a divulgação e desenvolvimento da IO no nosso país.

Rui Carvalho Oliveira
CESUR/DECivil - Instituto Superior Técnico - Universidade de Lisboa
Av. Rovisco Pais - 1049-001 Lisboa, Portugal
(roliv@tecnico.ulisboa.pt)

José Soeiro Ferreira
FEUP- Faculdade de Engenharia da Universidade do Porto
INESC TEC - INESC Tecnologia e Ciência
Rua Dr. Roberto Frias - 4200-465 Porto, Portugal
(jsf@fe.up.pt)

(Página deixada propositadamente em branco)

AGRADECIMENTOS AOS REVISORES

Ana Camanho | FEUP
Ana Carvalho | IST - UL
Ana Luísa Custódio | FCT - UNL
Ana Luísa Respício | FC - UL
Ana Moura | DEGI - UA
Ana Barbosa Póvoa | IST- UL
Ana Viana | INESC Porto
António Miguel Gomes | FEUP
António José Rodrigues | FC - UL
Bernardo Almada Lobo | FEUP
Carlos Ferreira | DEGI - UA
Carlos Henggeler Antunes | FCT - UC
Carlos Pestana Barros | ISEG - UL
Cláudia Sarrico | ISEG - UL
Cláudio Alves | DPS - UM
Dalila Fontes | FEP - UP
Fernando Martins | FEUP
Filipe Alvelos | DPS - UM
Helena Correia | UCP Porto
Helena Ramalhinho Lourenço | UPF - Barcelona
Isabel Horta Costa | FEUP
José Fernando Gonçalves | FEP - UP
José Fernando Oliveira | FEUP
João Paulo Costa | FE - UC
João Pedro Pedroso | FCUP
Luís Cavique | U. Aberta
Luís Dias | FE - UC
Luís Gouveia | FC - UL
Luís Valadares Tavares | IST - UL

Manuel Matos | FEUP
Margarida Cardoso | ISCTE - IUL
Margarida Vaz Pato | ISEG - UL
Maria Candida Mourão | ISEG - UL
Maria do Sameiro Carvalho | DPS - UM
Maria Eugénia Captivo | FC - UL
Maria Teresa Costa | ISEP
Marta Castilho Gomes | IST - UL
Mónica Oliveira | IST - UL
Paula Amaral | FCT - UNL
Pedro Mendes | IST - UL
Pedro Oliveira | ICBAS - UP
Rui Alves | FEP - UP
Rui Cunha Marques | IST - UL
Susana Baptista | FCT - UNL

Parte I

SETOR da
AGRICULTURA
E PESCAS

(Página deixada propositadamente em branco)

CASO 1

OTIMIZAÇÃO DE PLANOS DE EXPLORAÇÃO FLORESTAL EM BALDIOS DO NORTE DE PORTUGAL

Adelaide Cerveira

Departamento de Matemática
Universidade de Trás-os-Montes e Alto Douro
cerveira@utad.pt

Isabel Martins

Departamento de Ciências e Engenharia de Biosistemas
Instituto Superior de Agronomia
isabelinha@isa.ulisboa.pt

Artur Mota

Associação Florestal de Ribeira de Pena
eng.artur.mota@gmail.com

João Bento

Departamento de Ciências Florestais e Arquitetura Paisagista
Universidade de Trás-os-Montes e Alto Douro
j_bento@utad.pt

Teresa Fonseca

Departamento de Ciências Florestais e Arquitetura Paisagista
Universidade de Trás-os-Montes e Alto Douro
tfonseca@utad.pt

RESUMO

Este estudo descreve uma abordagem em programação linear inteira para a gestão de uma área florestal de pinheiro bravo localizada no Perímetro Florestal do Barroso, em Ribeira de Pena, na região Norte de Portugal. Esta área é composta por baldios, maioritariamente constituídos por parcelas de povoamentos florestais e matos. A gestão destes baldios, da responsabilidade das direções dos baldios e do Instituto de Conservação da Natureza e das Florestas, foi realizada até há bem pouco tempo de forma essencialmente empírica, com base na experiência e sensibilidade dos técnicos deste instituto. O problema da gestão dos baldios abordado neste estudo consiste em determinar onde e quando intervir e que tipo de intervenção fazer tal que, tendo em consideração um conjunto de restrições, o volume de madeira a retirar seja máximo. As intervenções estão sujeitas a restrições de natureza silvícola, operacional, organizacional e ambiental. Considera-se ainda um grupo de restrições que contribuem diretamente para a sustentabilidade da floresta. As restrições ambientais, de natureza espacial e mais complexas do que as restantes, consistem em limitar as áreas das clareiras. Estas restrições foram formuladas de diferentes formas de acordo com os modelos básicos *path*, *cluster* e *bucket* descritos na literatura para a gestão florestal com restrições de clareiras. Estes modelos têm um número muito elevado de variáveis e restrições para instâncias grandes, podendo a sua resolução por *branch-and-bound* ser difícil, o que não aconteceu com o caso de estudo, que é de dimensões médias. Os resultados computacionais indicam que a abordagem proposta é uma ferramenta útil para apoiar a gestão dos baldios, sendo facilmente adaptável a outras áreas florestais, com outras espécies e geridas por outras entidades. A recente atualização dos Planos de Gestão Florestal dos baldios de Ribeira de Pena foi feita com base em planos de gestão obtidos com a abordagem proposta.

PALAVRAS-CHAVE

Gestão florestal, Pinheiro bravo, Programação linear inteira, Restrições espaciais, Restrições não espaciais.

1. Introdução

O pinheiro bravo (*Pinus pinaster* Ait.) é uma das principais espécies florestais em Portugal, ocupando 23% da área continental e 193 mil hectares no norte do país. A maior mancha florestal desta espécie ocorre no vale do rio Tâmega, no concelho de Ribeira de Pena, no distrito de Vila Real. Esta mancha é formada maioritariamente por terrenos comunitários, designados por baldios, constituídos essencialmente por povoamentos florestais e vegetação arbustiva espontânea (matos). A gestão destes baldios, da responsabilidade conjunta dos órgãos gestores eleitos pelas Assembleias de Compartes¹ e do Instituto de Conservação da Natureza e das Florestas² (ICFN), foi feita até há bem pouco tempo de forma essencialmente empírica, com base na experiência e sensibilidade dos técnicos do ICFN. Embora os baldios possam fornecer uma diversidade de produtos, serviços e outros benefícios, historicamente, os rendimentos obtidos, que revertem a favor das entidades gestoras, estão diretamente relacionados com a madeira proveniente do corte das árvores.

Numa floresta para produção de madeira, distinguem-se dois principais tipos de corte: corte cultural e corte final. O corte cultural, também designado por desbaste, consiste na remoção das árvores mais defeituosas e debilitadas. Este corte tem por objetivo aumentar o espaço de crescimento das árvores mais vigorosas e antecipar a mortalidade das árvores mais fracas e, assim, incrementar a quantidade e a qualidade do material lenhoso que não é removido. No caso do pinheiro bravo, o corte cultural tem início entre os 15 e os 20 anos de idade, podendo prolongar-se, com uma periodicidade não inferior a cinco anos, até próximo da idade em que se realiza o corte final. O corte final, também designado por corte raso ou de realização, é um corte após o qual é necessário proceder a uma replantação ou sementeira. No caso do pinheiro bravo, tem-se adotado fazer este corte em povoamentos com cerca de 45 anos ou, em estações de melhor qualidade, com idades inferiores, de forma a garantir que a idade da explorabilidade absoluta (idade em que a produção de material lenhoso por unidade de área é máxima) já tenha sido atingida e a madeira retirada tenha dimensões adequadas para ser classificada

¹ A Assembleia de Compartes de um ou mais baldios é constituída pelos moradores das freguesias a que pertencem os baldios, que tenham direito ao seu uso e fruição.

² Ex-Autoridade Florestal Nacional.

como de serração. No entanto, recentemente, tem havido uma tendência para fazer o corte raso bastante mais cedo, por ser a produção de estilha (aparas de madeira) a principal utilização da madeira cortada. Numa floresta dividida em parcelas de gestão, é comum sujeitar cada parcela em toda a sua extensão à mesma decisão de intervenção (nomeadamente, desbaste, corte raso ou não intervenção).

A programação matemática tem sido uma ferramenta poderosa para modelar e resolver problemas de gestão florestal. De acordo com Bento 1994, as primeiras aplicações nesta área, com recurso à programação linear, surgiram nos anos 50 do século passado (Arimizu 1956, Yoho e Row 1958 cit. Bare 1971), coincidindo com o final dos anos 60 e o início da década de 70 o aparecimento de dois sistemas que, pela sua operacionalidade, rapidamente se expandiram: MAX-MILLION (Clutter 1968 cit. Bare 1971, Ware e Clutter 1971) e TIMBER RAM (Navon 1971). Nos anos 70, assistiu-se a um grande entusiasmo e houve um alargamento das utilizações destes sistemas para fora dos Estados Unidos, surgindo dois novos instrumentos, MAX2 (Clutter et al. 1978) e MUSYC (Johnson e Jones 1979 cit. Iverson e Alston 1986), e, nos anos 80, os Serviços Florestais Americanos adotaram o sistema FORPLAN (Kent et al. 1987). Com a disponibilização alargada dos Sistemas de Informação Geográfica e com o recurso a capacidades informáticas ininterruptamente em crescimento, abriu-se, a partir dos anos 90, um largo campo de desenvolvimento de aplicações suportadas em referenciais geográficos. Desde então recorreu-se com vantagem à programação linear inteira, nomeadamente para a modelação de restrições ambientais (veja-se Weintraub et al. 2000 e 2007 para uma revisão das principais e vastas abordagens publicadas).

As restrições ambientais têm vindo a ser consideradas para responder às novas exigências da sociedade associadas à proteção do ambiente, nomeadamente aos níveis da criação de *habitats* favoráveis à fauna e à flora, da promoção da diversidade, da manutenção da qualidade do solo e da água, e da preservação da beleza paisagística. Restringir as áreas das clareiras tem sido uma prática comum em muitos países para regular a atividade florestal sujeita a estas exigências. Uma clareira é uma região contínua de floresta que se encontra desarborizada ou com árvores de pequenas dimensões. Esta região

pode resultar de cortes rasos realizados no mesmo ano ou em anos diferentes desde que o intervalo de tempo entre estes anos não exceda o tempo de exclusão. No tempo de exclusão, a altura do coberto florestal que se desenvolve a seguir aos cortes rasos é ainda muito reduzida, podendo toda a região contínua intervencionada ser considerada uma clareira. As restrições que dependem das relações de adjacência entre as parcelas, como é o caso das restrições de clareiras, são habitualmente referidas como restrições espaciais. As restrições que não dependem destas relações são não espaciais.

Descrevem-se neste trabalho abordagens em programação linear inteira para apoiar a gestão de baldios para produção de madeira, em que se consideram as restrições ambientais que restringem as áreas das clareiras. A área de estudo encontra-se localizada no Perímetro Florestal do Barroso, no concelho de Ribeira de Pena, e é constituída pelos baldios de Canedo, Seirós, Melhe, Santa Marinha Norte e Santo Aleixo. A produção de estilha é a principal orientação da gestão. Para garantir intervenções mais assíduas, o plano de exploração é estabelecido para períodos de cinco anos, possibilitando uma gestão mais atenta e oportuna. Parte deste trabalho resultou de uma parceria estabelecida em 2004 entre o Departamento de Ciências Florestais e Arquitetura Paisagista (CIFAP) da Universidade de Trás-os-Montes e Alto Douro (UTAD) e a Associação Florestal de Ribeira de Pena (AFRP).

Na Secção 2 resume-se o processo de colaboração com as entidades gestoras dos baldios da área de estudo que decorreu durante a realização deste trabalho. Na Secção 3 apresenta-se o problema de otimização da gestão dos baldios, e nas Secções 4 e 5 descrevem-se os modelos que são usados para formular o problema. Na Secção 4 apresentam-se os modelos sem as restrições associadas às clareiras e na Secção 5 formulam-se estas restrições. Um exemplo de uma floresta com nove parcelas retirado do caso de estudo é usado para fins ilustrativos. Na Secção 6 descreve-se o caso de estudo. Na Secção 7 discutem-se os resultados computacionais e na Secção 8 estabelecem-se as conclusões.

2. Processo de colaboração com as entidades gestoras dos baldios

No âmbito da parceria estabelecida em 2004 entre o CIFAP da UTAD e a AFRP, fez-se o inventário das existências florestais dos povoamentos de pinheiro bravo cogерidos pelos técnicos do ICFN e da AFRP, com o objetivo de planejar a exploração da floresta com a caracterização atualizada dos povoamentos. Para o planeamento das atividades de exploração, verificou-se de imediato que a utilização de técnicas empíricas poderia não conduzir às soluções mais adequadas em termos de produção de madeira. As instituições parceiras, AFRP e CIFAP-UTAD, decidiram então ensaiar a utilização de técnicas de otimização. Os técnicos da Direção Regional das Florestas do Norte do ICFN diretamente ligados à gestão da área de estudo não só disponibilizaram informação como também ajudaram a fundamentar o problema da gestão dos baldios. Os órgãos cogestores, Presidentes dos Conselhos Diretivos dos Baldios e Presidentes das Juntas de Freguesia, foram também envolvidos no processo.

Quando, a partir de 2009, passou a ser obrigatória a apresentação de Planos de Gestão Florestal (PGF) nas áreas públicas e comunitárias (Decreto-Lei 16/2009), a AFRP, responsável pela elaboração dos PGF dos baldios, encontrou no trabalho entretanto desenvolvido, um suporte sólido para a revisão e a atualização das versões iniciais. Os PGF foram apresentados e explicados em sessões públicas aos Compartes (em 2012), que os aceitaram com entusiasmo, acabando por adotá-los para o planeamento da exploração dos respetivos baldios.

3. Descrição do problema

O problema da gestão dos baldios, abordado neste estudo, consiste em determinar onde (em que parcelas) e quando (em que anos do horizonte de planeamento) se deve intervir e que tipo de intervenção se deve fazer (corte raso, desbaste ou não intervir) tal que, tendo em consideração um conjunto de restrições espaciais e não espaciais, o volume de madeira a retirar no horizonte de planeamento seja máximo. Escolheu-se maximizar o volume de madeira, em vez de uma expressão financeira do resultado da exploração, porque, como o caso de estudo se refere a uma área geográfica restrita, com

grande homogeneidade ecológica, de exploração orientada para a produção de madeira de pequenas dimensões e para um período de referência muito curto (cinco anos), o preço da madeira mantém-se aproximadamente constante. Neste trabalho, considera-se que as intervenções estão sujeitas a restrições de natureza silvícola, operacional e organizacional. Consideram-se ainda restrições relacionadas com a persistência das áreas de floresta nos baldios (restrições de sustentabilidade) e restrições ambientais.

As restrições silvícolas impedem o desbaste ou o corte raso em parcelas cuja idade ou densidade são desadequadas à intervenção. Os desbastes não são feitos em parcelas com menos de 15 anos ou mais de 55 anos, ou com grau de ocupação florestal correspondente a valores do índice *SDI* inferiores a 60%. A ocupação do espaço pelo coberto florestal pode ser quantificada por este índice, do inglês *stand density index*, que relaciona o número de árvores com o diâmetro médio destas (Reineke 1933). Tem sido estipulado que o início da mortalidade das árvores ocasionada por competição intraespecífica ocorre com o valor de *SDI* de 60%. Neste trabalho, considera-se que os cortes rasos não são feitos em parcelas com menos de 15 anos. Os valores assumidos têm por base a silvicultura média praticada na região nas últimas décadas. Estas restrições impõem ainda, em cada parcela, intervalos mínimos de 5 anos entre intervenções, ou seja, no horizonte de planeamento, uma parcela não pode ser sujeita a duas intervenções.

As restrições operacionais impedem que, sempre que se intervém num baldio, o volume total de madeira a retirar seja inferior a um limite mínimo, que se considerou ser 250 m³, de acordo com informação disponibilizada por técnicos do ICFN.

As restrições organizacionais impõem, para as receitas obtidas em toda a floresta, distribuições equilibradas ao longo do horizonte de planeamento. Quanto às receitas obtidas em cada baldio, não é desejável a ocorrência de intervenções muito díspares. Estas restrições impõem também que se retire madeira de cada baldio em cada triénio.

As restrições de sustentabilidade garantem que o volume de madeira em pé e o acréscimo anual da produção de madeira no último ano não sejam inferiores aos valores observados no início do primeiro ano. Estas restrições,

que podem ser aplicadas globalmente a toda a floresta ou individualmente a cada baldio, impedem uma extração comprometedora de madeira no horizonte de planeamento.

Todas estas restrições são não espaciais.

As restrições ambientais, que são usadas para impedir uma redução significativa da qualidade ambiental, impõem um limite máximo nas áreas das clareiras de 10 ha, de acordo com as orientações regionais para a floresta (Plano Regional de Ordenamento Florestal). Considerou-se um tempo de exclusão de três anos, que corresponde ao período que a espécie necessita para regenerar e desenvolver-se até atingir uma altura média de 30 cm. As restrições ambientais são de natureza espacial.

Para reduzir os custos operacionais associados ao corte raso, pode ser importante limitar, para cada baldio, o número de clareiras a formar em cada ano, o que também foi considerado neste trabalho. Tal como as restrições ambientais, estas restrições operacionais são também de natureza espacial.

4. Modelos sem restrições de clareiras

Nesta secção, propõe-se formular o problema da gestão dos baldios sem as restrições espaciais (restrições de clareiras), mais complexas do que as restantes. Este problema exige soluções inteiras e, por isso, vai ser formulado em programação inteira (linear). O horizonte de planeamento é dividido em cinco períodos anuais e, para cada ano, cada parcela da floresta está sujeita a uma decisão (desbaste, corte raso ou não intervenção).

A notação usada é a seguinte: j – índice das parcelas da área de estudo; i – índice dos baldios; t – índice dos períodos (anos); V – conjunto das parcelas; B – conjunto dos baldios; B_i – conjunto das parcelas do baldio $i \in B$ ($B_i \subset V$); $T = \{1, \dots, 5\}$ – horizonte de planeamento constituído por cinco períodos consecutivos; C_j^t – volume de madeira em pé (m^3) na parcela $j \in V$ no período $t \in T$; D_j^t – volume de madeira para desbaste (m^3) na parcela $j \in V$ no período $t \in T$; Vp_j^t – volume de madeira em pé (m^3) na parcela $j \in V$ no último período, se esta parcela é sujeita ao desbaste no período $t \in T$; Ac_j^1 e Ac_j^5 – acréscimos anuais da produção de madeira ($m^3 \text{ano}^{-1}$) na parcela $j \in V$ nos períodos 1 e 5, respetivamente; $Ac_d_j^t$ – acréscimo anual da

produção de madeira ($\text{m}^3\text{ano}^{-1}$) na parcela $j \in V$ no último período, se esta parcela é sujeita ao desbaste no período $t \in T$; A_j – área (ha) da parcela $j \in V$; I_j – idade (anos) da parcela $j \in V$ no período 1; SDI_j^t – valor do índice da densidade SDI (%) da parcela $j \in V$ no período $t \in T$; t_j^1 – primeiro período $t \in T$ em que a parcela $j \in V$ tem $SDI_j^t \geq 60\%$; m_i – volume máximo de madeira (m^3) que se pode remover no baldio $i \in B$ no horizonte de planeamento T , $\sum_{j \in B_i} \max_{t \in T: I_j + t \geq 16} \{C_j^t\}$.

Os parâmetros C_j^t , D_j^t , Vp_j^t , Ac_j^1 , Ac_j^5 e Acd_j^t são calculados de acordo com modelos de crescimento e de produção adaptados ao caso de estudo (Secção 6). O parâmetro C_j^t coincide com o volume de madeira em pé que pode ser removido em corte raso na parcela j e no período t na ausência de intervenção prévia. O parâmetro D_j^t corresponde a uma fração de C_j^t , sendo dependente das características biométricas das parcelas (*e.g.* idade) e das prescrições silvícolas de controlo da densidade. Os parâmetros Vp_j^t e Acd_j^t são, respetivamente, o volume de madeira em pé e o acréscimo anual da produção de madeira previstos para a parcela j no período 5 se esta parcela é sujeita ao desbaste no período t . Os parâmetros Ac_j^1 e Ac_j^5 são os acréscimos anuais da produção de madeira esperados para a parcela j nos períodos 1 e 5, se não tiver havido nenhuma intervenção. O parâmetro SDI_j^t varia com a idade e o crescimento das árvores, sendo modificado com a realização de desbastes.

Considerem-se as seguintes variáveis de decisão: x_j^t (variável de corte raso) e z_j^t (variável de desbaste), que assumem o valor 1 se, respetivamente, o corte raso ou o desbaste é efetuado na parcela j no período t e 0 caso contrário; y_i^t (variável de intervenção), que assume o valor 1 se alguma intervenção (corte raso ou desbaste) é realizada no baldio i no período t e 0 caso contrário; w_i , que é um majorante do volume da madeira a retirar do baldio i , em qualquer período do horizonte de planeamento.

Para o problema sem as restrições de clareiras, propõe-se a formulação (1) – (18), em que Ω e Δ nas restrições (7) – (9) são parâmetros pertencentes a $[0,1]$. Nas restrições (3), $|B_i|$ é o número de parcelas do baldio i .

$$\max \sum_{i \in B} \sum_{j \in B_i} \sum_{t \in T} (C_j^t x_j^t + D_j^t z_j^t) \quad (1)$$

sujeito a:

$$\sum_{t \in T} (x_j^t + z_j^t) \leq 1 \quad \forall i \in B, j \in B_i \quad (2)$$

$$y_i^t \leq \sum_{j \in B_i} (x_j^t + z_j^t) \leq |B_i| y_i^t \quad \forall i \in B, t \in T \quad (3)$$

$$\sum_{t'=t}^{t+2} y_i^{t'} \geq 1 \quad \forall i \in B, t = 1, 2, 3 \quad (4)$$

$$\sum_{j \in B_i} (C_j^t x_j^t + D_j^t z_j^t) \geq 250 y_i^t \quad \forall i \in B, t \in T \quad (5)$$

$$\sum_{j \in B_i} (C_j^t x_j^t + D_j^t z_j^t) \leq w_i \quad \forall i \in B, t \in T \quad (6)$$

$$w_i \leq (1 + \Omega) \sum_{j \in B_i} (C_j^t x_j^t + D_j^t z_j^t) + m_i (1 - y_i^t) \quad \forall i \in B, t \in T \quad (7)$$

$$\sum_{i \in B} \sum_{j \in B_i} (C_j^{t+1} x_j^{t+1} + D_j^{t+1} z_j^{t+1}) \geq (1 - \Delta) \sum_{i \in B} \sum_{j \in B_i} (C_j^t x_j^t + D_j^t z_j^t) \quad \forall t \in T : t < 5 \quad (8)$$

$$\sum_{i \in B} \sum_{j \in B_i} (C_j^{t+1} x_j^{t+1} + D_j^{t+1} z_j^{t+1}) \leq (1 + \Delta) \sum_{i \in B} \sum_{j \in B_i} (C_j^t x_j^t + D_j^t z_j^t) \quad \forall t \in T : t < 5 \quad (9)$$

$$\sum_{i \in B} \sum_{j \in B_i} \sum_{t \in T} V p_j^t z_j^t + \sum_{i \in B} \sum_{j \in B_i} \left(1 - \sum_{t \in T} (x_j^t + z_j^t) \right) C_j^5 \geq \sum_{i \in B} \sum_{j \in B_i} C_j^1 \quad (10)$$

$$\sum_{i \in B} \sum_{j \in B_i} \sum_{t \in T} A c d_j^t z_j^t + \sum_{i \in B} \sum_{j \in B_i} \left(1 - \sum_{t \in T} (x_j^t + z_j^t) \right) A c_j^5 \geq \sum_{i \in B} \sum_{j \in B_i} A c_j^1 \quad (11)$$

$$\sum_{t \leq t_j^1 + 1} (x_j^t + z_j^t) = 1 \quad \forall i \in B, j \in B_i : t_j^1 < 5 \quad (12)$$

$$\sum_{t \leq 5} (x_j^t + z_j^t) = 1 \quad \forall i \in B, j \in B_i : t_j^1 = 5 \quad (13)$$

$$x_j^t = 0 \quad \forall i \in \mathbf{B}, j \in B_i, t \in T : I_j + t < 16 \quad (14)$$

$$z_j^t = 0 \quad \forall i \in \mathbf{B}, j \in B_i, t \in T : I_j + t < 16 \text{ ou } I_j + t > 56 \text{ ou } SDI_j^t < 60 \quad (15)$$

$$x_j^t, z_j^t \in \{0,1\} \quad \forall i \in \mathbf{B}, j \in B_i, t \in T \quad (16)$$

$$y_i^t \in \{0,1\} \quad \forall i \in \mathbf{B}, t \in T \quad (17)$$

$$w_i \geq 0 \quad \forall i \in \mathbf{B}. \quad (18)$$

A função objetivo (1) maximiza o volume da madeira a retirar dos baldios no horizonte de planeamento. As restrições (2) garantem que em cada parcela, em todo o horizonte de planeamento, ou se faz um corte raso ou se desbasta ou não se intervém.

As restrições (3) identificam, para cada baldio, os períodos em que se realizam intervenções. De notar que $\sum_{j \in B_i} (x_j^t + z_j^t)$ representa o número de parcelas do baldio i sujeitas ao corte raso ou a desbaste no período t . Se ocorrerem intervenções no baldio i no período t então, pela desigualdade $\sum_{j \in B_i} (x_j^t + z_j^t) \leq |B_i| y_i^t$, $y_i^t = 1$ (qualquer que seja o número das parcelas sujeitas a intervenção), caso contrário, por $y_i^t \leq \sum_{j \in B_i} (x_j^t + z_j^t)$, $y_i^t = 0$.

As restrições (4) obrigam a que haja, em cada baldio, pelo menos uma intervenção por triénio. As restrições (5) asseguram que em cada baldio se retira pelo menos 250 m³ de madeira em qualquer intervenção.

As restrições (6) definem, para cada baldio, um majorante do volume de madeira a retirar numa intervenção. Este majorante não deve exceder $(1 + \Omega)$ vezes o volume a retirar em cada período em que se faz uma intervenção (restrições (7)). Se há intervenção no baldio i no período t então, pelas restrições (3) e (17), $y_i^t = 1$ e, neste caso, w_i não pode exceder $(1 + \Omega)$ vezes o volume a retirar em t . Caso contrário, também pelas restrições (3) e (17), $y_i^t = 0$ e, assim, w_i não fica limitado pelo volume retirado em t , que é nulo, mas antes pelo volume máximo de madeira que se pode remover m_i . As restrições (6) e (7) garantem intervenções equilibradas em cada baldio em termos de volume de madeira a retirar e, portanto, em termos de receita.

As restrições (8) e (9) obrigam a que o volume total de madeira a retirar em cada período esteja entre $(1-\Delta)$ e $(1+\Delta)$ vezes o volume obtido no período anterior, assegurando assim uma regularidade da receita total obtida ao longo do horizonte de planeamento.

As restrições (10) garantem que o volume total de madeira em pé existente no final do horizonte de planeamento é pelo menos igual ao valor registado no início do horizonte, $\sum_{i \in B} \sum_{j \in B_i} C_j^1$. De notar que aquele volume nas parcelas sujeitas ao desbaste é dado por $\sum_{i \in B} \sum_{j \in B_i} \sum_{t \in T} V p_j^t z_j^t$ e nas parcelas sem qualquer intervenção por $\sum_{i \in B} \sum_{j \in B_i} \left(1 - \sum_{t \in T} (x_j^t + z_j^t) \right) C_j^5$. Relativamente às parcelas sujeitas a corte raso, o volume de madeira em pé no final do horizonte de planeamento é insignificante e, por isso, não é considerado.

A descrição das restrições (11) é similar. Estas restrições garantem que o acréscimo anual da produção de madeira no final do horizonte de planeamento não é inferior ao registado no início do horizonte, $\sum_{i \in B} \sum_{j \in B_i} A c_j^1$. O

acréscimo anual da produção de madeira no final do horizonte de planeamento nas parcelas sujeitas a corte raso não é considerado e nas parcelas sujeitas ao desbaste é $\sum_{i \in B} \sum_{j \in B_i} \sum_{t \in T} A c d_j^t z_j^t$. Nas parcelas sem intervenção

o acréscimo é dado por $\sum_{i \in B} \sum_{j \in B_i} \left(1 - \sum_{t \in T} (x_j^t + z_j^t) \right) A c_j^5$.

As restrições (12) asseguram que uma parcela que registre um valor de *SDI* superior ou igual a 60% é sujeita ao corte raso ou ao desbaste até ao período a seguir àquele em que este valor é observado pela primeira vez, ou até este período se for o último (restrições (13)). As restrições (14) e (15) estabelecem as idades limites de corte raso e desbaste. As restrições (15) também estabelecem que as parcelas com valores de *SDI* inferiores a 60% não estão sujeitas ao desbaste. As restrições (16) a (18) definem a natureza das variáveis. As restrições silvícolas correspondem às restrições (2), (12) a (15), as restrições operacionais às restrições (5), as restrições organizacionais às restrições (4), (6) a (9) e as restrições de sustentabilidade às restrições (10) e (11). A formulação (1) – (18) será referida por FA.

Alternativamente à função objetivo (1), propõe-se maximizar o menor volume relativo de madeira a retirar por baldio. Seja

$$p_i = \frac{\sum_{j \in B_i} \sum_{t \in T} (C_j^t x_j^t + D_j^t z_j^t)}{m_i} \text{ o rácio entre o volume total de madeira a remover}$$

e o volume máximo de madeira que pode remover-se no horizonte de planeamento, para cada baldio $i \in B$. A nova formulação maximiza o menor destes rácios ($\max \min \{p_i : i \in B\}$). Seja p a variável de decisão que é um minorante dos rácios p_i .

Considerem-se então as restrições (19) e (20)

$$\sum_{j \in B_i} \sum_{t \in T} (C_j^t x_j^t + D_j^t z_j^t) \geq m_i p \quad \forall i \in B \quad (19)$$

$$p \geq 0 \quad (20)$$

e a função objetivo (21)

$$\max P. \quad (21)$$

Com a formulação (21) e (2) – (20) pretende evitar-se, sempre que possível, planos de exploração ótimos em que a proporção do volume de madeira a retirar nuns baldios seja bastante inferior à de outros baldios. Esta formulação será referida por FB. Facilmente se conclui que o volume total ótimo obtido por FB é inferior ou igual ao obtido por FA.

Na formulação seguinte, designada por FAI, as restrições de sustentabilidade (10) e (11) relativas a toda a floresta são substituídas pelas restrições de sustentabilidade (22) e (23) relativas a cada baldio.

$$\sum_{j \in B_i} \sum_{t \in T} V p_j^t z_j^t + \sum_{j \in B_i} \left(1 - \sum_{t \in T} (x_j^t + z_j^t) \right) C_j^5 \geq \sum_{j \in B_i} C_j^1 \quad \forall i \in B \quad (22)$$

$$\sum_{j \in B_i} \sum_{t \in T} A c d_j^t z_j^t + \sum_{j \in B_i} \left(1 - \sum_{t \in T} (x_j^t + z_j^t) \right) A c_j^5 \geq \sum_{j \in B_i} A c_j^1 \quad \forall i \in B. \quad (23)$$

As restrições (8) – (9) são eliminadas, não sendo substituídas por restrições similares para cada baldio, porque não têm de ocorrer intervenções todos os anos em cada baldio. Note-se que as restrições (22) – (23) implicam as restrições (10) – (11), isto é, se a sustentabilidade de cada baldio é assegurada então a sustentabilidade de toda a floresta também o é. A implicação recíproca não é verdadeira. Como as restrições (8) – (9) estão ausentes em FAI, não se

pode estabelecer uma relação de ordem entre os volumes totais ótimos obtidos por FA e FAI. Note-se ainda que FAI não tem restrições que misturam variáveis associadas a baldios diferentes ou a parcelas de baldios diferentes. Seja FAI_i a formulação resultante de FAI em que a função objetivo e as restrições estão associadas apenas ao baldio i . O plano de gestão para cada baldio obtido com FAI pode também ser obtido com FAI_i e, por isso, é independente dos restantes baldios.

As formulações FA, FB e FAI (Tabela 1) têm $O(|V||T|)$ variáveis e restrições (Tabela A1 do Anexo).

Tabela 1: Função objetivo e restrições das formulações FA, FB e FAI

Formulação	Função objetivo	Restrições
FA	(1)	(2) – (18)
FB	(21)	(2) – (20)
FAI	(1)	(2) – (7), (12) – (18), (22) – (23)

5. Modelos com restrições de clareiras

Nesta secção, formulam-se as restrições associadas às clareiras. Na literatura tem-se discutido a abordagem designada por *Área Restriction Model* (ARM) (Murray 1999) para limitar as áreas das clareiras. Esta abordagem permite a agregação das parcelas para corte raso desde que a área de cada região contínua sujeita ao corte raso não exceda a área máxima imposta para uma clareira. A abordagem clássica designada por *Unit Restriction Model* (URM) requer uma prévia agregação das parcelas tal que a área de qualquer par de parcelas adjacentes excede esta área máxima (Murray 1999, Murray e Snyder 2000). Neste caso, a formulação impede que parcelas adjacentes sejam selecionadas simultaneamente para corte raso. Ambas as abordagens exigem soluções inteiras e, por isso, os problemas são formulados em programação linear inteira ou programação linear inteira mista. Para estes problemas, têm sido usados métodos exatos em programação inteira ou, para reduzir os obstáculos computacionais resultantes do uso destes métodos, têm sido propostas heurísticas. A abordagem ARM produz melhores soluções, dada a

sua maior flexibilidade (Murray e Weintraub 2002) mas, pode ser mais difícil de resolver.

Três formulações ARM básicas são descritas na literatura: a formulação *path* (Martins *et al.* 1999, McDill *et al.* 2002, Murray e Weintraub 2002, Crowe *et al.* 2003), com um número exponencial de restrições, a formulação *cluster* (Martins *et al.* 1999 e 2005, McDill *et al.* 2002, Goycoolea *et al.* 2005, Vielma *et al.* 2007), com um número exponencial de variáveis, e a mais recente formulação *bucket* (Constantino *et al.* 2008), com um número polinomial de variáveis e restrições. Provou-se teoricamente que a formulação *cluster* é mais forte do que a formulação *path* (Goycoolea *et al.* 2009) e a formulação *bucket* (Martins *et al.* 2012), ou seja, a relaxação linear da formulação *cluster* dá melhores majorantes do que as relaxações lineares das restantes formulações. Goycoolea *et al.* 2009 implementaram as três formulações num sistema comercial de *branch-and-bound* e obtiveram soluções próximas das ótimas para um conjunto de instâncias relativamente grandes. Estes autores mostraram que não há uma formulação mais forte do que a outra entre as formulações *path* e *bucket*.

Nas formulações *cluster* e *bucket*, uma clareira é constituída por parcelas sujeitas ao corte raso no mesmo período, enquanto na formulação *path*, uma clareira pode formar-se gradualmente no tempo de exclusão. Para impedir a ocorrência de clareiras no tempo de exclusão com área superior ao valor máximo, as primeiras formulações restringem as áreas das clareiras em cada período e não permitem a ocorrência de clareiras adjacentes a estas durante o tempo de exclusão. A formulação *path* restringe apenas as áreas das clareiras no tempo de exclusão. As formulações *cluster* e *bucket* podem ser assim mais restritivas do que a formulação *path*. Para uma descrição resumida das três formulações veja-se Goycoolea *et al.* 2009. Para uma descrição mais detalhada veja-se McDill *et al.* 2002 (formulação *path*), Goycoolea *et al.* 2005 (formulação *cluster*) e Constantino *et al.* 2008 (formulação *bucket*).

As relações de adjacência entre parcelas, necessárias para modelar as restrições de clareiras, podem ser descritas por um grafo $G = (V, E)$ cujos nós em V representam as parcelas da área de estudo e cujas arestas em E correspondem a pares de parcelas adjacentes. Neste trabalho, considera-se que

duas parcelas são adjacentes quando a fronteira comum é um conjunto de pontos não discreto (adjacência forte). Existem outras definições de adjacência, como por exemplo, quando a fronteira comum é pelo menos um ponto. No caso da adjacência forte, o grafo G é planar, isto é, pode ser representado por uma figura planar em que as arestas não se intercetam. Para exemplificar, na Figura 1, está representada uma sub-região dos baldios Melhe e Seirós e o grafo correspondente.

Exemplo

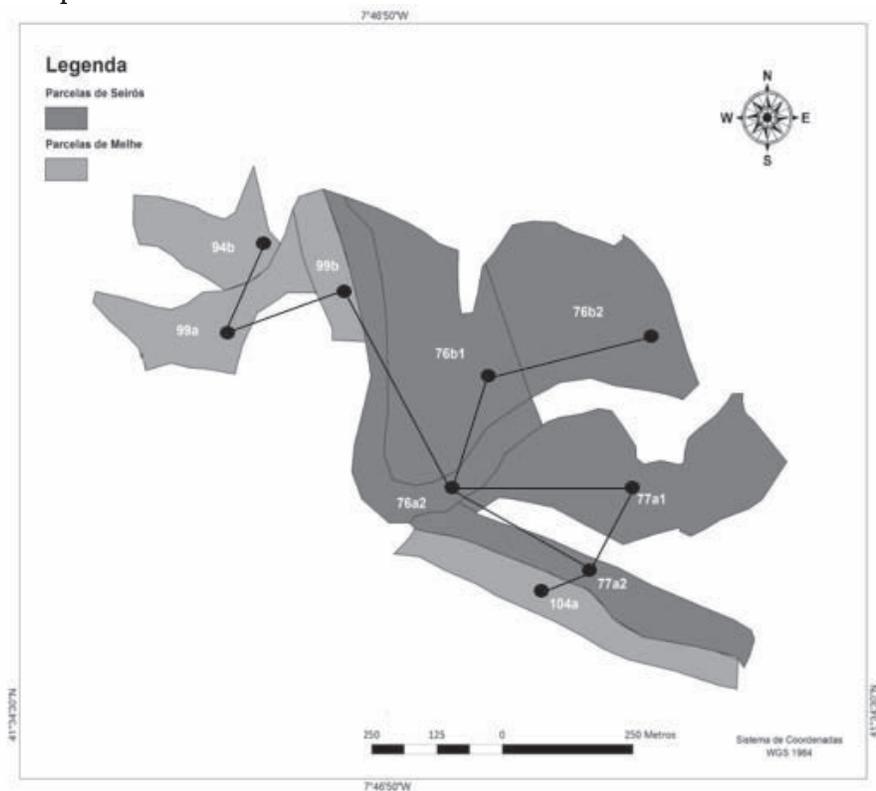


Figura 1: Sub-região dos baldios Melhe e Seirós e o grafo correspondente. As áreas das parcelas (em ha) são as seguintes: 76a2 - 4.88; 76b1 - 9.23; 76b2 - 8.93; 77a1- 8.81; 77a2 - 4.34; 94b - 3.13; 99a - 4.04; 99b - 1.92; 104a - 4.24. As parcelas têm mais de 15 anos no período 1

Usando os modelos descritos na secção anterior, a formulação *path* é a única que não necessita de novas variáveis. Para incluir as restrições (operacionais) do número de clareiras, a formulação *cluster* é a mais adequada, dado que permite contabilizar naturalmente este número. Decidiu-se

apresentar as três formulações, por qualquer uma delas constituir uma abordagem útil para apoiar o problema da gestão dos baldios.

Descrevem-se a seguir as variáveis e restrições que são necessárias para limitar as áreas das clareiras segundo as formulações *path*, *cluster* e *bucket*.

5.1 Formulações *Path* e *Cluster*

Seja A_{\max} a área máxima de uma clareira. Designe-se por *cluster* uma região contínua formada por parcelas com pelo menos a idade mínima para corte raso (15 anos). Um *cluster* é *admissível* se a sua área for inferior ou igual a A_{\max} . Caso contrário, é *não admissível*. Um *cluster* não admissível é *minimal* quando não contém outros *clusters* não admissíveis. Qualquer *cluster* corresponde a um subgrafo conexo do grafo G .

Exemplo (continuação) Considere-se a sub-região representada na Figura 1. Seja $A_{\max} = 10$ ha. O conjunto $\{76a2,77a2\}$ é um *cluster* admissível e $\{76a2,77a2,104a\}$ é um *cluster* não admissível minimal.

Formulação *path*

A formulação *path* garante que em qualquer *cluster* não admissível minimal haja pelo menos uma parcela que não seja sujeita ao corte raso durante o tempo de exclusão. Para cada período t , seja \mathfrak{R}^t o conjunto dos *clusters* não admissíveis minimais (veja-se Goycoolea *et al.* 2009 para um procedimento que determina \mathfrak{R}^t).

Exemplo (continuação) Seja $A_{\max} = 10$ ha. Para $t \in T$,

$$\mathfrak{R}^t = \{\{76a2,76b1\}, \{76a2,77a1\}, \{76b1,76b2\}, \{77a1,77a2\}, \{76a2,99a,99b\}, \\ \{76a2,77a2,99b\}, \{76a2,77a2,104a\}\}. \quad \bullet$$

Se r é o número de parcelas de um elemento R de \mathfrak{R}^t , então no máximo $(r - 1)$ parcelas podem ser cortadas em R no tempo de exclusão (que começa no período t e acaba em $t + 2$), o que é exigido pelas restrições (24).

$$\sum_{t'=t}^{t+2} \sum_{j \in R} x_j^{t'} \leq |R| - 1 \quad t = 1,2,3, \forall R \in \mathfrak{R}^t \quad (24)$$

Estas restrições asseguram assim que a área de cada clareira a formar no tempo de exclusão não excede A_{\max} .

O número de restrições (24) pode aumentar exponencialmente com o número de parcelas. No entanto, se NCN é o número máximo de parcelas dos *clusters* não admissíveis minimais, o número destas restrições é $O\left(T\|V\|^{NCN}\right)$ (Tabela A2 do Anexo). As formulações FA, FAI ou FB com as restrições (24) serão designadas por FA-*path*, FAI-*path* e FB-*path*, respetivamente.

Formulação cluster

A formulação *cluster* seleciona *clusters* admissíveis para corte raso em cada período e impede que *clusters* admissíveis adjacentes sejam selecionados durante o tempo de exclusão. Seja, para cada período t , o conjunto \mathfrak{S}^t constituído pelos *clusters* admissíveis (veja-se Goycoolea *et al.* 2009 para um procedimento que determina \mathfrak{S}^t).

Exemplo (continuação) Considere $A_{\max} = 10$ ha. Para $t \in T$,

$$\mathfrak{S}^t = \{\{76a2,76b1,76b2, 77a1,77a2,94b, 99a, 99b, 104a, \{76a2,77a2\}, \{76a2,99b\}, \{77a2,104a\}, \{94b, 99a\}, \{99a, 99b\}, \{94b,99a,99b\}\}. \bullet$$

Seja \aleph o conjunto das cliques maximais do grafo G . Um conjunto de parcelas adjacentes entre si corresponde a uma clique em G . Uma clique é maximal se não está contida em outra clique. Se o grafo G é planar, não contém cliques com cinco ou mais nodos (Diestel 2000) e, neste caso, \aleph pode ser determinado por um algoritmo polinomial.

Exemplo (continuação) Tem-se:

$$\aleph = \{\{76a2,76b1\}, \{76b1,76b2\}, \{76a2,99b\}, \{77a2,104a\}, \{94b,99a\}, \{99a,99b\}, \{76a2,77a1,77a2\}\}. \bullet$$

Considerem-se as seguintes variáveis de decisão: cl_s^t , que assume o valor 1 se a região $S \in \mathfrak{S}^t$ é sujeita ao corte raso no período t e 0 caso contrário. As restrições (25) – (27) são adicionadas às formulações FA, FAI ou FB.

$$x_j^t = \sum_{\substack{S \in \mathfrak{S}^t: \\ j \in S}} cl_s^t \quad \forall j \in V, t \in T : I_j + t \geq 16 \quad (25)$$

$$\sum_{t'=t}^{t+2} \sum_{\substack{S \in \mathfrak{S}^t: \\ S \cap C \neq \emptyset}} cl_s^{t'} \leq 1 \quad t = 1,2,3, \forall C \in \mathfrak{N} \quad (26)$$

$$cl_s^t \in \{0,1\} \quad \forall t \in T, S \in \mathfrak{S}^t. \quad (27)$$

As restrições (25) garantem que a parcela j é sujeita ao corte raso no período t quando um *cluster* que a contém é sujeito ao corte raso em t . As restrições (26) impedem que *clusters* adjacentes sejam selecionados para cortar no tempo de exclusão (as restrições (2) asseguram a seleção de *clusters* disjuntos entre si). Estas restrições garantem assim que a área de cada clareira não excede A_{\max} no tempo de exclusão. As restrições (27) indicam a natureza das novas variáveis.

O número de clareiras a formar no baldio i e no período t é dado por $\sum_{S \in \mathfrak{S}^t \cap B_i} cl_s^t$. Assim, se se pretender limitar este número, as seguintes restrições

são consideradas

$$\sum_{S \in \mathfrak{S}^t: S \cap B_i = S} cl_s^t \leq Ncl_{\max} \quad \forall B_i \in B, t \in T \quad (28)$$

em que Ncl_{\max} é o número máximo de clareiras a formar por período e por baldio.

O número de variáveis cl_s^t pode aumentar exponencialmente com o número de parcelas. No entanto, se NC é o número máximo de parcelas dos *clusters* admissíveis, o número destas restrições é $O(|T||V|^{NC})$ (Tabela A2 do Anexo). As formulações FA, FAI ou FB com as restrições (25) – (27) serão designadas por FA-*cluster*, FAI-*cluster* e FB-*cluster*, respetivamente.

5.2 Formulação *Bucket*

Como cada parcela é sujeita ao corte raso no máximo uma vez, a área a cortar na floresta no horizonte de planeamento é constituída por regiões contínuas que não se intercetam. Assim, cada uma destas regiões corresponde a um conjunto de parcelas que pode ser identificado por um dos seus elementos, por exemplo, o de menor índice. Considere para cada parcela k o conjunto $S_k = \emptyset$, designado por *bucket*. A formulação *bucket* atribui parcelas a S_k nalgum período (a parcela k e, eventualmente, parcelas $j > k$), transformando S_k numa região sujeita ao corte raso nesse período, ou deixa

S_k vazio. As parcelas são atribuídas a S_k de maneira a que os *buckets* não vazios não sejam adjacentes no tempo de exclusão e tenham área inferior ou igual a A_{\max} .

Considerem-se as seguintes variáveis de decisão: cl_j^{kt} , que assume o valor 1 se a parcela $j \in V$ pertence a S_k ($k \in V$) no período $t \in T$ e 0 caso contrário; v_C^{kt} , que assume o valor 1 se alguma parcela da clique $C \in \mathfrak{N}$ pertence a S_k ($k \in V$) no período $t \in T$ e 0 caso contrário. Note-se que se não existem regiões contínuas com j e k ($j \geq k$) cujas áreas são inferiores ou iguais a A_{\max} , então j não será atribuída a S_k em nenhum período e, neste caso, não vale a pena definir a variável cl_j^{kt} . Consequentemente, também não se definem as variáveis v_C^{kt} para as cliques $C \in \mathfrak{N}$ tais que a variável cl_j^{kt} não existe para qualquer nodo $j \in C$. Seja $El(k)$ o conjunto das parcelas $j \geq k$ tais que existem regiões contínuas que incluem ambas as parcelas (k e j) e cujas áreas são inferiores ou iguais a A_{\max} (veja-se Constantino *et al.* 2008 para um procedimento que determina $El(k)$).

Exemplo (continuação) Considere $A_{\max} = 10$ ha. Tem-se:

$$El(99b) = \{76a2,94b,99a,99b\}. \quad \bullet$$

As variáveis cl_j^{kt} não são definidas para os períodos t em que as parcelas k e j não têm a idade adequada para o corte raso, isto é, $I_k + t < 16$ e $I_j + t < 16$. Consequentemente, também não se definem as variáveis v_C^{kt} para as cliques C tais que não existem nodos $l \in C$ em que $I_l + t \geq 16$ e $l \geq k$. Para simplificar o texto, estas condições não são incluídas na descrição da formulação.

As restrições (29) – (36) são adicionadas às formulações FA, FAI ou FB.

$$x_j^t = \sum_{\substack{k \leq j; \\ j \in El(k)}} cl_j^{kt} \quad \forall j \in V, t \in T \quad (29)$$

$$cl_j^{kt} \leq cl_k^{kt} \quad \forall j \in V, k \leq j : j \in El(k), t \in T \quad (30)$$

$$cl_j^{kt} \leq v_C^{kt} \quad \forall C \in \mathfrak{N}, j \in C, k \leq j : j \in El(k), t \in T \quad (31)$$

$$v_C^{kt} \leq cl_k^{kt} \quad \forall C \in \mathfrak{N}, k \leq j \text{ e } k \in El(j) \text{ para algum } j \in C, t \in T \quad (32)$$

$$\sum_{t'=t}^{t+2} \sum_{\substack{k \leq j \text{ e } k \in El(j), \\ \text{para algum } j \in C}} v_C^{kt'} \leq 1 \quad t = 1, 2, 3, \forall C \in \mathfrak{N} \quad (33)$$

$$\sum_{j \in El(k)} A_j c l_j^{kt} \leq A_{\max} c l_k^{kt} \quad \forall k \in V, t \in T \quad (34)$$

$$c l_j^{kt} \in \{0, 1\} \quad \forall j \in V, k \leq j : k \in El(j), t \in T \quad (35)$$

$$v_C^{kt} \geq 0 \quad \forall C \in \mathfrak{N}, k \leq j \text{ e } k \in El(j) \text{ para algum } j \in C, t \in T. \quad (36)$$

As restrições (29) garantem que a parcela j é sujeita ao corte raso no período t quando é atribuída a algum S_k em t . As restrições (30) garantem que qualquer S_k não vazio contém k . As restrições (31) definem a relação entre as variáveis $c l_j^{kt}$ e v_C^{kt} . As restrições (32) forçam as variáveis v_C^{kt} a ser zero quando S_k é vazio. As restrições (33) asseguram que não existem *buckets* não vazios adjacentes no tempo de exclusão (as restrições (2) garantem que os *buckets* não vazios são disjuntos entre si). As restrições (34) garantem que as áreas dos *buckets* não vazios não excedem A_{\max} . Fica assim assegurado que a área de cada clareira não excede A_{\max} no tempo de exclusão. As restrições (35) e (36) definem a natureza das variáveis.

Note-se que a integralidade das variáveis $c l_j^{kt}$ juntamente com as restrições (31) implica a integralidade das variáveis v_C^{kt} em pelo menos uma solução ótima. Note-se também que os *buckets* não vazios podem corresponder a regiões descontínuas dado que não há restrições que garantam a sua conectividade. No entanto, cada uma destas regiões é constituída por regiões contínuas com áreas inferiores ou iguais a A_{\max} .

Esta formulação não permite contabilizar naturalmente o número de clareiras a formar em cada ano. Note-se que $\sum_{k \in V} c l_k^{kt}$ representa o número de regiões sujeitas ao corte raso no período t , contínuas ou descontínuas, sendo apenas clareiras aquelas que são contínuas.

As formulações FA, FAI ou FB com as restrições (29) – (36) serão designadas por *FA-bucket*, *FAI-bucket* e *FB-bucket*, respetivamente. O número de variáveis e restrições destas formulações é $O(|V|^2|T|)$ quando o grafo G é planar (Tabela A3 do Anexo).

Como já foi referido, FAI não tem restrições que misturam variáveis associadas a baldios diferentes ou a parcelas de baldios diferentes. No entanto, como existem baldios adjacentes na área de estudo (Figura 2), algumas restrições ambientais têm variáveis associadas a parcelas (versões *path* e *bucket*) ou a *clusters* (versão *cluster*) de vários baldios. Consequentemente, os planos de gestão ótimos obtidos por FAI-*path*, FAI-*cluster* e FAI-*bucket* para cada baldio não são independentes dos restantes baldios, exceto para Canedo, que não tem baldios adjacentes.

6. Caso de estudo

A área de estudo, que se localiza no concelho de Ribeira de Pena, ocupa cerca de 1393 ha e é constituída pelos baldios de Canedo (B_1), com 121 ha, Seirós (B_2), com 165 ha, Melhe (B_3), com 341 ha, Santa Marinha Norte (B_4), com 306 ha e Santo Aleixo (B_5), com 465 ha. A espécie predominante das parcelas destes baldios é o pinheiro bravo.

Os dados sobre o estado dos povoamentos foram obtidos com base na informação recolhida em inventários florestais realizados entre 2004 e 2010 (Mota, 2011). Para além da constituição da base de dados com as características dos povoamentos, criou-se também um sistema de informação, assente em cartografia digital, para delimitação das parcelas e posterior representação das áreas a submeter ao plano de cortes. A delimitação teve em conta as divisões de planeamento (talhões) assumidas pela Direção Regional de Florestas do Norte e a homogeneidade dos povoamentos. Como as orientações regionais para a floresta impõem um limite máximo para as clareiras de 10 ha, procedeu-se a nova divisão de forma a garantir parcelas com áreas não superiores a este valor. A divisão foi realizada na aplicação informática ArcGIS 9.3 (ESRI-*Environmental Systems Research Institute*), com base em ortofotomapas obtidos a partir de uma cobertura aerofotográfica efetuada em 2006 pelo ICFN e numa folha *raster* georreferenciada da Série M888 da Carta Militar do Instituto Geográfico do Exército. A área de estudo ficou constituída por 209 parcelas (Figura 2) com áreas compreendidas entre 0.61 ha e 10 ha (área média de 6.67 ha).

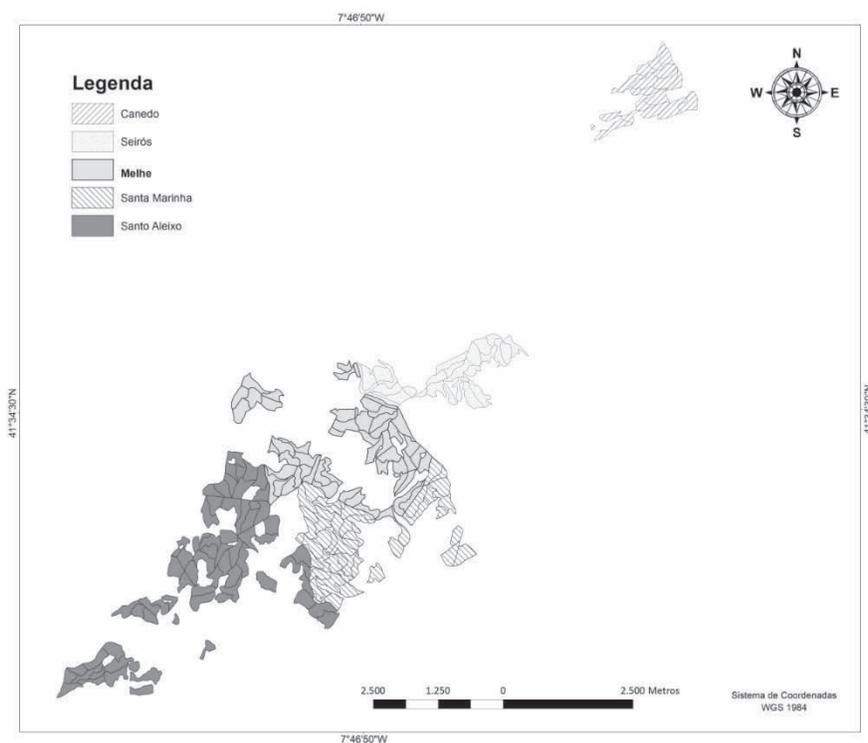


Figura 2: Enquadramento geográfico da área de estudo, com indicação dos baldios e das parcelas que os constituem

O horizonte de planeamento, com a duração de cinco anos, foi definido entre 2012 e 2016 (períodos 1 a 5). A projeção do crescimento de cada parcela para o horizonte de planeamento foi realizada de acordo com modelos de crescimento e de produção desenvolvidos para o pinheiro bravo (Marques 1991, Moreira e Fonseca 2002, Luís e Fonseca 2004, Fonseca *et al.* 2012a, Fonseca *et al.* 2012b).

Na tabela 2 descrevem-se sumariamente os baldios que constituem a área de estudo³. Para cada baldio i , apresenta-se o número de parcelas, $|B_i|$, e a

área média das parcelas, $\frac{\sum_{j \in B_i} A_j}{|B_i|}$. Apresentam-se também algumas medidas relativas ao início do período 1: a idade média ponderada pelas áreas da

³ A descrição individual das parcelas, que não é apresentada por ser demasiado extensa, pode ser fornecida pelos autores.

parcelas, $\frac{\sum_{j \in B_i} I_j A_j}{\sum_{j \in B_i} A_j}$; o índice de densidade SDI médio ponderado pelas áreas das parcelas, $\frac{\sum_{j \in B_i} SDI_j^1 A_j}{\sum_{j \in B_i} A_j}$; o volume em pé, $\sum_{j \in B_i} C_j^1$; e o acréscimo anual da produção de madeira, $\sum_{j \in B_i} Ac_j^1$.

Tabela 2: Descrição sumária dos baldios no início de 2012

Baldio	Nº parc.	Área média (ha)	Idade média (anos)	SDI Médio (%)	Volume em pé (m³)	Acréscimo anual da prod. madeira (m³ano ⁻¹)
Canedo	17	7.1	35.6	43.4	41161	2682
Seirós	25	6.6	38.5	25.7	29749	2266
Melhe	53	6.4	29.0	32.0	70262	6310
Sta. Marinha	46	6.5	32.0	26.5	61906	5029
St. Aleixo	68	6.8	32.4	36.4	124391	9544

Em 2012 não há parcelas com valores de SDI iguais ou superiores a 60% em qualquer baldio, pelo que não há madeira disponível para desbaste neste ano. De referir que, no horizonte de planeamento, valores de SDI superiores ou iguais a 60% ocorrem apenas em parcelas do baldio de St. Aleixo, a partir de 2013 inclusive.

Neste trabalho, consideraram-se os parâmetros $\Omega = \Delta = 0.2$ para as restrições (7) – (9), valor que permite variações aceitáveis nos volumes de madeira retirada globalmente em toda a floresta ou individualmente em cada baldio.

7. Resultados computacionais

Os vários modelos foram implementados e resolvidos no sistema comercial de *branch-and-bound Xpress Optimizer Version 22.01.0* (disponível em <http://www.fico.com/en/Products/DMTools/Pages/FICO-Xpress-Optimization->

Suite.aspx). Limitou-se o tempo de execução do algoritmo em 7200 segundos. Usou-se um computador i5-2500 3.3GHZ/3.7Turbo com 8GB de RAM.

A Tabela 3 apresenta o número de variáveis e restrições dos vários modelos, bem como informação sobre o desempenho do algoritmo de *branch-and-bound* (número de nós analisados, tempo de execução, *gap*) e o valor da função objetivo da solução obtida. O *gap* associado a uma solução é a diferença relativa entre o respetivo valor da função objetivo e o melhor majorante determinado até ao momento.

Tabela 3: Dimensão dos modelos, desempenho do *branch-and-bound* e valor da função objetivo da solução obtida. Os valores entre () são os volumes totais de madeira a retirar

Formulação	Dimensão		<i>Branch-and-bound</i>			F.O.
	Nº var.	Nº rest.	Nº nós	Tempo (s)	<i>Gap</i> (%)	
FA	2120	365	157073	66	0.0	49109m ³
FA- <i>path</i>	2120	1935	30274	14	0.0	48944
FA- <i>cluster</i>	3270	2043	23763	12	0.0	48944
FA- <i>bucket</i>	6050	7908	17871	9	0.0	48944
FB	2121	370	4130023	7200	1.07	0.1085 (47718m ³)
FB- <i>path</i>	2121	1940	3608835	7200	1.42	0.1075 (47003)
FB- <i>cluster</i>	3271	2048	5995822	7200	2.56	0.1075 (47022)
FB- <i>bucket</i>	6051	7913	1934406	7200	2.39	0.1059 (46248)
FAI	2120	365	10503	3	0.0	49070m ³
FAI- <i>path</i>	2120	1935	12395	4	0.0	48291
FAI- <i>cluster</i>	3270	2043	34465	9	0.0	48291
FAI- <i>bucket</i>	6050	7908	12211	5	0.0	48291

É de salientar que o número de variáveis e restrições não é elevado. Por um lado, o número total de parcelas (Tabela 2) e o número de períodos são relativamente baixos. Por outro lado, as parcelas são relativamente grandes tendo em conta que $A_{\max} = 10$ (Tabela 2) e, por isso, o número de *clusters* não admissíveis minimais (versão *path*) e o número de *clusters* admissíveis (versão *cluster*), 314 e 230, respetivamente, não são elevados.

O algoritmo terminou com a obtenção de uma solução ótima com os modelos FA e FAI e apresentou um *gap* superior a 1% ao fim de duas horas com os modelos FB. Ou seja, o plano de exploração ótimo é mais difícil de determinar quando se pretende maximizar o menor volume relativo de madeira a retirar por baldio. Os modelos FA conduziram a planos de exploração com volumes totais de madeira a remover superiores aos obtidos

pelos correspondentes modelos FAI. Assim, as restrições de sustentabilidade dos modelos FAI relativas aos baldios (restrições (22) e (23)) causaram uma redução dos volumes totais ótimos obtidos pelos modelos FA, tendo sido por isso mais restritivas do que as restrições similares relativas à floresta (restrições (10) e (11)) e as restrições de regularidade da receita total (restrições (8) e (9)). Os volumes totais de madeira a remover mais baixos foram obtidos com os modelos FB.

A inclusão das restrições ambientais não diminuiu significativamente o volume ótimo de madeira a remover. Este valor decresceu 0.3% com os modelos FA, 1.6% com FAI e cerca de 1.5% com os modelos FB. Estas pequenas diferenças são explicadas pelo facto de ocorrer um número reduzido de clareiras com área superior a 10 ha nos planos de exploração obtidos com os modelos sem as restrições ambientais. É de salientar também que aquelas restrições não dificultaram a resolução do problema, o que não surpreende, pois o número de variáveis e restrições associadas às versões *path*, *cluster* e *bucket* não é elevado. Apesar das restrições *path* poderem ser menos restritivas, os planos de exploração ótimos obtidos com FA-*path* e FAI-*path* são iguais aos obtidos com as versões correspondentes *cluster* e *bucket*. Os modelos FB não foram resolvidos até à otimalidade, havendo ligeiras diferenças entre os planos de exploração obtidos.

A seguir são apresentados resultados computacionais mais detalhados sobre os modelos FA e FAI com as restrições ambientais. Como estes resultados são comuns às versões *path*, *cluster* e *bucket*, os modelos FA-*path*, FA-*cluster* e FA-*bucket* serão referidos indistintamente por FA-, os modelos FAI-*path*, FAI-*cluster* e FAI-*bucket* por FAI-. Relativamente a FB, são apresentados resultados com uma versão, por não existirem diferenças relevantes entre as versões *path*, *cluster* e *bucket*. Selecionou-se a versão *cluster* para se analisar o impacto das restrições (28). Nas Tabelas 4 a 6 descrevem-se os planos de exploração obtidos com estes modelos, discriminando, por baldio, os tipos de intervenção a realizar (corte raso «CR» ou desbaste «Db»), as idades mínima e máxima das parcelas sujeitas a intervenção no período em que esta ocorre, a área sujeita a intervenção e o volume de madeira a remover em cada ano do horizonte de planeamento.

Tabela 4: Plano de exploração obtido com FA-*path*, FA-*cluster* e FA-*bucket*

Baldio	Intervenção			Volume de madeira a remover (m ³)					
	Tipo	Idades das parcelas (anos)	Área (ha)	2012	2013	2014	2015	2016	Total
Canedo	CR	45	8.4	0	0	3630	0	0	3630
Seirós	CR	46 a 48	24.7	0	2315	0	2746	0	5062
Melhe	CR	43 a 46	28.3	0	6029	0	0	5272	11301
Sta. Marinha	CR	41 a 44	6.9	1053	0	0	1101	0	2154
St. Aleixo	CR	39 a 46	49.0	6199	0	4054	5928	4898	21079
	Db	26 a 39	43.7	0	0	2158	1450	2110	5718
<i>Total</i>			161.0	7252	8344	9842	11225	12280	48944

Observa-se que o volume total de madeira a retirar aumenta ao longo do horizonte de planeamento com FA-. A amplitude total dos volumes de madeira a retirar anualmente em toda a floresta é de 5028 m³. Note-se que as parcelas são ainda bastante jovens em 2012 (Tabela 2) e, por isso, o volume de madeira em pé em cada período é em geral crescente no horizonte de planeamento. Os cortes rasos são responsáveis pela maior parte do volume de madeira a remover (88.3%). Este tipo de intervenção ocorre em 8.4% da área total, em parcelas com idades entre 39 e 48 anos. Como já foi referido, as parcelas que atingem 60% de *SDI* no horizonte de planeamento pertencem ao baldio de St. Aleixo e, por isso, os desbastes ocorrem apenas neste baldio. Este tipo de intervenção realiza-se em parcelas com idades entre 26 e 39 anos, em 3.1% da área total.

Tabela 5: Plano de exploração obtido com FAI-*path*, FAI-*cluster* e FAI-*bucket*

Baldio	Intervenção			Volume de madeira a remover (m ³)					
	Tipo	Idade das parcelas (anos)	Área (ha)	2012	2013	2014	2015	2016	Total
Canedo	CR	38 a 41	7.3	0	0	2452	0	0	2452
Seirós	CR	44	7.6	0	0	1943	0	0	1943
Melhe	CR	42 a 46	37.5	0	6808	0	0	7228	14036
Sta. Marinha	CR	43 a 45	33.4	0	0	5218	0	6217	11435
St. Aleixo	CR	41 a 46	27.3	0	0	3076	4928	4768	12772
	Db	26 a 39	43.7	0	0	2733	810	2110	5653
<i>Total</i>			156.8	0	6808	15422	5738	20323	48291

No plano de exploração ótimo obtido com FAI-, em que as restrições de regularidade da receita total (restrições (8) e (9)) não são consideradas, a distribuição do volume total de madeira a retirar ao longo do horizonte de planeamento é mais irregular do que a observada com FA-. A amplitude total dos volumes de madeira a retirar anualmente em toda a floresta é de 20323 m³. O volume total de madeira a retirar é nulo no primeiro ano, oscila acentuadamente nos anos seguintes, e o maior valor é obtido no último ano. Note-se que, em cada baldio, tem de haver receita em cada triénio (restrições (4)), mas quanto mais tarde se intervém numa parcela maior é o volume de madeira a retirar devido ao crescimento das árvores. Em termos de mercado, a solução proposta por este modelo pode ser uma opção menos adequada por corresponder a picos na oferta de madeira.

Comparando com FA-, observou-se um aumento do volume de madeira a remover de mais de 400% no baldio de Santa Marinha e de cerca de 24% em Melhe. Nos baldios de Canedo, Seirós e Santo Aleixo, o volume total de madeira a remover diminuiu, tendo esta diminuição sido mais acentuada em termos percentuais em St. Aleixo (cerca de 73%).

Tabela 6: Plano de exploração obtido com FB-*cluster*

Baldio	Intervenção			Volume de madeira a remover (m ³)					
	Tipo	Idade das parcelas (anos)	Área (ha)	2012	2013	2014	2015	2016	Total
Canedo	CR	38 a 45	14.9	0	0	5803	0	0	5803
Seirós	CR	47 a 52	18.8	1364	0	0	1191	1394	3949
Melhe	CR	41 a 46	27.4	3702	3690	0	0	3132	10524
Sta. Marinha	CR	41 a 43	28.3	2791	0	3182	3009	0	8981
St. Aleixo	CR	38 a 46	29.6	0	3545	0	4870	4898	13313
	Db	25 a 39	36.5	0	1907	0	1450	1094	4451
<i>Total</i>			155.5	7857	9142	8985	10520	10518	47022

No plano de exploração obtido com FB-*cluster*, há uma distribuição equilibrada da receita total no horizonte de planeamento com tendência para aumentar. A amplitude total dos volumes de madeira a retirar anualmente em toda a floresta é de 2663 m³. Comparando com FA-, observa-se que a quantidade de madeira a remover aumenta nos baldios de Santa Marinha e Canedo, sendo este aumento mais acentuado em termos percentuais em Santa

Marinha (mais de 300%) e diminui nos restantes baldios, com o decréscimo mais acentuado em St. Aleixo (cerca de 67%). Estas alterações causaram uma redução do volume total de madeira a remover de 3%, aproximadamente. A área sujeita ao desbaste diminuiu, pelo facto de uma das parcelas que atinge 60% de *SDI* estar sujeita ao corte raso em vez de ao desbaste.

É de salientar que a idade das parcelas no ano em que são sujeitas ao corte raso é semelhante em todos os modelos.

Na Tabela 7 são apresentados os rácios p_i do volume de madeira a retirar relativamente ao volume máximo que se pode retirar em cada baldio para os planos de exploração descritos nas Tabelas 4 a 7.

Tabela 7: Rácios p_i do volume total de madeira a retirar relativamente ao volume máximo que se pode retirar no horizonte de planeamento

Baldio	FA-	FAI-	FB- <i>cluster</i>
Canedo	0.074	0.050	0.118
Seirós	0.163	0.112	0.108
Melhe	0.138	0.053	0.108
Sta. Marinha	0.025	0.135	0.106
St. Aleixo	0.117	0.146	0.109

Os rácios p_i variam entre 0.106 e 0.118 com *FB-cluster*, entre 0.025 e 0.163 com *FA-* e entre 0.05 e 0.146 com *FAI-*. A variação da amplitude total destes valores (0.012, 0.096 e 0.138 para *FB-cluster*, *FAI-* e *FA-*, respetivamente) reflete a variação do volume total de madeira a remover em toda a floresta nos planos de exploração obtidos (maiores amplitudes estão associadas a maiores volumes).

Nas Tabelas 8 e 9 descrevem-se os baldios em termos de idade média das parcelas, volume em pé e acréscimo anual da produção de madeira no final de 2016, para os planos de exploração apresentados nas Tabelas 4 a 6.

Com os modelos em que as restrições de sustentabilidade são aplicadas globalmente a toda a floresta, *FA-* e *FB-cluster*, observa-se que, em cada baldio, o volume em pé no final do horizonte de planeamento é superior ao volume em pé inicial, o que não acontece com o acréscimo anual da produção de madeira, que é inferior nos baldios de Canedo, Seirós e Santo Aleixo (Tabela 2). Quanto à idade média das parcelas, os valores no final do horizonte de

planeamento que são inferiores aos iniciais estão mais próximos destes com os modelos FAI-, em que as restrições de sustentabilidade são aplicadas individualmente a cada baldio (Tabela 2).

Tabela 8: Descrição sumária dos baldios no final de 2016 para os planos de exploração ótimos obtidos com FA-*path*, FA-*cluster*, FA-*bucket*, FAI-*path*, FAI-*cluster* e FAI-*bucket*

Baldio	Modelos FA-			Modelos FAI-		
	Idade média (anos)	Volume em pé (m ³)	Acrésc. anual madeira (m ³ ano ⁻¹)	Idade média (anos)	Volume em pé (m ³)	Acrésc. anual madeira (m ³ ano ⁻¹)
Canedo	36.5	47674	2656	37.1	48896	2688
Seirós	35.1	31535	2111	40.5	34669	2271
Melhe	29.3	84042	6464	28.1	81154	6311
Sta. Marinha	35.1	80999	5511	31.2	71454	5029
St. Aleixo	31.9	136299	9090	33.8	145807	9544
<i>Total</i>	-	380549	25832	-	381980	25843

Tabela 9: Descrição sumária dos baldios no final de 2016 para o plano de exploração obtido com FB-*cluster*

Baldio	Modelo FB- <i>cluster</i>		
	Idade média(anos)	Volume em pé (m ³)	Acrésc.anual madeira (m ³ ano ⁻¹)
Canedo	34.3	45247	2527
Seirós	36.9	32599	2167
Melhe	29.5	84365	6480
Sta. Marinha	32.0	73273	5121
St. Aleixo	33.7	146425	9536
<i>Total</i>	-	381909	25831

A Figura 3 apresenta, como exemplo, o mapa do baldio de St. Aleixo com o plano de exploração ótimo obtido com os modelos FA-.

As restrições de sustentabilidade e as restrições ambientais estão ambas relacionadas com a proteção da floresta. Para avaliar a importância das restrições de sustentabilidade no problema da gestão dos baldios, foi feito um estudo com todos os modelos em que uma ou ambas as restrições de sustentabilidade foram eliminadas. A Tabela 10 sumariza os resultados obtidos com os modelos FAI-. Os resultados obtidos com as formulações FA- e FB-*cluster* são semelhantes aos de FAI- e, por isso, não são apresentados.

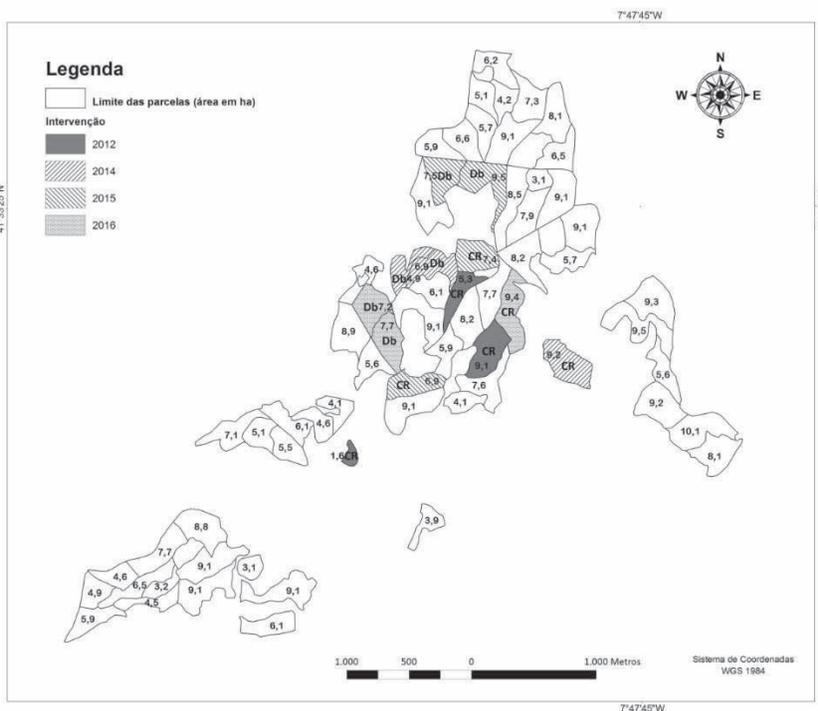


Figura 3: Plano de exploração ótimo para o baldio de St. Aleixo obtido com FA-*path*, FA-*cluster* e FA-*bucket*. As áreas das parcelas e os tipos de intervenção (CR – corte raso e Db - desbaste) são apresentadas

Tabela 10: Influência das restrições de sustentabilidade nos planos de exploração obtidos com FAI-*path*, FAI-*cluster*, FAI-*bucket*

Modelo FAI-	Volume removido (m ³)	Área desbaste (ha)	Área corte raso (ha)	Idade média final (anos)	Vol. em pé final (m ³)	Acrésc. anual madeira final (m ³ ano ⁻¹)
FAI-	48291	43.7	113.1	32.9	381979	25842
FAI- s/(23)	101598	35.9	299.7	28.2	327499	22181
FAI- s/(22)	48291	43.7	113.1	32.9	381979	25842
FAI- s/(22) e s/(23)	325204	14.1	1138.5	7.7	71615	5006

Observa-se que sem as restrições do acréscimo anual da produção de madeira (restrições (23)), as intervenções por corte raso acentuam-se e por desbaste diminuem, com consequências para a idade média, o volume em pé e o acréscimo anual no final do horizonte de planeamento, que diminuem.

Com as restrições do acréscimo anual e sem as restrições do volume em pé (restrições (22)), não há alterações. As restrições do acréscimo anual mostraram-se, assim, mais efetivas para garantir a persistência das áreas de floresta nos baldios. A diminuição mais acentuada da idade média das parcelas, do volume em pé e do acréscimo anual da produção de madeira no final do horizonte de planeamento ocorreu sem as duas restrições, parecendo assim que ambas as restrições são importantes para garantir uma floresta mais sustentável.

Como já foi referido, pode ser operacionalmente vantajoso restringir, para cada ano, o número de clareiras a formar, isto é, o número de regiões sujeitas ao corte raso. Na Tabela 11 apresenta-se, para cada baldio, o número de clareiras a formar no horizonte de planeamento, para os planos de exploração descritos nas Tabelas 4 a 6.

Tabela 11: Número de clareiras a formar no horizonte de planeamento, para os planos de exploração ótimos obtidos com FA-, FAI- e FB-cluster

Baldio	FA-					FAI-					FB-cluster				
	2012	2013	2014	2015	2016	2012	2013	2014	2015	2016	2012	2013	2014	2015	2016
Canedo	0	0	1	0	0	0	0	2	0	0	0	0	3	0	0
Seirós	0	2	0	2	0	0	0	1	0	0	1	0	0	1	1
Melhe	0	2	0	0	2	0	3	0	0	3	2	1	0	0	1
Sta. Marinha	1	0	0	1	0	0	0	3	0	3	1	0	2	1	0
St. Aleixo	3	0	1	2	1	0	0	1	2	1	0	2	0	2	1
<i>Total</i>	4	4	2	5	3	0	3	7	2	7	4	3	5	4	3

Apesar das soluções obtidas com os modelos FA-, FAI- e FB-cluster não apresentarem, em geral, muitas clareiras, efetuaram-se alguns testes computacionais com a versão *cluster* para analisar o impacto das restrições (28). Atendendo a que nas soluções obtidas o maior número de clareiras a formar por ano e por baldio é três, considerou-se $N_{cla}_{max} = 2$. Na Tabela 12 apresenta-se o volume total de madeira a remover no horizonte de planeamento e a Tabela 13 indica, para cada baldio, o número de clareiras a formar em cada ano, relativamente aos planos de exploração ótimos obtidos. O modelo FB-cluster não foi resolvido até à otimalidade, tendo o algoritmo de *branch-and-bound* terminado com um *gap* de 1.03% ao fim de 7200 segundos.

Tabela 12: Volume total de madeira a remover no horizonte de planejamento, em que

$$Ncla_{max} = 2$$

Baldio	Tipo de intervenção	Volume de madeira a remover (m ³)		
		FA-cluster + (28)	FAI-cluster + (28)	FB-cluster + (28)
Canedo	CR	3630	2452	5821
Seirós	CR	5062	1943	4048
Melhe	CR	11301	13682	10524
Sta. Marinha	CR	2209	11425	9140
St. Aleixo	CR	21047	12772	12401
	Db	5635	5653	5570
<i>Total</i>		48884	47927	47504

Tabela 13: Número de clareiras a formar no horizonte de planejamento, em que $Ncla_{max} = 2$

Baldio	FA-cluster + (28)					FAI-cluster + (28)					FB-cluster + (28)				
	2012	2013	2014	2015	2016	2012	2013	2014	2015	2016	2012	2013	2014	2015	2016
Canedo	0	0	1	0	0	0	0	2	0	0	0	0	2	0	0
Seirós	0	2	0	2	0	0	0	1	0	0	0	2	0	1	0
Melhe	0	2	0	0	2	2	0	0	2	2	2	1	0	0	1
Sta. Marinha	1	0	0	1	0	0	0	2	2	2	0	1	0	2	1
St. Aleixo	2	0	2	2	1	0	0	1	2	1	2	0	1	2	1
<i>Total</i>	3	4	3	5	3	2	0	6	6	5	4	4	3	5	3

Comparando com os resultados apresentados nas Tabelas 4 a 6, observa-se que o volume total de madeira a remover reduziu ligeiramente em relação aos modelos FA-cluster e FAI-cluster (cerca de 0.6%) e aumentou cerca de 1% em relação a FB-cluster. De notar que com FB-cluster, o algoritmo de *branch-and-bound* terminou com um gap inferior ao registado sem as restrições (28) (Tabela 3). Comparando com os resultados da Tabela 11, observa-se que em alguns baldios aparecem clareiras em anos em que não tinham sido propostos cortes rasos.

Foram também efetuados testes computacionais considerando $A_{max} = 20$ ha. Neste caso, em que há maior flexibilidade na formação das clareiras, obtiveram-se planos de exploração com menos e maiores clareiras por ano e por baldio.

Como seria de esperar, os planos de exploração obtidos com os modelos propostos têm características diferentes entre si e, por isso, podem satisfazer de

forma diferente os interesses e expectativas das entidades envolvidas na gestão dos baldios. Para descrever resumidamente o plano de exploração ótimo associado a cada modelo, selecionaram-se as seguintes seis características principais: 1) maximização da explorabilidade ao nível da floresta, 2) uniformização da explorabilidade ao nível dos baldios, 3) manutenção da sustentabilidade ao nível da floresta, 4) manutenção da sustentabilidade ao nível dos baldios, 5) regularização da receita no horizonte de planeamento e 6) limitação do número de clareiras a formar por ano e por baldio. As duas primeiras características estão associadas aos objetivos (1) e (21), respetivamente, e são avaliadas pelo volume de madeira a remover em toda a floresta no horizonte de planeamento (caraterística 1) e pela amplitude total dos valores do rácio entre o volume total de madeira a remover e o volume máximo de madeira que se pode remover no horizonte de planeamento para os diversos baldios (caraterística 2). As terceira e quarta características estão associadas às restrições de sustentabilidade (10) – (11) e (22) – (23), respetivamente. Um modelo tem cada uma destas características se garantir que as restrições correspondentes são satisfeitas. A quinta característica está associada às restrições (8) – (9) e pode ser avaliada pela amplitude total dos volumes anuais de madeira a remover em toda a floresta. A caraterística 6 está associada às restrições (28) e tem a ver com a possibilidade da formulação contabilizar e limitar o número de clareiras a formar por ano e por baldio.

A Tabela 14 descreve as características principais dos planos de exploração ótimos obtidos com os diversos modelos. As características associadas a cada modelo estão assinaladas com «✓». Não há nenhuma formulação que tenha todas as características. As formulações que maximizam a explorabilidade ao nível da floresta (modelos FA e FAI) não asseguram uma uniformização da explorabilidade ao nível dos baldios (modelos FB) e vice-versa. É evidente que as formulações que garantem a sustentabilidade ao nível dos baldios também asseguram a sustentabilidade ao nível da floresta (modelos FAI), mas o recíproco não é verdade (modelos FA e FB). Os modelos FAI não asseguram a regularidade da receita no horizonte de planeamento e os modelos nas versões *path* ou *bucket* não limitam o número de clareiras a formar por ano e por baldio.

Tabela 14: Características principais dos planos de exploração ótimos obtidos com os modelos FA, FB e FAI nas versões *path*, *cluster* e *bucket*. As características associadas a cada modelo estão assinaladas com «✓»

Caraterística	FA- <i>cluster</i>	FB- <i>cluster</i>	FAI- <i>cluster</i>	FA- <i>path/bucket</i>	FB- <i>path/bucket</i>	FAI- <i>path/bucket</i>
Maximização da explorabilidade ao nível da floresta	✓		✓	✓		✓
Uniformização da explorabilidade ao nível dos baldios		✓			✓	
Manutenção da sustentabilidade ao nível da floresta	✓	✓	✓	✓	✓	✓
Manutenção da sustentabilidade ao nível dos baldios			✓			✓
Regularização da receita no horizonte de planeamento	✓	✓		✓	✓	
Limitação do número de clareiras a formar por ano e por baldio	✓	✓	✓			

8. Conclusões

O objetivo deste trabalho consistiu em propor uma ferramenta útil para apoiar a gestão dos baldios de pinheiro bravo localizados no concelho de Ribeira de Pena, na região Norte de Portugal. A gestão destes baldios, da responsabilidade dos órgãos gestores eleitos pelas Assembleias de Compartes e do ICFN, foi feita até há bem pouco tempo de forma essencialmente empírica. A descrição do problema da gestão dos baldios resultou de um processo interativo com técnicos das entidades gestoras, que também colaboraram na recolha e tratamento dos dados. Durante este processo, apresentaram-se e discutiram-se soluções de problemas de otimização que iam sendo ajustados de forma a contemplar os interesses e as expetativas daquelas entidades.

A ferramenta proposta neste trabalho consiste em modelos de programação linear inteira que otimizam o plano de exploração dos baldios para períodos de cinco anos. Escolheu-se maximizar o volume de madeira a retirar, em vez de uma expressão financeira do resultado da exploração, porque, para o caso

de estudo, o preço da madeira mantém-se aproximadamente constante. Em alternativa, propôs-se maximizar o menor volume relativo de madeira a retirar por baldio. Consideraram-se restrições de natureza silvícola, operacional, organizacional, ambiental e, ainda, restrições de sustentabilidade, diretamente relacionadas com a persistência das áreas de floresta nos baldios. Propôs-se aplicar as restrições de sustentabilidade globalmente a toda a floresta ou individualmente a cada baldio. As restrições ambientais, que limitam as áreas das clareiras, foram formuladas de acordo com os modelos básicos *path*, *cluster* e *bucket* descritos na literatura. O plano de exploração ótimo é mais difícil de determinar quando se pretende maximizar o menor volume relativo de madeira retirada por baldio. As restrições ambientais não dificultaram a resolução dos modelos, dada a dimensão da área de estudo ser média e as parcelas terem áreas relativamente grandes.

Alguns dos planos de exploração obtidos pelos modelos propostos foram adotados pelas entidades gestoras dos baldios, com benefícios imediatos, nomeadamente em termos da exploração de madeira, não colocando em risco a sustentabilidade e as condições ambientais da floresta. Certamente que terá sido necessário fazer adaptações pontuais, para ultrapassar imprevistos próprios das atividades desenvolvidas em espaços abertos. Seria desejável no futuro generalizar o uso destas abordagens para fundamentar as opções de base dos Planos de Gestão Florestal nas áreas públicas e comunitárias. As Associações Florestais e os Gabinetes Técnicos Florestais das Câmaras Municipais, com responsabilidades na elaboração dos PGF, são constituídos maioritariamente por técnicos da área da Engenharia Florestal que saberão interpretar e ajustar as soluções obtidas.

Relativamente à área de estudo, seria desejável fazer uma divisão mais detalhada dos baldios, de forma a obter parcelas com menores dimensões e, assim, tornar a gestão daquela área mais flexível. Os princípios orientadores de gestão traduzidos neste trabalho revelam uma atitude cuidadosa relativamente à extração da madeira ao longo do horizonte de planeamento (devido às restrições de sustentabilidade e ambientais). Após a concretização dos cinco anos para os quais se fez o planeamento, poder-se-á colocar a questão da necessidade de contribuir para a regularização da distribuição das classes de

idade das parcelas, tornando mais efetiva a persistência das áreas de floresta nos baldios. As abordagens propostas não têm em conta riscos provenientes de fatores de incerteza como por exemplo os incêndios. O risco de incêndio também depende de fatores inerentes às características das parcelas, podendo por isso ser modificado através das decisões de gestão. Como os incêndios são cada vez mais frequentes e devastadores em Portugal, seria desejável incorporar o risco de incêndio nestas abordagens.

Agradecimentos

Os autores gostavam de salientar o apoio dado pelos técnicos do ICNF diretamente ligados à gestão da área de estudo, nomeadamente o Eng^o. Rogério Rodrigues (Diretor Regional das Florestas do Norte) e o Eng^o. António Martinho (Chefe da Unidade de Gestão Florestal do Tâmega); ao nível das Unidades de Baldio, o apoio dado pelos Órgãos cogestores, Presidentes dos Conselhos Diretivos dos Baldios de Melhe e de Seirós e Presidentes das Freguesias de Canedo, Santo Aleixo e Santa Marinha; e o apoio dado pela AFRP e pelo Município de Ribeira de Pena, cujo técnico (coautor desta publicação) participou no desenvolvimento deste trabalho, no âmbito de provas académicas; o apreço ao Prof. Cat. Carlos Pacheco Marques (UTAD), pela coordenação dos inventários florestais realizados na área de estudo, e o agradecimento aos alunos de Eng. Florestal da UTAD envolvidos na recolha de informação. Os autores gostavam de agradecer aos dois revisores pelos importantes comentários e sugestões que contribuíram para melhorar este documento.

ANEXO

A tabela A1 indica as dimensões dos modelos FA, FB e FAI. A tabela A2 indica o número máximo de restrições *path* e variáveis *cluster*, e a tabela A3 o número máximo de restrições e variáveis *bucket*. $|V|$, $|B|$, $|T|$ e $|N|$ denotam o número total de parcelas da área de estudo, o número de baldios, o número de períodos e o número de cliques maximais do grafo que representa a floresta, respetivamente.

Tabela A1: N° máximo de restrições e variáveis das formulações FA, FB e FAI ($|\mathbf{B}| \leq |V|$)

Formulação	N° max. restrições	N° max. variáveis
FA	$5 V + 5 \mathbf{B} T + 2 T $	$ V T + \mathbf{B} T + \mathbf{B} $
FB	$5 V + 6 \mathbf{B} T + 2 T $	$ V T + \mathbf{B} T + \mathbf{B} + 1$
FAI	$5 V + 5 \mathbf{B} T + 2$	$ V T + \mathbf{B} T + \mathbf{B} $

Tabela A2: N° máximo de restrições *path* e variáveis *cluster*. *NCN* - número máximo de parcelas dos *clusters* não admissíveis minimais; *NC* - número máximo de parcelas dos *clusters* admissíveis

N° max.	
Restrições <i>path</i>	$ T \sum_{k=1}^{NCN} \binom{ V }{k}$
Variáveis <i>cluster</i>	$ T \sum_{k=1}^{NC} \binom{ V }{k}$

Tabela A3: N° máximo de restrições e variáveis *bucket* ($|\mathcal{S}| \in O(|V|)$ num grafo planar, Diestel 2000)

Formulação	N° max.
Restrições <i>bucket</i>	$2 V T + \mathcal{S} T + 5 \mathcal{S} V T + V ^2 T $
Variáveis <i>bucket</i>	$ V ^2 T + \mathcal{S} V T $

BIBLIOGRAFIA

BARE, Bruce, Application of operation research and forest management: a survey, Quantitative science paper 26, Center for quantitative science in Forestry, Fisheries and Wildlife. University of Washington, 1971.

BENTO, João Santos, *Oferta sustentada de material lenhoso de pinheiro bravo, uma aplicação a nível nacional*. Tese de Doutoramento apresentada à Universidade de Trás-os-Montes e Alto Douro, 1994.

CLUTTER, Jerome L., Fortson, James e Pienaar, Leon, MAX MILLION II – *A computerized forest management planning system, user's manual*, School of Forest Resources, University of Georgia, 1978.

CONSTANTINO, Miguel, Martins, Isabel e Borges, José Calvão, "A new mixed integer programming model for harvest scheduling subject to maximum area restrictions", *Operations Research* 56(3), 2008, pp. 542-551.

CROWE, Kevin, Nelson, John, Boyland, Mark, "Solving the area-restricted harvest-scheduling model using the branch and bound algorithm", *Canadian Journal of Forest Research* 33(9), 2003, pp. 1804-1814.

DIESTEL, Reinhard, *Graph Theory. Graduate texts in Mathematics*, Edição 2, Springer, 2000.

FONSECA, Teresa Fidalgo, Parresol, Bernard, Marques, Carlos Pacheco e de Coligny, François, *Models to implement a sustainable forest management – an overview of the ModisPinaster model*, Sustainable Forest Management / Book 1.InTech - Open Access Publisher, 2012a.

FONSECA, Teresa Fidalgo, Cerveira, Adelaide e Mota, Artur, "An integer programming model for a forest harvest problem in *Pinus pinaster* stands", *Forest Systems* 21 (2), 2012b, pp. 272-283.

GOYCOOLEA, Marcos, Murray, Alan T., Vielma, Juan Pablo e Weintraub, Andres, "Evaluating approaches for solving the area restricted model in harvest scheduling", *Forest Science* 55(2), 2009, pp. 149-165.

GOYCOOLEA, Marcos, Murray, Alan T., Barahona, Francisco, Epstein, Rafael e Weintraub, Andres, "Harvest scheduling subject to maximum area restrictions: exploring exact approaches", *Operations Research* 53(3), 2005, pp. 90-500.

IVERSON, David, Alston, Richard, *The genesis of FORPLAN: a historical and analytical review of Forest Service planning models*, General Technical Report, INT-214. US Forest Service. Intermountain Research Station, 1986.

KENT, B., J. Kelly, W. Flowers, Jr, *Experience with the solution of USDA Forest Service FORPLAN models. The 1985 Symposium on Systems Analysis in Forest Resources*, Ed. P. Dress, R. Field, 1-19, The Georgia Center for Continuing Education, 1987.

LUIS, Jaime Sales e Fonseca, Teresa Fidalgo, "The allometric model in the stand density management of *Pinus pinaster* Ait. in Portugal", *Annals of Forest Science* 61, 2004, pp. 1-8.

MARQUES, Carlos Pacheco, "Evaluating site quality of even-aged maritime pine stands in northern Portugal using direct and indirect methods", *Forest Ecology Management* 41(3-4), 1991, pp.193-204.

MARTINS, Isabel, Alvelos, Filipe e Constantino, Miguel, "A branch-and-price approach for harvest scheduling subject to maximum area restrictions", *Computational Optimization and Applications* 51(1), 2012, pp. 363-385.

MARTINS, Isabel, Constantino, Miguel e Borges, José Calvão, "A column generation approach for solving a non-temporal forest harvest model with spatial structure constraints", *European Journal of Operational Research* 161(2), 2005, pp. 478-498.

MARTINS, Isabel, Constantino, Miguel e Borges, José Calvão, 1999. *Forest management models with spatial structure constraints*, Working Paper No. 2/1999, Centro de Investigação Operacional, Faculdade de Ciências de Lisboa, 1999.

MCDILL, Mark E., Rebain, Stephanie A. e Braze, Janis, "Harvest scheduling with area-based adjacency constraints", *Forest Science* 48(4), 2002, pp. 631-642.

MOREIRA, Adelina Mónica e Fonseca, Teresa, "Tabela de produção para o Pinhal do Vale do Tâmega", *Silva Lusitana* 10, 2002, pp. 63-71.

MOTA, Artur, *Plano de exploração para Pinus pinaster Ait. dos baldios de Ribeira de Pena, Perímetro Florestal do Barroso*. Tese de Mestrado apresentada à Universidade de Trás-os-Montes e Alto Douro, 2011.

MURRAY, Alan T. e Weintraub, Andrés, "Scale and unit specification influences in harvest scheduling with maximum area restrictions", *Forest Science* 48(4), 2002, pp. 779-789.

MURRAY, Alan T. e Snyder, Stephanie., "Spatial modeling in forest management and natural resource planning", *Forest Science* 46(2), 2000, pp. 153-156.

MURRAY, Alan T., "Spatial restrictions in harvest scheduling", *Forest Science* 45(1), 1999, pp. 45-52.

NAVON, Daniel I., 1971, *Timber RAM a long-range planning method for commercial timber lands under multiple-use management*, Research paper PSW-70, U.S. Forest Service. Pacific Southwest Forest Range Experiment Station, 1971.

REINEKE, Lester Henry, "Perfecting a stand-density index for even-aged forests", *Journal of Agricultural Research*, 46, 1933, pp. 627-638.

VIELMA, Juan Pablo, Murray, Alan T., Ryan, David M., Weintraub, Andrés, "Improving computational capabilities for addressing volume constraints in forest harvest scheduling problems", *European Journal of Operational Research* 176(2), 2007, pp. 1246-1264.

CASO 2

PLANEAMENTO DE EXPLORAÇÕES AGROPECUÁRIAS SUSTENTÁVEIS: UMA ABORDAGEM MULTICRITÉRIO

Ana Alexandra Marta-Costa

Departamento de Economia, Sociologia e Gestão da
Universidade de Trás-os-Montes e Alto Douro (UTAD)
Centro de Estudos Transdisciplinares para o Desenvolvimento (CETRAD)
amarta@utad.pt

Armando B. Mendes

Departamento de Matemática da Universidade dos Açores
Centro de Estudos de Economia Aplicada do Atlântico (CEEApIA)
amendes@uac.pt

RESUMO

O texto apresenta um modelo para o planeamento de explorações agropecuárias sustentáveis em áreas de Montanha, nomeadamente para as serras do Marão-Alvão-Padrela e para a serra de Montemuro, áreas de exploração das raças bovinas autóctones Maronesa e Arouquesa, respetivamente.

Dado o carácter multidimensional da sustentabilidade, como resultado da complementaridade e interação entre as dimensões económica, social e ambiental, o planeamento de explorações agropecuárias sustentáveis é realizado tendo por base a programação multiobjetivo (método NISE - *NonInferior Set Estimation Method*), complementado com a programação de compromisso. São considerados dois objetivos para a formulação matemática do modelo. Um objetivo económico – o Valor Acrescentado Bruto (VAB) – a maximizar, e o segundo objetivo é de âmbito ambiental – os custos energéticos – a minimizar. São ainda apresentadas variações do modelo atendendo aos diferentes cenários possíveis de incorporação ou não de subsídios à atividade corrente da exploração.

O modelo é instanciado com parâmetros obtidos por recolha direta e por recurso a fontes estatísticas. Apesar do carácter aberto das soluções encontradas, constata-se que a atividade leite está sempre presente quando se pretende aumentar o VAB. Também é certo que a atividade carne está sempre presente quando a tendência for para o menor custo energético. As atividades bovinas selecionadas induzem à conclusão de que, face aos recursos disponíveis, as áreas mais agrestes e menos produtivas, como os baldios, devem ser aproveitadas pelas raças rústicas (Maronesa e Arouquesa), transformadores naturais dos recursos intrínsecos das zonas de montanha, resultando numa mais-valia ambiental e também económica. As áreas mais produtivas, por outro lado, devem ser utilizadas para atividades mais rendíveis, como a exploração de bovinos leiteiros, cuja venda do vitelo deve ser realizada à nascença, devido não só ao maior rendimento económico daí resultante, mas também à ausência de recursos alimentares disponíveis para a sua criação.

Neste sentido, a implementação em ambiente real das soluções encontradas passa pelo incentivo à seleção e complementaridade das duas atividades, em zonas de montanha, para obter a sua sustentabilidade.

PALAVRAS-CHAVE

Modelação do agroecossistema; Sistemas agrários; Sustentabilidade; Zonas de montanha; Programação multiobjetivo.

1. Definição do problema

O processo de degradação ambiental causado pela agricultura tem registado um enfoque particular, essencialmente nas últimas décadas, devido à evolução tecnológica e à reestruturação agrária. Também a política agrária seguida inicialmente pela Comunidade Europeia, ao prosseguir objetivos evidentes de aumento das produções e das produtividades da terra e do trabalho, conduziu a uma degradação ambiental, com especial gravidade nos países do Norte e Centro da Europa nos quais a intensificação cultural atingiu uma maior expressão. Tal processo tem vindo a favorecer a concentração e especialização da produção, a par dos métodos intensivos na agricultura e pecuária, com vista ao aumento das produções. Uma tal revolução agrária não se tem feito com os cuidados devidos e, em muitos casos, foi responsável por problemas que continuam a degradar o ambiente, prejudicando a vida atual e comprometendo a produção futura.

Baseados em estudos anteriores, compilados em Marta-Costa (2001), destacam-se o acréscimo de poluição das águas do solo e da atmosfera pelo uso crescente de fertilizantes e produtos fitofarmacêuticos; o aumento das áreas agricultadas em detrimento das áreas florestais, com consequências desastrosas para a qualidade do ar; a excessiva exploração do solo com efeitos no esgotamento da sua fertilidade e com problemas de erosão e desertificação; o consumo intenso dos recursos hídricos que põe em causa a sua disponibilidade para as populações; e o excessivo uso da monocultura com consequente perda acentuada da diversidade vegetal. A estes fatores, que põem em causa a biodiversidade e as paisagens, acresce o risco de abandono das atividades agrárias, quando menos rendíveis, e o despovoamento humano do território.

Existem também determinadas práticas agrárias que são manifestamente positivas para o meio que nos rodeia, contribuindo para a preservação de certos ecossistemas e espécies vegetais e animais e, acima de tudo, a agricultura é responsável pela paisagem de muitas localidades que caracterizam a herança cultural dos nossos antepassados.

Constata-se, portanto, que a agricultura, quer por excesso quer por defeito, constitui um dos principais focos de degradação ambiental, com inevitáveis reflexos económicos e sociais.

Neste contexto, os sistemas agrários encontram-se sujeitos a um duplo compromisso, muitas vezes conflituoso, fundamental para o seu sucesso. Por um lado, há que rendibilizar o capital investido e maximizar o resultado económico alcançado. Por outro, face à conjuntura sócio-ambiental, é necessário preservar e proteger o ambiente e os recursos naturais e promover a equidade social. A solução passa, entre outras coisas, por um adequado compromisso com o mercado dos fatores (tais como fertilizantes e fitofármacos); pelo mercado dos produtos finais ao consumo; por uma readaptação das tecnologias utilizadas; pela poupança energética; sem, contudo, pôr em causa a segurança alimentar há muito requerida pela sociedade.

Neste quadro, o problema que atualmente se coloca centra-se no planeamento da exploração por forma a ser capaz de dar resposta a objetivos económicos e também a objetivos sócio-ambientais, numa perspetiva de futuro e de acordo com a implementação da Estratégia do Desenvolvimento Sustentável. Sendo assim, o problema de sustentabilidade das atividades agrícolas é um problema naturalmente multicritério ou multiobjetivo.

Na verdade, o princípio subjacente aos dois objetivos referidos está de todo relacionado com a temática do desenvolvimento sustentável. Por um lado, a manutenção da atividade agrícola é possível e desejável, para explorações economicamente viáveis, o que ajudará à sua reprodução e evitará problemas de desertificação humana. Por outro, o próprio conceito de desenvolvimento sustentável tem implícito que todas as atividades devem ser desenvolvidas segundo o princípio de proteção do ambiente, a usufruir também pelas gerações futuras. Daí o menor impacto ambiental negativo que é exigido.

Este trabalho pretende delinear um modelo de exploração agropecuária para áreas de montanha, nomeadamente para as serras do Marão-Alvão-Padrela e para a serra de Montemuro, áreas de exploração das raças bovinas autóctones Maronesa e Arouquesa, respetivamente. Pretende-se um modelo de exploração sustentável, atendendo a diferentes cenários possíveis, promotor da conservação de valores naturais através da gestão ativa dos sistemas tradicionais, economicamente viável, socialmente atrativo e gerador de mais-valias ambientais. A aplicabilidade deste modelo prende-se com a criação de

conhecimento sobre as melhores soluções a adotar, para a criação de normas e de boas práticas.

O modelo para as serras do Marão-Alvão-Padrela foi desenvolvido no âmbito do trabalho de doutoramento de Marta-Costa (2008) e publicado em Marta-Costa (2010). Com base neste, procedeu-se ao desenvolvimento do modelo de exploração agropecuária sustentável para a serra de Montemuro, com vista a dar resposta a um dos requisitos do “Plano de Gestão Integrada do Sítio PTCO0025”, requerido pela Associação de Municípios do Vale Douro Sul (AMVDS, 2008) à UTAD, cujos resultados foram divulgados em Marta-Costa *et al.* (2010) e posteriormente publicados em livro (Marta-Costa *et al.*, 2013).

2. O objeto de estudo

As aldeias de montanha, em geral muito afastadas dos centros urbanos, caracterizam-se, de acordo com Miranda (2000), por apresentarem magníficas paisagens, património riquíssimo e seculares tradições, sendo também o último refúgio de muitas raças autóctones, de fauna e de flora selvagens. Não usufruindo de fáceis acessibilidades, os povoamentos são concentrados, com ruas estreitas, alojamentos com fracas condições de habitabilidade, rigor climatérico, orografia agreste e acentuado despovoamento. Estas características revelam uma pobreza intrínseca às áreas de montanha, aferida pelo baixo valor dos parâmetros relativos à qualidade de vida das suas populações; pelos poucos rendimentos económicos, devido à falta de competitividade dos seus produtos nos mercados, o que impede a realização de benefícios económicos e a retirada de mais-valias; e pela pouca qualidade dos solos.

O abandono das atividades agropecuárias, nomeadamente das raças autóctones, implica o desaparecimento de recursos genéticos. A delapidação da diversidade biológica, em geral e dos recursos genéticos animais em particular, inviabiliza quaisquer imperativos económicos, culturais e científicos impostos por eventuais futuras necessidades humanas (Alves, 1993).

Por último, a degradação ambiental arrastará, inevitavelmente, a componente económica, podendo originar o despovoamento humano ainda mais acentuado.

Potenciar as áreas de montanha como fator de desenvolvimento, assente na humanização do território, com total respeito pelo espaço natural, é um desafio que urge vencer, visando salvaguardar a cultura e história que estas serras têm guardado através das gerações.

Algumas atividades económicas têm renascido como capazes de potenciar o valor destes territórios. É o caso, por exemplo, do aproveitamento das áreas mais inóspitas para a exploração de gado bovino e também de pequenos ruminantes. Nalguns casos, verifica-se que estas atividades são a principal fonte de receitas das famílias-explorações, através da venda de crias, acrescidas dos prémios e subsídios atribuídos aos sistemas produtivos (Silva, 2000).

A criação de gado bovino e também de pequenos ruminantes, nas áreas de montanha, essencialmente as raças autóctones, deve ser incentivada pelas seguintes razões:

- a rusticidade permite identificá-las como as mais bem adaptadas às difíceis condições destas áreas, capazes de suportar os rigores climáticos e a escassez do alimento de certas alturas do ano, sendo também, as mais bem preparadas para retirar o máximo de proveito das condições de pastoreio existentes (Miranda, 2000);
- os encargos comparativos de exploração destes animais são reduzidos, pois passam a maior parte do ano em pastoreio livre, na serra. A alimentação dos vitelos é à base de leite materno nos primeiros meses de vida. A utilização de ração é nula ou quase insignificante, apenas é administrada a vitelos e por vezes às vacas aleitantes, em baixa quantidade (Silva, 2000);
- devido à sua elevada rusticidade, pode também induzir-se que encargos da assistência veterinária e uso de fármacos tornam-se, também eles, muito reduzidos;
- pese embora os baixos encargos, os rendimentos são também baixos, mas constituem, nestes casos, em comparação com o abandono destas áreas, uma importante mais-valia económica;
- outros benefícios para a família são ainda imputados ao sistema, nomeadamente, o autoemprego (o custo de oportunidade da mão-de-obra, na maior parte dos casos é zero), o autoconsumo (leite e carne) e o

aproveitamento do estrume e da tração animal como fatores de produção (Silva, 2000);

- as questões de segurança alimentar humana e saúde pública, são também asseguradas pelo modo de produção utilizado na exploração das raças autóctones, respeitador do ambiente e da saúde, cuja alimentação baseia-se nos recursos naturais e da exploração;
- o contributo para a autossuficiência alimentar da comunidade humana em estudo. Trata-se de um aspeto chave para a sobrevivência e competitividade de qualquer área do globo terrestre, que vai de encontro à nova economia do desenvolvimento sustentável defendida por Robertson (2007), centrada nas pessoas e ecologicamente sustentável. Como projeta aquele autor, deverão ser as famílias e comunidades locais a produzir uma maior proporção, em comparação com a atual, das coisas de que necessitam, como por exemplo alimentos e energia. Dependerão mais de si próprias e umas das outras, e serão menos dependentes das grandes empresas e das instituições financeiras e serviços governamentais. Destaque-se que o nível de autossuficiência de carne em Portugal, para 1999, foi de 82% (CCE e DGA, 2004);
- o contributo para o controlo de matos e pousios através do pastoreio animal de raças de elevada rusticidade (autóctones), transformadores naturais dos recursos intrínsecos das zonas de montanha, valorizando recursos alimentares não utilizáveis diretamente pelo homem (Moreira, 2002);
- finalmente, o aproveitamento turístico das zonas de montanha poderá promover a melhoria das condições de habitabilidade das populações residentes, quer pela retirada dos animais das aldeias quer pela recuperação das casas tradicionais quer, ainda, pela criação de atividades complementares ao turismo (Miranda, 2000).

É neste quadro que se insere o estudo da “revitalização” e “rendibilização” dos sistemas de exploração das raças bovinas autóctones Maronesa e Arouquesa, complementados, quando possível, com a exploração de pequenos ruminantes. Refira-se, no entanto, que se trata de sistemas relativamente frágeis

e com um equilíbrio precário, pois os bovinos das raças autóctones, produtores de carne e/ou leite e prestadores de tração, têm sido substituídos por espécies mais produtivas e pela mecanização agrícola. Torna-se, por isso, cada vez mais necessário o desenvolvimento de ações no sentido de tornar sustentáveis os sistemas de produção baseados na exploração das raças autóctones.

Limitam-se, como áreas de estudo, as serras do Marão-Alvão-Padrela e a serra de Montemuro, áreas de exploração das raças bovinas Maronesa e Arouquesa, respetivamente (mais detalhes das investigações em Marta-Costa, 2008 e Azevedo *et al.*, 2009).

A base geográfica da exploração da raça bovina Maronesa (serras do Marão-Alvão-Padrela) engloba, em termos de divisão administrativa, os concelhos de Alijó; Mondim de Basto; Murça; Ribeira de Pena; Sabrosa; Vila Pouca de Aguiar; e Vila Real (Alves, 1993).

A raça Arouquesa encontra-se no distrito de Aveiro, tendo como centro Arouca e arredores. É explorada numa zona montanhosa de acessos difíceis, tendo-se delimitado, para este estudo a área da serra de Montemuro. A serra de Montemuro é um dos 60 sítios da Lista Nacional de Sítios da Rede Natura 2000 aprovados pela resolução do Conselho de Ministros nº 142/97. Estende-se por uma área aproximada de 39 mil hectares, abrangendo parte dos concelhos portugueses de Arouca, Castro Daire, Cinfães, Lamego e Resende. O Sítio Serra de Montemuro é uma das áreas classificadas de maior expressão na Região Norte de Portugal, não só pela sua dimensão territorial, mas sobretudo pela sua importância para a preservação do património natural, ambiental e paisagístico da região (AMVDS, 2008).

Ambos os territórios de montanha apresentam “fracas condições naturais” para o desenvolvimento agrícola. Apresentam um clima muito irregular; altitude média-alta com declives acentuados; solo, em alguns casos, exageradamente pedregoso; e recursos hídricos de difícil utilização. Os pontos mais elevados atingem os 1415 metros no Marão, 1283 metros no Alvão, 1148 metros na Padrela e os 1300 metros em Montemuro (Colaço-do-Rosário, 1998 e AMVDS, 2008).

Tratam-se de territórios despovoados e envelhecidos, à semelhança da generalidade das regiões interiores do país. A nível agrário destacam-se como traços mais marcantes os seguintes (Marta-Costa, 2008 e Azevedo *et al.*, 2008):

- A agricultura é uma atividade que se encontra em regressão nas áreas em estudo, apresentando uma evolução negativa do número de explorações com superfície agrícola útil (SAU) e da própria SAU;
- A estrutura fundiária é dominada por explorações de pequena dimensão, com elevado grau de parcelamento;
- Os matos e incultos ocupam grande parte da área, sendo a SAU cerca de 70% da superfície total;
- Atendendo às condições edafoclimáticas específicas destas zonas, as pastagens permanentes ocupam a maior parte da SAU da serra de Montemuro enquanto nas serras do Marão-Alvão-Padrela predominam as terras aráveis;
- Os cereais para grão, em especial o milho e o centeio, e, em menor grau, os prados temporários e culturas forrageiras, as leguminosas secas para grão e a batata são as culturas anuais mais importantes, tendo vindo, também, a regredir de forma acentuada, na serra de Montemuro. Por outro lado, no conjunto das restantes serras em estudo predominam as culturas forrageiras, seguidas pelos cereais para grão, e, com menor representatividade, a batata, prados temporários e leguminosas secas;
- A atividade pecuária, principal fonte de rendimento das populações residentes nas serras, está, também, em franco decréscimo. Os bovinos e os pequenos ruminantes são as espécies pecuárias predominantes nas áreas em estudo, sendo as carnes de bovino, das raças autóctones Arouquesa e Maronesa, e de cabrito produtos qualificados com Denominação de Origem Protegida (DOP) e Identificação Geográfica Protegida (IGP).

3. Metodologia

3.1. O Método NISE e a Programação Compromisso

Dado o carácter multidimensional da sustentabilidade, como resultado da complementaridade e interação entre as dimensões ambiental, económica e

social, o planeamento de explorações agropecuárias sustentáveis foi realizado através da aplicação da teoria da decisão multicritério – método NISE (*NonInferior Set Estimation Method*), complementado naturalmente com a Programação Compromisso, tendo por base o trabalho de Marta-Costa (2008).

Atendendo aos documentos de Cohon (1978); Romero e Rehman (1989); Romero (1993); Poeta (1994) e Marta-Costa (2008), é exposto de seguida um breve resumo do procedimento inerente a estas duas metodologias.

O método NISE é uma técnica de programação multiobjectivo, que permite uma rápida aproximação ao conjunto eficiente¹, quando são definidos dois objetivos, através da sua combinação numa função agregada e ponderada. A finalidade do método é reduzir, gradualmente, a zona de eficiência, através de um processo iterativo de ponderações (w_1 e w_2), relacionado com a inclinação da reta que une os dois pontos extremos eficientes (curvas *trade-off*) encontrados na iteração anterior, através da resolução do seguinte problema:

$$\begin{aligned} & \text{Max} Z(x_1, x_2, \dots, x_n; w_1, w_2) \\ & \text{Sujeito a: } x \in F \\ & w_1 = \frac{Z_1^*(A) - Z_{*1}(A)}{Z_{*2}(B) - Z_2^*(B)}; w_2 = 1 \end{aligned}$$

Sendo A e B as soluções eficientes; Z_i^* e Z_{*i} os valores ideais e anti-ideais de cada objetivo e o conjunto F formado pelas restrições impostas ao modelo.

A aproximação ao conjunto eficiente é controlada, neste método, através de um critério pré-fixado de máximo erro admissível e, posteriormente, comparado com o erro obtido em cada iteração. O erro máximo possível é a distância máxima entre as fronteiras inferior e superior, medida na direção perpendicular à fronteira inferior. O erro máximo admissível é uma percentagem, definida pelo decisor, do valor do erro máximo possível obtido na primeira iteração.

A programação compromisso propõe reduzir o conjunto eficiente, obtido pelo método NISE, baseando-se na noção de distância entre as soluções

¹ Solução Eficiente ou Ótimo de Pareto é definida como a solução melhor possível, não sendo possível melhorar um atributo sem piorar outro (Romero e Rehman, 1989).

(possíveis) e o ponto ideal. Este ideal (I^*) é uma solução utópica mas é, também, um ponto de referência para o agente decisor. Neste sentido, a programação compromisso assume que qualquer agente decisor procura uma solução tão perto quanto possível do ponto ideal (Romero e Rehman, 1989). Para encontrar a solução compromisso calcula-se a distância entre cada solução e o ponto ideal. O grau de proximidade d_j entre o j -ésimo objetivo e o seu ideal é dado pela expressão seguinte, quando o j -ésimo objetivo é, respetivamente, maximizado ou minimizado, sendo Z_j^* o valor ideal:

$$d_j = Z_j^* - Z_j(x) \quad \text{ou} \quad d_j = Z_j(x) - Z_j^*$$

Os graus de proximidade entre os diferentes objetivos e os seus valores ideais podem ser agregados numa função de distância conjunta. Como as unidades de medida dos vários objetivos são, muitas vezes, diferentes, estes devem ser normalizados da forma seguinte, sendo Z_{*j} o valor anti-ideal:

$$d_j = \frac{|Z_j^* - Z_j(x)|}{|Z_j^* - Z_{*j}|}$$

Por vezes é necessário, também, ponderar os objetivos, afetando cada um com um determinado peso (W_j), para que o nível de realização desejado seja diferente.

A programação compromisso utiliza a família de distâncias métricas L_p para encontrar o subconjunto do conjunto eficiente, utilizando o conceito de medidas de distância seguinte:

$$L_p(W) = \left[\sum_{j=1}^n W_j^p \left| \frac{Z_j^* - Z_j(x)}{Z_j^* - Z_{*j}} \right|^p \right]^{\frac{1}{p}} \quad \text{ou} \quad L_p(W) = \left[\left(\sum_{j=1}^n w_j d_j \right)^p \right]^{\frac{1}{p}}$$

Para diferentes conjuntos de valores de j e W geram-se diferentes soluções compromisso. Admitindo, no entanto, que todos os objetivos têm igual

importância, o conjunto compromisso fica definido, para as métricas L_1 e L_∞ , quando p é igual a 1 e ∞ , respectivamente. Isto porque as métricas $p = 1$ e $p = \infty$ definem os dois limites, L_1 e L_∞ das soluções compromisso. Por outras palavras, todas as outras soluções compromisso caem entre elas. Desta forma, será suficiente resolver estes dois problemas de programação linear, para cada conjunto de pesos preferenciais (W_i).

Para a métrica L_1 (isto é $p = 1$), o melhor compromisso ou a solução mais próxima ao ponto ideal é obtida pela resolução do seguinte problema:

$$\text{Min} L_1(W) = \sum_{j=1}^n W_j \left| \frac{Z_j^* - Z_j(x)}{Z_j^* - Z_{*j}} \right|$$

Sujeito a: $x \in F$

Para a métrica L_∞ ($p = \infty$), na qual o desvio individual máximo é minimizado, a melhor solução compromisso pode ser obtida, resolvendo o seguinte modelo de programação linear:

$$\begin{aligned} \text{Min } d_\infty &= d \\ W_1 \frac{Z_1^* - Z_1(x)}{Z_1^* - Z_{*1}} &\leq d \\ \dots & \\ W_n \frac{Z_n^* - Z_n(x)}{Z_n^* - Z_{*n}} &\leq d \end{aligned}$$

Sendo d a maior distância. Os pontos L_1 e L_∞ definem, assim, o conjunto compromisso e correspondem aos pontos eficientes mais próximos do ponto ideal.

3.2. Justificação da metodologia adotada

O uso conjugado de diferentes métodos de análise de decisão multicritério, com a elaboração de modelos de programação matemática, integrando dados técnico-económicos característicos de atividades regionais revelam ser “ferramentas” de grande alcance para o desenvolvimento de sistemas de apoio

à tomada de decisão de gestores e agricultores (Carvalho, 2007). Neste contexto, o processo de tomada de decisão na agricultura é um procedimento complexo que deve tomar em consideração os diferentes objetivos, frequentemente em conflito, dos vários atores envolvidos no processo, tais como agricultores, planeadores, políticos, consumidores.

O conceito de sustentabilidade da agricultura, integrando a dimensão ambiental, económica e social, veio aumentar, significativamente, a complexidade dos processos de decisão, atendendo à multiplicidade dos objetivos envolvidos e às situações de conflito frequentemente geradas na sua otimização (Carvalho, 2006). Consideraram-se como objetivos primordiais do modelo a delinear, a competitividade económica e a sustentabilidade ambiental, pelas razões que se expõem no ponto seguinte. Estas dimensões, como refere Müller (1996), podem considerar-se no curto prazo, conflituosas, reconhecendo-se, no longo prazo, a sua interdependência, mais ou menos complementar.

De acordo com os vários fatores que condicionam o problema que se pretende resolver, nomeadamente os seus objetivos e recursos disponíveis, bem como as vantagens e inconvenientes das várias técnicas da decisão multicritério, foi selecionada a programação multiobjetivo, nomeadamente o método NISE. Sintetizam-se, de seguida, os aspetos mais favoráveis e as principais debilidades da metodologia adotada.

Como **vantagens** enumeram-se as seguintes:

- Permite, entre outras coisas, diferenciar os objetivos (pesos relativos) como reflexo das aspirações e prioridades que o agente decisor (empresário) atribui a cada um deles;
- Tem as vantagens inerentes à programação multiobjetivo, tais como a redução de informação necessária, relativamente à programação por metas, não sendo necessário identificar os níveis de aspiração, os pesos de cada variável de desvio nem ordenar as preferências do centro de decisão. Como indica Romero e Rehman (1989), este tipo de informação é muito difícil de obter, devendo ser o modelo a sugerir ao centro de decisão em vez de usada como um *input* requerido pelo modelo;

- O uso deste tipo de modelos permite reduzir a escolha das soluções a um subconjunto do conjunto eficiente (não dominado), o que facilita uma melhor apreciação das alternativas admissíveis;
- A informação fornecida pelo modelo permite conhecer o quanto é necessário sacrificar de um objetivo para conseguir um incremento unitário de outro (*trade-off*);
- O modelo permite gerar soluções não dominadas e articular as preferências do agente decisor através dos *trade-offs*, bem como conhecer o nível de realização de cada objetivo;
- O uso combinado com a programação compromisso reduz significativamente o uso computacional, relativamente ao uso isolado da programação multiobjetivo, pois só se torna necessário resolver dois problemas de programação linear (L_l e L_∞), para cada um dos pesos determinados.

Os principais **inconvenientes**, alguns dos quais comuns a outros métodos de programação multiobjetivo, podem resumir-se da seguinte forma, de acordo com Poeta (1994) e Silva (2001):

- A linearidade e um certo determinismo apenas existem em absoluto nos modelos teóricos, e a prática demonstra que, nas empresas, isto apenas ocorre num período curto de vida;
- A quantificação de alguns determinantes do ato produtivo, quer a relacionada com os objetivos ou com o enquadramento da exploração (face aos recursos disponíveis), não pode, em muitas situações, ser determinada;
- O modelo é construído com um conjunto de pressupostos, situados no tempo e no espaço, de modo que este torna-se estático na medida em que as ruturas não se refletem nos resultados do modelo;
- Os modelos são representações da realidade e, por isso, não reproduzem a realidade em toda a sua complexidade, pelo que se tornam necessárias contínuas operações de atualização e validação.

Do ponto de vista da aplicação dos resultados dos modelos, o reconhecimento destas limitações por parte de todos os atores envolvidos no

processo de decisão é essencial para o sucesso de todo o processo. Como exemplo de um estudo onde estes aspetos foram aprofundados pode citar-se Mendes e Cardoso (2006).

O método NISE foi usado, neste trabalho, para gerar as curvas *trade-off* entre os diferentes objetivos, o que permite uma melhor apreciação das alternativas (Cohon, 1978). A programação compromisso foi utilizada como instrumento para determinar o conjunto de soluções admissíveis mais próximas do ponto ideal. Este ideal (I^*) é uma solução utópica mas é, também, um ponto de referência para o agente decisor (Romero e Rehman, 1989).

Foi utilizado o programa LINDO - *Linear Interactive aNd Discrete Optimizer* (LINGO 10), tendo por base os aspetos operacionais do método NISE e da Programação Compromisso, indicados nos documentos de Cohon (1978); Romero e Rehman (1989); Romero (1993); Poeta (1994) e Marta-Costa (2008).

3.3. Formulação matemática do modelo

Apesar dos objetivos comuns e de algumas características muito similares entre as duas áreas geográficas em estudo, foram delineados dois modelos distintos devidos, essencialmente, às diferenças de fatores de produção existentes (por exemplo quanto às raças bovinas autóctones exploradas) e às suas disponibilidades. Além disso, foram identificadas algumas particularidades que interessavam respeitar em cada modelo, fruto dos contextos em que os estudos foram realizados. Esta situação foi considerada como vantajosa, pois permitiu testar o mesmo modelo (ou pressupostos muito próximos), a duas realidades semelhantes, com algumas condições ou restrições distintas².

Para as serras do Marão-Alvão-Padrela, o modelo assentou na informação recolhida de uma amostra de 112 explorações, correspondentes a cerca de 30% do total de explorações com cinco ou mais animais adultos, cuja atividade principal é a exploração de bovinos. Para a serra de Montemuro, a recolha de informação baseou-se, preferencialmente, em estatísticas oficiais.

Atendendo às variáveis e parâmetros que se apresentam no apêndice deste capítulo, explica-se de seguida as funções objetivo e restrições identificadas na

² Para melhor perceção das diferenças de cada modelo, será utilizado SMAP e SM, respetivamente, junto do número das expressões, das variáveis e parâmetros utilizadas exclusivamente no modelo das serras do Marão-Alvão-Padrela e Montemuro.

formulação matemática dos modelos. Os diversos parâmetros e coeficientes técnicos foram definidos em função das características médias das explorações. Outros basearam-se nos valores publicados na literatura da especialidade (INRA, 1988; GPPAA, 2001; Moreira *et al.*, 2001).

Os modelos obtidos foram constituídos por 129 variáveis e 98 restrições, para as serras do Marão-Alvão-Padrela, e por 135 variáveis e 103 restrições, para a serra de Montemuro. A definição das variáveis e parâmetros utilizados encontra-se em apêndice.

3.3.1. As funções objetivo

Foram consideradas duas medidas de desempenho na formulação do modelo, que se descrevem de seguida.

A) Objetivo VAB (€)

O primeiro objetivo (Z_1) é centrado no proveito económico, tendo sido escolhido como descritor de impacto o VAB, isto é, a diferença entre a venda de produtos e a compra de bens e serviços, expressa em euros, que importa maximizar.

A seleção deste objetivo teve em conta o facto das explorações para sobreviverem terem que apresentar maiores receitas monetárias conseguidas através da participação mais ativa no mercado, com a venda de produtos. A escolha deste resultado económico baseou-se em discussões com alguns gestores e agricultores, tendo-se identificado duas razões. Por um lado, a impossibilidade de quantificar os encargos fixos por unidade de variável de decisão, como é o caso das amortizações com o equipamento de ordenha, associado aos bovinos de raça leiteira. A este propósito, refira-se que a sua ausência no modelo vem penalizar a adoção das raças locais (Arouquesa e Maronesa), através da valorização do resultado económico dos bovinos produtores de leite.

Por outro lado, é princípio base do modelo que as explorações possuem um determinado aparelho de produção, correspondente à média das unidades de produção presentes na área em estudo, com os seus respetivos encargos fixos. Pretende-se, por isso, atribuir-lhe o melhor uso no sentido do equilíbrio

entre os resultados económicos e os efeitos ambientais. Além disso, o peso dos encargos referentes aos impostos, seguros, amortizações, salários pagos, rendas e juros de capital alheio é considerado baixo. Desta forma, com a maximização do VAB pretende-se a melhor remuneração do trabalho e do capital investido.

Identificaram-se como atividades vegetais principais as observadas nas explorações, nomeadamente: batata, milho grão, milho forragem, centeio, prados temporários, lameiros de feno e pasto, e ainda milho silagem para as serras do Marão-Alvão-Padrela.

A atividade pecuária considerada é relativa à exploração de bovinos, ovinos e caprinos de carne e/ou de leite, de diferentes grupos étnicos. Relativamente aos primeiros considera-se vacas adultas das raças locais Maronesa (para as serras Marão-Alvão-Padrela) e Arouquesa (serra de Montemuro) e da raça Frísia, mas com situações de descendência (F1) de animais puros ou resultantes de cruzamentos com raças de carne, que se designam, neste texto, animais do tronco frísio. A exploração de ovinos e caprinos tem como objetivo a produção de carne, leite e/ou queijo, para a serra de Montemuro. Para as serras do Marão-Alvão-Padrela foi considerada a exploração de caprinos de raça Bravia, para a produção exclusiva de carne.

O nível de preços dos fatores de produção e dos produtos no produtor foram obtidos alguns por inquirição direta e outros por consulta das cotações disponíveis na página *web* do Sistema de Informação de Mercados Agrícolas (<http://www.gppaa.min-agricultura.pt/sima.html>) e pelas estatísticas disponíveis para 2004 (INE, 2005).

Para além dos custos variáveis específicos de cada atividade, foram ainda considerados os custos relativos ao arrendamento de terra (apenas para a serra de Montemuro), de contratação de mão-de-obra temporária, de aluguer de tração mecânica, de compra de elementos fertilizantes e de alimentos para o gado, de modo a que estes fatores não fossem restritivos à expansão do processo produtivo.

$$\begin{aligned}
Z_1 = & -\sum_i \sum_j \sum_n l_{jn} \cdot pF_n \cdot XA_{ij} - \sum_p \sum_n l_{pn} \cdot pF_n \cdot XP_p - \sum_b \sum_n l_{bn} \cdot pF_n \cdot XB_b - \sum_h pT_h \cdot X_h - \sum_r \sum_f m_{rf} \cdot pF_f \cdot XV_r \\
& - \sum_s \sum_f m_{sf} \cdot pF_f \cdot XC_s + \sum_v pV_v \cdot VA_v + \sum_r \sum_x pC_{rx} \cdot VC_{rx} + \sum_s \sum_y pC_{sy} \cdot VC_{sy} + pLVL + \sum_s pL_s \cdot VL_s \\
& + \sum_s pQ_s \cdot VQ_s + \sum_r pEVE_r + \sum_s pEVE_s - \sum_z \sum_t n_z \cdot pTP_z \cdot UTM_t - \sum_t pMO \cdot MO_t - \sum_t pTM \cdot TM_t \\
& - pFN \cdot FN - pFP \cdot FP - pFK \cdot FK - \sum_c \sum_t pAl_c \cdot Al_{ct}
\end{aligned} \tag{1}$$

A formulação matemática dos modelos de exploração agrosustentável foi realizada para dois cenários distintos, que consistem na integração ou não, nos seus resultados económicos, das ajudas monetárias à atividade corrente (subsídios) auferidas pelas explorações, no âmbito da Política Agrícola Comum. No contexto socioeconómico real os subsídios existem e, por isso, interessa integrá-los no modelo. Por outro lado, pressupõe-se que a atribuição de subsídios pode não ser sustentável no longo prazo, pelo que interessa, também, fornecer informações sobre a situação em que aquelas ajudas monetárias sejam inexistentes. Para a situação em que são considerados os apoios monetários à atividade corrente, sob a forma de subsídios, incluiu-se o seu valor na função objetivo (Z_1').

$$Z_1' = Z_1 + \sum_i \sum_j sXA_{ij} \cdot XA_{ij} + \sum_p sXP_p \cdot XP_p + \sum_r sXV_r \cdot XV_r + \sum_s sXC_s \cdot XC_s + \sum_t sTP_t \cdot n_t \cdot UTM_t \tag{2}$$

B) Objetivo Custos energéticos (MJ)

A segunda medida de desempenho é o impacto ambiental das soluções testadas, cujo descritor são os custos energéticos dos bens e serviços comprados para o ato produtivo, expressos em megajoules (MJ), que importa minimizar.

Relativamente a este objetivo (Z_2), foi intenção que ele obedecesse a pressupostos ambientais. Desta forma foram abordados vários descritores como passíveis de serem escolhidos, tais como: minimização do consumo de água, minimização do consumo de fatores de produção poluentes (produtos fitossanitários e fertilizantes), minimização do uso de máquinas e equipamentos no solo. O descritor de minimização dos custos energéticos com os fatores de produção variáveis foi selecionado dada a maior facilidade na sua quantificação por fator de produção variável utilizado, em unidades energéticas (MJ), sem

qualquer relação com o seu custo económico. Note-se que até o consumo de água se encontra implícito nesta medida, tendo em conta a energia associada ao combustível necessário para a rega por motobomba ou por aspersão.

Este objetivo vem de encontro ao facto da eficiência energética constituir um importante fator a otimizar, na economia global, constituindo um indicador direto da sustentabilidade. Investigações sobre aspetos de energia ilustram que o seu uso está geralmente correlacionado com as emissões de gases com efeito estufa e com a depleção de recursos naturais. De forma a reduzir ambos os efeitos, têm que ser identificadas formas de poupança potenciais de energia nas atividades agrárias (Moerschner e Lücke, 2002), sendo este o principal fator que induziu à identificação da poupança energética como o segundo objetivo da exploração a planear.

Na prática muitos agricultores parecem não ter em conta condições ambientais, se não forem obrigados nem incentivados economicamente através de ajudas monetárias, por não verem uma relação direta entre a componente ambiental e as receitas da sua exploração. Porém, esta relação começa já a ser percebida pelos atores do meio agrário, como confirmam os estudos identificados no primeiro ponto deste trabalho. Os sistemas de produção, nomeadamente os mais intensivos, originam efeitos prejudiciais no ambiente, não só ao nível da conservação da natureza e proteção dos recursos naturais, mas também na diminuição da fertilidade dos solos, na indução de resistência nos inimigos das culturas/animais, na redução da qualidade dos produtos, na quebra das produtividades dos fatores de produção usados, na diminuição da produção total, se não no curto, pelo menos no médio e longo prazo. Estas situações têm revelado já consequências diretas nas receitas auferidas pelo agricultor. Por essa razão, eles próprios têm necessidade de introduzir o objetivo ambiental no planeamento das suas unidades de produção. Começa a ser percebido que o sacrifício de uma unidade económica hoje, poderá valer a sua quantidade total no futuro. Também a Comunidade Europeia tem vindo a desenvolver várias medidas de incentivos a práticas ambientalmente mais sustentáveis, nomeadamente as medidas agroambientais. Acresce aqui a necessidade de um planeamento que obedeça, não só a parâmetros de competitividade económica, através da maximização das receitas, contemplado

no primeiro objetivo, mas também a pressupostos ambientalmente sustentáveis, que deverão passar pela utilização de fontes de energia menos poluentes. Só assim se poderá assegurar por mais tempo a produtividade (ambiental e económica) do sistema agrário.

Os coeficientes energéticos utilizados foram obtidos do referencial para a análise energética adotados no âmbito da metodologia “PLANETE – *Méthode Pour L’ANalyse Energétique de l’Exploitation*” (*Établissement National d’Enseignement Supérieur Agronomique de Dijon - ENESAD e ADEME, 2002*).

$$Z_2 = \sum_i \sum_j \sum_n l_{ijn}.cE_n.XA_{ij} + \sum_p \sum_n l_{pin}.cE_n.XP_p + \sum_b \sum_n l_{bin}.cE_n.XB_b + \sum_r \sum_f m_{rf}.cE_f.XV_r + \sum_s \sum_f m_{sf}.cE_f.XC_s + \sum_z \sum_t n_z.cET_z.UTM_t + \sum_z \sum_t o_z.cET_z.TM_t + cEFN.FN + cEFP.FP + cEFK.FK + \sum_c \sum_t cEAl_c.Al_{ct} \quad (3)$$

De acordo com a perspetiva exposta, espera-se, assim, com o modelo desenvolvido, melhorar as condições económico-ambientais das explorações observadas, através de duas atitudes deliberadamente assumidas: a competitividade no mercado com produtos que apresentem maior VAB e a minimização dos seus custos energéticos. O compromisso a assumir entre os dois objetivos estará sempre presente na procura da melhor solução para o modelo. O primeiro objetivo tenderá para a eleição de soluções mais eficientes do ponto de vista económico e o segundo propenderá para a escolha de soluções energéticas menos poluentes, isto é, energeticamente mais eficientes, independentemente do seu preço.

Outros objetivos, diretamente ligados com a temática em curso, encontram-se no modelo, não de uma forma direta como os dois anteriores, mas impostos sob a forma de restrições.

3.3.2. As restrições

O sistema de restrições do modelo foi agrupado em quinze categorias principais, que se enumeram de seguida. A sua conceção exprime o que a unidade de exploração pode fazer, isto é, o seu campo de escolha e a forma como esta se relaciona com as características tecnológicas, os recursos disponíveis, a procura a satisfazer e os limites do mercado, tendo como base as condições ambientais e a autonomia da exploração.

a) Restrições na utilização da terra (Ha)

Estas restrições garantem que a utilização do espaço físico da área disponível não é ultrapassada.

Dividiu-se a SAU em terrenos de regadio e terrenos de sequeiro, e em solos de uso agrícola e pascícola (lameiros³), de acordo com a sua aptidão produtiva (expressões 4 a 7).

As áreas de baldio foram também consideradas, pois os pastos nestas áreas são uma fonte complementar de alimentos a baixo custo. Para a serra de Montemuro foram consideradas duas situações distintas de aproveitamento/limpeza dos baldios (expressão 9). Na primeira considerou-se a prática de pastoreio local pelos animais da exploração, pelo que, neste caso, as pastagens do baldio são apenas utilizadas/limpas pelos bovinos de raça local Arouquesa e pelos pequenos ruminantes e, apenas, no período estival como indica Moreira *et al.* (2001). Esta situação coincidiu com a considerada para as serras do Marão-Alvão-Padrela (expressão 8). Na segunda situação considerou-se que o baldio é limpo com recurso à mecanização, nomeadamente pela roçadora de mato, sendo a forragem cortada dada em verde aos animais de qualquer uma das espécies presentes na exploração e durante todo o ano.

De acordo com as práticas frequentemente desenvolvidas na área em estudo, e tendo em atenção os pressupostos ambientais, foram ainda estabelecidas as restrições com os números (10) e (11), relativas à sucessão de culturas, e as restrições (12) a (16), que obrigam à rotação de culturas. Nas primeiras considerou-se que a cultura intercalar⁴ sucede ou antecede, no mesmo exercício, a cultura de batata e/ou milho. Para as segundas adotam-se as rotações de centeio com batata e prados temporários e, também, milho em rotação com os prados temporários.

³ Os lameiros sob a designação de “lameiros de pasto” destinam-se unicamente e durante todo o ano ao pastoreio direto, enquanto os designados de “lameiros de feno” são impedidos de ser pastoreados durante um espaço de tempo, geralmente durante a Primavera, para permitir o crescimento da erva.

⁴ Geralmente ferrã, isto é, cevada ou centeio cortados enquanto verdes para alimento de gado.

Refira-se que este tipo de práticas culturais tem uma grande importância ambiental, mas também económica, principalmente por razões de fertilização e sanidade das culturas. Em culturas arvenses, principalmente em cereais, como indica Ferreira *et al.* (2002), a rotação é especialmente importante com vista a aumentar a fixação biológica de azoto e a matéria orgânica do solo, dada a dificuldade em obter e aplicar fertilizantes orgânicos em grandes áreas.

Considerou-se a hipótese de arrendamento de terra (apenas para a serra de Montemuro), limitada a 25% da área por conta própria (expressão 17).

$$\begin{aligned} \sum_j XA_{1j} &\leq AR + \sum_h X_h & (4) & \quad \sum_j XA_{2j} \leq AS + \sum_h X_h & (5) & \quad XP_1 \leq LF + \sum_h X_h & (6) \\ XP_2 &\leq LP + X_h & (7) & \quad XB \leq BA & (8 \text{ SMAP}) & \quad \sum_b XB_b \leq BA & (9 \text{ SM}) \\ XA_{17} - XA_{11} - XA_{12} - XA_{13} - XA_{14} &\leq 0 & (10) & \quad XA_{27} - XA_{21} - XA_{22} - XA_{23} - XA_{24} &\leq 0 & (11) \\ XA_{15} - XA_{11} &\leq 0 & (12) & \quad XA_{15} - XA_{16} &\leq 0 & (13) & \quad XA_{25} - XA_{26} &\leq 0 & (14) \\ XA_{12} + XA_{13} + XA_{14} - XA_{16} &\leq 0 & (15) & \quad XA_{22} + XA_{23} + XA_{24} - XA_{26} &\leq 0 & (16) \\ \sum_h X_h &\leq 0,25.(AR + AS + LF + LP) & (17 \text{ SM}) \end{aligned}$$

b) Restrições na utilização de mão-de-obra (H)

As restrições da mão-de-obra garantem que as necessidades para as atividades vegetais e animais são iguais ou inferiores às disponibilidades provenientes da mão-de-obra familiar e contratada temporariamente. De acordo com os dados obtidos no trabalho de campo, não foi considerada mão-de-obra contratada permanentemente, pois a sua existência verificou-se em muitos poucos casos.

Devido à flutuação que se verifica nas necessidades em mão-de-obra com as diferentes operações culturais, de atividade para atividade, em diferentes momentos do tempo, considerou-se mais adequado dividir o ano em quatro períodos de tempo.

Dada a dificuldade em discriminar, pelos inquiridos, as necessidades deste fator de produção, por atividade vegetal e por período, foram utilizados os dados, em suporte informático, facultados pelo GPPAA (2001), relativos às

“Contas de Cultura das Actividades Vegetais – Modelo de Base Microeconómico”.

$$\sum_t \sum_j a_{jt} \cdot XA_{jt} + \sum_p a_{pt} \cdot XP_p + \sum_b a_{bt} \cdot XB_b + \sum_r b_{rt} \cdot XV_r + \sum_s b_{st} \cdot XC_s + \sum_q q_{st} \cdot VQ_s - MO_t \leq DMO_t, \forall t \quad (18)$$

c) Restrições na utilização de tração (H)

À semelhança da mão-de-obra, esta restrição garante que, em cada período, as necessidades para as actividades vegetais são iguais ou inferiores às disponibilidades provenientes da tração mecânica própria e/ou alugada.

O consumo de tração mecânica por cada uma das actividades vegetais, em cada período, foi também baseado no documento do GPPAA (2001).

$$\sum_t \sum_j c_{jt} \cdot XA_{jt} + \sum_p c_{pt} \cdot XP_p - TM_t - UTM_t \leq 0, \forall t \quad (19) \quad UTM_t \leq DTM_t, \forall t \quad (20)$$

d) Restrições na utilização dos fertilizantes (Kg)

No âmbito da fertilização racional, isto é, fertilização por medida, indispensável à obtenção da melhor rendibilidade económica da produção agrícola e à preservação da qualidade do ambiente, nomeadamente a protecção das águas superficiais e das águas subterrâneas contra a poluição (eutrofização) com nutrientes minerais veiculados pelos fertilizantes (MADRP, 1997), procedeu-se à elaboração de restrições que garantem que os consumos dos principais nutrientes necessários às actividades vegetais sejam iguais ou inferiores à quantidade veiculada pelos estrumes animais incorporados no solo juntamente com os fertilizantes químicos sintéticos comprados ao exterior (expressões 20, 22 e 23). Foram considerados o nitrolusal 20,5%; o foskamónio 7-14-14 e o superfosfato de cálcio 18% como os adubos mais utilizados, decompostos nos seus elementos (azoto, fósforo e potássio).

Dada a dificuldade em discriminar as necessidades em elementos nutritivos das culturas vegetais, resultantes de inúmeros fatores, tais como o pH do solo, o seu teor em matéria orgânica, entre outros, foram utilizados os dados facultados pelo GPPAA (2001). Teve-se o cuidado, no entanto, de fornecer informação sobre a possibilidade de uso de estrumes, cuja disponibilidade foi

calculada dependente da espécie e raça animal e do tipo de sistema de estabulação respetivo.

Impôs-se, ainda, um uso razoável de azoto que não exceda a quantidade especificada por hectare no âmbito da Diretiva dos Nitratos, emanada pela CE (1991), com o objetivo de proteger a água subterrânea da excessiva contaminação dos nitratos da agricultura, e, em particular, dos estrumes (expressão 22). A quantidade especificada por hectare é a quantidade de estrume que contem 170 quilogramas de azoto. O teor em nutrientes e o coeficiente de utilização pelas culturas do azoto do estrume proveniente das unidades pecuárias foi obtido do “Código de Boas Práticas Agrícolas para a Protecção da Água contra a Poluição com Nitratos de Origem Agrícola”, publicado pelo MADRP (1997).

$$\sum_i \sum_j eN_{ij} \cdot XA_{ij} + \sum_p eN_p \cdot XP_p - \sum_r \%N_r \cdot UE_r - \sum_s \%N_s \cdot UE_s - FN \leq 0 \quad (21)$$

$$\sum_r \%N_r \cdot UE_r + \sum_s \%N_s \cdot UE_s + FN \leq 170 \cdot (AR + AS + LF + LP + \sum_h X_h) \quad (22)$$

$$\sum_i \sum_j eP_{ij} \cdot XA_{ij} + \sum_p eP_p \cdot XP_p - \sum_r \%P_r \cdot UE_r - \sum_s \%P_s \cdot UE_s - FP \leq 0 \quad (23)$$

$$\sum_i \sum_j eK_{ij} \cdot XA_{ij} + \sum_p eK_p \cdot XP_p - \sum_r \%K_r \cdot UE_r - \sum_s \%K_s \cdot UE_s + FK \leq 0 \quad (24)$$

e) Restrições ao equilíbrio alimentar e à venda dos produtos fornecidos pelas atividades culturais (Kg)

Das atividades vegetais praticadas na área em estudo considera-se, de acordo com a informação recolhida no trabalho de campo, que todas as suas produções são destinada à alimentação animal e/ou vendidas, incluindo a destinada ao autoconsumo pela família, bem como a reempregue como semente (uma vez que esta é incluída como encargo variável no cálculo dos coeficientes da função objetivo).

Procura-se, assim, atingir um equilíbrio entre as disponibilidades produtivas da exploração, incluindo a do baldio, e as necessidades alimentares do efetivo pecuário ao longo do ano, essencialmente para os produtos forrageiros não destinados ao mercado (milho-forragem, erva verde e pastos). Para os restantes produtos, resultantes das atividades batata; milho-grão; centeio; e também silagem de milho e feno, considera-se, para além do consumo animal, a

possibilidade de venda. Esta situação foi observada na maioria das explorações visitadas, sendo a silagem de milho e o feno apenas transacionados, como já referido, nas situações em que a sua produção excede as necessidades alimentares dos animais. Neste último caso, as restrições são estabelecidas para permitirem que as quantidades vendidas não ultrapassem a diferença entre as quantidades produzidas e as consumidas pelos animais. Considerou-se a existência de escoamento assegurado no mercado dos produtos principais obtidos na exploração.

Pretende-se, ainda, assegurar que a quantidade de alimentos armazenados, transferidos para os períodos em que são necessários, não ultrapassem a disponibilidade existente dos mesmos.

Para a elaboração destas restrições, dividiu-se o ano em quatro períodos de tempo.

Dada a dificuldade de, através de inquérito, obter valores das produções de ferrejos, erva verde e pastagens dos prados temporários e dos lameiros, em cada período, utilizaram-se valores encontrados em publicações da especialidade (Klapp, 1986 e Moreira, 2002). Com base nos mesmos documentos definiu-se a taxa de aproveitamento do pasto – eficiência do pastoreio⁵, bem como as perdas com os processos de fenação e de ensilagem.

$$\sum_j CA_{jt} + \sum_p CP_{pt} + CB_t + \sum_b CB_{bt} + \sum_a TF'_{at} + \sum_v VA_v - \sum_i \sum_j g_{ji} \cdot XA_{ij} + \sum_p g_{pi} \cdot XP_p + g_t \cdot XB + \sum_b g_{bt} \cdot XB_b \leq 0, \forall t \quad (25)$$

f) Restrições à alimentação das atividades pecuárias (UF)

Estas restrições estão em Unidades Forrageiras Leite (UFL⁶), dado ser esta a unidade recomendada para exprimir as necessidades das fêmeas, em lactação, em gestação, secas e, também, para as novilhas. Apenas, nos casos das necessidades energéticas dos animais provenientes dos cruzamentos das raças em estudo com raças de carne, seria mais adequado a sua expressão em

⁵ Eficiência do pastoreio - proporção de erva crescida que é consumida pelos animais (Klapp, 1986 e Moreira, 2002).

⁶ Uma UFL é a quantidade de energia líquida (1700 kcal) fornecida por um quilograma de cevada média distribuída acima da manutenção de uma vaca em lactação. Esta energia é exportada no leite ou fixada sob a forma de reservas corporais (INRA, 1988).

Unidades Forrageiras Carne (UFV⁷), recomendadas para os ruminantes em crescimento rápido ou em engorda. Porém, para uniformização das unidades usadas optou-se pela utilização da primeira, por ser mais abrangente (INRA, 1988).

Estas restrições destinam-se a permitir que as disponibilidades de unidades forrageiras (UF) produzidas pela exploração e obtidas do exterior não sejam inferiores às necessidades das atividades pecuárias, garantindo deste modo a sua satisfação (expressão 26).

Além disso, fornece-se informação sobre as forragens permitidas para alimentação do gado Maronês, no sentido da produção de “carne Maronesa DOP”, indicando a interdição no uso de silagens e concentrados compostos sintéticos, como informa o seu caderno de especificações (Cooperativa Agrícola de Vila Real, 2001) (expressão 27).

A restrição (28) pretende impor a não utilização do baldio, pelos bovinos leiteiros, como se constatou no trabalho de campo. Não se considerou, também, a possibilidade de alimentá-los com “concentrado Maronês”, produzido especificamente para o gado bovino indicado na sua designação.

Para maior ajustamento das disponibilidades forrageiras e necessidades alimentares, dividiu-se o ano em quatro períodos de tempo.

$$\sum_r f_{rr} \cdot XV_r + \sum_s f_{sr} \cdot XC_s - \sum_j cUFA_j \cdot CA_{jt} - \sum_p cUFP_p \cdot CP_{pt} - cUFB \cdot CB_t - \sum_b cUFB \cdot CB_{bt} - \sum_a cUFC_a \cdot TF_{at}^t - \sum_c cUFAl_c \cdot Al_{ct} \leq 0, \forall t \quad (26)$$

$$f_{lt} \cdot XV - \sum_j cUFA_j \cdot CA_{jt} - \sum_p cUFP_p \cdot CP_{pt} - cUFB \cdot CB_t - \sum_b cUFB \cdot CB_{bt} - \sum_{a=1}^4 cUFC_a \cdot TF_{at}^t - \sum_{c=1}^7 cUFAl_c \cdot Al_{ct} \leq 0, \forall t \quad (27 \text{ SMAP})$$

$$\sum_{r=3}^6 f_{rr} \cdot XV_r - \sum_j cUFA_j \cdot CA_{jt} - \sum_p cUFP_p \cdot CP_{pt} - \sum_a cUFC_a \cdot TF_{at}^t - \sum_{c=1}^6 cUFAl_c \cdot Al_{ct} - \sum_{c=8}^9 cUFAl_c \cdot Al_{ct} \leq 0, \forall t \quad (28)$$

g) Restrições à alimentação das atividades pecuárias (Kg PDI)

A satisfação das necessidades alimentares em unidades forrageiras nem sempre garante a satisfação das necessidades proteicas. Impôs-se, por isso, ao

⁷ Uma UFV é a quantidade de energia líquida (1820 kcal) fornecida por um quilograma de cevada média para a manutenção e engorda de um ruminante (INRA, 1988).

modelo as mesmas condições colocadas no grupo de restrições anteriores, mas em unidades de proteína digestível no intestino (PDI⁸).

$$\begin{aligned} & \sum_r h_{rt}.XV_r + \sum_s h_{st}.XC_s - \sum_j cPDIA_j.CA_{jt} - \sum_p cPDIP_p.CP_{pt} - cPDIB.CB_t - \sum_b cPDIB.CB_{bt} \\ & - \sum_a cPDIC_a.TF_{at}' - \sum_c cPDIAI_c.AI_{ct} \leq 0, \forall t \end{aligned} \quad (29)$$

$$\begin{aligned} & h_{1t}.XV - \sum_j cPDIA_j.CA_{jt} - \sum_p cPDIP_p.CP_{pt} + cPDIB.CB_t - \sum_b cPDIB.CB_{bt} - \sum_{a=1}^4 cPDIC_a.TF_{at}' \\ & - \sum_{c=1}^7 cPDIAI_c.AI_{ct} \leq 0, \forall t \end{aligned} \quad (30 \text{ SMAP})$$

$$\begin{aligned} & \sum_{r=3}^6 h_{rt}.XV_r - \sum_j cPDIA_j.CA_{jt} - \sum_p cPDIP_p.CP_{pt} - \sum_a cPDIC_a.TF_{at}' - \sum_{c=1}^6 cPDIAI_c.AI_{ct} - \sum_{c=8}^9 cPDIAI_c.AI_{ct} \leq 0, \forall t \end{aligned} \quad (31)$$

h) Restrição à capacidade máxima de ingestão de alimento das atividades pecuárias (Kg de MS)

Para o estabelecimento do equilíbrio alimentar do gado presente nas unidades de produção, torna-se necessário tomar em consideração a quantidade de forragem (feno, silagens, palhas, erva) que os animais podem ingerir, pois a sua capacidade de ingestão, variando com o estado fisiológico em que o animal se encontra, é limitada.

Com as restrições relativas à ingestão de matéria seca pretende-se que a quantidade ingerida da mesma seja sempre inferior à capacidade de ingestão das unidades pecuárias.

$$\begin{aligned} & \sum_j cMSA_j.CA_{jt} + \sum_p cMSP_p.CP_{pt} + cMSB.CB_t + \sum_b cMSB.CB_{bt} + \sum_a cMSC_a.TF_{at}' + \sum_c cMSAI_c.AI_{ct} \\ & - \sum_r i_{rt}.XV_r - \sum_s i_{st}.XC_s \leq 0, \forall t \end{aligned} \quad (32)$$

$$\begin{aligned} & \sum_j cMSA_j.CA_{jt} + \sum_p cMSP_p.CP_{pt} + cMSB.CB_t + \sum_b cMSB.CB_{bt} + \sum_{a=1}^4 cMSC_a.TF_{at}' \\ & + \sum_{c=1}^7 cMSAI_c.AI_{ct} - i_{1t}.XV_1 \leq 0, \forall t \end{aligned} \quad (33 \text{ SMAP})$$

$$\begin{aligned} & \sum_j cMSA_j.CA_{jt} + \sum_p cMSP_p.CP_{pt} + \sum_a cMSC_a.TF_{at}' + \sum_{c=1}^6 cMSAI_c.AI_{ct} + \sum_{c=8}^9 cMSAI_c.AI_{ct} \\ & - \sum_{r=3}^6 i_{rt}.XV_r \leq 0, \forall t \end{aligned} \quad (34)$$

⁸ O sistema PDI determina o valor azotado de cada alimento relativamente à quantidade de ácidos aminados realmente absorvidos no intestino e que são fornecidos pelas proteínas alimentares não degradadas no rúmen, ou pelas proteínas microbianas (INRA, 1988).

i) Restrição à ingestão mínima de alimentos produzidos na exploração (Kg de MS)

As necessidades alimentares respeitantes às atividades pecuárias devem ser fundamentalmente satisfeitas pelos alimentos produzidos nas explorações, valorizando-se, assim, as culturas forrageiras através da produção animal. Tendo como objetivo conseguir o equilíbrio forragem-pecuária obtido na exploração, mas também de forma a assegurar a sua autonomia, incluiu-se no modelo a restrição relativa à quantidade mínima de alimento produzido pelas atividades vegetais, destinadas ao consumo animal. Esta informação vem obedecer, também, ao estipulado na “Norma para Carne Bovino”, definida no âmbito do Projeto *Extensity* – Sistemas de Gestão Ambiental e de Sustentabilidade na Agricultura Extensiva (Domingos *et al.*, 2005), ao indicar-se que as unidades de produção devem ser autossuficientes em pelo menos cinquenta por cento da alimentação, em termos de matéria seca, calculada numa base anual.

$$0,5 \cdot \sum_r \sum_t i_{rt} \cdot XV_r + 0,5 \cdot \sum_s \sum_t i_{st} \cdot XC_s - \sum_j \sum_t cM_{SA_j} \cdot CA_{jt} - \sum_p \sum_t cM_{SP_p} \cdot CP_{pt} - \sum_t cM_{SB} \cdot CB_t - \sum_b \sum_t cM_{SB} \cdot CB_{bt} - \sum_a \sum_t cM_{SC_a} \cdot TF_{at}' \leq 0 \quad (35)$$

j) Restrição à venda de carne (Kg)

Impõe-se ao modelo, com estas restrições, que a quantidade de venda de carne, proveniente das diversas unidades pecuárias, seja igual ou inferior à sua produção. Não se coloca qualquer outro limite à venda de carne, pois verifica-se que a procura deste produto é superior à oferta existente na região.

$$\sum_r \sum_x VC_{rx} - \sum_r \sum_x j_{rx} \cdot XV_r \leq 0 \quad (36) \qquad \sum_s \sum_y VC_{sy} - \sum_s \sum_y j_{sy} \cdot XC_s \leq 0 \quad (37)$$

k) Restrição à venda de leite (Kg)

Impõe-se ao modelo, com esta restrição, que a quantidade de venda de leite seja igual ou inferior à sua produção.

Acrescenta-se, ainda, a expressão (39), que tenta simular o efeito das quotas leiteiras, impostas às explorações da área em estudo, tendo-se feito coincidir o

limite à produção de leite com a quantidade média de leite produzido nas explorações visitadas.

$$VL - \sum_{r=3}^4 k_r . XV_r \leq 0 \quad (38)$$

$$VL \leq \overline{VL} \quad (39)$$

$$\sum_s VL_s - \sum_s k_s . XC_s \leq 0 \quad (40 \text{ SM})$$

D) Restrição à venda de queijo (Kg):

Esta restrição impõe que a quantidade de venda de queijo, apenas na Serra de Montemuro, seja igual ou inferior à sua produção.

$$\sum_s VQ_s - \sum_s ch_s . XC_s \leq 0 \quad (41 \text{ SM})$$

m) Restrição à venda de estrume (Kg)

Esta restrição é estabelecida para permitir a venda de estrume de origem bovina, em quantidades não superiores à diferença entre as quantidades produzidas e as utilizadas nas culturas vegetais.

Desconhecendo-se as necessidades do mercado em absorver este produto, acrescentou-se a restrição (44) que impõe um limite à venda de estrume coincidente com a quantidade média vendida na área em estudo.

Atendendo à preservação dos sistemas agropecuários e a sua compatibilização com o ambiente, com as expressões definidas limita-se, também, a produção de estrumes às necessidades das culturas e à venda possível. Isto porque, quando a área do solo é insuficiente para a aplicação dos estrumes, não se observa o encaminhamento dos efluentes em excesso para outro destino, para além da transação verificada, como é o caso de estações de tratamento de águas residuais (ETAR) individuais ou coletivas, cuja ausência é perceptível nas áreas em estudo. Nestas condições, o armazenamento, processamento e eliminação inadequados poderiam trazer consequências várias sobre o ambiente e os seres vivos

$$\sum_r VE_r + \sum_r UE_r - \sum_r d_r . XV_r = 0 \quad (42) \quad VE + UE - d . XC = 0 \quad (43)$$

$$\sum_r VE_r + VE \leq \overline{VE} \quad (44)$$

n) Restrição ao encabeçamento animal (CN)

No sentido de garantir um número de animais compatível com os recursos naturais existentes, considerou-se que o encabeçamento da unidade de produção deve ser inferior ou igual a três Cabeças Normais⁹ (CN) por hectare de SAU, em zonas de montanha, de acordo com as “Boas Práticas Agrícolas” (MADRP, 2003).

Na situação média do conjunto de explorações estudadas, não se verifica restrição ao encabeçamento animal, imposta pela capacidade de alojamento, pois esta é superior à imposta pela área disponível.

$$\sum_r CN_r \cdot XV_r + \sum_s CN_s \cdot XC_s \leq 3 \left(\sum_i \sum_j XA_{ij} + \sum_p XP_p + XB + \sum_b XB_b + \sum_h X_h \right) \quad (45)$$

o) Restrição ao efetivo bovino (CN)

Para as serras do Marão-Alvão-Padrela incluiu-se a restrição (46) relativa ao número mínimo de CN de gado bovino (cinco), porque foi a condição considerada base da sustentabilidade económica, de acordo com o trabalho de Marta-Costa (2008), para as explorações sustentáveis daquela área. Neste modelo foi ainda imposto que o efetivo de gado Maronês deveria ser, pelo menos, vinte e cinco por cento dos bovinos totais adotados (expressão (47)). As razões para esta condição prendem-se com o importante contributo ambiental dos bovinos de raça autóctone Maronesa comparativamente com as diversas raças bovinas em estudo, demonstrada no mesmo trabalho e já referidas anteriormente.

Para a serra de Montemuro não foi considerado limite mínimo ao número de animais.

$$\sum_r XV_r \geq 5 \quad (46 \text{ SMAP}) \qquad XV_1 + XV_2 - 0,25 \sum_r XV_r \geq 0 \quad (47 \text{ SMAP})$$

⁹ Medida pecuária que relaciona os efetivos, convertidos em cabeças normais, em função das espécies e das idades, através de uma tabela de conversão, e, em que, um animal adulto da espécie bovina corresponde a 1 C.N. (INE, 2005).

4. Resultados da otimização do modelo

Neste ponto apresentam-se os resultados dos modelos delineados para as serras do Marão-Alvão-Padrela e para a serra de Montemuro, atendendo aos pressupostos identificados anteriormente.

O método utilizado – NISE – inicia-se com a otimização individual de cada objetivo, sujeito ao conjunto de restrições apresentadas para os modelos, sendo usado o valor de cada um para a obtenção de cada solução ótima. Este procedimento conduziu à matriz de *pay-off* que se encontra no Quadro 1 e, consequentemente, à obtenção dos dois primeiros pontos extremos eficientes, para as situações sem e com apoios monetários à atividade corrente, respetivamente.

Da matriz de compromisso (*pay-off*) confirma-se o forte grau de conflito entre os dois objetivos considerados, em ambos os cenários, nas duas regiões em estudo. Quando o VAB é maximizado, os custos energéticos alcançam o seu pior valor ou anti-ideal e vice-versa. Os elementos da diagonal principal representam o *ponto ideal* (I^*), isto é, a solução na qual todos os objetivos encontram os seus valores ótimos. Os piores valores em cada coluna são chamados *ponto anti-ideal* (I).

Quadro 1: Matriz de *pay-off* para os objetivos Z1 e Z2 apresentando os pontos P1 e P2 para cada um dos modelos e cenários.

		Cenário sem subsídios		Cenário com subsídios	
		VAB (Euros)	Custos energéticos (MJ)	VAB (Euros)	Custos energéticos (MJ)
Serras do Marão-Alvão-Padrela	MAX Z ₁	12 513	239 852	20 860	188 845
	MIN Z ₂	1 821	56 142	6 749	56 142
Serra de Montemuro	MAX Z ₁	17 817	298 770	20 349	414 012
	MIN Z ₂	387	10 185	1 085	10 185

Fonte: Resultados próprios.

No sentido de reduzir o tamanho da zona de eficiência, isto é, o triângulo $P_1P_2I^*$, procedeu-se, seguidamente, com o método NISE, através da otimização de uma única função objetivo, composta pelas funções – VAB e custos energéticos – sendo estas ponderadas pelo valor de inclinação da reta que une

os dois pontos extremos eficientes encontrados (pontos ideais e anti-ideais da matriz de compromisso).

As iterações continuaram até o erro encontrado ter sido inferior ao fixado. O erro máximo admissível foi 7% do valor do erro máximo possível, sendo este a distância máxima entre as fronteiras inferior e superior da zona de eficiência, obtidas na primeira iteração, medida na direção perpendicular à fronteira inferior. Utilizando o método NISE foram identificados sete e cinco pontos eficientes no cenário sem e com subsídios, respectivamente, para cada uma das áreas em estudo.

Dado que, no espaço dos objetivos considerados, os resultados encontrados têm uma amplitude de variação bastante grande, seguiu-se a determinação do conjunto compromisso, utilizando o procedimento descrito no ponto 3. Este corresponde ao conjunto de soluções eficientes que estão mais próximas da solução ideal, a distância compreendida entre L_1 e L_∞ , considerados pesos idênticos para cada um dos objetivos considerados.

Salienta-se, porém, que dada a impossibilidade de existência de um número não inteiro de animais nas explorações, fornecido pelas soluções do modelo inicial, procuraram-se novas soluções compromisso (soluções compromisso alteradas). Para isso, impôs-se aos modelos iniciais a condição de que cada atividade pecuária deveria ser igual ao número inteiro mais próximo ao obtido com as primeiras soluções compromisso.

As soluções compromisso (alteradas) obtidas ao modelo nos cenários sem e com subsídios à atividade corrente encontram-se no Quadro 2. A sua análise permite enumerar as seguintes observações:

- As atividades selecionadas pelo modelo demonstram uma acentuada utilização das áreas ocupadas por culturas vegetais ligadas à atividade pecuária. Por exemplo, o prado temporário (de regadio) é uma atividade sempre presente. A terra arável de sequeiro é, na generalidade, ocupada com batata e/ou prado temporário (apenas nas serras do Marão-Alvão-Padrela);
- As áreas de lameiros de feno e pasto são integralmente utilizadas, como imposto aos modelos, sendo necessário arrendar pastagens de regadio, nas

situações em que se consideram subsídios à atividade corrente, na serra de Montemuro;

- As pastagens de baldio são integralmente utilizadas para pastoreio animal nas serras do Marão-Alvão-Padrela. Também na serra de Montemuro elas são utilizadas/limpas diretamente pelo pastoreio animal, em detrimento da limpeza mecânica, mas são, na generalidade, consumidas numa relação proporcional ao efetivo bovino identificado nas soluções;
- De entre as várias hipóteses fornecidas ao modelo, relativamente à atividade bovina, verifica-se em todas as soluções obtidas a seleção de animais do tronco frísio, em número que varia dos 3-4 aos 4-6 animais, nas serras do Marão-Alvão-Padrela e Montemuro, respetivamente, com venda ao nascimento das suas crias cruzadas. Os bovinos maroneses estão sempre presentes em número entre 7-9 e os arouqueses devem, também, estar presentes nas explorações, com exceção de uma solução, em número entre 1 e 2 unidades pecuárias, em cada uma das áreas em estudo. Destaque-se que, em todas as soluções obtidas para os dois modelos considerados para as serras do Marão-Alvão-Padrela, o efetivo bovino Maronês é manifestamente superior a 25% do efetivo bovino total, limite imposto nos modelos, com valores mais acentuados na situação com subsídios à atividade corrente. Os descendentes puros de raça Frísia são apenas considerados em duas soluções da serra de Montemuro, que coincidem com aquelas em que o objetivo económico assume valores superiores (pontos L_i). Nestas soluções, denota-se também, de modo geral, um número mais elevado de bovinos do tronco Frísio, relativamente às restantes soluções;
- No âmbito dos pequenos ruminantes, o modelo considera, em todas as soluções expostas para a serra de Montemuro, a existência de ovinos para carne e leite, sendo este transformado, na sua totalidade, em queijo. Para as serras do Marão-Alvão-Padrela, não foram selecionados caprinos da raça Bravia;
- Observa-se a venda dos produtos vegetais obtidos na exploração, incluindo o feno, na situação em que as disponibilidades excedem as necessidades alimentares dos animais. Esta situação verifica-se, principalmente, pela substituição daqueles por alimentos mais ricos em

energia e em proteína e com volumes de matéria seca inferiores (concentrado comercial e silagem de milho);

- Também o estrume é vendido até ao limite imposto, excetuando em duas das soluções da Serra de Montemuro, cujo objetivo ambiental é melhorado, isto é, os custos energéticos são inferiores, devido ao menor número de animais explorados;
- A contratação de mão-de-obra temporária foi relativamente reduzida, sendo necessária, apenas, em duas situações que coincidem com a existência de maior encabeçamento animal. Refira-se que, apesar da minimização da mão-de-obra temporária não ser um objetivo claramente definido no modelo formalizado, identificado indiretamente, no objetivo da maximização do VAB através do custo que acarreta, era importante a sua redução devido à escassez em mão-de-obra temporária disponível que se verifica nas áreas em estudo. No entanto, é ainda de salientar que, em todas as soluções obtidas se verificam períodos em que a mão-de-obra própria é excedentária, com distribuição não uniforme pelos períodos considerados;
- Também com a tração se verifica que a disponibilidade existente é mais do que suficiente para as necessidades verificadas, não sendo necessário recorrer ao seu aluguer;
- Os modelos escolhem, para todas as soluções obtidas, compra ao exterior de fertilizantes azotados e fosfatados, em detrimento da utilização de quantidades superiores de estrume, que são vendidas;
- Os concentrados comerciais e a silagem de milho comprados são identificados numa relação proporcional com o efetivo bovino das raças do tronco Frísio, devendo ser distribuído, de forma regular, ao longo do ano.

Os pontos obtidos permitem, ainda, conhecer as curvas de transformação (*trade-off*) e avaliar, na perspetiva do conceito de marginal, qual o sacrifício dos vários objetivos quando se deseja melhorar, de uma unidade adicional, um deles. Este exercício é apresentado, nas Figuras 1 e 2, para serras do Marão-Alvão-Padrela, na situação sem e com subsídios, respetivamente.

Quadro 2: Soluções compromisso nos cenários sem e com subsídios à atividade corrente, para cada um dos modelos e cenários.

Pontos extremos	Serras do Marão-Alvão-Padrela				Serra de Montemuro			
	Sem subsídios		Com subsídios		Sem subsídios		Com subsídios	
	L_1	L_∞	L_1	L_∞	L_1	L_∞	L_1	L_∞
Objetivos								
VAB (€)	10 915	9 708	18 672	16 036	16 289	9 588	18 748	13 222
Custos energéticos (MJ)	117 185	104 354	119 352	101 516	246 827	146 435	243 666	159 572
Variáveis de decisão principais								
Regadio (Hectare - Ha)								
Batata	0	0	0	0	0,2040	0	0	0,3075
Centeio	0	0	0	0	0,2040	0	0	0
Milho silagem	0,012	0	0	0	0	0	0	0
Prado temporário	3,430	3,442	3,442	3,442	0,2040	0,612	0,612	0,3045
Intercalar (ferrã)	0	0	0	0	0	0	0	0,1754
Sequeiro (Ha)								
Batata	0,9354	0,9019	0,5689	0,1918	0,6178	0,7379	0,8128	1,2644
Milho silagem	0	0,0154	0,0455	0	0	0	0	0
Prado temporário	0,0386	0,0567	0,3596	0,7822	0	0	0	0
Lameiros e baldio (Ha)								
Feno	2,7511	2,7511	2,7511	2,7511	0,7841	0,7841	1,4832	1,0317
Pasto	5,4369	5,4369	5,4369	5,4369	1,5919	1,5919	1,5919	1,5919
Baldio (com pastoreio)	14,526	14,526	14,526	14,526	5,1492	0,9092	2,9613	1,4012
Bovinos (Unidade Pecuária - UP)								
Maronesa (F1 pura)	8	7	9	8	0	0	0	0
Arouquesa (F1 pura)	0	0	0	0	1	0	1	2
Frísia (F1 pura - 0m)	0	0	0	0	1	0	1	1
Tronco Frísio (F1 cruz. - 0m)	4	3	4	3	6	4	6	5
Ovinos e caprinos (UP)								
Ovinos (carne e queijo)	0	0	0	0	30	17	28	9
Venda de produtos de origem vegetal e animal (Quilograma - kg)								
Batata	10 795	10 408	6 565	2 213	9 704	8 515	9 380	18 473
Feno	30 838	30 889	30 889	30 889	5 513	7 241	11 387	7 406
Leite de vaca	0	0	0	0	44 551	25 458	44 551	31 822
Queijo de ovelha	0	0	0	0	1 080	612	1 008	324
Estrume bovino	70 000	70 000	70 000	70 000	70 000	39 545	70 000	48 131

Quadro 2: Soluções compromisso nos cenários sem e com subsídios à atividade corrente, para cada um dos modelos e cenários (continuação).

Pontos extremos	Serras do Marão-Alvão-Padrela				Serra de Montemuro			
	Sem subsídios		Com subsídios		Sem subsídios		Com subsídios	
	L_1	L_∞	L_1 ,	L_∞ ,	L_1	L_∞	L_1 ,	L_∞ ,
Aquisição de fatores de produção ao exterior (kg)								
Terra arável sequeiro Ha	0	0	0	0	0,0058	0,1259	0,2008	0,6524
Pastagens regadio (Ha)	0	0	0	0	0	0	0,6992	0,2476
Mão-de-obra (H)	13,062	0	33,608	0	47,878	0	58,239	0
N	168,87	191,58	148,08	164,40	31,391	56,619	57,156	86,609
P2O5	188,66	234,50	142,42	172,17	35,403	81,617	42,059	101,08
Milho grão	0	0	0	0	473,96	0	0	1,446
Silagem de milho	0	0	0	0	33 177	27 281	35 659	18 520
Concentrado Maronesa	9 430	8 965	11 827	10 570	0	0	0	0
Concentrado comercial	5 816	3 188	4 201	2 892	37 145	20 148	35 740	20 411
Outros alimentos (kg)	79,301	0	82,163	0	0	0	0	0

Fonte: Resultados próprios publicados em Marta-Costa (2010) e Marta-Costa *et al.* (2013).

Observações: (1) Apresentam-se, apenas, as variáveis principais cujo resultado é diferente de zero; (2) para os produtos das atividades com destinos diversos, coloca-se somente a fração vendida, sendo a restante reempregue na exploração (casos do estrume e produtos vegetais); (3) com exceção dos animais destinados à substituição, os restantes nascidos destinam-se à venda, bem como os animais de refugio; (4) os valores encontram-se arredondados aos dígitos apresentados, com exceção do efetivo animal.

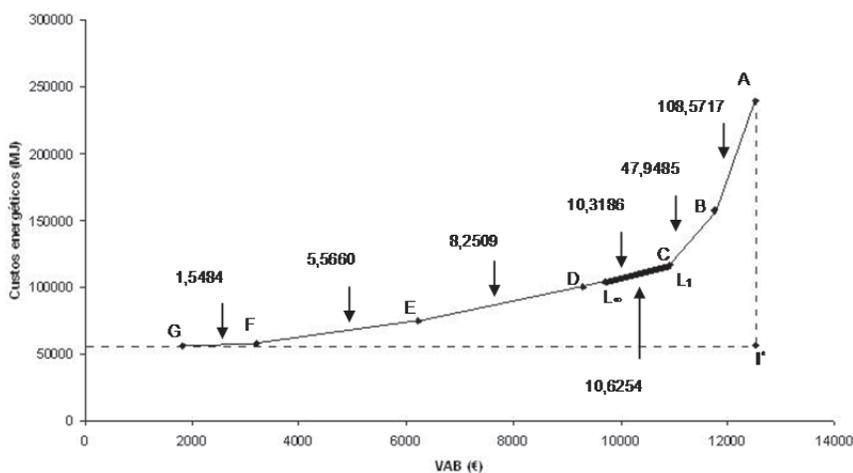


Figura 1: Curvas de *trade-off* entre os objetivos VAB e Custos energéticos para o cenário sem subsídios à atividade corrente, para as serras do Marão-Alvão-Padrela (Resultados próprios).

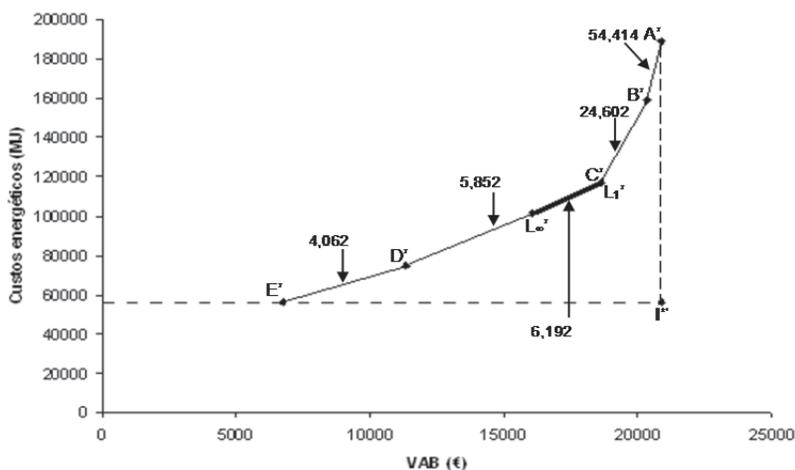


Figura 2: Curvas de *trade-off* entre os objetivos VAB e Custos energéticos para o cenário com subsídios à atividade corrente, para as serras do Marão-Alvão-Padrela (Resultados próprios).

Analisando as Figuras 1 e 2, observa-se que entre os pontos A e B a inclinação é, nesta parte da curva de *trade-off*, de 108,572 e 54,414 respetivamente. Isto significa que a cada incremento de um euro de VAB, corresponde um incremento de 108,572 e 54,414 MJ de custos energéticos, para as situações sem e com subsídios, respetivamente. O custo de oportunidade de cada Euro de VAB pode, assim, ser medido como o sacrifício de 108,572 e 54,414 MJ de custo energético, respetivamente.

Por outro lado, observa-se, também, que entre as várias soluções possíveis o custo energético cresce a uma taxa crescente (a inclinação vai aumentando), sendo a mesma acompanhada pela subida do custo de oportunidade do VAB. Se, por hipótese, a escolha for o ponto B em vez do ponto A da Figura 1, então, uma diminuição de aproximadamente 760 euros no VAB compensam a redução de 82 469 MJ de custos energéticos.

Nas soluções compromisso um aumento de um euro no VAB conduz a um acréscimo de 10,625 (Figura 1) e 6,192 MJ (Figura 2) nos custos energéticos, para os cenários sem e com subsídios à atividade correntes, respetivamente, o que significa um aumento mais que proporcional dos custos energéticos.

5. Considerações finais e conhecimento adquirido

Ao longo deste trabalho apresentaram-se dois modelos de exploração agropecuária para as serras do Marão-Alvão-Padrela e de Montemuro, no contexto da sustentabilidade, usando cenários sem e com subsídios. As técnicas utilizadas permitiram encontrar um conjunto de soluções, as quais terão de ser avaliadas por cada decisor. Isto significa que não se pode excluir, *a priori*, qual dos modelos ou qual das soluções se deveria eleger na orientação da tomada de decisão. Todos apresentam vantagens e inconvenientes, quando os resultados são analisados dentro de possíveis cenários alternativos com conjunturas socioeconómicas diversas e onde os objetivos possam ter mais ou menos importância.

Tal como na maioria dos problemas de multicritério ou multiobjetivo não existe uma solução ótima, mas um conjunto de soluções eficientes que poderão ser ótimas considerando diferentes valorações dos critérios ou objetivos. Assim, apenas o decisor poderá avaliar a relevância das diferentes soluções apresentadas. Neste caso, os modelos desenvolvidos não têm por objetivo uma aplicação direta. Estes modelos são muito úteis para os decisores poderem manipular e, deste modo, adquirir conhecimento que será utilizado na tomada de decisão nomeadamente sobre políticas e normas a implementar. Assim, o objetivo é essencialmente adquirir conhecimento sobre o sistema em estudo e, deste modo, perceber relações entre componentes do sistema anteriormente desconhecidas ou, mesmo, desenvolver algumas regras de decisão que possam ser aplicadas em políticas de gestão do território futuras.

Nos parágrafos seguintes enumeram-se algum do conhecimento reconhecido pelos analistas como resultados deste projeto e transmitido aos decisores, nomeadamente à Associação de Criadores do Maronês e à Associação de Municípios do Vale Douro Sul, desconhecendo-se a sua aplicação direta ou indireta no terreno.

Tendo por base as soluções compromisso extremas obtidas, considerando pesos idênticos para os objetivos considerados, verifica-se que o equilíbrio resulta da seleção das atividades batata e prados temporários (regadio e sequeiro); lameiros de feno e de pasto; e utilização de baldio. Como atividades pecuárias são sempre selecionados animais do tronco frísio, para produção de

carne e leite; e ovinos para produção de carne e leite (serra de Montemuro), devendo este último ser transformado, na íntegra, em queijo, no sentido de assegurarem o rendimento económico da exploração. Os animais de raça Maronesa (serras do Marão-Alvão-Padrela) e Arouquesa (serra de Montemuro) são também uma opção a considerar em quase todas as situações.

Todas aquelas soluções incluíram, assim, as atividades pecuárias carne e leite, sendo certo que a atividade leite está sempre presente quando a tendência for para um maior VAB. Também é certo que a atividade carne está sempre presente, quando a tendência for para o mínimo custo energético.

As atividades bovinas selecionadas permitem concluir que, face aos recursos disponíveis, as áreas mais agrestes e menos produtivas, como os baldios, devem ser aproveitadas pelas raças rústicas (Maronesa e Arouquesa), transformadores naturais dos recursos intrínsecos das zonas de montanha, resultando numa mais-valia ambiental e económica. Para estas situações, a dimensão do efetivo bovino autóctone pode ainda ser reforçado, com consequências económicas, se atribuídas ajudas monetárias à sua exploração. As áreas mais produtivas, por outro lado, devem ser utilizadas para atividades mais rendíveis, como a exploração de bovinos leiteiros, cuja venda do vitelo deve ser realizada à nascença, devido não só ao maior rendimento económico daí resultante, mas também à ausência de recursos alimentares disponíveis para a sua criação. A seleção e complementaridade das duas atividades, nos seus extremos parece ser, assim, a chave para o futuro do setor agropecuário, em zonas de montanha, no sentido da sustentabilidade:

Bovinos de carne/raça mais rústica (autóctone)/menor custo energético/menor VAB

Bovinos de leite/raças exóticas/maior custo energético/maior VAB

Pode concluir-se, portanto, que é possível planear as atividades agropecuárias por forma a conciliar alguns objetivos essenciais no contexto da promoção da agricultura sustentável em áreas de montanha. Por um lado é possível obedecer aos requisitos ambientais da manutenção da biodiversidade genética e do território sem degradar o ambiente, através da exploração de raças animais autóctones, com elevada rusticidade e promotores de benefícios

económicos, apesar da sua baixa produtividade. Por outro lado, a exploração de raças exóticas, apesar de originar consequências ambientais negativas, fortalece os benefícios económicos totais da exploração, promovendo maior qualidade de vida, essencial para a manutenção da população e para o desenvolvimento rural.

A elaboração dos modelos definidos e as soluções obtidas suscitaram ainda algumas reflexões, tendo presente o enquadramento dos objetivos que se pretendem alcançar e que se passam a expor:

- A área das explorações é uma das principais limitações dos resultados obtidos. Aumentando o limite de área arrendada, por exemplo, as soluções diferem das alcançadas inicialmente, melhorando-se os resultados económicos. Além disso, na hipótese das necessidades alimentares respeitantes às atividades pecuárias serem fundamentalmente satisfeitas pelos alimentos produzidos nas explorações, em pelo menos 50%¹⁰, verifica-se a inexistência de qualquer atividade pecuária, com penalização nos seus resultados económicos;
- Verifica-se um grande excedente de mão-de-obra, em diversos períodos, bem como número de horas de tração sobranes. Neste sentido, para a mão-de-obra não utilizada, indicada pelos modelos, é necessário encontrar formas alternativas de emprego. Da mesma forma, é preciso ponderar a venda ou a utilização de algumas máquinas e equipamentos em trabalhos para o exterior, com custos horários muito elevados devido à sua baixa/nula utilização, para rendibilizar este fator de produção;
- O nível de subsídios concedidos aos animais de raças autóctones não se mostra suficiente para incentivar a sua exploração, nem para cobrir as receitas obtidas com os animais de raças exóticas.

Note-se que os modelos construídos foram fornecidos aos decisores - principalmente órgãos de gestão pública - para que possam ser utilizados no processo de decisão. Tratam-se de modelos dinâmicos, pois as múltiplas

¹⁰ Condição definida na “Norma para Carne Bovino” no âmbito do Projeto *Extensivity* – Sistemas de Gestão Ambiental e de Sustentabilidade na Agricultura Extensiva, tendo como objetivo conseguir o equilíbrio forragem-pecuária obtido na exploração, mas também assegurar a autonomia da unidade de produção, condição fundamental para a sua sustentabilidade (Domingos *et al.*, 2005).

variáveis e coeficientes encontram-se em constante mutação. É o caso, por exemplo, dos preços dos produtos e dos fatores de produção que poderão influenciar, dentro de alguns limites, os modelos delineados. No entanto, a complementaridade entre as várias atividades será (sempre) um desafio que urge praticar.

BIBLIOGRAFIA

AMVDS (Associação de Municípios do Vale Douro Sul), *Plano de Gestão Integrada do Sítio PTCO0025 Montemuro, Memória descritiva complementar*. Resende: AMVDS, 60 pp. 2008.

ALVES, V. C. *Estudo sobre a "Raça Bovina Maronesa". Situação Actual e Perspectivas Zootécnicas*. Tese de Doutoramento. Vila Real: UTAD, 172 p. 1993.

AZEVEDO, J.; Tibério, L; Torres, F.; Fonseca, C.; Almeida, L.; Marta, A.; Mascarenhas, A.; Mesquita, M.; Magalhães, M.; E Ferreira, P. *Relatório Intercalar I – Parte I (Diagnóstico de Síntese) do Projeto "Plano de Gestão Integrada do Sítio PTCO0025 Montemuro"*. Vila Real: UTAD, 63 p. 2008.

AZEVEDO, J.; Tibério, L; Torres, F.; Fonseca, C.; Almeida, L.; Capapé, A.; Marta, A.; Mascarenhas, A.; Mesquita, M.; Magalhães, M.; Ferreira, P.; Silva, J.; Costa, F.; Gonçalves, H.; Resende, D.; E Bento, J. *Relatório Intercalar II – Plano de Acção do Projeto "Plano de Gestão Integrada do Sítio PTCO0025 Montemuro"*. Vila Real: UTAD, 258 p + apêndices. 2009.

CARVALHO, N. S. *Agricultura Sustentável em Regadio e o seu Planeamento Face aos Desafios Económicos, Sociais e Ambientais (O Caso da Futura Zona Irrigada de Intervenção de Alqueva)*. In Actas do 5º Congresso Ibérico "Gestão e Planeamento de Água". 4-8 Dezembro 2006, Universidade do Algarve - Faro, 10 p. 2006.

CARVALHO, N. S. *O Planeamento Agrícola Face aos Actuais Desafios Ambientais, Sociais e Económicos (Uma Aplicação da Programação por Metas Lexicográficas ao Caso da Futura Zona de Regadio de Intervenção de Alqueva)*. In Actas do V Congresso Nacional da Associação Portuguesa de Economia Agrária "Globalização, Agricultura e Áreas Rurais". 4-6 Outubro 2007, UTAD - Vila Real, 10 pp. 2007.

CCE (Comissão das Comunidades Europeias) e DGA (DIREÇÃO-GERAL DA AGRICULTURA. *O Sector da Carne da União Europeia*. Fact sheet. Bruxelas, 23 p. 2004.

CE (Comunidade Europeia). Directiva 91/676/CE do Conselho, de 12 de Dezembro de 1991, relativa à protecção das águas contra a poluição causada por nitratos de origem agrícola. *Jornal Oficial das Comunidades Europeias* N.º L 375 de 31.12.1991, pp. 1-8. 1991.

COHON, J. L. *Multiobjective Programming and Planning. Mathematics in Science and Engineering* Vol. 140. San Diego, California: Academic Press, 333 p. 1978.

COLAÇO-DO-ROSÁRIO, M. F. *Ensaio de Caracterização Agrária Integrada do Norte de Portugal no Contexto do Continente – Estudo com vista à Orientação do Agros na Perspectiva da Modernidade no Equilíbrio*. Tese de Doutoramento. Vila Real: UTAD, 638 p. 1998.

COOPERATIVA AGRÍCOLA DE VILA REA. *Carne Maronesa - Caderno de Especificações*. Vila Real, 14 p. 2001.

DOMINGOS, T.; Rodrigues, O.; Avelar, T.; Brito, A.; Piçarra, A.; Sendim, A.; Ferreira, F.; Dias, N.; Crespo, D.; Crespo, J.; Lopes, A.; Belo, C.; Alcazar, R.; Sarmiento, N. E Sequeira, E. *Sustentabilidade Garantida - Norma para Carne Bovino*. Projecto Extensivity – Sistemas de Gestão Ambiental e de Sustentabilidade na Agricultura Extensiva LIFE ENV/P/505, 27 p. 2005.

ENESAD (Établissement National d'Enseignement Supérieur Agronomique de Dijon) e ADEME (Agence de l' Environnement et de la Maitrise de l'Énergie). *Analyse Énergétique d'Exploitations Agricoles et Pouvoir de Réchauffement global. Méthode et Résultats sur 140 Fermes Française*. Document du Groupe Planète. 2002.

FERREIRA, J.; Strecht, A.; Ribeiro, J.; Soeiro, A. E Cotrim, G. *Manual de Agricultura Biológica. Fertilização e Protecção das Plantas para uma Agricultura Sustentável*. Lisboa: Associação Portuguesa de Agricultura Biológica (AGROBIO), 435 p. 2002.

GPPAA (Gabinete de Planeamento e Política Agro-Alimentar). *Contas de Culturas das Actividades Vegetais – Modelo de Base Microeconómica*. Lisboa, CD-ROM. 2001.

INE. *Estatísticas Agrícolas 2004*. <www.ine.pt> 2005.

INRA. *Alimentation des Bovins, Ovins & Caprins*. Paris: INRA, 471 p. 1988.

KLAPP, E. *Prados e Pastagens*. 2ª Edição. Lisboa: Fundação Calouste Gulbenkian, 872 p. 1986.

MADRP *Código de Boas Práticas Agrícolas para a Protecção da Água contra a Poluição com Nitratos de Origem Agrícola*. Lisboa, 52 p.+anexos. 1997.

MADRP "Portaria N.º 1212/2003, aprova o Regulamento de Aplicação da Intervenção «Medidas Agro-Ambientais», do Plano de Desenvolvimento Rural (RURIS)". *Diário da República* N.º 240, I-B Série, de 16 de Outubro de 2003, pp. 6894-6931. 2003.

MARTA-COSTA, A. A. *Reflexão sobre o Programa de Medidas Agro-Ambientais Aplicado no Período 1994/1999, como Contributo para Futuros Programas Agro-Ambientais a Aplicar em Trás-os-Montes*. Trabalho de Síntese para prestação de Provas de Aptidão Pedagógica e Capacidade Científica. Vila Real: UTAD, 301 p. 2001.

MARTA-COSTA, A. A. *A Tomada de Decisão no Planeamento da Exploração Agrária no Contexto da Sustentabilidade: O Caso da Produção de Carne Bovina Maronesa*. Dissertação de Doutoramento. Vila Real: UTAD, Vol. I e II. 2008.

MARTA-COSTA, A. A. "Application of Decision Support Methods for Sustainable Agrarian Systems". *New Medit – Mediterranean Journal of Economics, Agriculture and Environment*, Vol. IX – N.º 2, pp. 42-49. 2010.

MARTA-COSTA, A.; Manso, F.; Tibério, L. E Fonseca, C. *Assay of a Sustainable Farming Model for the Montemuro Mountain (Portugal)*. On: Proceedings of 9th European IFSA Symposium "Building Sustainable Rural Futures. The Added Value of Systems Approaches in Times of Change and Uncertainty". I. Darnhofer e M. Grötzer Eds., Vienna (Austria), 4-7 July 2010, p. 835-844, ISBN: 978-3-200-01908-9. 2010.

MARTA-COSTA, A. A.; Manso, F.; Tibério, L.; Fonseca, C. *Ecological-economic modelling for farming systems of Montemuro Mountain (Portugal)*. In: Methods and procedures for building sustainable farming systems: Application in the European context. Marta-Costa, A. and Silva, E. Eds. Springer publication, pp. 207-217. 2013.

MENDES, A. B. e Cardoso, M. G. M. S. Clustering supermarkets: The role of experts. *Journal of Retailing and Consumer Services*, 13 (13), pp. 231-247. 2006.

MIRANDA J. *Contributos para a Definição de uma Estratégia de Intervenção em Zonas de Montanha*. In: Cadernos da Montanha – Peneda-Soajo II. Arcos de Valdevez: Francisco Rodrigues de Araújo, pp. 101-107. 2000.

MOERSCHNER, J. e Lücke, W. *Energy Investigations of Different Intensive Rape Seed Rotations – A German Case Study*. In *Economics of Sustainable Energy in Agriculture*, Chapter 3. Netherlands: Kluwer Academic Publishers, pp. 27-40. 2002.

MOREIRA, N. *Agronomia das Forragens e Pastagens*. Vila Real: UTAD, 183 p. 2002.

MOREIRA, N.; Aguiar, C. E Pires, J. *Medidas Agro-Ambientais: 3.3 Lameiros e outros prados e pastagens de elevado valor florístico - Pastagens de Montanha*. Lisboa: DGDR, 47 p. 2001.

MÜLLER, S. *¿Como Medir la Sostenibilidad? Una Propuesta para el Area de la Agricultura y de los Recursos Naturales*. Serie Documentos de Discusión sobre Agricultura Sostenible y Recursos Naturales N.º 1. Costa Rica: IICA/BMZ/GTZ, 55 p. 1996.

POETA, A. M. S. D. *A Tomada de Decisão no Planeamento da Exploração Agrícola num Contexto de Objectivos Múltiplos*. Tese de Doutoramento. UTAD: Vila Real, 182 p. 1994.

ROBERTSON, J. "Transformar a Economia – Desafio para o Terceiro Milénio". *Cadernos Schumacher para a Sustentabilidade*, 10ª Edição. Águas Santas: Edições Sempre-em-Pé, 77 p. 2007.

ROMERO, C. *Teoría de la Decisión Multicriterio: Conceptos, Técnicas y Aplicaciones*. Madrid: Alianza Editorial, 195 p. 1993.

ROMERO, C. e Rehman, T. *Multiple Criteria Analysis for Agricultural Decisions*. *Developments in Agricultural Economics* 5. Amsterdam: Elsevier, 257 p. 1989.

SANTOS, J. M. L. *Sistemas de Aproveitamento de Recursos Naturais no Alto Barroso*. Tese de Doutoramento. Lisboa: Universidade Técnica de Lisboa, ISA, 466 p. 1991.

SILVA, E. L. D. G. S. *Análisis Multicriterio da la Eficiência Económica de las Explotaciones Agroganaderas de las Azores (Portugal)*. Tesis Doctoral. Universidad de Córdoba, 295 p. 2001.

SILVA, M. A Valorização do Gado Bovino. In: *Cadernos da Montanha – Peneda-Soajo II*. Arcos de Valdevez: Francisco Rodrigues de Araújo. http://www.ardal.pt/modules/xt_conteudo/index.php?id=12. 2000.

APÊNDICE

1. Abreviaturas e símbolos usados para variáveis

XA_{ij} – atividades culturais em terras aráveis (Ha); i – tipo de terreno; j – tipo de cultura

XP_p – atividades de pastagens permanentes (Ha); p – tipo de pastagem

XB – atividade baldio (Ha) (SMAP)

XB_b – atividade baldio (Ha); b – tipo de aproveitamento/limpeza dos baldios (SM)

X_h – terra arrendada (Ha); h – tipo de terra (SM)

XV_r – atividade bovina, vacas (UP); r – raça bovina

XC_s – atividade caprina/ovina, cabras/ovelhas (UP); s – espécie animal

VA_v – atividade de venda do produto fornecido pelas atividades culturais (Kg); v – tipo de produto vegetal

VC_{rx} – atividade de venda de carne proveniente das atividades bovinas “r” (Kg); x – categoria etária animal

VC_{sy} – atividade de venda de carne proveniente da atividade caprina/ovina “s” (Kg); y – categoria etária animal

VL – atividade de venda de leite de origem bovina (Kg)

VL_s – atividade de venda de leite proveniente da atividade caprina/ovina “s” (Kg)

VQ_s – atividade de venda de queijo proveniente da atividade caprina/ovina “s” (Kg) (SM)

- VE_r – atividade de venda de estrume proveniente das atividades bovinas “r” (Kg)
 VE_s – atividade de venda de estrume proveniente da atividade caprina/ovina “s” (Kg)
 CA_{jt} – consumo de alimento proveniente das culturais “j”, no período “t” (Kg); t – período
 CP_{pt} – consumo de alimento proveniente das pastagens “p”, no período “t” (Kg)
 CB_t – consumo de alimento proveniente do baldio, no período “t” (Kg) (SMAP)
 CB_{bt} – consumo de alimento proveniente do baldio “b”, no período “t” (Kg) (SM)
 UE_r – quantidade de estrume incorporado no solo, proveniente das atividades bovinas “r” (Kg)
 UE_s – quantidade de estrume incorporado no solo, proveniente da atividade caprina/ovina “s” (Kg)
 UTM_t – quantidade de tração mecânica própria utilizada, no período “t” (H)
 TF'_{at} – transferência de alimentos do período “t” para “t” ($t \neq t$) (Kg); a – tipo de alimento
 MO_t – contratação de mão-de-obra, no período “t” (H)
 TM_t – aluguer de tração mecânica, no período “t” (H)
 FN – compra de fertilizante N (Kg de N)
 FP – compra de fertilizante P_2O_5 (Kg de P_2O_5)
 FK – compra de fertilizante K_2O (Kg de K_2O)
 Al_c – compra de alimentos no período “t” (Kg); c – tipo de alimento

2. Abreviaturas e símbolos usados para parâmetros

- AR – disponibilidade de área total de terra arável de regadio (Ha)
 AS – disponibilidade de área total de terra arável de sequeiro (Ha)
 LF – disponibilidade de área total de lameiro de feno (Ha)
 LP – disponibilidade de área total de lameiro de pasto (Ha)
 BA – disponibilidade de área total de baldio (Ha)
 a_{ijt} – necessidade de mão-de-obra para as atividades culturais “j”, no terreno “i”, no período “t” (H/Ha)
 a_{pt} – necessidade de mão-de-obra para as atividades de pastagens “p”, no período “t” (H/Ha)
 a_{bt} – necessidade de mão-de-obra para as atividades de baldio “b”, no período “t” (H/Ha) (SM)
 b_{rt} – necessidade de mão-de-obra para a atividade bovina “r”, no período “t” (H/UP)
 b_{st} – necessidade de mão-de-obra para a atividade caprina/ovina “s”, no período “t” (H/UP)
 q_{st} – necessidade de mão-de-obra para o queijo proveniente da atividade caprina/ovina “s”, no período “t” (H/Kg)
 DMO_t – disponibilidade de mão-de-obra familiar no período “t” (H)
 c_{ijt} – necessidade de tração para as atividades culturais “j”, no terreno “i”, no período “t” (H/Ha)
 c_{pt} – necessidade de tração para as atividades de pastagens “p”, no período “t” (H/Ha)
 DTM_t – disponibilidade de tração mecânica no período “t” (H)
 d_r – quantidade de estrume animal fornecido pela atividade bovina “r” (Kg/UP)
 d_s – quantidade de estrume animal fornecido pela atividade caprina/ovina “s” (Kg/UP)
 $\%N_r$ – teor em N disponível do estrume fornecido pela atividade bovina “r” (%)
 $\%N_s$ – teor em N disponível do estrume fornecido pela atividade caprina/ovina “s” (%)
 $\%P_r$ – teor em P_2O_5 do estrume fornecido pela atividade bovina “r” (%)
 $\%P_s$ – teor em P do estrume fornecido pela atividade caprina/ovina “s” (%)

$\%K_r$ - teor em K_2O do estrume fornecido pelas atividades bovinas "r" (%)
 $\%K_s$ - teor em K do estrume fornecido pela atividade caprina/ovina "s" (%)
 eN_{ij} - necessidade de fertilizante N, para as atividades culturais "j", no terreno "i" (Kg/Ha)
 eN_p - necessidade de fertilizante N, para as atividades de pastagens "p" (Kg/Ha)
 eP_{ij} - necessidade de fertilizante P_2O_5 , para as atividades culturais "j", no terreno "i" (Kg/Ha)
 eP_p - necessidade de fertilizante P_2O_5 , para as atividades de pastagens "p" (Kg/Ha)
 eK_{ij} - necessidade de fertilizante K_2O , para as atividades culturais "j", no terreno "i" (Kg/Ha)
 eK_p - necessidade de fertilizante K_2O , para as atividades de pastagens "p" (Kg/Ha)
 f_{rt} - necessidade de unidades de alimentação da atividade bovina "r", no período "t"
(UF/UP)
 f_{st} - necessidade de unidades de alimentação da atividade caprina/ovina "s", no período "t"
(UF/UP)
 g_{ijt} - quantidade de produto fornecido pelas atividades culturais "j", no terreno "i", no período "t" (Kg/Ha)
 g_{pt} - quantidade de produto fornecido pelas atividades de pastagens "p", no período "t"
(Kg/Ha)
 g_t - quantidade de produto fornecido pelo baldio, no período "t" (Kg/Ha) (SMAP)
 g_{bt} - quantidade de produto fornecido pelo baldio "b", no período "t" (Kg/Ha) (SM)
 $cUFA_j$ - coeficiente de conversão para unidades forrageiras da unidade de alimentação do produto resultante das atividades culturais "j" (UF/Kg)
 $cUFP_p$ - coeficiente de conversão para unidades forrageiras da unidade de alimentação do produto resultante das pastagens "p" (UF/Kg)
 $cUFB$ - coeficiente de conversão para unidades forrageiras da unidade de alimentação do produto resultante do baldio (UF/Kg)
 $cUFC_a$ - coeficiente de conversão para unidades forrageiras dos alimentos conservados "a"
(UF/Kg)
 $cUFAL_c$ - coeficiente de conversão para unidades forrageiras dos alimentos comprados "c"
(UF/Kg)
 h_{rt} - necessidade de unidades de alimentação da atividade bovinas "r", no período "t" (kg PDI/UP)
 h_{st} - necessidade de unidades de alimentação da atividade caprina/ovina "s", no período "t"
(kg PDI/UP)
 $cpDIA_j$ - coeficiente de conversão para PDI da unidade de alimentação do produto resultante das atividades culturais "j" (Kg PDI/Kg)
 $cpDIP_p$ - coeficiente de conversão para PDI da unidade de alimentação do produto resultante das pastagens "p" (Kg PDI/Kg)
 $cpDIB$ - coeficiente de conversão para PDI da unidade de alimentação do produto resultante do baldio (Kg PDI/Kg)
 $cpDICA_a$ - coeficiente de conversão para PDI dos alimentos conservados "a" (Kg PDI/Kg)
 $cpDIAL_c$ - coeficiente de conversão para PDI dos alimentos comprados "c" (Kg PDI/Kg)
 i_{rt} - capacidade máxima de ingestão de unidades de alimentação da atividade bovina "r", no período "t" (Kg MS/UP)
 i_{st} - capacidade máxima de ingestão de unidades de alimentação da atividade caprina/ovina "s", no período "t" (Kg MS/UP)

- cMSA_j - coeficiente de conversão para unidades de matéria seca da unidade de alimentação do produto resultante das atividades culturais “j” (Kg MS/Kg)
- cMSP_p - coeficiente de conversão para unidades de matéria seca da unidade de alimentação do produto resultante das pastagens “p” (Kg MS/Kg)
- cMSB - coeficiente de conversão para unidades de matéria seca da unidade de alimentação do produto resultante do baldio (Kg MS/Kg)
- cMSC_a - coeficiente de conversão para unidades de matéria seca dos alimentos conservados “a” (Kg MS/Kg)
- cMSAL_c - coeficiente de conversão para unidades de matéria seca dos alimentos comprados “c” (Kg MS/Kg)
- j_{rx} - quantidade de carne “x” fornecida pela atividade bovina “r” (Kg/UP)
- j_{sy} - quantidade de carne “y” fornecida pela atividade caprina/ovina “s” (Kg/UP)
- k_r - quantidade de leite fornecida pela atividade bovina “r” (Kg/UP)
- \overline{VL} - quantidade média de leite de vaca vendido nas explorações em estudo (Kg) (SMAP)
- k_s - quantidade de leite fornecida pela atividade caprina/ovina “s” (Kg/UP)
- ch_s - quantidade de queijo fornecida pela atividade caprina/ovina “s” (Kg/UP)
- \overline{VE} - quantidade média de estrume vendido nas explorações em estudo (Kg)
- CN_r - fator de conversão para CN da atividade bovina “r” (Unidade/UP)
- CN_s - fator de conversão para CN da atividade caprina/ovina “s” (Unidade/UP)
- l_{ijn} - necessidade de bens e serviços para as atividades culturais “j”, realizadas no terreno “i” (Unidade/Ha); n – tipo de bem e serviço
- l_{pn} - necessidade de bens e serviços “n” para as atividades de pastagens “p” (Unidade/Ha)
- l_{bn} - necessidade de bens e serviços “n” para as atividades de baldio “b” (Unidade/Ha) (SM)
- pF_n - preço unitário dos bens e serviços “n” (Euros/Unidade)
- pT_h - preço unitário da terra arrendada “h” (Euros/Ha) (SM)
- m_{rf} - necessidade de bens e serviços “f” para a atividade bovina “r” (Unidade/UP); f – tipo de bem e serviço
- m_{sf} - necessidade de bens e serviços “f” para a atividade caprina/ovina “s” (Unidade/UP)
- pF_f - preço unitário dos bens e serviços “f” (Euros/Unidade)
- pV_v - preço unitário dos produtos “v” fornecidos pelas atividades vegetais (Euros/Kg)
- pC_{rx} - preço unitário da carne “x” fornecida pela atividade bovina “r” (Euros/Kg)
- pC_{sy} - preço unitário da carne “y” fornecida pela atividade caprina/ovina “s” (Euros/Kg)
- pL - preço unitário do leite de vaca (Euros/Kg)
- pL_s - preço unitário do leite proveniente da atividade caprina/ovina “s” (Euros/Kg)
- pQ_s - preço unitário do queijo proveniente da atividade caprina/ovina “s” (Euros/Kg)
- pE - preço unitário do estrume (Euros/Kg)
- n_z - necessidade de bens e serviços com a tração mecânica própria (Unidade/Hora); z – tipo de bem e serviço
- pTP_z - preço unitário dos bens e serviços “z” (Euros/Unidade)
- pMO - preço unitário da mão-de-obra (Euros/H)
- pTM - preço unitário da tração mecânica (Euros/H)
- pFN - preço unitário do fertilizante N (Euros/Kg)
- pFP - preço unitário do fertilizante P₂O₅ (Euros/Kg)
- pFK - preço unitário do fertilizante K₂O (Euros/Kg)
- pAl_c - preço unitário dos alimentos comprados “c” (Euros/Kg)

sXA_{ij} - apoio monetário atribuído às atividades culturais “j”, realizadas no terreno “i”
(Euros/Ha)

sXP_p - apoio monetário atribuído às atividades de pastagens permanentes “p” (Euros/Ha)

sXV_r - apoio monetário atribuído à atividade bovina “r” (Euros/UP)

sXC_s - apoio monetário atribuído à atividade caprina/ovina (Euros/UP)

sTP_1 - apoio monetário atribuído ao combustível (gasóleo verde) (Euros/L)

cE_n - coeficiente de conversão para unidades energéticas dos bens e serviços “n”
(MJ/Unidade)

cE_f - coeficiente de conversão para unidades energéticas dos bens e serviços “f”
(MJ/Unidade)

cET_z - coeficiente de conversão para unidades energéticas dos bens e serviços “z”
(MJ/Unidade)

o_z - necessidade de bens e serviços “z” da tração mecânica alugada (Unidade/H)

$cEFN$ - coeficiente de conversão para unidades energéticas do fertilizante N (MJ/Kg)

$cEFP$ - coeficiente de conversão para unidades energéticas do fertilizante P_2O_5 (MJ/Kg)

$cEFK$ - coeficiente de conversão para unidades energéticas do fertilizante K_2O (MJ/Kg)

$cEAl_c$ - coeficiente de conversão para unidades energéticas dos alimentos comprados “c”
(MJ/Kg)

(Página deixada propositadamente em branco)

CASO 3

UM MODELO PARA A PRODUÇÃO DE LEITE NOS AÇORES

Emiliana Silva

Departamento de Ciências Agrárias da Universidade dos Açores
Centro de Estudos de Economia Aplicada do Atlântico
emiliana@uac.pt

Armando B. Mendes

Departamento de Matemática da Universidade dos Açores
Centro de Estudos de Economia Aplicada do Atlântico (CEEApIA)
amendes@uac.pt

RESUMO

Neste trabalho, utilizam-se técnicas de investigação operacional para modelar o sistema de produção de leite bovino, em regime intensivo, da Região Autónoma dos Açores. Este modelo foi desenhado para apoiar decisões estratégicas a nível político, tendo-se, no entanto, revelado igualmente útil para decisões de prazo mais curto e técnico.

Embora as decisões dos agricultores dos Açores sejam influenciadas por vários objetivos em conflito, como foi demonstrado por um trabalho prévio de Silva (2001), o modelo apresentado opta pelo objetivo clássico de maximização do rendimento agrícola. As restrições consideradas mais marcantes estão relacionadas com o modo de produção de leite nos Açores: área de pastagem, alimentação, trabalho agrícola e risco agrícola.

Obteve-se, deste modo, um modelo de programação linear com um objetivo e 55 restrições. Os dados económicos utilizados no modelo foram recolhidos junto da Secretaria Regional da Agricultura e Florestas da Região Autónoma dos Açores.

Os resultados obtidos para a rentabilidade da atividade agrícola na RAA mostram que se obtêm rentabilidades sustentáveis graças à mão-de-obra familiar e ao alimento verde (pastagem) produzido pelos terrenos agrícolas da RAA. Note-se, no entanto, que o rendimento agrícola apresentado não seria possível de alcançar se se verificasse uma redução dos subsídios atualmente disponibilizados pela União Europeia e pelo Governo Regional dos Açores.

PALAVRAS-CHAVE

Açores; Sistemas agrários; Leite; Suporte à decisão, Programação linear.

1. Modelos de suporte à decisão na Agricultura

A utilização de modelos de otimização é muito comum no sector primário. A título ilustrativo, apresentam-se alguns modelos ligados às práticas de produção animal, em regiões geográficas díspares. Mourits *et al.* (2000) utilizam um modelo de programação dinâmica para determinar a taxa ótima de reposição de um rebanho na Pensilvânia, Estados Unidos da América do Norte, incluindo algumas características relacionadas com a produtividade: idade, peso corporal, fase e taxa de crescimento “prepubertal” de gravidez e sazonalidade. A prática ótima obtida para este modelo, indica uma idade média dos animais de 20,5 meses e um peso corporal que ronda os 563 kg, em que os rendimentos por novilha foram mais altos nos animais nascidos em Dezembro do que nos nascidos em Maio.

Na Costa Rica, Herrero *et al.* (1999) usam um sistema de apoio à decisão tendo por base um modelo de programação matemática, para representar um sistema de produção leiteiro à base de pastagem em altitude. Consideram os aspetos biológicos, tais como crescimento da erva, pastoreio, digestão e metabolismo, e produtividade animal, que permitiram simular os comportamentos dos animais em diferentes regimes. Este trabalho discute os efeitos e implicações de vários cenários e políticas agrícolas na produtividade bioeconómica dos animais. Na produtividade bioeconómica, adicionam-se aos aspetos económicos os aspetos ecológicos: ambientais, biológicos ou orgânicos. A produtividade é obtida pela valorização dos valores económicos e ambientais em função do número dos animais.

Outro trabalho relevante é apresentado por Tozer e Stokes (2001) onde é apresentado um modelo para otimização da dieta alimentar de vacas leiteiras, de modo a minimizar os custos do alimento sem deixar de disponibilizar os nutrientes adequados às necessidades alimentares dos animais. Sendo assim, o modelo minimiza os custos mas também os desperdícios de fósforo e azoto.

Tendo em conta a importância das quotas leiteiras para a economia regional o trabalho de Ramsden *et al.* (1999) assume um interesse particular. Estes autores apresentam um modelo de programação linear para avaliar o impacto na produção leiteira da variação de preços devidos à existência do sistema de quotas leiteira que limitam a produção. O modelo incorpora as necessidades

em proteína e energia de acordo com o nível de produção das vacas leiteiras, e ainda uma taxa de substituição alimentar entre a forragem e os concentrados. Os resultados mostram que o sistema de quotas é pernicioso uma vez que se verifica um forte incentivo financeiro para diminuir o nível de *inputs* e transferir a produção leiteira para animais menos produtivos com menor uso de alimentos: produção de pastagem (adubos azotados) e rações (concentrados). No entanto, com a Política Agrícola Comum (PAC) atual e com os níveis de eficiência das explorações, era ainda rentável alimentar as vacas muito produtivas com grandes quantidades de rações.

Num trabalho relacionado com o anterior, Oliveira *et al.* (2010) concluem que o fim do sistema de quotas leiteiras poderá ter consequências muito negativas a curto prazo com a redução do preço do leite à produção, mas são igualmente previsíveis oportunidades a longo prazo, principalmente para os produtores dos Açores.

Um trabalho pioneiro de planeamento agrícola em Portugal foi o de Barros e Estácio (1972) apresentado em Ramalhete *et al.* (1984). Este modelo conta com onze atividades: sete vegetais, três atividades intermédias relativas à alimentação e uma atividade pecuária. Trata-se de um modelo de programação linear que maximizava, como objetivo, o resultado de uma exploração agrícola limitado pelos recursos disponíveis, tais como a área agrícola (quatro restrições), a mão-de-obra (quatro restrições) e o equilíbrio alimentar (quatro restrições).

Depois de um longo interregno surgiu o trabalho de Poeta (1994) aplicando a metodologia multicritério, nomeadamente o método NISE e a programação compromisso, a três tipos de explorações representativas de Trás-os-Montes. Para testar a aderência do modelo utilizou a programação por metas, considerando como metas os valores médios de cada atividade. Como objetivos, considera a maximização do valor acrescentado bruto, a minimização dos custos variáveis e a racionalização do trabalho, de tração animal e do rendimento da família. Como restrições apresenta a sazonalidade da mão-de-obra, a tração animal, o uso de terra, os adubos e fatores alimentares. Como atividades considera as culturas anuais e plurianuais, a atividade pecuária (de

carne e de leite), o aluguer de tração animal, a transferência e consumo de alimentos, e ainda, a aquisição de adubos.

No Alentejo, Carvalho (1999) avaliou o efeito da variabilidade da produção de pastagens e de forragens no rendimento dos agricultores e na capacidade de produção pecuária. O mesmo trabalho pretendeu ainda identificar ajustamentos a fazer na alimentação animal, através de alimentos conservados e de concentrados, e determinar épocas de vendas do produto animal/carne ajustadas àquela variabilidade. Para estimar o risco na variabilidade do rendimento dos agricultores opta pela utilização do MOTAD (*Minimization of Total Absolute Deviation*), pressupondo que os agricultores são adversos ao risco. O objetivo considerado foi o de maximização do lucro, incluindo ainda vinte e duas restrições relativas à utilização da terra, da utilização da tração mecânica, da mão-de-obra das atividades vegetais e animais, do equilíbrio alimentar (disponibilidades, necessidades e capacidade de ingestão), restrições relativas às alternativas de comercialização, aquisição e custo de concentrados e relativas às receitas das diferentes alternativas de comercialização.

Nos Açores, Silva (2001 e 2006) desenvolve modelos multicritério aplicados a três sistemas tipo de explorações pecuárias: sistema intensivo, extensivo e intermédio. Considera quatro objetivos em conflito: maximização da margem bruta das atividades e minimização do risco, da sazonalidade e das necessidades de mão-de-obra assalariada. Todos estes objetivos foram identificados no processo de tomada de decisão dos agricultores. O modelo inclui ainda setenta e duas restrições relativas à área, à rotação e intensificação do sistema, à mão-de-obra, variabilidade do risco e relacionadas com a alimentação animal; além de dezasseis variáveis não negativas. Uma das conclusões mais interessantes refere-se à utilização do tradicional objetivo de maximização do rendimento, o qual se verifica que nem sempre é o mais importante a condicionar a decisão dos agricultores açorianos, dependendo, no entanto, do sistema de produção considerado. Este objetivo surge com maior importância no sistema de produção com menor grau de intensificação (Silva, 2006).

Marta-Costa (2008) desenvolve um modelo para a região transmontana concentrando-se nas atividades pecuárias de produção de carne de bovino. O

modelo apresenta cento e vinte e nove variáveis relativas às atividades culturais em terras aráveis, em atividades vegetais, atividades de baldio, atividades pecuárias de raça maronesa e frísio pura ou cruzadas, atividades de raça bravia e noventa e oito restrições relativas à área utilizada, mão-de-obra, utilização de tração, utilização de fertilizantes, equilíbrio alimentar das atividades pecuárias; venda de produtos fornecidos pela atividade pecuária e relativas ao encabeçamento animal. Considera ainda dois objetivos em conflito: os económicos, medidos pela maximização do valor acrescentado bruto - VAB, e ambientais, estimados pela minimização dos custos energéticos.

2. A Região Autónoma dos Açores e a identificação do problema

O arquipélago dos Açores é uma região autónoma política e administrativa de Portugal, constituída após a revolução de 1974. Os Açores integram a União Europeia, desde 1986, por este facto, a política agrícola regional é, na sua maior parte, dirigida pelos regulamentos comunitários que nalguns casos são suavizados pelo estatuto de região ultraperiférica.

O Produto Interno Bruto (PIB) açoriano no ano de 2009 subiu para 3,7 mil milhões de euros (representando 2,8% do PIB nacional), registando um PIB *per capita* de 15,2 milhares de euros. Nesse mesmo ano, 8,6% do Valor Acrescentado Bruto (VAB) açoriano concentrou-se no sector “Agricultura, produção animal, caça e silvicultura”, contra os 2,2% que este sector representa a nível nacional (INE, 2011).

A Região Autónoma dos Açores (RAA) como região ultraperiférica da União Europeia (EU) apresenta algumas particularidades que são relevantes na descrição do sistema a modelar. Uma das principais características, a descontinuidade territorial, resulta de ser um arquipélago constituído por nove ilhas. A área da RAA corresponde aproximadamente a 2,5% do Estado Português, ou seja 2 322 km² (SREA, 2009) com uma Superfície agrícola útil (SAU) de 48,2%. A superfície agrícola útil em 2009 era, na sua maior parte (89,3%), usada para pastagem permanente e apenas 0,4% da área para hortas familiares. A área utilizada nas explorações de terra arável ocupava 8,4% da superfície agrícola útil, sendo a parte restante (1,9%) ocupada pelas culturas permanentes.

Outros fatores reconhecidos como limitantes são apontados pelo índice médio de envelhecimento nos Açores que, em 2007, foi de 65,1 sendo muito mais baixo do que o verificado no Continente Português (116,2) ou a média nacional (113,6) (SREA, 2010). Ainda que a RAA apresente menos problemas de rejuvenescimento e sustentabilidade dos sistemas produtivos, os valores apresentados são considerados muito elevados e com forte tendência para piorarem. Em 2007, apenas 3,1% e 1,7% da população açoriana e continental portuguesa, respetivamente, estavam empregadas no sector primário correspondente a 11 493 indivíduos, nos Açores. Sabe-se ainda que a mão-de-obra nas explorações agrícolas é predominantemente familiar, constituindo este outro fator limitante às explorações agrícolas açorianas.

Na atualidade, a produção leiteira açoriana contribui com cerca de 30% da produção portuguesa, sendo a área de atividade que mais contribui para o PIB regional e a mais relevante na sustentabilidade económica das ilhas. Apesar dos limites impostos pelo regime de quotas a produção de leite de bovino tem vindo a aumentar ligeiramente ao longo dos anos. O leite de vaca entregue nas fábricas dos Açores, em 2003, era de 492 211 174 litros. Esta produção de leite tem aumentado gradualmente ao longo dos últimos anos, alcançando o valor de 540 199 225 litros. Este valor da produção de leite é ligeiramente superior ao montante da quota de leite atribuída à Região Autónoma dos Açores, 500 milhões de litros. O leite é igualmente transformado pela indústria regional. O queijo produzido nos Açores é principalmente de pasta semidura (97,3%). Em 2006, a produção de manteiga cifrou-se em 7 489 toneladas e a produção de iogurtes 310 toneladas. A produção de natas rondou os 623 525 litros.

Em 2007, existiam 13 155 explorações agrícolas nos Açores, com uma superfície agrícola útil (SAU) de 112 054 hectares *i.e.*, 4,8% do total de explorações agrícolas do país. Além destes fatores de pequenez é ainda muito importante realçar que enquanto a maior parte das explorações em Portugal (50,9%) tem uma dimensão entre 1 e 5 hectares, nos Açores, 43,8% das explorações tem área inferior a 1 hectare. A pequena dimensão das explorações agrícolas açorianas é outro fator limitante à produção agrícola. A superfície agrícola por Unidade de Trabalho Agrícola (UTA) em Portugal toma um valor de 10,1 o que é aproximadamente o dobro da verificada na Região

Autónoma dos Açores (4,9). No entanto, o rendimento por hectare nos Açores (1 306 €) é mais do dobro do obtido em Portugal. Este facto pode ser explicado pelo maior encabeçamento médio de 1,67 animais por hectare nos Açores, mais do dobro do número de animais por hectare no Continente (0,58).

Dada a relevância e impacto do sector leiteiro na economia regional, a utilização de modelos de otimização que permitam criar conhecimento sobre o sistema, atores envolvidos, relações entre variáveis e concluir sobre políticas ótimas de funcionamento do sistema são de primordial importância no apoio a decisões.

Na secção seguinte é descrito o processo que levou ao desenho de um modelo de decisão para as atividades ligadas à produção de leite na região autónoma dos Açores. Apresenta-se em seguida os resultados obtidos do modelo e da discussão com os decisores e técnicos que se seguiram. Por fim, apresentam-se conclusões e perspectivas para trabalhos futuros.

3. Modelo de decisão

Nesta secção descreve-se o modelo de decisão utilizado para modelar as atividades envolvidas no sector leiteiro da RAA. Além de descrever o sistema, considera-se que o processo que levou ao desenvolvimento do modelo é igualmente digno de nota, assim como os pressupostos e aproximações adotadas.

3.1. Processo e pressupostos

A relevância do sector leiteiro açoriano no contexto nacional e a consequente importância na economia açoriana leva a um problema de tomada de decisão técnica e política fundamental, que será mais consistente e adaptada à realidade se for feita tendo por base estudos científicos. A importância deste estudo foi reconhecida pela Secretaria Regional da Agricultura e Florestas da Região Autónoma dos Açores, gestora do sector agrícola nos Açores, que disponibilizou dados confidenciais e que permitiram a realização deste modelo.

Ainda que se tenha optado por modelos normativos, a aproximação utilizada no desenvolvimento do modelo é prescritiva, uma vez que foram

utilizados dados e conhecimento tácito para induzir as normas incluídas no modelo.

Os dados recolhidos junto da Rede de Informação de Contabilidade Agrícola para os Açores (RICA-A), da Secretaria Regional da Agricultura e Florestas, permitiram o cálculo da margem bruta para o ano em estudo (1996) e estimar a variabilidade do rendimento agrícola (1990 a 1996). Os coeficientes técnicos relativos à alimentação do gado, unidades forrageiras de leite, proteína digestível animal (energia e nitrogénio limitantes), o cálcio, fósforo e matéria seca disponível foram calculados com base em valores do INRA (1988). Os dados relativos à racionalização da mão-de-obra foram retirados do trabalho de Avillez (1991) e de Berbel e Barros (1993) e confirmados recorrendo à elaboração de inquéritos realizados posteriormente e no âmbito deste trabalho. O cálculo de valores dos coeficientes técnicos utilizados no modelo de decisão para os agricultores açorianos está disponível em Silva (2001). Os dados encontram-se disponíveis para consulta.

Deste processo e do tratamento dos dados anteriores resultou um conjunto de pressupostos que se passa a expor.

Os parâmetros do modelo refletem apenas o regime intensivo, correspondente a encabeçamentos por hectare superiores a 2,4, uma vez que é o sistema que mais penaliza o rendimento dos agricultores açorianos nas medidas agroambientais da Política Agrícola Comum e revelou-se especialmente relevante para os decisores. Note-se, no entanto, que foram identificados por Silva e Berbel (2007) mais dois sistemas de produção leiteira nos Açores, sistema de produção extensivo (inferior a 1,4 cabeças normais por hectare) e semi-intensivo, com encabeçamentos entre os 1,4 e 2,4 cabeças normais por hectare.

De trabalhos anteriores de Silva (2001 e 2006) concluiu-se que as decisões dos agricultores açorianos relativamente ao encabeçamento, tipo de alimento e outras práticas agrícolas são tomadas com base não num objetivo mas em vários objetivos conflituosos, tais como a maximização do rendimento, a minimização da sazonalidade da mão-de-obra e minimização do risco agrícola. No entanto, optou-se por um modelo uniobjetivo tendo em conta a opinião dos decisores políticos e ainda por se considerar que a utilização do objetivo

mais racional do ponto de vista económico (segundo a teoria neoclássica) daria indicações mais precisas de como deveria funcionar o sistema, tal como considera Tauer (1995).

O descritor de impacto utilizado inclui os subsídios à produção de leite. A pertinência de se utilizar os subsídios deve-se à importância que os mesmos representam para a agricultura açoriana, cerca de 20% dos rendimentos das explorações pecuárias dos Açores são devidas aos subsídios (Silva, 2012).

As restantes medidas de desempenho foram introduzidas no modelo como restrições. Além dessas, foram ainda consideradas um elevado número de restrições correspondentes a condicionalismos das atividades. A definição dessas restrições resultou de um processo de consulta que envolveu decisores, dirigentes de cooperativas agrícolas e técnicos.

No caso do risco agrícola, e dada a verificada aversão ao risco dos agricultores, recorreu-se ao rendimento da produção leiteira ao longo de vários anos, ajudando a quantificar a influência que as condições edafoclimáticas têm num sistema de produção à base de pastagem.

No que concerne às restrições de área, e sabendo-se que a maior parte da produção de leite é à base de pastagem, são consideradas diferentes altitudes e com diferente manejo da pastagem (Matos, 1997). Esta diferenciação foi considerada muito relevante para o nível de decisão que se pretendia apoiar.

Dada a importância da alimentação na produção de leite, introduziram-se ainda várias restrições de modo a avaliar se as disponibilidades do alimento em verde (pastagem, silagem e feno) complementada com alimento a seco (ração) são suficientes para superar as necessidades dos animais. Considerou-se ainda que os animais das explorações eram apenas para produção leite. Estes animais têm um peso vivo médio de 650 Kg e produzem 25 litros de leite por dia em média. Considerou-se que as vacas estão em lactação 305 dias por ano, e que estão “secas” os restantes 60 dias (no oitavo e nono mês de gestação). Também se considerou que a cada vaca estava associada a 20% de novilhas de substituição e a 50% de vitelas de substituição. As vacas são inseminadas com um ano para parirem com menos de dois anos.

A mão-de-obra agrícola nos Açores é considerada apenas familiar, composta pelo produtor agrícola (agricultor) e restantes membros da família com

atividade na exploração familiar e não implica um pagamento efetivo. No entanto, a sua disponibilidade e necessidade é diferente ao longo do ano, com maiores necessidades na Primavera e Outono, que coincide com a data de parição dos rebanhos leiteiros nos Açores. Assim, foi necessário considerar esta sazonalidade dividindo o ano em seis períodos, cada um formado por dois meses acompanhando o ciclo de produção de leite na região Açores.

Como se pretendia dar maior importância à alimentação dos animais, devido ao seu custo, decidiu-se dividir as atividades em animais e vegetais: pastagens, silagens, fenos. Como esta informação não estava disponível na base de dados utilizada, optou-se por ajustá-la à base de dados de Berbel e Barros (1993) usando o conhecimento de investigadores e técnicos disponíveis.

Note-se que se opta, no modelo a que se chega, por modelar o risco agrícola e a mão-de-obra utilizando restrições que têm uma estrutura de metas. Tal como na programação por metas, as restrições usadas neste modelo, permitem que os valores variem para cima ou para baixo do nível de aspiração na estrutura da meta, usando variáveis auxiliares de desvio positivas e negativas, o que permite que se ultrapassem ou não os limites previamente estipulados. No entanto, as variáveis auxiliares não são minimizadas como na programação por metas, o que faz com que as restrições funcionem de forma informativa. Assim, é possível considerar a possibilidade, ou não, de mão-de-obra agrícola familiar ser superior à existente correspondendo a um maior esforço ou não do trabalho familiar. Estes tipos de restrições são descritas e utilizadas, por exemplo, em Romero *et al.* (1989).

3.2. O modelo analítico

O modelo pretende otimizar os valores de um conjunto de variáveis decisionais relacionadas com as atividades agrícolas consideradas relevantes. Os tipos de variáveis e níveis considerados são resumidos no Quadro 1.

Quadro 1: Atividades, tipos e níveis utilizados no modelo.

ATIVIDADES	TIPO	NÍVEIS
Pastagem	Altitude Tipo de Aproveitamento	Alta, Média, Baixa, Direto, Silagem, Feno
Milho Ensilado	Semente Altitude	Híbrida Média, Baixa
“Outonos” (Avena sativa e Vicia fava)	Semente Altitude	Híbrida, Regional Média, Baixa
Leite	Dimensão do Efetivo Ração	Grande, Médio, Pequeno Quantitativo

Da combinação do tipo de atividades com os diferentes níveis do Quadro 1 resultam as seguintes quinze variáveis decisoriais:

- X_1 – Superfície de pastagem alta com pastoreio direto (em ha)
- X_2 – Superfície de pastagem média com pastoreio direto (em ha)
- X_3 – Superfície de pastagem média com pastoreio direto e silagem de erva (em ha)
- X_4 – Superfície de pastagem média com pastoreio direto e feno (em ha)
- X_5 – Superfície de pastagem baixa com pastoreio direto (em ha)
- X_6 – Superfície de pastagem baixa com pastoreio direto e silagem de erva (em ha)
- X_7 – Superfície de pastagem baixa com pastoreio direto e feno (em ha)
- X_8 – Superfície de milho silagem de média altitude (em ha)
- X_9 – Superfície de milho silagem de baixa altitude (em ha)
- X_{10} – Superfície de “Outono” (*Avena sativa* e *Vicia fava*) regional de média altitude (em ha)
- X_{11} – Superfície de “Outono” regional de baixa altitude (em ha)
- X_{12} – Superfície de “Outono” melhorado (*Lolium perene*) de média altitude (em ha)
- X_{13} – Superfície de “Outono” melhorado de baixa altitude (em ha)
- X_{14} – Quantidade em quilogramas de ração (soja e milho) (em Kg)
- X_{15} – Número de animais (produtores de leite) (em n°).

A medida de desempenho escolhida é o rendimento proveniente da atividade leiteira. Como descritor de impacto optou-se pela margem bruta da exploração (MB). A margem bruta da exploração é obtida pelo somatório das margens brutas das várias atividades (neste caso o leite) acrescidas dos prémios e subsídios, produção de imobilizado, prestação de serviços e percas e ganhos da produção. Por sua vez, a margem bruta das atividades é calculada pela diferença entre a receita de produção bruta do leite e os encargos variáveis e fixos específicos referentes à atividade leiteira. Como receitas de produção total das explorações leiteiras consideraram-se a venda de leite, a venda de carne das vacas de refugio e os subsídios à atividade leiteira. Nos encargos variáveis incluíram-se as rações, as pastagens, silagens e fenos produzidos na exploração e os gastos da pecuária, tais como: medicamentos, veterinário, detergentes, entre outros. No caso das explorações amostradas e para o cálculo das margens brutas, apenas existiam prémios e subsídios, não havendo portanto, a produção do imobilizado, a prestação de serviços e os ganhos ou percas extraordinárias.

Os coeficientes - margens brutas - da função objetivo, foram considerados positivos, no caso de se ter vendido os produtos agrícolas ao exterior e negativos no caso de terem sido utilizados na própria exploração ou adquiridos ao exterior. No caso das atividades vegetais (pastagem, milho e “outonos”), uma vez que estes não são comercializados, valoriza-se a margem bruta pelo valor dos encargos variáveis utilizados em cada uma das atividades vegetais: sementes, adubos, aluguer de máquinas, entre outros específicos das atividades vegetais.

Com o objetivo de maximizar a margem bruta, obteve-se finalmente a expressão analítica (1) em euro.

$$\begin{aligned} \text{Max MB} = & -17,92 X_1 - 50,98 X_2 - 68,385X_3 - 65,986X_4 - 53,9049X_5 - \\ & - 69,582X_6 - 70,181X_7 - 55,486X_8 - 55,486X_9 - 10,794X_{10} - 10,794X_{11} - 35,09X_{12} - \\ & - 35,09X_{13} - 0,23693X_{14} + 1166,189 X_{15} \quad (\text{€}) \end{aligned} \quad (1)$$

A área total de pastagem da exploração agrícola está dividida em alta, superior a 800 metros, média, de 300 a 800 metros, e baixa, inferior a 300

metros. Os valores utilizados foram adaptados do DREPA (2010) e são: 49,8% para pastagem de baixa altitude, 45% para pastagem de média altitude e 5,2% para pastagem de alta altitude.

Considerou-se que as áreas das culturas existentes nas superfícies: alta, média e baixa terão de ser inferiores ou iguais, às áreas disponíveis em alta (S_A), média (S_M) e baixa altitude (S_B), em hectares, respetivamente. Tal resultou nas três restrições da equação (2)

$$\begin{aligned} X_1 &\leq S_A \text{ (ha)} \\ X_2 + X_3 + X_4 + 1/2X_8 + 1/4 (X_{10} + X_{12}) &\leq S_M \text{ (ha)} \\ X_5 + X_6 + X_7 + 1/2X_9 + 1/4 (X_{11} + X_{13}) &\leq S_B \text{ (ha)} \end{aligned} \quad (2)$$

Finalmente a área ocupada por todas as atividades terá de ser inferior à área total da exploração (S_T), obtendo-se a expressão analítica (3).

$$X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + 1/4 (X_8 + X_9) + 1/8 (X_{10} + X_{11} + X_{12} + X_{13}) \leq S_T \text{ (ha)}. \quad (3)$$

Na sequência da restrição anterior as superfícies de alta, média e de baixa altitude teriam que ser igualmente inferiores à área total da exploração (S_T) (em hectares).

$$S_A + S_M + S_B \leq S_T \text{ (ha)} \quad (4)$$

A restrição seguinte refere-se à renovação da pastagem de média e baixa altitude (em hectares). Considera-se que esta se efetua em cinco anos e que todos os anos se renovam 20% da área total da exploração. A renovação da pastagem é feita à base de milho para silagem e de “outonos” regionais e melhorados.

$$X_8 + X_9 \leq 0,2 (S_M + S_B) \text{ (ha)} \quad (5)$$

A renovação da pastagem faz-se com milho, a que sucede o “Outono” regional ou melhorado. A área de “Outono” tem que ser sempre igual à do milho que só se faz em altitude média ou baixa. O milho fica no solo cerca de 5 a 6 meses e os “Outonos” a restante parte do ano. Deste modo obtêm-se mais duas restrições indicando que as áreas cultivadas (em hectares) pelos “Outono” regional e melhorado têm de ser iguais ou superiores às áreas do milho, uma vez que o milho precede os “Outonos”.

$$X_{10} + X_{12} \leq X_8 \text{ (ha)}$$

$$X_{11} + X_{13} \leq X_9 \text{ (ha)} \tag{6}$$

Quanto à sazonalidade da Mão-de-Obra disponível por Período (MOP) consideraram-se os períodos: I - Janeiro e Fevereiro; II – Março e Abril; III – Maio e Junho; IV – Julho e Agosto; V – Setembro e Outubro; e VI – Novembro e Dezembro. Neste conjunto de restrições incluiu-se a possibilidade de utilizar mão-de-obra (horas/período/hectare) a mais (p) ou a menos (n) do que a disponível, considerando como nível de aspiração a quantidade de mão-de-obra média disponível para cada um dos seis períodos. Por exemplo, teríamos uma disponibilidade por período de 600 horas, resultante do trabalho anual de 1 trabalhador (2 400 horas) e família (1 200 horas) distribuído pelos 6 períodos, a que se acrescenta um aumento de esforço por período de 60 horas. Obtém-se, assim, um tempo (T) de 660 horas, que pode ser ou não totalmente utilizado. As 60 horas correspondem a um incremento do esforço de 10% da mão-de-obra por dia, que corresponde a cerca de 12 minutos diários. O tempo de trabalho utilizado na distribuição da ração foi incluído no tempo despendido nas vacas, por ser difícil atribuir um tempo a esta operação.

$$\text{MOPI} \rightarrow 0,72X_2 + 0,63X_3 + X_4 + 0,72X_5 + 0,63X_6 + X_7 + 14,13X_{15} + n_{eI} - p_{eI} = 660 \text{ (horas/ano)}$$

$$\text{MOPII} \rightarrow 4,09X_1 + 3,21X_2 + 3,14X_3 + 3,88X_4 + 2,21X_5 + 3,14X_6 + 3,88X_7 + 2,53X_{12} + 2,53X_{13} + 14,13X_{15} + n_{eII} - p_{eII} = 660 \text{ (horas/ano)}$$

$$\text{MOPIII} \rightarrow 2,49X_2 + 6,5X_3 + 6,62X_4 + 2,49X_5 + 6,83X_6 + 6,62X_7 + 2,5X_8 + 2,5X_9 + 14,13X_{15} + n_{eIII} - p_{eIII} = 660 \text{ (horas/ano)}$$

MOPIV --> $3,99 X_3 + 2,74X_4 + 3,99X_6 + 2,74X_7 + 14,13X_{15} + n_{eIV} - p_{eIV} = 660$
(horas/ano)

MOPV --> $1,5X_1 + 2,49X_2 + 2,51X_3 + 5,38X_4 + 2,49X_5 + 2,51X_6 + 5,38X_7 + 28,06X_8 + 28,06X_9 + 7,91X_{10} + 7,91X_{11} + 5,95X_{12} + 5,95X_{13} + 14,13X_{15} + n_{eV} - p_{eV} = 660$ (horas/ano)

MOPVI --> $14,13X_{14} + n_{eVI} - p_{eVI} = 660$ (horas/ano) (7)

Para se medir o risco associado à atividade agrícola, utilizou-se o método MOTAD (*Minimization of Total Absolute Deviation*), definido por Hazell e descrito em Carvalho (1999). Esta medida engloba um conjunto de sete restrições que se referem à variabilidade de rendimento (desvio da margem bruta da atividade – MB – em relação à média da margem bruta da atividade), em que N_j é o desvio negativo em relação à média para o ano j , e o P_j é o desvio positivo. O período utilizado na medição do risco foi de 1990 a 1996 e foi elaborado a partir dos dados da RICA-A a preços constantes de 1990.

$$MOTAD = \sum_{i=1}^{15} MB_{ij} X_i - \overline{MBX} + N_j - P_j = 0 \text{ para } j = 1...7$$

MOTAD 96 --> $-0,005 X_1 - 0,014X_2 - 0,019X_3 - 0,018X_4 - 0,015X_5 - 0,019X_6 - 0,019X_7 - 0,015X_8 - 0,015X_9 - 0,003X_{10} - 0,003X_{11} - 0,010X_{12} - 0,0110X_{13} + 0,001 X_{14} - 28,086X_{15} + N_1 - P_1 = 0$

MOTAD 95 --> $0,414 X_1 + 1,174X_2 + 1,575X_3 + 1,520X_4 + 1,243X_5 - 1,603X_6 + 1,616X_7 + 1,278X_8 + 1,278X_9 + 0,249X_{10} + 0,249X_{11} - 0,808X_{12} + 0,808X_{13} + 0,002X_{14} + 19,614X_{15} + N_2 - P_2 = 0$

MOTAD 94 --> $-0,192 X_1 - 0,545X_2 - 0,731X_3 - 0,705X_4 - 0,577X_5 - 0,744X_6 - 0,750X_7 - 0,593X_8 - 0,593X_9 - 0,115X_{10} - 0,115X_{11} - 0,375X_{12} - 0,375X_{13} + 0,002 X_{14} - 11,814X_{15} + N_3 - P_3 = 0$

MOTAD 93 --> $0,183 X_1 + 0,518X_2 + 0,695X_3 + 0,671X_4 + 0,549X_5 - 0,708X_6 + 0,714X_7 + 0,564X_8 - 0,564X_9 + 0,110X_{10} + 0,110X_{11} + 0,357X_{12} + 0,357X_{13} - 0,002X_{14} + 10,114X_{15} + N_4 - P_4 = 0$

$$\text{MOTAD 92} \rightarrow 0,247 X_1 + 0,699X_2 + 0,938X_3 + 0,905X_4 + 0,741X_5 + 0,955X_6 + 0,963X_7 + 0,761X_8 + 0,761X_9 + 0,148X_{10} + 0,148X_{11} + 0,481X_{12} + 0,481X_{13} + 0,003X_{14} + 29,314X_{15} + N_5 - P_5 = 0$$

$$\text{MOTAD 91} \rightarrow -0,086 X_1 - 0,242X_2 - 0,325X_3 - 0,314X_4 - 0,257X_5 - 0,331X_6 - 0,334X_7 - 0,264X_8 - 0,264X_9 - 0,051X_{10} - 0,051X_{11} - 0,167X_{12} - 0,167X_{13} - 0,003 X_{14} - 17,086X_{15} + N_6 - P_6 = 0$$

$$\text{MOTAD 90} \rightarrow -0,562 X_1 - 1,591X_2 - 2,134X_3 - 2,059X_4 - 1,685X_5 - 2,171X_6 - 2,190X_7 - 1,731X_8 - 1,731X_9 - 0,337X_{10} - 0,337X_{11} - 1,095X_{12} - 1,095X_{13} - 0,004X_{14} - 25,686X_{15} + N_7 - P_7 = 0 \quad (8)$$

Foram efetuados testes em que ao minimizar o risco agrícola se obtinha uma solução trivial. Assim, houve a necessidade de atribuir um rendimento mínimo anual, que neste caso foi de 15 497,65 €. Este valor é cerca do dobro do salário mínimo nacional em vigor, o que foi considerado um valor aceitável para remunerar a mão-de-obra familiar e ainda atribuir uma remuneração adicional aos fatores de produção próprios (terra própria que poderia ter um custo de oportunidade no mercado).

No que se refere às necessidades alimentares dos animais da exploração obtiveram-se seis restrições para cada conjunto. Considerou-se que as necessidades de energia (UFL), gramas de proteína (PDIE e PDIN), cálcio (CA) e fósforo (P) disponíveis nos alimentos têm de ser suficientes para cobrir as necessidades dos animais, em cada ano, nestes mesmos nutrientes (unidades: grama/nutriente/animal/ano). Os coeficientes técnicos relativos à alimentação animal tiveram em conta os pressupostos descritos na subsecção anterior e em dados do INRA (1988).

Os componentes dos alimentos que se utilizaram foram:

1. As unidades forrageiras de leite (UFL):

$$\text{UFLPI} \rightarrow 372 X_2 + 439 X_3 + 366 X_4 + 855X_5 + 951 X_6 + 790X_7 + 2 338 X_{10} + 2 338X_{11} + 3 465X_{12} + 3 465X_{13} + 187,5X_{14} \geq 953X_{15}$$

$$\text{UFLPII} \rightarrow 1 232 X_1 + 2 470 X_2 + 2 914 X_3 + 2 434 X_4 + 3 059X_5 + 3 407 X_6 + 2 827X_7 + 2 337 X_{10} + 2 337X_{11} + 3 465X_{12} + 3 465X_{13} + 187,5X_{14} \geq 953X_{15}$$

$$\text{UFLPIII} \rightarrow 1\,232 X_1 + 5\,785 X_2 + 6\,823 X_3 + 5\,701 X_4 + 4\,527 X_5 + 5\,042 X_6 + 4\,183 X_7 + 187,5 X_{14} \geq 953 X_{15}$$

$$\text{UFLPIV} \rightarrow 1\,232 X_1 + 2\,178 X_2 + 2\,569 X_3 + 2\,147 X_4 + 377 X_5 + 5\,042 X_6 + 4\,183 X_7 + 187,5 X_{14} \geq 953 X_{15}$$

$$\text{UFLPV} \rightarrow 462 X_1 + 131 X_2 + 1\,556 X_3 + 1\,302 X_4 + 2\,345 X_5 + 2\,613 X_6 + 2\,167 X_7 + 4\,050 X_8 + 4\,050 X_9 + 187,5 X_{14} \geq 953 X_{15}$$

$$\text{UFLPVI} \rightarrow 462 X_1 + 357 X_2 + 419 X_3 + 350 X_4 + 957 X_5 + 1\,067 X_6 + 885 X_7 + 4\,050 X_8 + 4\,050 X_9 + 187,5 X_{14} \geq 953 X_{15} \quad (9)$$

2. As proteínas digestivas no intestino com energia e azoto limitante (PDIE e PDIN):

$$\text{PDIEI} \rightarrow 37\,948 X_2 + 31\,005 X_3 + 39\,783 X_4 + 87\,984 X_5 + 66\,591 X_6 + 83\,895 X_7 + 223\,750 X_{10} + 223\,750 X_{11} + 353\,500 X_{12} + 353\,500 X_{13} + 15\,450 X_{14} \geq 99\,298,6 X_{15} \text{ (g/ano)}$$

$$\text{PDIEII} \rightarrow 123\,200 X_1 + 251\,750 X_2 + 205\,855 X_3 + 264\,241 X_4 + 315\,016 X_5 + 238\,518 X_6 + 300\,390 X_7 + 223\,750 X_{10} + 223\,750 X_{11} + 353\,500 X_{12} + 353\,500 X_{13} + 15\,450 X_{14} \geq 99\,298,6 X_{15} \text{ (g/ano)}$$

$$\text{PDIEIII} \rightarrow 123\,200 X_1 + 589\,572 X_2 + 482\,040 X_3 + 618\,728 X_4 + 466\,128 X_5 + 352\,926 X_6 + 444\,465 X_7 + 15\,450 X_{14} \geq 99\,298,6 X_{15} \text{ (g/ano)}$$

$$\text{PDIEIV} \rightarrow 123\,200 X_1 + 221\,964 X_2 + 181\,480 X_3 + 233\,002 X_4 + 38\,792 X_5 + 29\,421 X_6 + 36\,975 X_7 + 15\,450 X_{14} \geq 99\,298,6 X_{15} \text{ (g/ano)}$$

$$\text{PDIEV} \rightarrow 46\,200 X_1 + 134\,408 X_2 + 109\,915 X_3 + 141\,421 X_4 + 241\,488 X_5 + 182\,889 X_6 + 230\,265 X_7 + 288\,000 X_8 + 288\,000 X_9 + 15\,450 X_{14} \geq 99\,298,6 X_{15} \text{ (g/ano)}$$

$$\text{PDIEVI} \rightarrow 46\,200 X_1 + 36\,358 X_2 + 29\,640 X_3 + 38\,092 X_4 + 98\,592 X_5 + 74\,655 X_6 + 94\,010 X_7 + 288\,000 X_8 + 288\,000 X_9 + 15\,450 X_{14} \geq 99\,298,6 X_{15} \text{ (g/ano)}$$

$$\text{PDIN I} \rightarrow 47.256 X_2 + 40.545 X_3 + 39.336 X_4 + 114.210 X_5 + 82.446 X_6 + 77.973 X_7 + 230.000 X_{10} + 230.000 X_{11} + 353.500 X_{12} + 353.500 X_{13} + 24.450 X_{14} \geq 99.298,6 X_{15} \text{ (g/ano)}$$

$$\text{PDIN II} \rightarrow 118.400 X_1 + 313.896 X_2 + 269.195 X_3 + 261.272 X_4 + 408.915 X_5 + 295.308 X_6 + 279.186 X_7 + 230.000 X_{10} + 230.000 X_{11} + 416.500 X_{12} + 416.500 X_{13} + 24.450 X_{14} \geq 99.298,6 X_{15} \text{ (g/ano)}$$

$$\text{PDIN III} \rightarrow 118.400X_1 + 734.184X_2 + 630.360X_3 + 611.776X_4 + 605.070X_5 + 436.956X_6 + 413.091X_7 + 24.450X_{14} \geq 99.298,6X_{15} \text{ (g/ano)}$$

$$\text{PDIN IV} \rightarrow 118.400X_1 + 276.408X_2 + 237.320X_3 + 230.384X_4 + 50.355X_5 + 36.426X_6 + 34.365X_7 + 24.450X_{14} \geq 99.298,6X_{15} \text{ (g/ano)}$$

$$\text{PDIN V} \rightarrow 44.400X_1 + 167.376X_2 + 143.735X_3 + 139.832X_4 + 313.470X_5 + 226.434X_6 + 214.011X_7 + 238.500X_8 + 238.500X_9 + 24.450X_{14} \geq 99.298,6X_{15} \text{ (g/ano)}$$

$$\text{PDIN VI} \rightarrow 44.400X_1 + 45.276X_2 + 38.760X_3 + 37.664X_4 + 127.980X_5 + 92.430X_6 + 87.374X_7 + 238.500X_8 + 238.500X_9 + 24.450X_{14} \geq 99.298,6X_{15} \text{ (g/ano).}$$

(10)

3. O cálcio (CA):

$$\text{CAI} \rightarrow 2\ 900X_2 + 3\ 600X_3 + 3\ 400X_4 + 6\ 400X_5 + 6\ 900X_6 + 6\ 400X_7 + 22\ 500X_{10} + 22\ 500X_{11} + 22\ 800X_{12} + 22\ 800X_{13} + 500X_{14} \geq 7\ 547X_{15}$$

$$\text{CAII} \rightarrow 12\ 000X_1 + 19\ 000X_2 + 23\ 800X_3 + 22\ 300X_4 + 22\ 700X_5 + 24\ 600X_6 + 23\ 000X_7 + 22\ 500X_{10} + 22\ 500X_{11} + 22\ 700X_{12} + 22\ 700X_{13} + 500X_{14} \geq 7547X_{15}$$

$$\text{CAIII} \rightarrow 12\ 000X_1 + 44\ 500X_2 + 55\ 600X_3 + 52\ 100X_4 + 33\ 600X_5 + 36\ 400X_6 + 34\ 000X_7 + 500X_{14} \geq 7547X_{15}$$

$$\text{CAIV} \rightarrow 12\ 000X_1 + 16\ 800X_2 + 20\ 900X_3 + 19\ 600X_4 + 2\ 800X_5 + 3\ 000X_6 + 2\ 800X_7 + 500X_{14} \geq 7\ 547X_{15}$$

$$\text{CAV} \rightarrow 4\ 500X_1 + 10\ 100X_2 + 12\ 700X_3 + 11\ 900X_4 + 17\ 400X_5 + 18\ 900X_6 + 17\ 600X_7 + 13\ 500X_8 + 13\ 500X_9 + 500X_{14} \geq 7\ 547X_{15}$$

$$\text{CAVI} \rightarrow 4\ 500X_1 + 2\ 700X_2 + 3\ 400X_3 + 3\ 200X_4 + 7\ 100X_5 + 7\ 700X_6 + 7\ 200X_7 + 13\ 500X_8 + 13\ 500X_9 + 500X_{14} \geq 7\ 547X_{15}$$

(11)

4. O fósforo (P):

$$\text{PI} \rightarrow 1\ 800X_2 + 1\ 900X_3 + 1\ 800X_4 + 3\ 800X_5 + 3\ 700X_6 + 6\ 400X_7 + 10\ 000X_{10} + 10\ 000X_{11} + 17\ 500X_{12} + 17\ 500X_{13} + 470X_{14} \geq 4\ 271,5X_{15}$$

$$\text{PII} \rightarrow 4\ 800X_1 + 11\ 900X_2 + 12\ 700X_3 + 11\ 800X_4 + 13\ 600X_5 + 13\ 300X_6 + 23\ 000X_7 + 10\ 000X_{10} + 10\ 000X_{11} + 17\ 500X_{12} + 17\ 500X_{13} + 470X_{14} \geq 4\ 271,5X_{15}$$

$$\text{PIII} \rightarrow 4\ 800X_1 + 27\ 800X_2 + 29\ 600X_3 + 27\ 800X_4 + 20\ 200X_5 + 19\ 600X_6 + 34\ 000X_7 + 470X_{14} \geq 4\ 271,5X_{15}$$

$$\text{PIV} \rightarrow 4\ 800X_1 + 10\ 500X_2 + 11\ 200X_3 + 10\ 500X_4 + 1\ 700X_5 + 1\ 600X_6 + 2\ 800X_7 + 470X_{14} \geq 4\ 271,5X_{15}$$

$$\text{PV} \rightarrow 1800X_1 + 6300X_2 + 6800X_3 + 6400X_4 + 10400X_5 + 10200X_6 + 17600X_7 + 9000X_8 + 9000X_9 + 470X_{14} \geq 4\ 271,5X_{15}$$

$$\text{PVI} \rightarrow 1800X_1 + 1700X_2 + 1800X_3 + 1700X_4 + 4300X_5 + 4100X_6 + 7200X_7 + 9000X_8 + 9000X_9 + 470X_{14} \geq 4\ 271,5X_{15} \quad (12)$$

Finalmente há que considerar a ingestão de matéria seca (MS), em grama de matéria seca/ano, pelos animais. Assim, a matéria seca proporcionada pelos alimentos, em verde e ração tem de ser inferior à capacidade de ingestão dos animais.

$$- 6000 X_1 - 12000 X_2 - 16000X_3 - 15000X_4 - 12000X_5 - 15000X_6 - 14000X_7 - 9000X_8 - 9000X_9 - 5000X_{10} - 5000X_{11} - 7000X_{12} - 7000X_{13} - 900X_{14} < 7466 \quad (\text{g/ano}) \quad (13)$$

Também se considerou que os animais comiam 1,5 Kg de ração por dia, o que ao ano corresponde a 547,7 quilogramas de ração (Kg de ração/ano/animal):

$$X_{14} - 547,7X_{15} = 0 \quad (14)$$

Considerou-se, como já referido, o encabeçamento superior a 2,4 cabeças normais por hectare. Houve, no entanto, a necessidade de limitar o número a 3,4 cabeças normais, de modo a que o sistema de produção não se afastasse do que é corrente na região, isto é, alimentar o gado principalmente à base de erva.

$$2,4 \sum_{i=1}^{13} X_i \leq X_{15} \leq 3,4 \sum_{i=1}^{13} X_i \quad (15)$$

Finalmente o último conjunto de restrições corresponde a um conjunto de restrições de não negatividade, que evita que as variáveis de decisão alcancem valores negativos.

O modelo analítico genérico é o indicado nas equações (16), na página seguinte.

4. Resultados do Modelo

Este modelo foi otimizado usando as ferramentas fornecidas pela *Solver Foundation* (Solver Foundation, 2012). Nomeadamente utilizou-se o MS Excel como interface, tendo sido o modelo programado em OML - *Optimization Modeling Language* e resolvido pelo algoritmo de otimização (*solver*) incluído no pacote. Os valores ótimos obtidos são descritos nas linhas seguintes.

A margem bruta ronda os 55 721 €/ano e as atividades que proporcionam este rendimento são primeiro 48,6 animais produtores de leite (X_{15}); que se alimentariam à base de pastagem média com pastoreio direto (X_2) que ocuparia 7 hectares de terra agrícola; à base de pastagem alta com pastoreio direto (X_1) ocupando 0,8 hectares de superfície; à base de pastagem baixa com pastoreio direto e silagem de erva (X_6) e que ocuparia 7,7 hectares de terra agrícola. Esta alimentação em verde seria complementada com 3 877 quilogramas de ração (X_{14}).

Este valor parece elevado, mas é necessário ter em conta que os resultados apresentados flexibilizam as restrições de mão-de-obra, o que majora o valor real alcançado. A flexibilização resulta de um acréscimo de 10% proposto (60 horas) relativo a mão-de-obra familiar. Este acréscimo de horas possibilitou um acréscimo ligeiro da produção, uma vez que a mão-de-obra era limitante. Note-se ainda que a flexibilização resulta também de se ter utilizado variáveis de desvio positivas e negativas permitindo que as 660 horas por período sejam superadas ou não sejam alcançadas. Além disso, são valores para uma família e não para um indivíduo, ainda que se admita que a família possa ter outros rendimentos.

$$\max MB = \max \sum_{i=1}^{15} X_i MB_i$$

s.a.:

$$X_1 \leq S_A$$

$$X_2 + X_3 + X_4 + \frac{1}{2}X_8 + \frac{1}{4}(X_{10} + X_{12}) \leq S_M$$

$$X_5 + X_6 + X_7 + \frac{1}{2}X_9 + \frac{1}{4}(X_{11} + X_{13}) \leq S_D$$

$$S_A + S_M + S_D \leq S_T$$

$$X_8 + X_9 \leq 0,2(S_M + S_D)$$

$$X_{10} + X_{11} \leq X_8$$

$$X_{11} + X_{13} \leq X_9$$

$$\sum_{i=1}^{15} MO_j X_i + n_j - p_j = \overline{MO}_{d_j} \text{ com } j=1,\dots,6$$

$$\sum_{i=1}^{15} (MB_{ik} X_i - \overline{MB}_{ik}) + N_k - P_k = 0 \text{ com } k=1,\dots,7$$

$$\sum_{i=1}^{15} MB_i X_i \geq 15\,498$$

$$\sum_{i=1}^{14} UFL_{ij} MS_{ij} X_i \geq UFL_{15j} X_{15} \text{ com } j=1,\dots,6$$

$$\sum_{i=1}^{14} PDIE_{ij} MS_{ij} X_i \geq PDIE_{15j} X_{15} \text{ com } j=1,\dots,6$$

$$\sum_{i=1}^{14} PDIN_{ij} MS_{ij} X_i \geq PDIN_{15j} X_{15} \text{ com } j=1,\dots,6$$

$$\sum_{i=1}^{14} CA_{ij} MS_{ij} X_i \geq CA_{15j} X_{15} \text{ com } j=1,\dots,6$$

$$\sum_{i=1}^{14} P_{ij} MS_{ij} X_i \geq P_{15j} X_{15} \text{ com } j=1,\dots,6$$

$$\sum_{i=1}^{14} MS_i X_i \geq MS_{15} X_{15} \text{ com } j=1,\dots,6$$

$$X_{14} - 547,7X_{15} = 0$$

$$X_{15} - 3,4 \sum_{i=1}^{13} X_i \leq 0$$

$$X_{15} - 2,4 \sum_{i=1}^{13} X_i \geq 0$$

$$X_i \geq 0 \text{ com } i=1,2,\dots,15$$

(16)

A mão-de-obra é um fator condicionante da produção leiteira em qualquer período considerado. Ou seja, para rebanhos com cerca de 48 animais, as 660 horas disponíveis serão integralmente usadas, e ainda teriam de recorrer à mão-de-obra familiar (variáveis de desvio positivas) em todos os períodos: MPOI, MPOII, III, IV, V, VI, sendo respetivamente em horas, 37; 135; 63; 26,4; 71,8 e 26,4. Sendo assim, no primeiro período necessitaria no total de 697 horas em mão-de-obra: as 660 horas disponíveis, mais as 37 horas proporcionadas pelo maior esforço do trabalho familiar. Os períodos mais críticos seriam o II (de Março a Abril) em que seriam necessárias 135 horas adicionais às 660 disponibilizadas e que totalizariam 795 horas e o V (Setembro e Outubro), que necessitaria de 731,8 horas.

No que refere à variabilidade do rendimento agrícola, estimada pelo MOTAD (ou risco agrícola medido pelos desvios à margem bruta da exploração) verifica-se que os períodos 1996, 1991 e 1990, não se alcançam os valores médios da margem bruta, ou seja, obtêm-se desvios negativos de 1 363, 828, e 1 225 € respetivamente. De um modo geral, nos restantes períodos, o rendimento é superior à média.

No que se refere ao uso da terra, toda a terra disponível (15,5 hectares) é utilizada para pastagem (alta, média e baixa altitude) e silagem de erva no caso da pastagem em baixa altitude.

As necessidades em unidades forrageiras de leite, em proteínas digestivas no intestino com energia e azoto limitante (PDIE e PDIN), em cálcio e fósforo são asseguradas pelos alimentos disponibilizados (pastagem, silagem de erva e ração) na quase totalidade dos períodos. Apenas no período I é completamente usado o valor do cálcio disponibilizado pelos alimentos considerados. A matéria seca proporcionada pelos alimentos em verde (pastagem e silagem de erva) e ração surge como inferior à capacidade de ingestão dos alimentos.

5. Conclusões do Modelo

O rendimento mensal obtido nas explorações leiteiras, considerando um ano de catorze meses, em regime intensivo, é de cerca de 3 980 €. Assim, a solução ótima obtida pelo modelo é um valor superior ao verificado, em média, nos Açores em 2007, 2 197 € (Silva, 2012), sendo opinião dos especialistas e

decisores que este resultado indica que há espaço para melhoria do rendimento das explorações agrícolas na RAA. De lembrar que o rendimento é majorado pelo aumento parcial da mão-de-obra agrícola familiar que em todos os períodos excede o tempo disponível (660 horas).

No entanto, este rendimento médio por vaca de 1 146,5 €, é alcançado com um encabeçamento médio de 3,13 animais por hectare, valor superior ao pretendido pela política agroambiental da União Europeia. Estas orientações de política europeia têm incentivado a extensificação dos sistemas de produção, ou seja, encabeçamentos inferiores a 1,4. Conclui-se, assim, que se a União Europeia quer optar por este tipo de encabeçamentos, certamente mais apelativos a nível de impacto ambiental, poderá ter de compensar os agricultores pela perda de rendimento correspondente se pretender a sustentabilidade das explorações agrícolas nos Açores.

Nota-se igualmente dos resultados que a sustentabilidade e/ou viabilidade económica e financeira das explorações só é possível devido à existência dos subsídios auferidos quer pela União Europeia quer pelo Governo Regional dos Açores. Estes são parte integrante do rendimento e não algo atribuído a título de prémio ou um qualquer extra.

A superfície agrícola é integralmente utilizada, bem como, a mão-de-obra disponível, indicando mais uma vez que se consideram as atividades agrícolas rentáveis, nas condições e com os pressupostos incluídos no modelo. Estes valores confirmam as opiniões de especialistas de que mesmo as pastagens de mais elevada altitude, menos produtivas, podem ser rentáveis nas condições consideradas.

Estes rendimentos também exigem um esforço de mão-de-obra superior às 330 horas mensais consideradas pelos decisores. Assim, seria necessário recorrer a um maior esforço da mão-de-obra familiar, em qualquer dos seis períodos considerados.

As necessidades em unidades forrageiras de leite, em proteínas digestivas no intestino com energia e azoto limitante (PDIE e PDIN), em cálcio e fósforo são asseguradas pelos alimentos disponibilizados (pastagem, silagem de erva e ração) na quase totalidade dos períodos. Apenas no período I é completamente usado o valor do cálcio disponibilizado pelos alimentos considerados. A

ingestão da matéria seca não é um fator limitante ao contrário do que era esperado pelos especialistas.

A variabilidade do rendimento agrícola, estimada pelo MOTAD, nos sete períodos considerados, não é muito relevante, variando apenas em percentagens relativamente reduzidas.

As fontes de informação agrícola utilizadas (RICA-A, INRA e SREA) embora nos tenham proporcionado os elementos principais para a elaboração deste trabalho, foram, de uma forma geral, insuficientes para a informação necessária à construção do modelo, tendo de se recorrer a adaptações à realidade açoriana.

Um dos resultados mais relevantes deste trabalho é o reconhecimento de que as autoridades terão de investir em bases de dados ajustadas à realidade açoriana, de onde se possa obter informação técnica e económica relativa às atividades existentes na exploração pecuária, e ainda, relativa à constituição dos alimentos produzidos e utilizados nas explorações pecuárias dos Açores.

A conclusão mais relevante é, no entanto, a inferência de que se houver uma política agroambiental mais agressiva por parte da União Europeia e que obrigue à diminuição do encabeçamento, o rendimento das explorações leiteiras em regime intensivo baixará muito, o que poderá pôr em questão a sustentabilidade de algumas explorações.

De um modo global, pode dizer-se que o conhecimento do processo de decisão dos agricultores açorianos é escasso. Há que desenvolver investigação nesta área, procurar informação complementar à existente na Região, analisá-la de forma a permitir o desenvolvimento do sector leiteiro e em última instância, contribuir para o desenvolvimento económico e agrícola dos Açores.

Outras técnicas de IO que poderão ser utilizadas em estudos semelhantes podem passar por técnicas mais descritivas ou prescritivas como é o caso da simulação combinada com metodologias de estruturação de problemas como os mapas cognitivos. Um outro resultado desse interesse será o estudo sobre eficiência de explorações agrícolas usando DEA e SFA.

Agradecimentos

Os autores agradecem aos profissionais da Secretaria Regional da Agricultura e Florestas da Região Autónoma dos Açores que disponibilizaram os dados e o seu conhecimento e tempo na validação do modelo e discussão de resultados. Os autores agradecem ainda aos dois revisores que ao fazerem uma revisão de grande qualidade permitiram a melhoria significativa do trabalho apresentado.

BIBLIOGRAFIA

AVILLETZ, F., *Estudo de base microeconómica sobre as perspectivas de desenvolvimento da agricultura dos Açores*, ISA, 1991.

BARROS, H. e F. Estácio. *Economia da Empresa Agrícola*, Universidade de Luanda, 1972.

BERBEL, J. e M. Barros, "Planificación Multicritério de Empresas Agroganaderas con Bovino de Leche de Islas Azores (Portugal)", *Invest.agr.:Econ.*, 8 (2), 1993: 197-208.

ROMERO, C. e T. Rehman, Multiple Criteria Analysis for Agricultural Decisions, *Development in Agricultural Economics*, 5, Elsevier, Amsterdam, 1989.

CARVALHO, M.L.S., *Efeitos da Variabilidade das Produções Vegetais na Produção Pecuária, aplicação em explorações Agro-pecuárias do Alentejo: Situações Actual e Decorrentes da Nova PAC*, 1º prémio, III Edição do Prémio de Estudos de Economia Agrícola e Agro-Alimentar, APDEA, Associação Portuguesa de Economia Agrária, Lisboa, 1999.

DREPA, Direcção Regional de Estudos e Planeamento, (<http://www.drepa.raa.pt>, consultado em 19 de Maio de 2010).

HERRERO, M., R.H. Fawcet, J.B. Dent "Bio-economic evaluation of dairy farm management scenarios using integrated simulation and multiple-criteria models" *Agricultural Systems* 62 (1999) 169-1888.

INE. 2001. Instituto Nacional de Estatística. Estatísticas Agrícolas 2000. (<http://www.ine.pt/>, consultado em 17 de Janeiro de 2013).

INRA (Institut National de la Recherche Agronomique), *Alimentation des bovins, ovins et caprins*, R. Jarrige ed., Paris, 1988.

MARTA-COSTA, A. *A tomada de decisão no planeamento da Exploração Agrária no contexto da sustentabilidade: o caso da produção de carne bovina maronesa*. Tese de doutoramento. Universidade de Trás-os-Montes e Alto-Douro, 2008.

MATOS, J., *A Pecuária Açoriana. Uma perspectiva à luz de novos critérios e desafios*, Lição proferida no âmbito das Provas de Agregação, Departamento de Ciências Agrárias, Universidade dos Açores, Angra do Heroísmo, 1997.

MOURITS, M.C.M., D.T. Galligan, A.A. Dijkhuizen, R.B.M. Huime "Optimization of Dairy Heifer Management: Decisions based on Production Condition of Pennsylvania", *J Dairy Sci* 83:1989-1997, 2000.

OLIVEIRA, C.; Mendes, A. e E. Silva. . *Quotas leiteiras: Aplicação do método Delphi*. In Silva, Emiliana et al. (Eds.) *Proceedings do congresso da APDEA e Gestão e Conservação da Natureza*. Univ. dos Açores: Ponta Delgada. (2010) p. 15-17.

POETA, A. M.S. *A Tomada de Decisão no Planeamento da Exploração Agrícola num Contexto de Objectivos Múltiplos*, Dissertação de doutoramento, Universidade de Trás-os-Montes e Alto Douro, Vila Real, 1994.

RAMALHETE, M., J. Guerreiro e A. Magalhães. *Programação Linear*, vol. I, Mc-Graw – Hill, 1984.

RAMSDEN, S., J. Gibbons, P. Wilson "Impacts of changing relative prices on farm level dairy production in the UK" *Agricultural Systems* 62 (1999) 201- 215.

SILVA, E., *Análisis Multicriterio de la Eficiencia Económica de las Explotaciones Agroganaderas de las Azores (Portugal)*, Tesis doctoral, Escuela Técnica Superior de Ingenieros Agrónomos y de Montes, Universidad de Córdoba, 2001.

SILVA, E. *Os objectivos dos Agricultores dos Açores: uma abordagem multicritério*, DRCT e DRDA. ISBN 978-989-0422-8, 2006.

SILVA, E.; J. Berbel. "An Azorean Farms Typology", *New Medit*, vol.VI. nº1. pg. 51-54. 2007.

SILVA, E. "The agriculture of an ultraperipheral region of European Union: the Azores archipelago". Review on agriculture and rural development. *Scientific Journal of the University of Szeged, Faculty of Agriculture*, vol.1 (1), 2012: 20-25.

SOLVER Foundation (<http://msdn.microsoft.com/en-us/devlabs/hh145003>, consultado em 8 de Dezembro de 2011).

SREA - Secretaria Regional de Estatística dos Açores, Censos 2001, Publicações não Periódicas, Serviço Regional de Estatística dos Açores. (<http://srea.ine.pt>, consultado em 14 de Fevereiro de 2010).

SREA, *Açores em números*. Serviço Regional de Estatística dos Açores (SREA). Região Autónoma dos Açores. 2009.

TAUER, L.W., "Do New York Dairy Farmers Maximize profits or Minimize Cost?" *American Journal Agricultural economics*, 1995: 421-429.

TOZER, P.R. e J.R. Stokes "A multi objective programming approach to feed ration balancing and nutrient management" *Agricultural systems* 67 (2001) 201-215.

(Página deixada propositadamente em branco)

CASO 4

MODELO DE APOIO À GESTÃO DA PESCARIA DE BIVALVES COM GANCHORRA NO ALGARVE BASEADO EM DINÂMICA DE SISTEMAS

Ana S. Camanho

Faculdade de Engenharia da Universidade do Porto
acamanho@fe.up.pt

Joana H. Martins

Faculdade de Engenharia da Universidade do Porto
joana.hora@gmail.com

Manuela M. Oliveira

Instituto Português do Mar e da Atmosfera
moliveira@ipma.pt

Miguel B. Gaspar

Instituto Português do Mar e da Atmosfera
Centro de Ciências do Mar (CCMAR), Universidade do Algarve
mbgaspar@ipma.pt.

RESUMO

Este trabalho consistiu no desenvolvimento de um modelo de gestão baseado em Dinâmica de Sistemas com o objetivo de melhorar a gestão da pesca artesanal de bivalves com ganchorra na costa sul de Portugal, considerando as dimensões biológica, social e económica. O modelo pretende simular o comportamento da pescaria com ganchorra e inclui as quatro principais espécies alvo e os dois segmentos de frota, local e costeira, que operam no Algarve. Foram estudados diversos cenários para avaliar o impacto da implementação de medidas de gestão sobre a sustentabilidade do sistema. A principal contribuição do modelo desenvolvido é permitir apoiar os agentes de decisão na seleção das medidas mais adequadas para atingir os seus objetivos, sem comprometer a sustentabilidade das frotas, dos ecossistemas e dos recursos explorados.

PALAVRAS-CHAVE

Pesca artesanal, Dinâmica de Sistemas, Simulação, DPSIR, Sustentabilidade das pescas.

1 Introdução

De acordo com a *Food and Agriculture Organization of the United Nations* (FAO, 2010, p.64), o consumo total e per capita de peixe aumentou significativamente nas últimas cinco décadas. Desde 1961, o consumo total de peixe cresceu a uma taxa anual de 3,1%, enquanto a população mundial, no mesmo período, cresceu a uma taxa anual de 1,7%. A pesca artesanal desempenha um papel importante nos mercados locais e regionais, representando mais de metade do peixe capturado para consumo humano (FAO, 2010, p.70).

Nas zonas costeiras do Arco Atlântico Europeu, as pescarias artesanais são uma tradição antiga e uma fonte de identidade incorporada nas culturas das diferentes regiões, que deve ser preservada. O desenvolvimento sustentável da pesca artesanal é essencial para garantir a preservação dos ecossistemas marinhos e a estabilidade socioeconómica das comunidades piscatórias.

Para assegurar a sustentabilidade da pesca artesanal, é necessário recolher informação atualizada sobre o sistema em análise. De um modo geral, a informação sobre a pesca artesanal é escassa, limitando assim o desenvolvimento de estudos analíticos que permitam sustentar cientificamente a implementação de estratégias e iniciativas de gestão.

A definição de medidas apropriadas a uma boa gestão das pescas é um processo complexo. O problema da pesca excessiva, que leva ao colapso dos recursos naturais, tem sido observado mesmo em países com comunidades científicas dedicadas exclusivamente à gestão das pescas (Dudley, 2008). O colapso dos *stocks* biológicos e consequente interdição da pesca do bacalhau no Canadá (Mason, 2002) e no Mar do Norte (Malakoff and Stone, 2002) são exemplos deste fenómeno. Entre as causas mais frequentemente apontadas para o insucesso dos modelos de gestão são: i) a grande variabilidade das condições climáticas que afetam os *stocks* biológicos (King and McFarlane, 2006; Rothschild et al., 2005), ii) atrasos na implementação das medidas regulatórias devido a pressões socioeconómicas (Caddy, 2002, Jackson et al., 2001), e iii) a dificuldade de visualização holística da complexidade de inter-relações entre os elementos biofísicos e socioeconómicos do sistema (Dudley, 2008).

No contexto da pesca artesanal em Portugal, as medidas regulatórias são tomadas tendo por base a monitorização do estado dos recursos naturais, mas não é possível quantificar o seu impacto nos vários elementos que compõem o sistema de pesca. Assim, este trabalho tem por objetivo desenvolver um modelo que represente as inter-relações entre os principais componentes de um sistema de pesca, incluindo aspetos biológicos, sociais e económicos. Assim, será possível prever o impacto das medidas regulatórias sobre o sistema, e melhorar o processo de tomada de decisão.

Neste trabalho apresenta-se um conjunto de indicadores utilizados para estudar um sistema de pesca artesanal em Portugal. Os indicadores foram selecionados utilizando a metodologia DPSIR (*Driving forces, Pressure, State, Impact, Responses*) proposta pela Agência Europeia do Ambiente (EEA, 1999). Os indicadores foram também categorizados em três dimensões de sustentabilidade: social, económica e biológica, às quais se adicionou uma quarta, institucional, que tem em vista regular as outras dimensões.

O modelo de gestão desenvolvido, baseado em Dinâmica de Sistemas, tem por objetivo apoiar a introdução de medidas regulamentares para promover a sustentabilidade da pesca artesanal de bivalves com ganchorra na costa sul de Portugal, considerando as dimensões biológica, social e económica do sistema. O modelo de simulação desenvolvido permite (i) simular o comportamento do sistema em função dos cenários a explorar, (ii) registar a evolução das variáveis mais relevantes do sistema, tendo em conta as suas interações e (iii) avaliar a reação do sistema a alterações da estratégia de gestão.

O sistema em análise é constituído pelas quatro espécies comerciais de bivalves exploradas pela frota de ganchorra: a conquinha (*Donax trunculus*), a amêijoia-branca (*Spisula solida*), o pé-de-burrinho (*Chamelea gallina*) e a navalha (*Ensis siliqua*); e por dois segmentos de frota: a frota costeira, com maior poder de pesca e autonomia, e a frota local, constituída por barcos de menores dimensões. O modelo de Dinâmica de Sistemas construído para simular o comportamento desta pescaria inclui variáveis biológicas e ecológicas que permitem descrever a evolução de cada espécie (e.g. taxa de crescimento natural, área de ocorrência de cada espécie), variáveis socioeconómicas (e.g., preços de venda e lucro dos pescadores), e variáveis que induzem efeitos

regulatórios (e.g. quotas diárias de pesca, época de defeso). Esta metodologia permitir simular cenários com o objetivo de avaliar o impacto da implementação de determinadas medidas de gestão sobre a rentabilidade das embarcações e estado dos recursos, a médio-longo prazo.

O modelo desenvolvido representa uma mais-valia no processo de tomada de decisão. A divulgação dos resultados obtidos com modelos de simulação pode ser útil para auxiliar os pescadores a compreender e aceitar as opções seguidas pela Administração em matéria de regulação da pesca. Assim, pode ser uma ferramenta fundamental para a evolução da gestão da pesca artesanal de um modelo *top-down* para um modelo de cogestão onde os pescadores terão uma maior responsabilização na gestão dos recursos. Por exemplo, é essencial a colaboração dos pescadores no cumprimento dos limites máximos de capturas e dos tamanhos mínimos de captura, bem como motivar a cooperação entre os pescadores no sentido de criar melhores formas de comercialização do pescado. Finalmente, poderá permitir cativar o interesse da população em geral sobre as pescarias artesanais.

A estrutura deste capítulo é a seguinte. A secção 2 inclui a descrição dos conceitos de Dinâmica de Sistemas e faz uma revisão da sua aplicação à pesca. A secção 3 descreve o estudo de caso da pesca da ganchorra. A secção 4 descreve o modelo de Dinâmica de Sistemas desenvolvido para o caso de estudo. A secção 5 apresenta os resultados das simulações efetuadas e discute as suas implicações para a gestão da pescaria. A secção 6 apresenta as conclusões.

2 Dinâmica de Sistemas aplicada a sistemas de pesca

Segundo Sterman (2000), a Dinâmica de Sistemas baseia-se na teoria do Pensamento Sistémico (ver Checkland, 1999; Rosenhead, 1989), uma vez que aplica uma visão ao nível do sistema global que permite a modelação da estrutura causal do sistema (incluindo processos de causa-efeito, ciclos de retroalimentação e interações entre variáveis).

O Pensamento Sistémico implica o reconhecimento de que o sistema em análise é complexo e integra interdependências entre as variáveis que o compõem. Uma alteração feita num determinado momento temporal numa

determinada variável repercute-se em todo o sistema, causando alterações noutras variáveis instantaneamente ou ao longo do tempo, uma vez que todas as variáveis estão interligadas de alguma forma. A elaboração de um modelo de Dinâmica de Sistemas é normalmente precedida de uma fase do Pensamento Sistémico, que identifica as variáveis que compõem o sistema e as ligações possíveis entre elas. Segundo Forrester (2007), a fase de Pensamento Sistémico permite muitas vezes julgar que se apreendeu muito sobre o funcionamento de um sistema. No entanto, o conhecimento aprofundado de um sistema depende essencialmente da fase de estruturação de modelos e análise dos resultados obtidos com base nesses modelos. É apenas a partir das simulações do comportamento dos sistemas reais que as inconsistências dos nossos modelos mentais são reveladas. O Pensamento Sistémico deve, assim, ser visto como um primeiro passo para a compreensão da dinâmica de um problema complexo, que deve ser complementado, sempre que possível, pela construção de modelos de Dinâmica de Sistemas.

A modelação com Dinâmica de Sistemas permite estudar a interação das variáveis consideradas no sistema ao longo do tempo, assim como analisar de uma forma rápida e eficaz a resposta do sistema a estímulos cujo impacte se pretende estudar (variando parâmetros ou relações entre variáveis). Como explicado por Sterman (2000), não se pretende que um modelo de Dinâmica de Sistemas gere resultados precisos sobre a evolução futura dos recursos ou resultados do sistema. Estes modelos devem ser vistos como ferramentas que permitem explorar e avaliar configurações alternativas do sistema.

Entre os elementos que integram os modelos de Dinâmica de Sistemas, destacam-se as variáveis de armazenamento (*stock*), os fluxos contínuos entre as variáveis, e as relações dinâmicas de realimentação (*feedback*) entre as variáveis constituintes do sistema (Forrester, 1991). Uma descrição detalhada sobre os diferentes tipos de elementos que integram os modelos de Dinâmica de Sistema pode ser consultada em Few (1996). Sterman (2000) apresenta uma descrição pormenorizada dos conceitos de Dinâmica de Sistemas.

Esta metodologia tem sido amplamente aplicada em áreas bastante diferentes, incluindo a gestão de recursos hídricos (Wang et al., 2011), produção animal (Tedeschi et al., 2011) ou ecossistemas costeiros (Arquitt e

Johnstone, 2008). O trabalho desenvolvido por Ford (1999) inclui uma descrição esclarecedora sobre a aplicação da Dinâmica de Sistemas à modelação e simulação de sistemas ambientais.

Aplicações anteriores da metodologia de Dinâmica de Sistemas a sistemas de pesca tiveram por objetivo a análise da pesca de espécies migratórias em várias zonas (Mchich et al., 2000), o estudo das implicações de políticas desenhadas para revitalizar a indústria da pesca (Otto e Struben, 2004), a comparação entre a modelação de sistemas complexos e simplificados (Moxnes, 2005), a avaliação do impacte biológico de alterações da camada sedimentar (Lazar et al., 2010), a avaliação da regulação da pescaria através da imposição de quotas (Garrity, 2011), entre outros. Descrevem-se em seguida com mais pormenor três estudos particularmente relevantes para o presente trabalho, desenvolvidos por Yndestad e Stene (2002), Bueno e Basurto (2009) e BenDor et al. (2009), dado que se focaram na gestão de pescarias específicas, analisando o impacte de medidas regulatórias tendo em conta a evolução do ecossistema e a atuação dos pescadores.

Yndestad e Stene (2002) modelaram a pesca de capelim no Mar de Barents considerando que a evolução natural do capelim (definida pelo crescimento, reprodução e mortalidade) depende principalmente da predação natural por outras espécies e das capturas realizadas pela pesca. Este estudo concluiu que a taxa de captura deve ser ajustada tendo em conta a evolução dos ciclos climáticos e a sustentabilidade biológica das espécies.

Bueno e Basurto (2009) estudaram as pescas de pequena escala de moluscos bivalves no México. O modelo de simulação foi desenvolvido considerando variáveis institucionais e ecológicas, cujos resultados demonstraram que mesmo pequenas alterações nas relações entre as variáveis podem transformar um sistema aparentemente robusto num sistema insustentável.

BenDor et al. (2009) estudaram as interações entre as dimensões económica e ecológica de uma pescaria multi-espécie explorada por vários métiers. Este estudo considerou a capacidade de regeneração das populações de peixe como um indicador da condição dos recursos naturais. Os indicadores socioeconómicos considerados incluíram o lucro, o nível de empregabilidade e

a coesão social. Os cenários analisados incluíram alterações no comportamento dos pescadores, cujos resultados demonstraram que o planeamento de quotas de pesca e a cooperação entre as várias entidades operantes no mercado são essenciais para a resolução de conflitos associados à sobrepesca.

Pode concluir-se que a aplicação de modelos de Dinâmica de Sistemas às pescas potencia o estudo dos impactos associados a políticas de gestão alternativas, permitindo escolher as que mais se adequam aos objetivos pretendidos.

3 Contextualização do caso de estudo

Este estudo foca-se na pescaria artesanal de bivalves com ganchorra, que ocorre na costa sul de Portugal. As embarcações dirigem o esforço de pesca para a captura de quatro espécies (Figura 1): a conquitilha (*Donax trunculus*), a amêijoia-branca (*Spisula solida*), o pé-de-burrinho (*Chamelea gallina*) e a navalha (*Ensis siliqua*). O habitat dos bivalves pode ser caracterizado pelas profundidades de ocorrência dos bancos de cada espécie: a amêijoia-branca entre os 3 e os 12m, a conquitilha entre os 0 e 5m, a navalha entre os 3 e os 11m e o pé-de-burrinho entre os 3 e os 15m.



Figura 1. Espécies alvo da pesca da ganchorra da costa sul Portuguesa. (a) Amêijoia-branca. (b) Conquitilha. (c) Pé-de-burrinho. (d) Navalha.

As embarcações da frota da ganchorra podem ser consideradas, de acordo com as suas características, de pesca local (EPL) ou de pesca costeira (EPC) (Figura 2). As EPL são caracterizadas por uma baixa motorização e pela escassez de equipamentos de alagem. O seu comprimento fora a fora (Cff) é igual ou inferior a 9m e/ou a potência máxima do motor não é superior a 75 kW (=100 Hp) ou 45 kW (=60 Hp), consoante se tratem de embarcações de convés fechado ou de convés aberto. Na sua maioria estas embarcações apresentam um casco de fibra. As EPC apresentam um Cff superior a 9m, tonelagem inferior a 180 GT, motor com potência superior a 25 kW (=35 Hp) e

um casco normalmente em madeira. Atualmente a frota de ganchorra que opera na costa sul de Portugal é constituída por 56 embarcações, 23 das quais pertencem à frota costeira e as restantes 33 à frota local. A capacidade de pesca desta frota é de 333 GT e de 3771 Hp de potência. Estas embarcações desenvolvem a sua atividade principalmente entre Olhos d'Água e Vila Real de Santo António.

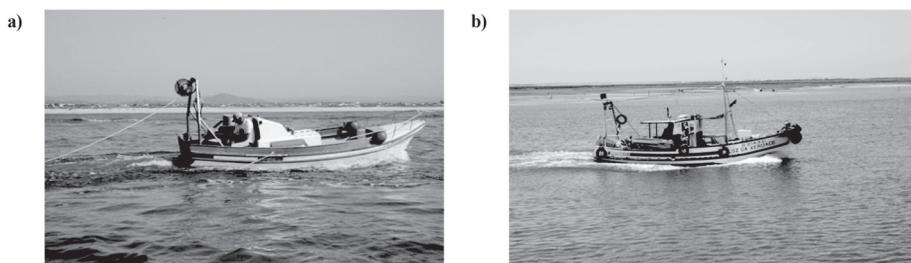


Figura 2. Ilustração das embarcações da frota Algarvia de ganchorra. (a) Embarcação de pesca local. (b) Embarcação de pesca costeira.

A pesca é realizada com ganchorra que é uma arte de arrasto pelo fundo. A ganchorra é uma estrutura metálica, constituída por um aro em semicircunferência, que na posição de diâmetro tem a barra onde está implantado um pente de dentes. O tamanho dos dentes, dependendo da espécie a que se destina, varia entre os 10cm no caso da conquilha e os 60 cm no caso da navalha. A este aro está soldada uma outra estrutura metálica, formada por três hastes em forma de tripé, onde se amarra o cabo de tração e o cabo da «rabeira» ou do «vira». O marisco que entra na ganchorra é retido por um saco de rede (ganchorra tradicional) ou por uma grelha metálica (ganchorra de grelha) (Figura 3). Atualmente, a ganchorra tradicional é utilizada apenas na pesca da navalha, enquanto a ganchorra de grelha é utilizada para capturar as outras espécies.

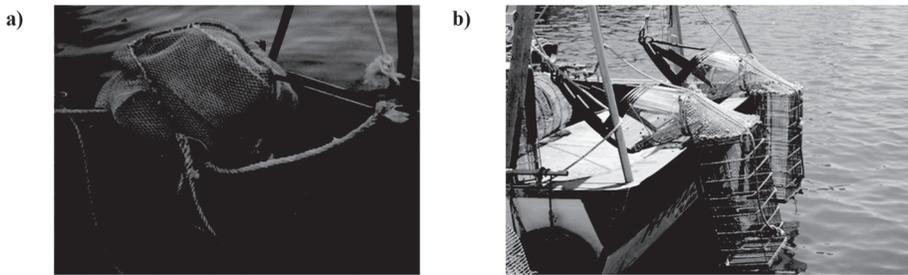


Figura 3. Ilustração da arte da ganchorra. (a) Ganchorra tradicional. (b) Ganchorra de grelha.

A pescaria de bivalves com ganchorra é regulada por um período de defeso (de 1 de maio a 15 de junho), por tamanhos mínimos de captura, por quotas diárias por espécie e embarcação e por restrições à atividade, podendo as embarcações operar seis dias por semana (entre domingo e sexta-feira) e fazer uma única maré diária. As características técnicas da ganchorra encontram-se também regulamentadas. Extraordinariamente podem ser decretadas interdições temporais da pesca como resultado de episódios de biotoxinas ou no caso de sobre-exploração dos recursos.

Os principais portos de pesca na costa do Algarve para a frota de ganchorra são quatro, nomeadamente Faro, Olhão, Tavira e Vila Real de Santo António. Os pescadores escolhem o seu porto de pesca em função da localização dos bancos de bivalves de modo a reduzir o tempo de viagem.

4 Descrição do modelo de Dinâmica de Sistemas

4.1 Identificação dos indicadores

Esta secção apresenta os indicadores selecionados para a construção do modelo de Dinâmica de Sistemas. Os indicadores são apresentados de acordo com a metodologia DPSIR e classificados segundo as quatro dimensões de sustentabilidade.

A Organização para a Cooperação e Desenvolvimento Económico propôs a metodologia PSR (*Pressure, State, Response*) para a identificação de indicadores em 1993 (OECD, 1993). Esta metodologia foi posteriormente aperfeiçoada pela EEA (1999), dando origem à metodologia DPSIR (*Driving Force, State, Pressure, Impact, Response*), através da inclusão de duas novas categorias: forças motrizes (*Driving Force: D*) e impacto (*Impact: I*). A utilização da metodologia

DPSIR para a seleção de indicadores das pescas foi proposta pela FAO (1999) e pela EEA (2002). No entanto, o número de estudos que aplicaram a metodologia DPSIR a sistemas de pesca é ainda reduzido. Neste âmbito, destacam-se os estudos realizados por Mangi et al. (2007) e Knudsen et al. (2010).

Os indicadores associados à categoria *Driving Force* (D) descrevem as necessidades e motivações socioeconómicas que impulsionam a existência da atividade humana. Os indicadores associados à categoria *Pressure* (P) referem-se a ações humanas que podem induzir alterações no ecossistema. Os indicadores classificados na categoria *State* (S) visam descrever o estado dos ecossistemas naturais, o desenvolvimento socioeconómico das populações associadas ao sistema em análise, ou uma combinação de ambos. A categoria *Impact* (I) engloba os efeitos negativos no ecossistema ou na sociedade causados pela atividade humana. A classe *Response* (R) inclui todas as medidas adotadas com o objetivo de melhorar o estado do sistema, podendo incluir ações preventivas, adaptativas ou curativas.

As diretrizes propostas pela FAO (1999) também sugerem que os indicadores a utilizar na análise de sistemas de pesca sejam classificados dentro das quatro principais dimensões da sustentabilidade: económica, social, ambiental e institucional. Seguindo os critérios sugeridos pela FAO (1999), os indicadores ambientais visam medir as características dos ecossistemas naturais, incluindo aspetos biológicos das espécies envolvidas e parâmetros físicos e químicos do ecossistema. No presente estudo, dado não se ter utilizado parâmetros físico-químicos na definição do modelo, apenas se considerou a componente biológica. Os indicadores económicos descrevem aspetos relacionados com os barcos, mecanismos de pesca, a procura do setor, o consumo, os lucros da atividade, a descrição do esforço de pesca e o valor económico dos desembarques. Os indicadores sociais relacionam-se com a caracterização da comunidade envolvida na atividade de pesca, abrangendo informação sobre o nível de emprego, número de pescadores, tradições, cultura e nível educacional. Os indicadores classificados na dimensão institucional referem-se às ações tomadas pelo governo ou outras instituições com o objetivo de melhorar/regular a atividade de pesca.

A Tabela 1 apresenta o resumo dos indicadores selecionados para a construção do modelo de Dinâmica de Sistemas para a pesca de bivalves com ganchorra na costa do Algarve, organizados de acordo com a metodologia DPSIR e classificados nas quatro dimensões de sustentabilidade propostas pela FAO (1999). Estes dados foram utilizados para especificar as relações entre as variáveis incluídas no modelo de Dinâmica de Sistemas.

Os dados para a maioria dos indicadores foram fornecidos pela Direção Geral de Recursos Naturais, Segurança e Serviços Marítimos (DGRM) e pelo Instituto Português do Mar e da Atmosfera (IPMA), e são relativos aos anos entre 1996 e 2009. Os dados relativos ao preço do combustível (gasóleo verde) foram fornecidos pela GALP. Como o IPMA não tem informação sobre o crescimento natural de cada espécie, estes valores foram estimados a partir de outros indicadores (i.e. *stock* e desembarques).

4.2 Modelação das relações entre as variáveis

Esta secção descreve o modelo de Dinâmica de Sistemas, desenvolvido com o software Vensim PLE Plus, para a pesca de bivalves com ganchorra no Algarve. O modelo de simulação contempla a frota costeira (FC), a frota local (FL) e as quatro espécies alvo separadamente, i.e., amêijoia-branca (Ss), conquilha (Dt), navalha (Es) e pé-de-burrinho (Cg). A evolução do sistema foi representada com uma periodicidade anual. A Figura 4 apresenta o esquema simplificado do modelo de Dinâmica de Sistemas desenvolvido.

Na Figura 4 as variáveis com índice i são replicadas para cada espécie e as variáveis com índice j são replicadas para cada frota. Os círculos a branco referem-se às variáveis DPSIR descritas na Tabela 1 e os círculos sombreados representam variáveis auxiliares.

A especificação de um modelo de Dinâmica de Sistemas implica a representação da evolução de cada variável ao longo do tempo, bem como a relação entre as várias variáveis. Assim, as expressões matemáticas incluídas em cada um dos blocos representados na Figura 6 são descritas detalhadamente em seguida.

Tabela 1. Indicadores incluídos no modelo de Dinâmica de Sistemas.

Indicador	Dimensão	Descrição	Unidades
Driving Forces			
Custos fixos	Económica	Custo anual da renovação da licença de pesca, manutenção do barco e dos mecanismos de pesca. • CF_j^t – custo fixo de um barco da frota j no ano t .	Euro /barco
Custos variáveis	Económica	Taxas de desembarques anuais (comissão para as autoridades piscatórias e contribuição para a segurança social), custo de combustível, seguro de tripulação e contribuições para as organizações de produtores. • CV_j^t – custo variável de um barco da frota j no ano t .	Euro /barco
Preço de combustível	Económica	Preço de combustível para a atividade de pesca. • PC^t – preço do combustível no ano t .	Euro / litro
Consumo de combustível	Económica	Consumo médio de combustível por viagem para barcos costeiros e locais. • CC_j^t – consumo médio de combustível por viagem de um barco da frota j no ano t .	Litros / viagem
Preço de primeira venda	Económica	Preço médio de cada espécie alvo em lota. • PPV_i^t – Preço de primeira venda da espécie i no ano t .	Euro /kg
Pressures			
Viagens de mar	Económica	Número anual de viagens de mar realizadas por barco de cada frota. • VM_j^t – Viagens de mar de um barco da frota j no ano t .	Viagem
Número de barcos	Económica	Número de barcos ativos em cada frota. • NB_j^t – número de barcos a operar na frota j no ano t .	Barcos
Tripulantes	Social	Número médio de pescadores constituintes da tripulação, por barco, em cada frota. • T_j^t – número médio de pescadores num barco da frota j no ano t .	Pessoa
Desembarques	Económica	Desembarques anuais por espécie (em peso) por barco de cada frota. • D_{ij}^t – desembarques da espécie i por cada barco da frota j no ano t .	toneladas
State			
Stock	Ambiental	Stock biológico disponível no sistema para cada espécie. • S_i^t – stock biológico da espécie i no ano t .	toneladas
Crescimento natural	Ambiental	Crescimento natural de cada espécie. • CN_i^t – Crescimento natural da espécie i no ano t .	toneladas
Impact			
Salário por pescador	Económica	Salário médio anual por pescador de cada frota. • SP_j^t – salário anual por pescador da frota j no ano t .	Euros
Responses			
Quotas por espécie	Institucional	Limite diário de capturas por espécie. • QE_i^t – quota diária da espécies i no ano t .	kg / viagem
Quotas para o total de capturas	Institucional	Limite diário de capturas (incluindo todas as espécies) por barco e por frota. • QT_j^t – quotas diárias para o total de capturas por barco da frota j no ano t .	kg / viagem

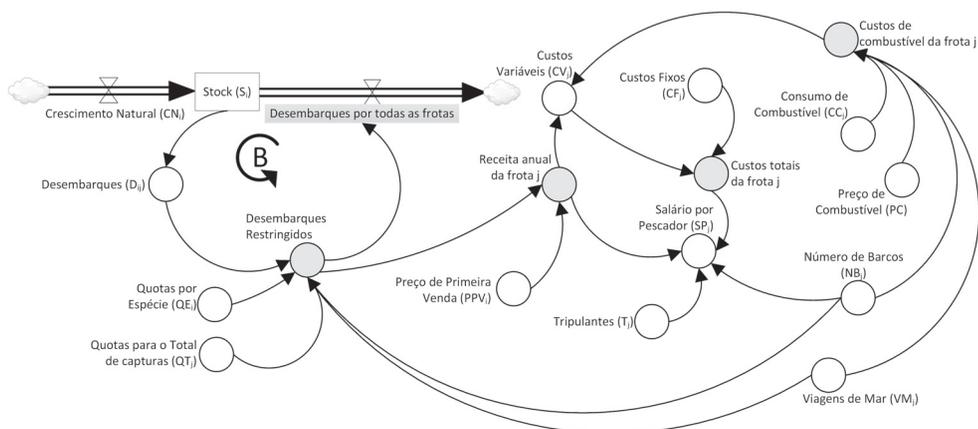


Figura 4. Visualização simplificada do modelo de Dinâmica de Sistemas desenvolvido.

O modelo calcula a evolução do *stock* biológico de cada espécie como a diferença entre o crescimento natural e os desembarques de ambas as frotas.

Como o IPMA não tem não estimativas relativas ao crescimento natural de cada espécie, este foi estimado com base nos dados históricos do período de 1996 a 2009. Para cada ano, o crescimento natural de cada espécie foi calculado como a soma entre a diferença dos *stocks* em anos consecutivos e os desembarques de ambas as frotas, como se demonstra na expressão (1).

$$CN_i^t = (S_i^{t+1} - S_i^t) + \sum_{j=FL,FC} D_{ij}^t \quad (1)$$

A Figura 5 apresenta os resultados obtidos para a evolução do crescimento natural entre 1996 e 2009 para cada uma das espécies alvo consideradas no modelo.

O crescimento natural da amêijoia-branca foi representado no modelo como sendo igual à média observada entre 1996 e 2009 (520 toneladas), excluindo os anos onde se registaram valores muito elevados (2001 e 2003). A ocorrência de valores muito elevados no crescimento natural para esta espécie foi modelada considerando o aparecimento de um pico a cada 14 anos (com o valor de 2336 toneladas, estimado pela média dos valores observados nos anos 2001 e 2003).

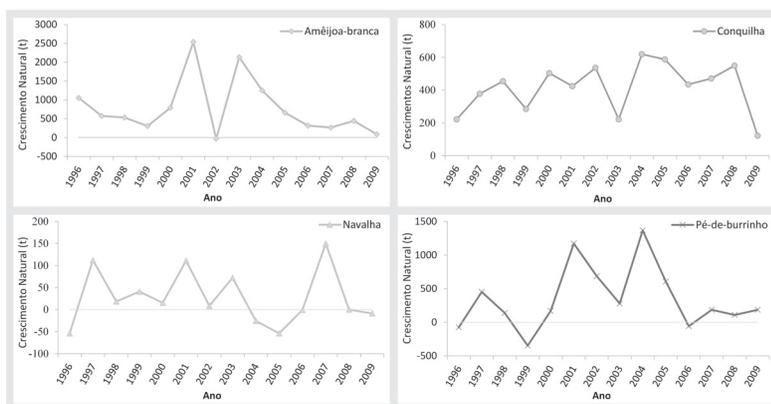


Figura 5. Crescimento natural estimado para as espécies alvo.

No caso da conchilha, como os anos anteriores a 2003 mostraram um comportamento muito variável do crescimento natural, optou-se por especificar esta variável no modelo como sendo igual ao valor médio observado entre 2003 e 2009 (429 toneladas).

O crescimento natural do pé-de-burrinho foi definido pela média dos valores observados (194 toneladas) excluindo os anos onde se observaram picos (2001 e 2004). À semelhança do procedimento adotado para a amêijo-branca, a ocorrência de picos nesta espécie foi modelada com uma frequência de 14 anos (1270 toneladas, calculado como o valor médio observado em 2001 e 2004).

Relativamente à navalha, o seu crescimento natural foi modelado considerando a média de todos os anos observados (34 toneladas).

Para modelar os desembarques de cada espécie, foi investigada a relação entre os desembarques de cada espécie para cada frota e os níveis de *stock* das espécies-alvo, bem como os desembarques de outras espécies. Verificou-se que a evolução dos desembarques para as espécies conchilha, navalha e pé-de-burrinho é explicada pelos níveis de *stock* respetivos (em ambas as frotas). Para a amêijo-branca, os desembarques são explicados por duas variáveis: o estado de conservação do *stock* e os desembarques de pé-de-burrinho. A

Tabela 2 apresenta as equações das regressões utilizadas para modelar a evolução dos desembarques de cada espécie em cada frota.

Tabela 2. Desembarques de cada espécie por cada barco (em toneladas).

Frota local		Significância	
		R ²	p-value
Amêijoa-branca	$D_{Ss,FL}^t = 0,317 + 0,008 \cdot S_{Ss}^t - 0,585 \cdot D_{Cg,FL}^t$	0,854	0,000
Conquilha	$D_{Dt,FL}^t = 4,917 + 0,013 \cdot S_{Dt}^t$	0,202	0,107
Navalha	$D_{Es,FL}^t = -0,169 + 0,002 \cdot S_{Es}^t$	0,910	0,000
Pé-de-burrinho	$D_{Cg,FL}^t = -0,249 + 0,005 \cdot S_{Cg}^t$	0,777	0,000
Frota costeira			
Amêijoa-branca	$D_{Ss,FC}^t = 12,575 + 0,034 \cdot S_{Ss}^t - 1,397 \cdot D_{Cg,FC}^t$	0,678	0,002
Conquilha	$D_{Dt,FC}^t = 1,837 + 0,036 \cdot S_{Dt}^t$	0,832	0,011
Navalha	$D_{Es,FC}^t = 0,065 + 0,007 \cdot S_{Es}^t$	0,718	0,000
Pé-de-burrinho	$D_{Cg,FC}^t = 1,667 + 0,011 \cdot S_{Cg}^t$	0,565	0,002

As equações apresentadas na Tabela 2 representam as capturas anuais (em toneladas) por barco de cada frota. Os desembarques são condicionados pela legislação através de quotas diárias que limitam as capturas totais por barco e por espécie. Para tal é necessário incluir no modelo informação sobre os limites das quotas (totais e por espécie). Finalmente, para obter o total de capturas para cada frota, o modelo inclui também informação sobre o número de barcos de cada frota e o número de viagens (dias de mar) efetuadas anualmente por cada barco.

A Tabela 3 apresenta o número de barcos a operar em cada frota considerando o período compreendido entre 1996 e 2009. Para as simulações de Dinâmica de Sistemas considerou-se que o número de barcos, para cada frota, seria constante e igual ao valor médio do histórico observado: 25 barcos na frota local e 22 na costeira.

Tabela 3. Número de barcos em cada ano.

	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Número de barcos locais (NB _{FL} ^t)	16	18	19	23	23	27	28	28	31	28	28	29	28	26
Número de barcos costeiros (NB _{FC} ^t)	20	20	21	21	21	23	24	24	26	25	24	25	24	14

O histórico de observações relativo ao número anual de viagens por barco é descrito na Tabela 4. Neste estudo explorou-se o impacto dos níveis de *stock* de cada espécie no número de viagens de mar, tendo-se encontrado significância estatística na explicação da evolução do número de viagens por barco costeiro em função dos níveis de *stock* da amêijoia-branca.

Tabela 4. Número anual de viagens de mar por barco de cada frota.

	2001	2002	2003	2004	2005	2006	2007	2008	2009
Viagens de mar por barco local (VM_{FL}^t)	109	133	128	114	138	94	113	107	101
Viagens de mar por barco costeiro (VM_{FC}^t)	146	156	162	165	155	107	132	114	118

O modelo de regressão usado para prever as viagens por barco costeiro é apresentado na expressão (2), com $R^2=0,616$ ($p=0,012$).

$$VM_{FC}^t = 121,771 + 0,02 \times S_{SS}^t \quad (2)$$

Uma vez que não foi encontrada nenhuma relação com significância estatística entre a evolução do número de viagens da frota local e as potenciais variáveis explicativas incluídas no modelo, considerou-se que esta variável seria igual à média dos valores históricos observados (115 viagens).

A tonelagem de arqueação bruta (GT) média de cada frota foi adotada como referência para a modelação de barcos «tipo» de cada frota (i.e. considerou-se que cada barco local tem um GT de 3,89 toneladas e cada barco costeiro tem um GT de 9,25 toneladas). O modelo pressupõe ainda que o valor das quotas observadas na última revisão da legislação (realizada em 2005) seria mantido nos anos futuros. As quotas entre 1996 e 2009 estão apresentadas na Tabela 5.

As variáveis económicas apresentadas no retângulo do lado direito da Figura 5 são explicadas de seguida. As receitas anuais de cada frota são obtidas como o produto dos desembarques de cada espécie, após ajuste dos seus valores de forma a não ultrapassar as quotas, e o seu preço de primeira venda. Os preços de primeira venda observados para cada espécie entre 1996 e 2009 são apresentados na Tabela 6.

Tabela 5. Quotas diárias por espécie e por barco incluías no modelo (em kg/viagem).

Quotas	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005 – 2009
Amêijoa-branca (QE_{Ss}^t)	--	--	--	--	--	200	200	400	400	225
Conquilha (QE_{Dt}^t)	--	--	--	--	--	220	150	150	150	150
Navalha (QE_{Es}^t)	--	--	--	--	--	100	50	50	50	30
Pé-de-burrinho (QE_{Cg}^t)	--	--	--	--	--	100	200	200	200	250
Quotas para o total de capturas na frota local (QT_{FL}^t)	200	200	120	120	200	300	300	600	600	390
Quotas para o total de capturas na frota costeira (QT_{FC}^t)	200	200	200	200	200	300	300	600	600	390

Tabela 6. Preço de primeira venda por espécie.

Preço de primeira venda	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Amêijoa branca (PPV_{Ss}^t)	0,86	1,05	1,01	0,83	0,71	0,73	1,00	0,79	0,67	0,67	0,67	0,67	1,18	1,33
Conquilha (PPV_{Dt}^t)	2,45	2,16	1,54	1,69	1,92	2,04	2,07	2,09	2,09	2,08	2,07	2,12	3,06	3,42
Navalha (PPV_{Es}^t)	1,27	1,68	1,28	2,65	2,84	2,51	2,66	0,00	2,00	2,53	2,66	n.d.	n.d.	n.d.
Pé de burrinho (PPV_{Cg}^t)	1,97	1,24	0,94	1,54	1,83	1,98	1,99	1,91	2,00	1,99	2,00	1,99	2,01	2,00

No modelo de Dinâmica de Sistemas, considerou-se que os preços de primeira venda aumentariam 2,5% por ano, usando os preços de 2009 como valores de base. Este aumento está próximo do valor médio da inflação observada em Portugal entre 1996 e 2009, sendo por isso razoável usar este valor como estimativa do ajuste dos preços no futuro.

O custo global foi estimado considerando tanto os custos variáveis como os custos fixos. Os custos fixos incluem a manutenção dos equipamentos e a renovação anual das licenças, e assumiu-se que eram iguais a 420€ anuais por barco. Os custos variáveis contemplam o custo de combustível, o seguro da tripulação, comissões pagas às autoridades portuárias, contribuições pagas à segurança social e a associações de produtores. O custo de combustível é obtido pela multiplicação entre o combustível consumido por viagem, o número de barcos, o número de viagens realizadas por barco e o preço do combustível. O consumo de combustível assumiu-se como igual ao valor médio estimado dos registos existentes, com o valor de 81 litros para uma viagem de um barco da frota local e de 122 litros para uma viagem de um

barco costeiro. O histórico de registos dos preços dos combustíveis e do consumo de cada barco por viagem, para cada frota, é apresentado na Tabela 7.

Tabela 7. Preço de combustível (€/litro) e consumo de combustível (litros/viagem) implementados no modelo em cada ano.

	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Preço de combustível (PC^t)	0,268	0,273	0,268	0,268	0,325	0,246	0,298	0,312	0,373	0,493	0,546	0,550	0,678	0,492
Consumo de combustível na FL (CC_{FL}^t)	n.d.	n.d.	n.d.	n.d.	n.d.	93	65	67	80	59	102	92	n.d.	92
Consumo de combustível na FC (CC_{FC}^t)	n.d.	n.d.	n.d.	n.d.	n.d.	143	116	114	131	118	147	122	n.d.	84

A evolução futura do preço de combustível foi modelada através de uma regressão linear ($R^2=0,748$; $p=0,000$), apresentada na expressão (3).

$$PC^t = 0,1693 + 0,0288 \times t \quad (3)$$

As restantes componentes dos custos variáveis foram estimadas como uma proporção da receita anual dos barcos (20%).

Por fim, o lucro por barco foi definido como a diferença entre as receitas e os custos. O salário por pescador pode ser obtido pela divisão do lucro de cada barco pelo número de membros da tripulação respetiva. O número de membros da tripulação de cada barco foi considerado constante ao longo dos anos: um barco local é operado por 2 pescadores e um barco costeiro por 3 pescadores.

5 Resultados das simulações

5.1 Validação

Para verificar se a especificação do modelo de Dinâmica de Sistemas permite representar adequadamente a evolução real do sistema, fez-se a validação do modelo comparando os valores obtidos para duas variáveis (*stock* biológico e salário anual por pescador) pela simulação do modelo com os dados históricos respetivos, no período compreendido entre 1996 e 2009.

Relativamente ao crescimento natural da amêijoia-branca e do pé-de-burrinho, o modelo simulou anos de pico de forma coincidente com os anos em que efetivamente foram observados picos (i.e. 2001 e 2003 para a amêijoia-branca; 2001 e 2004 para o pé-de-burrinho). No caso da conquinha, como os dados anteriores a 2002 foram considerados atípicos para a estimação dos valores futuros do crescimento natural, na validação do modelo por comparação com os dados históricos esta variável foi representada pela equação (4) para modelar a tendência crescente do crescimento natural entre 1996 e 2002.

$$CN_{Dt}^t = 244,953 + 38,837 \cdot t \quad (4)$$

As quotas por espécie e as quotas para o total de capturas foram incluídos no modelo seguindo a legislação respetiva em vigor (ver Tabela 3).

O preço de primeira venda foi definido como a média dos valores observados entre 1996 e 2009. Note-se que o aumento de 2,5% no preço de primeira venda apenas foi incluído no modelo a partir de 2010, inclusive.

A Figura 6 apresenta os resultados obtidos com a simulação relativamente ao *stock* biológico das quatro espécies e os respetivos dados históricos.

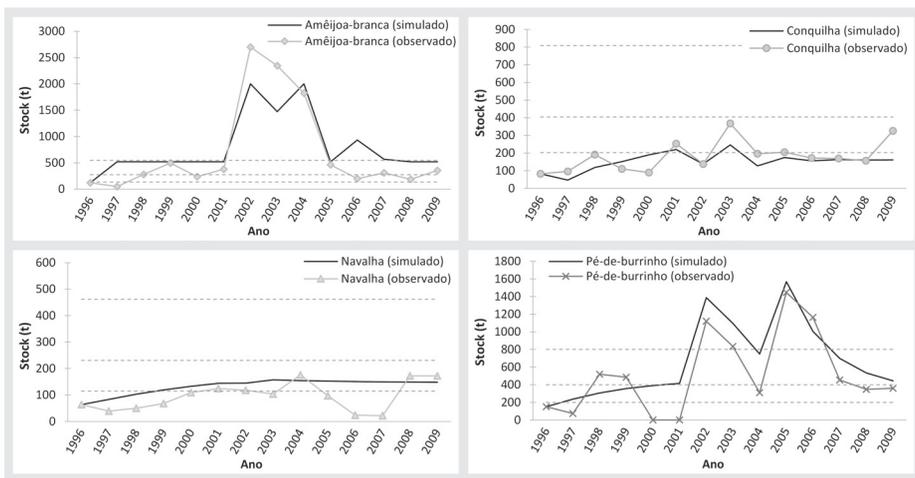


Figura 6. Evolução do *stock* das espécies - alvo.

Em relação à evolução do *stock* da amêijoa-branca, os resultados da simulação mostram que o *stock* se mantém estável entre 1996 e 2001, com um valor próximo das 500 toneladas, com a ocorrência de picos de crescimento natural em 2002 e 2004. A partir de 2005, o *stock* retoma os valores observados nos anos anteriores à ocorrência dos picos, permanecendo estável a partir daí.

Os resultados da simulação do *stock* biológico para a conchilha indicam que o modelo replica de forma bastante satisfatória a evolução desta variável, mostrando que o *stock* se mantém estável, com valores típicos inferiores a 200t.

Em relação à navalha, a simulação também obteve valores bastante próximos dos observados, com exceção dos anos de 2006 e 2007, em que o *stock* observado foi excepcionalmente baixo.

Finalmente, a simulação do *stock* de pé-de-burrinho evidencia que os resultados obtidos são bastante semelhantes aos observados para todos os anos analisados.

Da análise da Figura 6 pode concluir-se que os resultados do modelo seguem bem as tendências observados na evolução do *stock*, sendo o melhor ajuste observado no caso do *stock* da amêijoa-branca e do pé-de-burrinho. Em relação à magnitude dos resultados do modelo comparados com os valores observados, o ajuste também é aceitável, sendo o melhor ajuste correspondente ao *stock* da amêijoa-branca e conchilha. Para estas espécies, a diferença percentual entre os valores observados e as previsões do modelo é cerca de 15%. Os resultados da navalha são os piores em termos de comparação dos valores médios de *stock* observado e previsto, uma vez que a diminuição do *stock* entre 2005 e 2007 não foi prevista pelo modelo de Dinâmica de Sistemas.

A Figura 7 apresenta os resultados obtidos com o modelo de simulação para o salário anual por pescador, assim como os salários observados em cada frota. Esta Figura inclui a tracejado a indicação do valor do salário mínimo nacional, como valor de referência. Note-se que como não foi possível obter dados históricos do consumo de combustível anteriores a 2001, não é possível calcular o valor observado da rentabilidade das embarcações e salário médio por pescador nesse período.

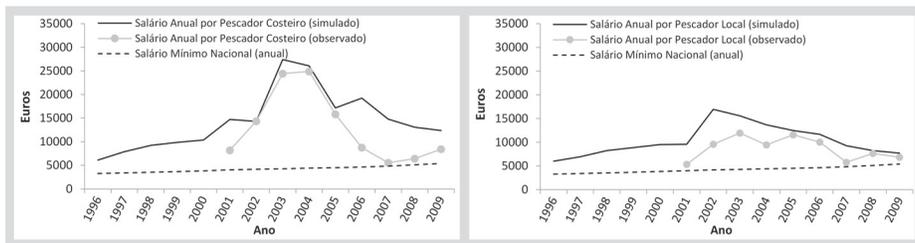


Figura 7. Evolução do salário anual por pescador de cada frota. Salário mínimo nacional anual incluído como referência (linha a tracejado).

Na Figura 7 pode-se observar que o ajustamento entre os valores simulados e observados foi particularmente bom entre 2002 e 2005 para a frota costeira, e entre 2005 e 2009 para a frota local. O ajuste é aceitável em relação à previsão das tendências de evolução do *stock*, embora em termos de magnitude os resultados previstos sejam, em média, cerca de 30% superiores aos observados.

5.2 Simulação de cenários futuros

5.2.1 Cenário base

O cenário base considera que o sistema evolui segundo as relações descritas na secção 4.2. Este cenário foi simulado para o período compreendido entre 2009 e 2020. Os resultados obtidos são utilizados para a avaliação da sustentabilidade biológica, económica e social do sistema, e consideram que as condições de regulação desta pescaria observadas em 2009 se manteriam nos anos subsequentes.

A Figura 8 apresenta o *stock* simulado para cada espécie alvo. Esta Figura inclui ainda três linhas a tracejado, que separam quatro zonas distintas relativas ao estado do *stock* para cada espécie (níveis mau, médio, bom e muito bom).

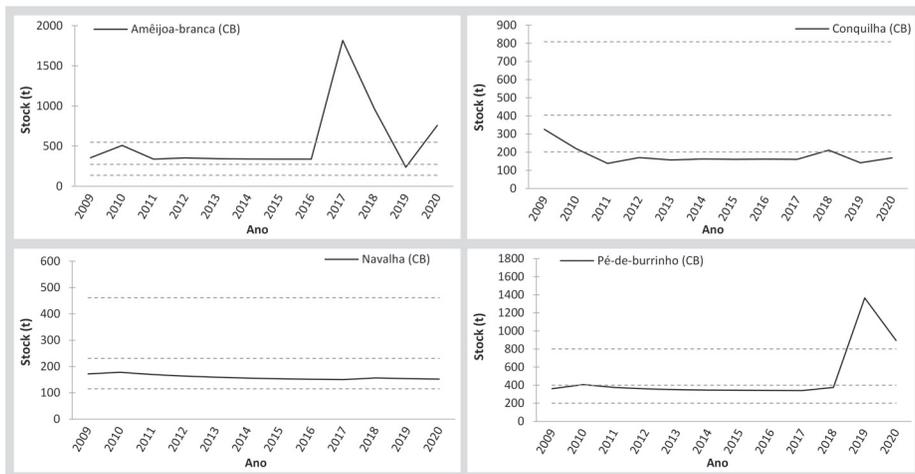


Figura 8. *Stock* simulado para cada espécie alvo no cenário base (CB).

Para a espécie amêijoia-branca, o modelo prevê um nível de *stock* estável e bom entre 2009 e 2016. Foi simulada a ocorrência de um pico do crescimento natural desta espécie em 2017, que originou um aumento significativo no nível de *stock* levando a que o mesmo permanecesse num nível muito bom durante dois anos.

O *stock* biológico da conchilha permanece baixo durante todo o período de simulação, com uma tendência estável. Apesar do *stock* se encontrar num nível baixo, a estabilidade observada nesta variável evidencia que a sustentabilidade biológica da conchilha não está comprometida.

Para a espécie navalha, o *stock* biológico permanece num nível médio ao longo dos anos simulados. Quanto ao pé-de-burrinho, prevê-se um *stock* de nível médio para o mesmo período. Para esta espécie simulou-se um pico de crescimento natural no ano de 2019, o que levou ao aumento do *stock* respetivo (atingindo um nível muito bom).

Em conclusão, o cenário base sugere que o sistema terá uma evolução sustentável ao nível biológico nos anos futuros, contendo três espécies em níveis de *stock* médio e bom, e uma espécie num nível baixo mas com evolução estável de *stock*.

Relativamente à evolução do salário anual por pescador de cada frota, os resultados obtidos com a simulação do modelo estão apresentados na Figura 9.

A linha a tracejado representa o salário mínimo nacional anual em 2009, que é tido como referência para a avaliação da sustentabilidade económica do sistema.

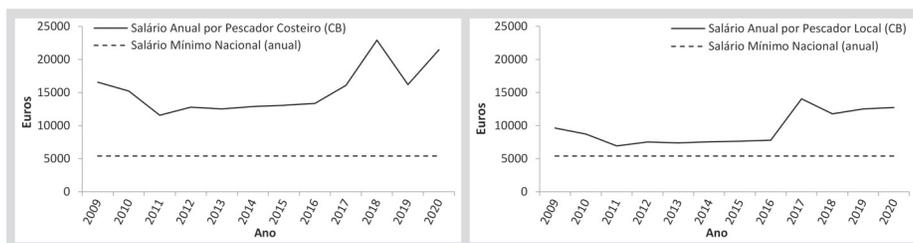


Figura 9. Salário anual por pescador de cada frota simulado no cenário base (CB).

A Figura 9 mostra que se espera que ambas as frotas sejam economicamente sustentáveis, uma vez que o salário anual por pescador simulado é sempre superior ao salário mínimo nacional anual.

De seguida apresenta-se a análise de dois cenários desenvolvidos para testar o impacto de medidas regulatórias pensadas para melhorar a sustentabilidade do sistema.

5.2.2 Cenário 1

O presente cenário pretende avaliar o impacto da redução do esforço de pesca nas populações das espécies alvo, assim como na remuneração média anual dos pescadores. O objetivo é promover o aumento de biomassa explorável disponível de modo a que se reflita, positivamente, no salário dos pescadores. A redução do esforço de pesca é feita à custa do abate de 10 embarcações (5 para cada segmento de frota) e implica a deterioração da sustentabilidade social do sistema com a perda 25 postos de trabalho (10 pescadores locais e 15 pescadores costeiros), o que representa aproximadamente 20% do número total de postos de trabalho.

Os resultados obtidos relativamente à evolução do *stock* de cada espécie são apresentados na Figura 10, que também inclui os resultados homólogos obtidos no cenário base para permitir uma análise comparativa.

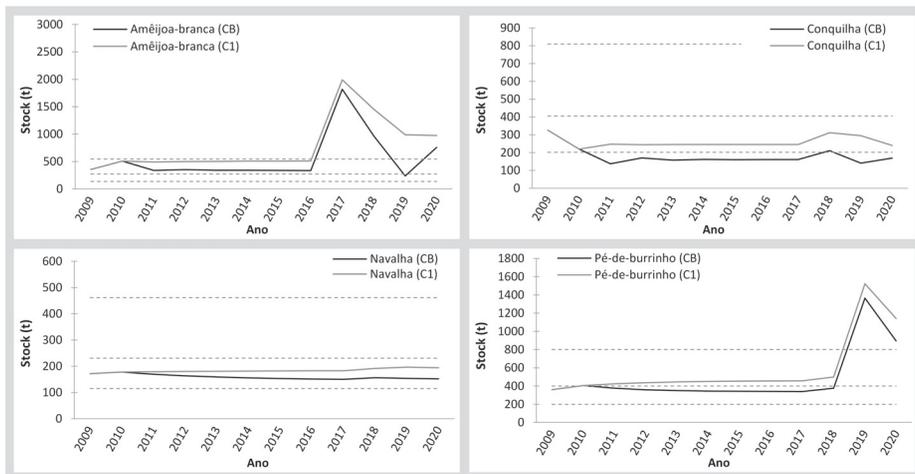


Figura 10. *Stock* simulado de cada espécie alvo no cenário base (CB) e no cenário 1 (C1).

No âmbito deste cenário todas as espécies aumentaram o *stock* biológico, como já era esperado devido à redução do esforço de pesca. Os resultados relativos ao salário anual obtido por pescador estão apresentados na Figura 11.

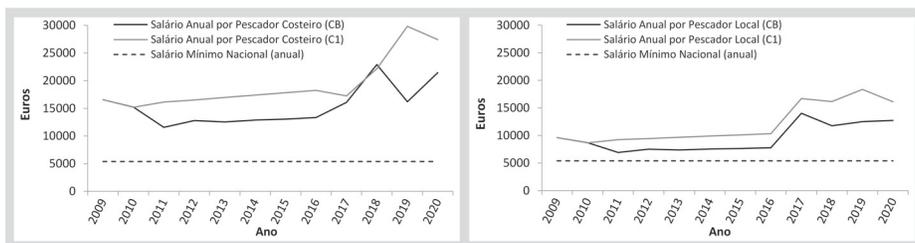


Figura 11. Salário anual por pescador de cada frota simulado no cenário base (CB) e no cenário 1 (C1).

O lucro anual por pescador obtido no cenário 1 é maior do que o observado no cenário base (o aumento de *stock* biológico observado em todas as espécies permite que cada barco consiga capturar volumes maiores de pescado). Este cenário assegura a sustentabilidade biológica e económica, em detrimento de uma redução induzida na sustentabilidade social.

5.2.3 Cenário 2

Neste cenário simula-se a interdição da exploração da amêijoia-branca durante um ano (em 2010), com o objetivo de promover o aumento da abundância desta espécie através da criação de condições para o sucesso da desova e recrutamento da espécie. Este cenário surge como consequência do baixo número de juvenis estimados em 2009 o que poderá colocar em causa a sustentabilidade do recurso, caso o esforço de pesca sobre a biomassa explorável não for reduzido significativamente. Neste sentido, procura-se avaliar o efeito daquela medida nos *stocks* das espécies comerciais, bem como na remuneração média anual dos pescadores. Este cenário pressupõe que a interdição de captura de uma espécie levaria a que o preço de primeira venda respetivo aumentasse em 10% no ano de interdição, e que este novo preço se manteria nos anos seguintes. Neste cenário considera-se ainda que o esforço de pesca direcionado às restantes espécies aumentaria, originando um aumento dos desembarques respetivos de 15%.

A evolução do *stock* biológico obtida com a simulação deste cenário está apresentada na Figura 12.

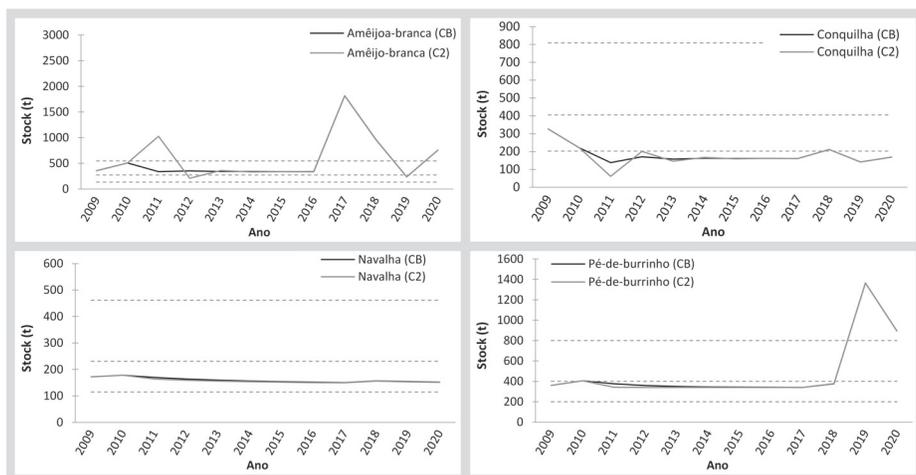


Figura 12. *Stock* de cada espécie alvo simulado no cenário base (CB) e no cenário 2 (C2).

Os resultados obtidos com a simulação sugerem que o *stock* de amêijoia-branca aumentaria em 2011 devido ao período de interdição imposto em 2010.

Esta recuperação de *stock* seria perdida em 2012 devido à reabertura da pesca direcionada à amêijoia-branca. Os *stocks* das espécies navalha e pé-de-burrinho apresentariam descidas marginais em 2011, ao passo que a conquilha desceria significativamente em 2011. Todas as espécies tendem a alcançar os valores observados no cenário base poucos anos depois do período de interdição (i.e. aproximadamente em 3 anos). A Figura 13 apresenta o lucro por pescador para ambas as frotas.

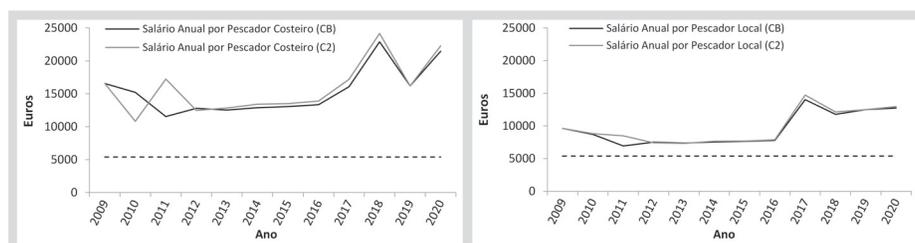


Figura 13 . Salário anual por pescador de cada frota simulado no cenário base (CB) e no cenário 2 (C2).

O cenário 2 leva à redução do lucro anual por pescador, particularmente para a frota costeira no ano de 2010, indicando que a interdição da pesca da amêijoia-branca não seria compensada economicamente pelo esforço de pesca redirecionado às restantes espécies. O aumento de receita observado após 2010 está relacionado com o aumento do preço de primeira venda da amêijoia-branca assumido no modelo. Além disso, a interdição de pesca de amêijoia-branca em 2010 potenciará a recuperação do seu nível de *stock*, o que permitiria um aumento nas capturas desta espécie findo o período de interdição e um conseqüente aumento no lucro obtido. Este cenário revela-se uma alternativa interessante para a abordagem de questões de sustentabilidade biológica com impacte no curto prazo.

6 Conclusões

Neste trabalho a metodologia de Dinâmica de Sistemas foi aplicada ao estudo da evolução da pesca artesanal de bivalves com ganchorra na costa sul Portuguesa.

A modelação de sistemas de pesca com recurso à Dinâmica de Sistemas potencia o conhecimento do sistema, assim como a avaliação dos impactes de diferentes medidas regulatórias introduzidas no mesmo. Esta avaliação pode ser realizada através da análise de indicadores de sustentabilidade ambiental, económica e social.

As variáveis seleccionadas para construir o modelo de simulação foram organizadas de acordo com a metodologia DPSIR. As relações estabelecidas entre as variáveis seleccionadas foram baseadas na análise do histórico de informação existente.

A validação do modelo foi realizada através da comparação dos resultados da simulação com os valores observados para o período compreendido entre 1996 e 2009.

A simulação do cenário base mostra a evolução prevista do sistema, assumindo que as condições regulatórias se manteriam como observado no passado. Esta análise não antecipa problemas de sustentabilidade, embora o *stock* de conquinha deva ser acompanhado com especial atenção, visto que se prevê que se mantenha num nível baixo, apesar de estável. Os cenários 1 e 2 permitem avaliar os impactes de diferentes medidas regulatórias pensadas para melhorar a sustentabilidade ecológica das espécies, bem como as condições económicas e sociais da comunidade piscatória.

A formulação e análise de cenários fornece informação relevante para a avaliação do impacto de potenciais medidas regulatórias, pelo que os modelos de simulação são ferramentas importantes a considerar nos processos de tomada de decisão. Como complemento, a divulgação dos resultados obtidos pode ser usada para atrair o interesse da população em geral sobre as pescas artesanais e para fomentar debates sobre a regulação desta atividade. Além disso, pode ajudar os pescadores a entender e aceitar as opções seguidas pelas autoridades administrativas relativamente à regulação das pescas artesanais. Assim, este tipo de modelos pode ser uma ferramenta fundamental para o setor já que poderá ser utilizada para implementar um sistema de cogestão onde o setor assumiria uma maior responsabilidade na gestão da pescaria. Importa, contudo, referir a importância de obter um maior conjunto de informação de modo a aumentar a robustez das previsões.

Agradecimentos

Este estudo foi realizado no âmbito do projeto «Desarrollo sostenible de las pesquerías artesanales del Arco Atlántico» (PRESPO) do Programa INTERREG IV B - Atlantic Area, cofinanciado pela União Europeia (ERDF Programme).

REFERÊNCIAS

Arquitt, S. e Johnstone, R. (2008). Use of system dynamics modelling in design of an Ambiental restoration banking institution. *Ecological Economics* 65(1), 63-75.

BenDor, T., Scheffran, J. e Hannon, B. (2009). Ecological and Economic sustainability in fishery management: A multi-agent model for understanding competition and cooperation. *Ecological Economics* 68(4), 1061-1073.

Bueno, N. e Basurto, X. (2009). Resilience and collapse of artisanal fisheries: a system dynamics analysis of a shellfish fishery in the Gulf of California, Mexico. *Sustainability science* 4(2), 139-149.

Caddy, J. F. (2002). Limit reference points, traffic lights, and holistic approaches to fisheries management with minimal stock assessment input. *Fisheries Research*, 56(2), 133-138.

Checkland, P. (1999). *Systems Thinking, Systems Practice: Includes a 30-Year Retrospective*. Wiley. New York.

Dudley, R. G. (2008). A basis for understanding fishery management dynamics. *System Dynamics Review*, 24(1), 1-29.

EEA – European Ambiental Agency (1999). *Ambiental indicators: Typology and overview*. EEA Technical report no. 25, Copenhagen.

EEA – European Ambiental Agency (2002). *An indicator-based approach to assessing the Ambiental performance of European marine fisheries and aquaculture*. EEA Technical report no. 87, Copenhagen.

FAO - Food and Agriculture Organization of the United Nations (1999). *Indicators for Sustainable Development of Marine Capture Fisheries*. FAO technical Guidelines for Responsible Fisheries, No. 8, Rome.

FAO - Food and Agriculture Organization of the United Nations (2010). *The State of World Fisheries and Aquaculture*. FAO Fisheries and Aquaculture Department, 197 p., Rome.

Few, A.A. (1996). *System Behavior and System Modeling*. University Science Books, Sausalito, California (USA). Third Edition.

Ford, A. (1999). *Modeling the Environment: an Introduction to System Dynamics Modeling of Ambiental Systems*. Island Press, Washington, D.C.

Forrester, J.W. (1991). System dynamics and the lessons of 35 years. *The systemic basis of policy making in the 1990s*, 29.

Forrester, J.W. (2007). System dynamics—a personal view of the first fifty years. *System Dynamics Review*, 23(2 - 3), 345-358.

Garrity, E.J. (2011). System Dynamics Modeling of Individual Transferable Quota Fisheries and Suggestions for Rebuilding Stocks. *Sustainability* 3(1), 184-215.

Jackson, J.B., Kirby, M.X., Berger, W.H., Bjorndal, K.A., Botsford, L.W., Bourque, B.J., Bradbury, R.H., Cooke, R., Erlandson, J., Estes, J.A., Hughes, T.P., Kidwell, S., Lange, C.B., Lenihan, H.S.,

Pandolfi, J.M., Peterson, C.H., Steneck, R.S., Tegner, M.J. e Warner, R.R. (2001). Historical overfishing and the recent collapse of coastal ecosystems. *Science*, 293(5530), 629-637.

King, J.R. e McFarlane, G.A. (2006). A framework for incorporating climate regime shifts into the management of marine resources. *Fisheries Management and Ecology*, 13(2), 93-102.

Knudsen, S., Zengin, M. e Koçak, M.H. (2010). Identifying drivers for fishing pressure. A multidisciplinary study of trawl and sea snail fisheries in Samsun, Black Sea coast of Turkey. *Ocean & Coastal Management* 53(5), 252-269.

Lazar, A.N., Butterfield, D., Futter, M.N., Rankinen, K., Thouvenot-Korppoo, M., Jarritt, N., Lawrence, D.S.L., Wade, A.J. e Whitehead P.G. (2010). An assessment of the fine sediment dynamics in an upland river system: INCA-Sed modifications and implications for fisheries. *Science of the Total Environment* 408(12), 2555–2566.

Malakoff, D. e Stone, R. (2002). Scientists recommend ban on North Sea cod. *Science*, 298(5595), 939-939.

Mangi, S.C., Roberts, C.M. e Rodwell, L.D. (2007). Reef fisheries management in Kenya: Preliminary approach using the driver – pressure – state – impacts – response (DPSIR) scheme of indicators. *Ocean & Coastal Management* 50(5), 463–480.

Mason, F. (2002). The Newfoundland cod stock collapse: a review and analysis of social factors. *Electronic Green Journal*, 1(17).

Mchich, R., Auger, P., & Raissi, N. (2000). The dynamics of a fish stock exploited in two fishing zones. *Acta Biotheoretica*, 48(3), 207-218.

Moxnes, E. (2005). Policy sensitivity analysis: simple versus complex fishery models. *System Dynamics Review* 21(2), 123–145.

OECD – Organisation for Economic Cooperation and Development (1993). OECD core set of indicators for Ambiental performance reviews. OECD Ambiental Monographs no. 83, Paris.

Otto, P. e Strubenb, J. (2004). Gloucester Fishery: insights from a group modeling intervention. *System Dynamics Review* 20(4), 287–312.

Rosenhead, J. (1989). *Rational Analysis for a Problematic World: Problem Structuring Methods for Complexity, Uncertainty and Conflict*. Wiley, New York.

Rothschild, B.J., Chen, C. e Lough, R.G. (2005). Managing fish stocks under climate uncertainty. *ICES Journal of Marine Science: Journal du Conseil*, 62(7), 1531-1541.

Sterman, J.D. (2000). *Business dynamics: systems thinking and modeling for a complex world*. New York: McGraw-Hill.

Tedeschi, L.O., Nicholson, C.F. e Rich, E. (2011). Using System Dynamics modelling approach to develop management tools for animal production with emphasis on small ruminants. *Small Ruminant Research*, 98(1-3), 102-110.

Wang, X., Zhang, J., Liu, J., Wang, G., He, R., Elmahdi, A. e Elsayah, S. (2011). Water resources planning and management based on system dynamics: a case study of Yulin city. *Environment, Development and Sustainability* 13(2), 331 – 351.

Yndestad, H. e Stene, A. (2002). System dynamics of the Barents Sea capelin. *ICES Journal of Marine Science: Journal du Conseil*, 59(6), 1155-1166.

Parte

II

SETOR da

INDÚSTRIA

(Página deixada propositadamente em branco)

CASO 5

MODELAÇÃO E SIMULAÇÃO DE UMA LINHA DE PRODUÇÃO DE PAINÉIS SOLARES

Ana Raquel Xambre

CIDMA – Centro de Investigação e Desenvolvimento em Matemática e Aplicações
Departamento de Economia, Gestão e Engenharia Industrial (DEGEI)
Universidade de Aveiro
raquelx@ua.pt

Ana Luísa Ramos

GOVCOPP – Unidade de Investigação em Governança,
Competitividade e Políticas Públicas
Departamento de Economia, Gestão e Engenharia Industrial (DEGEI)
Universidade de Aveiro
aramos@ua.pt

João Miguel Nicolau

Departamento de Economia, Gestão e Engenharia Industrial (DEGEI)
Universidade de Aveiro
nicolau@ua.pt

Helena Alvelos

CIDMA – Centro de Investigação e Desenvolvimento em Matemática e Aplicações
Departamento de Economia, Gestão e Engenharia Industrial (DEGEI)
Universidade de Aveiro
helena.alvelos@ua.pt

RESUMO

O estudo de caso “Modelação e Simulação de uma Linha de Produção de Painéis Solares” foi desenvolvido em ambiente industrial e resultou de uma colaboração entre uma PME portuguesa da área das energias renováveis e a Universidade de Aveiro (Departamento de Economia, Gestão e Engenharia Industrial). O seu principal objetivo foi o de analisar uma linha de produção de painéis solares fotovoltaicos no sentido de apoiar decisões operacionais relacionadas com um desejado aumento da capacidade produtiva da referida linha.

Os decisores (elementos da equipa de gestão da PME) pretendiam dar resposta a um aumento verificado na procura de painéis solares e, como tal, desejavam averiguar se um conjunto de alterações a introduzir na respetiva linha de produção, que já discutiam há algum tempo, poderia aumentar o seu volume de produção. Para tal foi inicialmente realizada a modelação do sistema usando a linguagem gráfica SysML (Systems Modeling Language) sendo, de seguida, desenvolvido um modelo de simulação por forma a testar e analisar dinamicamente o desempenho do sistema produtivo face à introdução das alterações pretendidas.

Após validação do modelo de simulação do sistema atual definiu-se o cenário alternativo a testar (correspondente às alterações a introduzir na linha) e analisaram-se os resultados associados a cada um destes cenários. Todos os modelos de simulação desenvolvidos foram acompanhados pelos respetivos modelos de animação que permitiram visualizar, de um modo apelativo, o comportamento dinâmico da linha de produção e comunicar os resultados de um modo mais eficiente.

Os resultados obtidos permitiram apoiar o processo de decisão e evidenciaram, perante os decisores, as vantagens da utilização de modelos de engenharia (modelos gráficos e de simulação) para analisar situações reais complexas. A empresa identificou-se com os modelos gerados e com os resultados obtidos. Esses resultados estão a ser utilizados para fazer pequenos ajustes aos investimentos inicialmente previstos e para introduzir as modificações na linha de produção de painéis solares fotovoltaicos. Os objetivos da Universidade foram também alcançados uma vez que o projeto foi completado com sucesso e os resultados do mesmo estão a ser utilizados pela empresa para melhorar o seu desempenho.

PALAVRAS-CHAVE.

Modelação, Simulação, Apoio à Decisão

1. Introdução

O caso de aplicação apresentado foi desenvolvido em ambiente industrial e resultou de uma colaboração, no âmbito de um projeto de mestrado, entre uma PME portuguesa da área das energias renováveis e a Universidade de Aveiro (Departamento de Economia, Gestão e Engenharia Industrial - DEGEI). O estudo incidiu sobre a análise de uma linha de produção de painéis solares fotovoltaicos com o objetivo de apoiar decisões operacionais relacionadas com um desejado aumento da capacidade produtiva da referida linha.

A empresa, recentemente instalada, viu-se perante um acréscimo significativo da procura e, para fazer face a esse aumento, precisava de elevar a sua capacidade produtiva e pretendia assim implementar algumas alterações no seu processo de fabrico. Uma vez que as alterações em questão (nomeadamente, substituição de um equipamento da linha e modificação de um procedimento operativo) envolviam investimentos consideráveis, era desejável avaliar o impacto dessas propostas antes da sua implementação e aferir se, de facto, poderiam conduzir ao resultado esperado.

A dificuldade principal era determinar se as alterações a implementar teriam ou não o resultado pretendido em termos do aumento da capacidade instalada, antes de realizar os investimentos e de forma a validar, ou não, essas mesmas despesas.

Numa primeira fase, a equipa de projeto (elementos da PME e elementos do DEGEI) decidiu efetuar a modelação do sistema em causa. Esta tarefa, levada a cabo através de uma linguagem gráfica de modelação particularmente adequada à representação de sistemas com considerável dimensão e complexidade (SysML - *Systems Modeling Language*), possibilitou um melhor entendimento do problema em análise, a clara definição das fronteiras do estudo, a identificação das variáveis a contemplar, a clarificação da linguagem de projeto e uma comunicação mais eficiente entre todos os membros da equipa.

Numa segunda fase, decidiu-se desenvolver um estudo de simulação de modo a analisar dinamicamente o sistema e a tentar apoiar as decisões em causa. Optou-se pela simulação como “a ferramenta ideal” para dar resposta ao problema, uma vez que a modelação analítica apresentava algumas limitações,

nomeadamente: (i) a linha de produção de painéis apresentava um conjunto de diferentes problemas interrelacionados cuja resolução sequencial inviabilizava um processo global de decisão adequado, (ii) a avaliação da linha de produção através de um conjunto alargado de medidas de desempenho limitava a aplicabilidade das técnicas analíticas que, normalmente, somente consideram um único fator de desempenho ou um número muito limitado destes fatores, (iii) a inclusão de parâmetros estocásticos, tais como, tempos de processamento e de transporte ou avarias de máquinas, era difícil de incorporar na modelação analítica e (iv) as técnicas analíticas não refletiam o comportamento dinâmico dos sistemas o que dificultava, por exemplo, o dimensionamento das áreas de armazenamento intermédias.

Deste modo, o domínio e os objetivos do estudo foram claramente definidos (com as partes interessadas), as medidas de desempenho a analisar foram criteriosamente estabelecidas (por exemplo, volume de produção da linha de produção para um determinado período de planeamento, taxas de utilização dos recursos, dimensão das áreas de armazenamento intermédias), desenvolveu-se um modelo conceptual detalhado (com a inerente recolha de dados) que foi devidamente validado, implementou-se o modelo operacional no *software* de simulação Arena[®], verificou-se e validou-se este modelo, definiram-se as condições experimentais, analisaram-se os resultados e efetuou-se a apresentação e discussão dos mesmos.

Numa fase seguinte do estudo de simulação, definiu-se o cenário alternativo a testar (correspondente às alterações a introduzir na linha e a uma versão modificada do modelo de simulação do sistema real em funcionamento, previamente validado) e analisaram-se os resultados associados a cada um destes cenários. Todos os modelos de simulação desenvolvidos foram acompanhados pelos respetivos modelos de animação que permitiram visualizar, de um modo apelativo, o comportamento dinâmico da linha de produção e comunicar os resultados de um modo mais eficiente.

Os resultados obtidos através da modelação e simulação da linha de produção foram discutidos com a equipa de projeto e com os decisores e validados pelos mesmos. Nesta fase de apresentação/discussão de resultados, os modelos de animação desenvolvidos possibilitaram uma clara identificação,

por parte dos elementos da empresa, do sistema em análise, contribuindo muito positivamente para a aceitação dos resultados e conseqüente sucesso do estudo. As alterações a introduzir na linha encontram-se em fase de implementação.

Na sua globalidade o projeto teve a duração de 9 meses, tendo sido cerca de 5 meses dedicados à modelação conceptual e recolha de dados, 2 meses utilizados na modelação operacional e outros 2 meses em experimentação e análise de resultados.

Esta experiência, que reflete um caso de sucesso de uma parceria Universidade/Empresa, evidenciou as vantagens da utilização da modelação e simulação na análise de sistemas e na melhoria do seu desempenho.

Por questões de confidencialidade, alguma informação relacionada com a empresa, com o seu produto e processo produtivo e com os resultados obtidos foi omitida, contudo acredita-se que este facto não compromete a descrição clara e detalhada do caso de aplicação.

Na próxima secção apresentar-se-á o produto e o processo de fabrico para que seja mais clara a complexidade do problema. Seguidamente será explicada a modelação do sistema em análise e o desenvolvimento do modelo de simulação. Por fim serão apresentados os resultados e algumas considerações finais.

2. Produto e Processo de Fabrico

Na ampla gama de energias renováveis destaca-se a energia solar que pode ser dividida em dois grandes grupos: solar térmica e solar fotovoltaica. Apesar de ambas terem como finalidade a transformação de energia solar, no primeiro caso o objetivo é obter calor através do aquecimento de água ou de um outro fluido, enquanto que no caso da energia solar fotovoltaica se pretende produzir eletricidade a partir da luz do Sol.

A empresa em estudo, situada na região de Aveiro, possui uma nave industrial com 3825 m² e dedica-se somente à produção de painéis solares fotovoltaicos em três gamas de potência: 40w, 42w e 44w. Na altura em que o trabalho foi desenvolvido contava com a colaboração de cerca de 150 trabalhadores, 120 dos quais na laboração por turnos, garantindo o

funcionamento durante 24 horas por dia e 7 dias por semana. A empresa tinha ainda uma capacidade instalada de 7.4 MW, com a laboração em contínuo, 353 dias por ano, correspondendo a 12 lotes diários, com perdas e rejeições na ordem dos 5%, sendo que cada lote compreende 48 painéis, o que representa cerca de 490 unidades embaladas por dia. Um painel incorpora diferentes materiais tais como vidro, alumínio, polímeros, materiais semicondutores, caixas elétricas, entre outros (Figura 1).



Figura 1: O produto: painéis solares fotovoltaicos

A implantação fabril da unidade industrial está organizada por processo (Stevenson, 2007) e encontra-se esquematizada na Figura 2. Os produtos

percorrem as diferentes secções (por exemplo, deposição, laminação), de maneira a que as operações requeridas para o seu processamento sejam efetuadas.

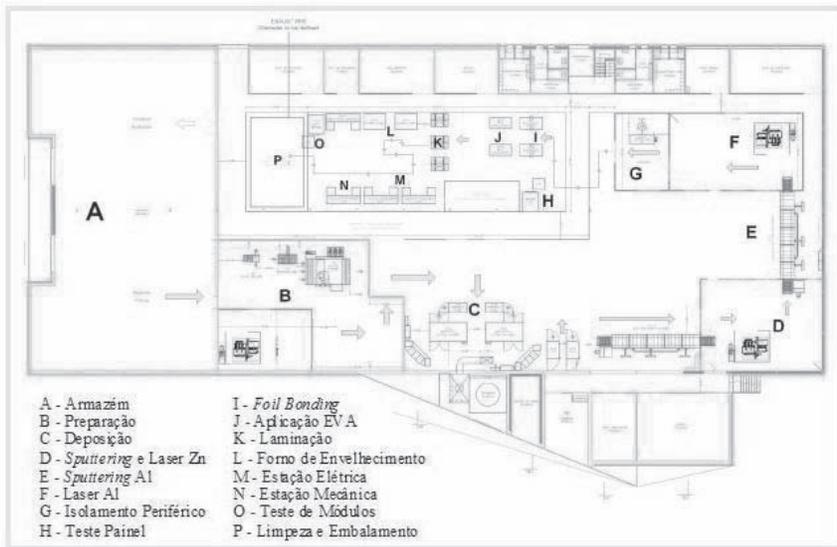


Figura 2: Implantação da empresa

A primeira fase do processo de fabrico é realizada na zona de preparação onde todos os vidros necessários à fabricação do produto são preparados. São realizadas operações como o boleamento das arestas no vidro com TCO (*Transparent Conductive Oxides*), para poder ser manuseado sem risco de corte para os operadores, lavagem do mesmo, corte a laser da camada condutora e carregamento das caixas de deposição. O vidro de fecho é apenas lavado e carregado em carros.

O vidro TCO segue depois para a deposição onde é feita a deposição do material semiconductor sobre o vidro. De seguida temos a pulverização catódica de zinco (*Sputtering Zn*) onde é depositada uma camada de zinco sobre a superfície magnetizada. É, então, feito um segundo corte a laser imediatamente ao lado do corte efetuado na zona de preparação e até à primeira camada de óxido condutor. Segue-se a pulverização catódica de alumínio (*Sputtering Al*) onde é depositada a última camada, de alumínio, que será cortada logo depois,

também a laser. O isolamento periférico serve para isolar o material semiconductor na periferia do painel.

Na zona de acabamentos realiza-se o teste de painel, que é a operação que define se o painel segue o processo até ao final, garantindo que se obterá a potência pretendida. É, também, feita a aplicação de fita de alumínio, que é o que fará os contactos elétricos do painel, e a aplicação do EVA (*Ethylene Vinyl Acetate*) que é o polímero que fica entre o painel e o vidro de fecho que fará o isolamento. Segue-se a laminação dos dois vidros, com esse polímero, e o envelhecimento no respetivo forno, que faz a recombinação do material semiconductor, estabilizando-o e dando mais potência ao painel. Finalmente, na estação elétrica, é montada a caixa elétrica e são cravados os seus terminais e, na estação mecânica, enche-se a caixa elétrica com uma mistura de silicone que também é utilizada para a colagem dos perfis de alumínio, que serão os apoios de fixação do painel.

Um outro aspeto relevante do funcionamento do sistema de produção é a forma de transporte, entre secções, dos painéis. Todos os transportes são realizados com recurso a carros (25 carros) que são movimentados pelos trabalhadores dos postos de trabalho. Estes carros são também utilizados como zona de suporte ao carregamento e descarregamento dos produtos nos diferentes equipamentos.

3. Modelação Gráfica

A modelação é uma atividade básica da atividade científica. Um modelo constitui uma representação de uma determinada parte do mundo (com ou sem existência física) que corresponde ao domínio de interesse. Essa representação pretende realçar os aspetos importantes (de um determinado ponto de vista) e simplificar ou omitir os aspetos irrelevantes. De acordo com Rumbaugh *et al.* (1999) os modelos são importantes para: (i) identificar e definir o domínio em análise de modo a que todas as partes interessadas o entendam; (ii) pensar sobre o sistema; (iii) produzir elementos de trabalho; (iv) organizar, examinar, manipular, filtrar e editar informação de sistemas de considerável dimensão; e (v) explorar soluções sob diversas dimensões (por exemplo, operacional, económica, ambiental). Sussman (2000) resume estas

funções do seguinte modo: “modelamos para entender, modelamos para explicar, modelamos para prever e modelamos para melhorar”.

Existem na literatura diversas taxonomias para a classificação de modelos. A taxonomia proposta por Buede (2009), bastante simples e concisa, considera quatro categorias de modelos: físicos, quantitativos, qualitativos e mentais. Os físicos abordam a modelação através de representações tridimensionais, à escala. Os quantitativos ou matemáticos fornecem respostas numéricas e subdividem-se em analíticos, de simulação e de opinião. Os qualitativos fornecem representações simbólicas, textuais ou gráficas. Os mentais (que podem ser considerados como estando na base de todos os outros) constituem representações mentais individualizadas do problema e das soluções. Pode dizer-se que, antes de desenvolver um modelo físico, quantitativo ou qualitativo, qualquer engenheiro/modelador tem o seu próprio modelo mental que precisa de ser “traduzido” para uma representação não ambígua que possa ser claramente partilhada e discutida por toda a equipa de trabalho (Figura 3).

A modelação gráfica ou visual enquadra-se, então, na categoria dos modelos qualitativos e utiliza elementos da teoria dos grafos ou elementos gráficos gerais para representar, por exemplo, a estrutura, o comportamento e os fluxos de um sistema. Neste estudo de caso, a modelação gráfica foi utilizada para especificar, descrever e comunicar uma visão partilhada do sistema em análise, e foi concretizada através de uma linguagem de modelação gráfica para sistemas, a linguagem SysML (*Systems Modeling Language*).

3.1 SysML – Systems Modeling Language

A linguagem de modelação gráfica SysML foi lançada em 2007, pelo OMG (Object Management Group), e corresponde a uma extensão da conhecida linguagem UML (*Unified Modeling Language*). Esta extensão ou perfil, dedicada à engenharia de sistemas, suporta a análise, especificação, desenho, verificação e validação de sistemas complexos que podem incluir *hardware*, *software*, dados, pessoas, procedimentos, instalações, equipamentos e outros elementos característicos de sistemas criados pelo homem (Friedenthal *et al.*, 2008). De acordo com David *et al.* (2010), esta linguagem permite lidar com as especificidades da modelação de sistemas (por exemplo, requisitos,

interdisciplinaridade, decomposição hierárquica, análises custo/benefício) evitando o vocabulário de *software* encontrado na linguagem UML.

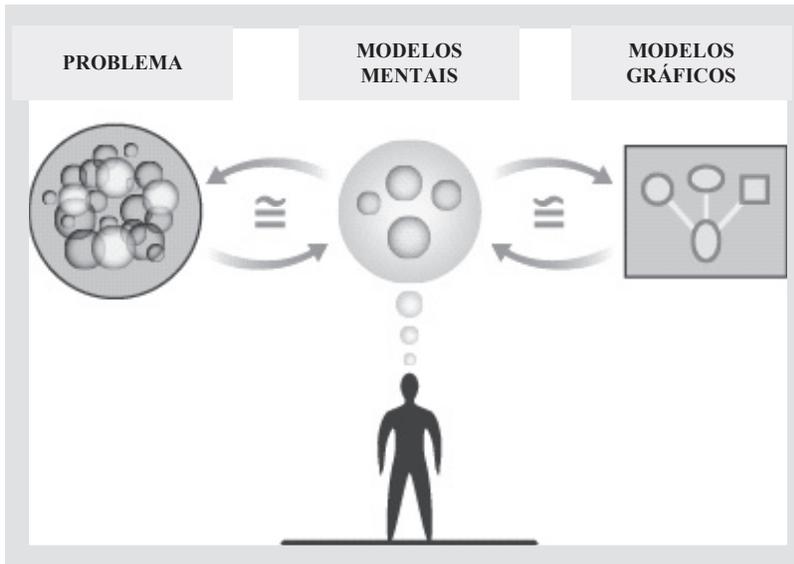


Figura 3: “Tradução” de modelos mentais para modelos gráficos (adaptado de www.idiagram.com, 2011, com permissão do autor)

As ferramentas SysML disponíveis no mercado (*Artisan Studio*, *Magic Draw+SysML plugin*, *Rational Rhapsody*, etc.) armazenam o modelo num repositório de modelos e a troca de informação do utilizador com o modelo faz-se através das representações gráficas ou diagramas. Os diagramas SysML são nove e estão organizados em quatro blocos principais que representam quatro dimensões de modelação: requisitos, estrutura, comportamento e relações paramétricas. Cada um destes blocos inclui um ou mais diagramas. Por exemplo, os requisitos são representados através de um diagrama de requisitos (req); a estrutura do sistema pode ser representada através de diagramas de pacotes (pkg), diagramas de definição de blocos (bdd) e/ou diagramas de blocos internos (ibd); o comportamento do sistema pode ser modelado através de diagramas de casos de utilização (uc), diagramas de atividade (act), diagramas de sequência (sd) e/ou diagramas de transição de estados (smd); as relações paramétricas são modeladas através de um diagrama paramétrico (par). A linguagem dispõe, também, de um conjunto de representações não gráficas

tais como tabelas de alocação e matrizes. A troca de dados e de modelos é assegurada via XMI (*XML Metadata Interchange*). Uma descrição detalhada da linguagem pode ser encontrada em Friedenthal *et al.* (2008) e Ramos (2011).

3.2 Modelação do sistema em análise

A modelação gráfica do sistema em análise (a linha de produção de painéis solares fotovoltaicos) foi levada a cabo para possibilitar a identificação e clara definição do domínio de estudo e para produzir um elemento de trabalho que pudesse ser entendido e partilhado por todos os membros da equipa (uma visão partilhada do sistema em análise). Tratou-se, fundamentalmente, de uma modelação estrutural do sistema. Os diagramas SysML foram desenvolvidos na ferramenta Artisan Studio.

Numa primeira fase, utilizou-se o diagrama de pacotes (pkg) para identificar, claramente, o sistema em análise (a linha de produção de painéis), para refletir o principal objetivo do estudo (aumentar a capacidade produtiva da linha), as principais entradas (por exemplo, materiais, equipamentos, energia) e saídas (painéis solares) do sistema e os principais clientes/atores (a equipa de projeto que incluiu a equipa da UA, os gestores de operações e supervisores de linha e os decisores da empresa) do sistema. Este diagrama (Figura 4) fornece uma visão geral do domínio de estudo, centrado no *package* 'Linha de Produção', e serviu de base para iniciar as discussões de trabalho, a um nível macro.

Após esta fase inicial definiu-se um diagrama de requisitos (req) para ilustrar quais as necessidades que os gestores da empresa manifestavam e quais as propriedades (funcionalidades, medidas de desempenho, etc.) que a linha de produção e o estudo deveriam exibir para dar resposta a essas necessidades. Pode dizer-se que o sucesso de um projeto pode ser avaliado em termos de satisfação dos requisitos manifestados pelas partes interessadas e pela realização das suas expectativas. A gestão de requisitos num estudo de simulação é fundamental para identificar as necessidades dos clientes, estabelecer prioridades, identificar conflitos, etc. Tipicamente, e à semelhança do que acontece em projetos de desenvolvimento de *software*, os maiores problemas deste tipo de estudo advêm de falhas nos requisitos.

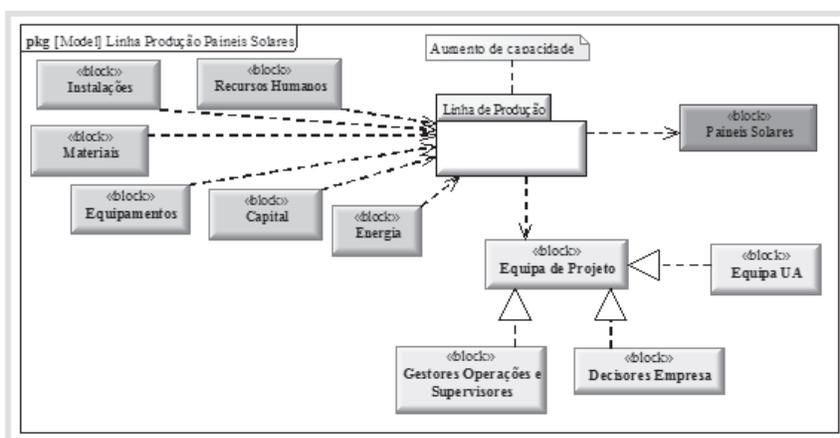


Figura 4: Diagrama de pacotes (SysML pkg) com visão geral do domínio de estudo

Este diagrama (req) é uma ferramenta de grande utilidade para a gestão de requisitos possibilitando a sua identificação, classificação, descrição, rastreabilidade, etc. A Figura 5 apresenta uma parte do diagrama de requisitos elaborado para o projeto podendo observar-se diversos requisitos e diferentes tipos de relacionamentos hierárquicos. Por exemplo, o requisito 'Medidas de Desempenho' está contido no requisito de mais alto nível 'Estudo de Simulação' e diz respeito à identificação dos principais resultados que o modelo de simulação deverá fornecer. Este diagrama foi particularmente importante para criar uma base inicial de necessidades e para identificar prioridades.

O requisito 'Medidas de Desempenho' identificado na Figura 5 foi detalhado (ligação «refine») através de um diagrama de casos de utilização (uc) que é apresentado na Figura 6. Pode ver-se que a criação de medidas de desempenho do sistema implica a identificação dessas medidas (por parte da equipa de projeto), a execução dos modelos de simulação, a verificação e validação (usualmente designado por V&V) dos resultados e a criação de gráficos e relatórios para os decisores. Este tipo de diagrama é indicado para detalhar os requisitos funcionais e permitiu, para casos particulares, definir com maior clareza a interação dos atores com o sistema de modo a alcançar determinados objetivos.

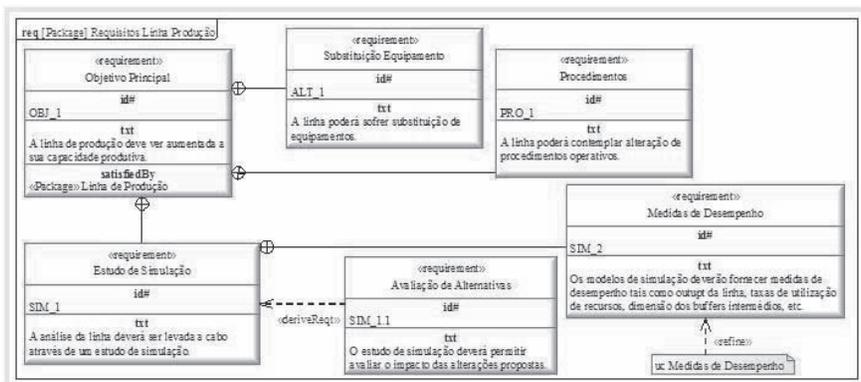


Figura 5: Diagrama de requisitos (SysML req) para a linha de produção

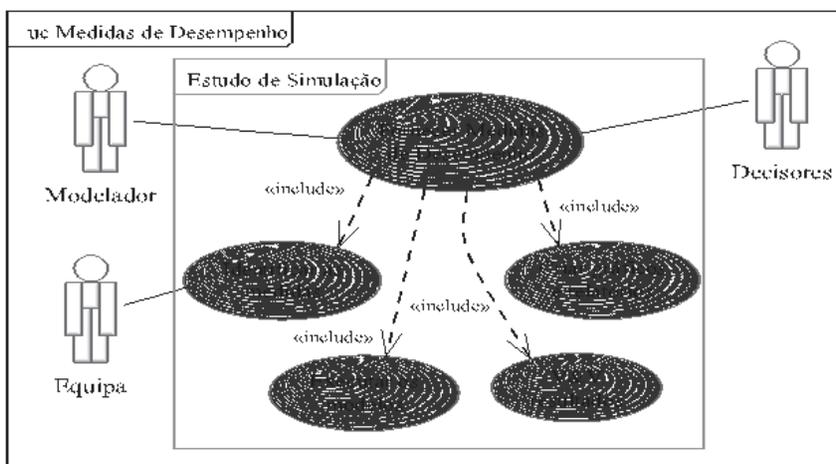


Figura 6: Diagrama de casos de utilização (SysML uc) para o requisito 'Medidas de Desempenho'

A modelação estrutural do sistema foi, ainda, complementada com um diagrama de definição de blocos (bdd). Este tipo de representação gráfica permite identificar a estrutura do sistema em análise em termos das suas relações hierárquicas e interrelações. O bloco é a unidade de estrutura fundamental da linguagem SysML. Neste caso particular o diagrama de blocos (Figura 7) foi utilizado para modelar a sequência de operações associada à linha de produção e identificar, com clareza, as operações a considerar no modelo de simulação. Este tipo de diagrama permite a inclusão de detalhes associados a cada bloco (por exemplo, a operação 'Preparação Vidro TCO')

inclui tarefas de boleamento, lavagem e corte a laser) que podem ser importantes para o estudo de simulação.

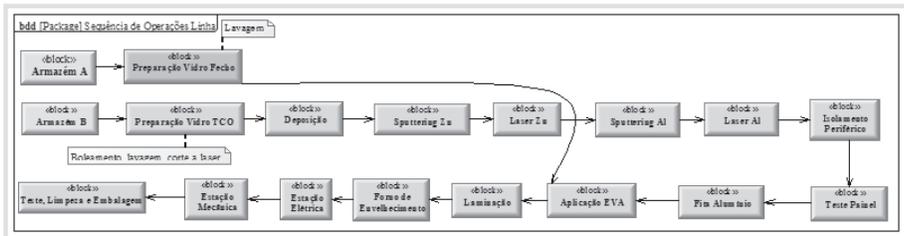


Figura 7: Diagrama de definição de blocos (SysML bdd) para a sequência de operações da linha de produção

A estrutura interna de alguns dos blocos criados foi definida em diagramas de blocos internos (ibd) que ilustram o modo como os componentes internos de um bloco se relacionam e quais os fluxos entre eles. A Figura 8 mostra um diagrama ibd para o bloco 'Preparação Vidro TCO' com inclusão de algumas fotografias que permitem representar, de um modo mais visível, os componentes presentes no diagrama. O diagrama mostra o conjunto de máquinas associadas à operação (por exemplo, máquina de lavar, máquina de lavar) e os materiais que circulam entre elas (por exemplo, vidro TCO boleado, vidro TCO lavado).

Foram, ainda, definidos outros diagramas SysML (essencialmente, diagramas bdd e diagramas ibd) para detalhar alguns aspectos estruturais críticos do sistema. A modelação gráfica do comportamento do sistema não foi desenvolvida nesta fase, uma vez que iria ser desenvolvido um estudo de simulação capaz de dar uma melhor resposta à modelação/avaliação dinâmica do sistema em análise.

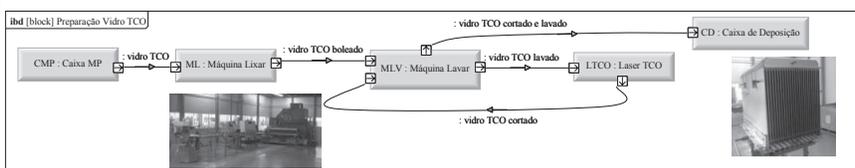


Figura 8: Diagrama de blocos internos (SysML ibd) para o bloco 'Preparação Vidro TCO'

4. Desenvolvimento do Modelo de Simulação

Dada a complexidade do sistema de produção e da associada movimentação de materiais, e considerando a variabilidade inerente às diferentes operações, a simulação surgiu, naturalmente, como a ferramenta mais adequada para responder aos objetivos do estudo, pois permite fornecer uma visão detalhada, integrada e dinâmica do sistema. Adicionalmente, a simulação permite, através da construção e análise de um modelo quantitativo, avaliar as necessidades de recursos e os procedimentos operacionais, o que se adequa perfeitamente à situação que a empresa pretendia averiguar/testar.

Neste ponto procurar-se-á descrever o processo de desenvolvimento do modelo de simulação, incluindo a definição das diferentes fases associadas ao estudo (definidas consensualmente pela maioria dos autores de referência na área tais como Law e Kelton (2000)). A descrição inclui, num primeiro ponto, a modelação conceptual do sistema (formulação do problema, definição dos objetivos, recolha de dados, etc.) e, seguidamente, a modelação operacional (implementação do modelo na plataforma de simulação, verificação, validação, etc.).

De notar que, normalmente, os estudos de simulação não seguem rigidamente a execução sequencial das diferentes fases, sendo necessário, por vezes, recolher mais dados, redefinir o modelo e/ou reformular os objetivos do estudo. No entanto, por uma questão de clareza, a informação será apresentada tendo em conta as fases referidas.

4.1 Modelação Conceptual

A primeira fase de um estudo de simulação inclui a clara definição dos objetivos do projeto, de modo a que possa ser definido o nível de detalhe apropriado para o modelo. Outras tarefas a executar durante esta etapa incluem (Law, 2009): (i) a listagem dos aspetos específicos a contemplar no modelo; (ii) a definição das fronteiras e do domínio do estudo; (iii) a seleção das medidas de desempenho do sistema relevantes para os objetivos do estudo e (iv) o planeamento do estudo em termos de recursos necessários, tempo, custos e canais de comunicação.

Na execução destas tarefas deve-se procurar garantir, por um lado, que no estudo de simulação não se irão consumir recursos na análise de cenários (ou partes de cenários) que ficam fora do âmbito do trabalho e, por outro, que todos os aspetos relevantes são incluídos, desde o início, no projeto, uma vez que, à medida que o projeto decorre, se torna cada vez mais difícil e dispendiosa a sua inclusão.

Assim, definiram-se: (i) as fronteiras e o domínio do estudo, (ii) o nível de detalhe de modelação requerido e (iii) as medidas de desempenho relevantes para a avaliação operacional do sistema. De realçar que a utilização dos modelos gráficos produzidos na fase anterior (modelos SysML) promoveu uma discussão clara e célere destes aspetos facilitando a sua definição.

Ficou determinado que o projeto tinha como objetivo principal o estudo da capacidade de produção da linha de produção da empresa, bem como das alterações com vista ao aumento da capacidade produtiva. O nível de detalhe pretendido circunscreveu-se à simulação da linha de produção com inclusão de todos os postos de trabalho, das respetivas filas de espera, dos recursos associados (máquinas, pessoas, meios de transporte de materiais, utilização de recursos humanos, etc.) e das operações de movimentação de materiais.

As seguintes medidas de desempenho, utilizadas na avaliação dinâmica do sistema, foram selecionadas em conjunto com os decisores da empresa:

- Número médio de painéis embalados por dia (N);
- Taxa de ocupação dos diversos recursos (Tr);
- Número médio de lotes completados por dia (R);
- Recursos gargalo que limitam a capacidade total do sistema.

Após esta definição inicial do problema em análise passou-se à modelação conceptual do sistema de produção e à inerente recolha de dados. Normalmente, estas duas tarefas são realizadas paralelamente uma vez que existe uma grande interação entre o desenvolvimento do modelo e a recolha dos dados necessários para definir os seus parâmetros.

A especificação do modelo conceptual de um sistema de produção requer, na maior parte dos casos, um profundo conhecimento dos detalhes do processo de fabrico que lhe estão associados, bem como das regras de controlo desse sistema (as quais, muitas vezes, não só não estão documentadas,

como não estão bem definidas). É, por isso, necessário obter a colaboração de todos atores que na empresa detêm o conhecimento detalhado destes aspetos. Por outro lado, cabe ao(s) especialista(s) de modelação/simulação (modeladores), filtrar e validar a informação que lhe(s) é fornecida, por forma a distinguir o essencial do acessório. Assim, o desenvolvimento do modelo conceptual é um trabalho de equipa, o qual é fundamental ser conduzido de forma a que o modelador fique com um conhecimento claro de todos os aspetos relevantes do processo de fabrico e das regras de controlo desse processo, a incluir no projeto de simulação. A "arte" da simulação corresponde a este processo de extração e validação do conhecimento detido pelos especialistas de diferentes setores industriais, que o modelador tem que realizar, sem que, normalmente, seja um especialista em cada uma destas áreas. Este processo de extração do conhecimento sobre o sistema de produção a analisar, traduz-se num conjunto de pressupostos que deverão ser analisados e validados pelos especialistas antes de o modelo ser implementado. Mais uma vez, a modelação gráfica produzida na fase anterior foi de grande valia neste processo de modelação conceptual uma vez que grande parte das especificidades e requisitos do sistema já estavam bem definidos, discutidos e validados pela equipa.

De acordo com os objetivos do estudo e com as fronteiras de modelação definidas, efetuou-se a recolha da informação adicional necessária ao desenvolvimento do modelo conceptual, nomeadamente:

- Implantação do sistema;
- Procedimentos operacionais (que muitas vezes não estão formalizados);
- Pressupostos.

A implantação fabril da unidade industrial foi apresentada anteriormente e encontra-se esquematizada na Figura 2. Para além da informação obtida acerca da mesma, foram também recolhidas informações acerca de procedimentos operacionais, parte deles informais, tendo-se procurado incorporar no modelo os aspetos considerados mais relevantes e que se passam a descrever:

- Apesar de haver um número de carros de transporte de painéis na ordem das 25 unidades, 8 estão destinados ao uso apenas na área da

preparação, para evitar que os stocks intermédios a montante da deposição impeçam o correto abastecimento da mesma por parte da área anterior.

- Há 4 carros destinados ao transporte de vidro de fecho, devidamente assinalados. Esta informação não foi considerada porque este transporte não influencia diretamente a produção do painel.
- Nas secções Laminação e Estação Mecânica, os carros de transporte deverão ser totalmente cheios, para maximizar a utilização do forno, bem como para diminuir o número de movimentações de carros de transporte.
- O stock intermédio anterior ao laser TCO não deverá ter mais do que 4 lotes, para permitir a operação na área da preparação.
- Não deverá haver mais do que 10 lotes em espera na operação de colocação de fita de alumínio e EVA.
- Depois da aplicação do EVA e antes da laminação não deverá haver mais que 10 painéis em espera, por questões de segurança, situação que poderia ocorrer devido ao excesso de capacidade deste posto de trabalho.

Foram também estabelecidos alguns pressupostos que permitiram, por um lado, simplificar o modelo em alguns pontos que não seriam relevantes para o estudo em causa e, por outro, modelar eficazmente a realidade. Foram então tidos em conta os seguintes aspetos:

- O número de caixas de deposição existentes é de 9, o que corresponde à realidade.
- O número de carros de painéis disponíveis para a produção é de 15. Em determinadas operações são utilizados dois carros no processo (um para carregar e outro para descarregar).
- As operações *Sputtering Zn*, Laser Zinco, *Sputtering Al* e Laser Alumínio processam-se em contínuo entre si.
- As operações Lixar e Lavar e Carregar Carro (realizadas na zona de preparação) processam-se em contínuo entre si.
- As operações Fita de Alumínio (*Foil Bonding*), Aplicação de EVA e Laminação processam-se em contínuo entre si.

- As operações Estação Elétrica e Estação Mecânica processam-se em contínuo entre si.
- As operações Teste de Módulos e Limpeza e Embalamento processam-se em contínuo entre si.
- Os operadores estão retidos e são considerados ocupados enquanto está a ser processado um painel no seu posto de trabalho e existem 23 operadores.
- Não há rejeição de painéis ao longo da linha. Todas as rejeições são consideradas no posto de teste considerando-se que a taxa de rejeição é de aproximadamente 5% (da deposição até ao embalamento), aproximação feita para facilitar a programação do modelo.
- Não foram considerados tempos de paragem para manutenção de equipamentos, nem por avarias nem por intervenções preventivas (planeadas para momentos de não laboração da empresa).
- Foi considerado o trabalho em contínuo 24h, 7 dias por semana, que corresponde à realidade atual da empresa. Não foram contempladas as mudanças de turno pois estes funcionam todos da mesma forma e com o mesmo número de trabalhadores, tendo sido os tempos de mudança assumidos como negligenciáveis.
- Não foram considerados consumos de matérias-primas.
- Os recursos humanos são libertados quando acabam o trabalho no seu posto.
- Todos os carros têm uma velocidade média de transporte de 0,7 m/s.

Os principais pressupostos, inerentes à modelação do sistema de produção analisado neste estudo, foram devidamente validados pelo gestor de produção e por outros colaboradores da empresa que auxiliaram o projeto. Este conjunto de pressupostos caracteriza o sistema a simular e define as regras de decisão normalmente utilizadas na sua operação.

Paralelamente à definição dos vários pressupostos de modelação efetuou-se a recolha dos dados necessários.

A necessidade de recursos humanos em cada operação foi feita de acordo com a Tabela 1.

Tabela 1: Distribuição de operadores pelos postos de trabalho

Operação	Nº operadores	Operação	Nº operadores
Lixar		Isolamento Periférico	
Lavar / Carregar Carro		Teste Painel	
Laser TCO		Colocação Fita Alumínio	
Lavar / Carregar Caixa		Aplicação EVA	
Forno Pré-Aquecimento	-	Laminação	
Deposição		Forno Envelhecimento	-
Arrefecimento / Desc. Caixa		Ligações Elétricas	
<i>Sputtering</i> (Zinco)		Enchimento com Silicone	
Laser (Zinco)		Teste Módulos	
<i>Sputtering</i> (Alumínio)		Limpeza/Embalamento	
Laser (Alumínio)			

A determinação dos tempos de processamento foi feita considerando a existência de operações que se realizam em contínuo (em que o lote de transferência é unitário) e outras em que se acumulam existências em curso de fabrico entre as mesmas (não contínuo).

Para o caso de operações não contínuas recorreu-se ao *software* de controlo da produção utilizado na empresa. Esse *software* faz o registo de todos os tempos de produção em todos os sectores, bem como das matérias-primas utilizadas, matérias-primas disponíveis em cada sector da linha e operadores associados a cada operação. Assim, foram recolhidos os tempos de 50 lotes ao longo de todo o processo tendo-se, contudo, excluído alguns desses dados por corresponderem a situações fora do normal, tais como avarias nos equipamentos, lançamentos incorretos no sistema e erro humano no lançamento dos lotes.

Nas operações contínuas o tempo registado no sistema de informação da empresa refere-se à duração da última operação. Para estes casos, os tempos correspondentes às diversas operações foram recolhidos manualmente em amostras de dimensão nunca inferior a 20. A seleção das distribuições teóricas mais adequadas foi efetuada através da utilização do módulo de análise *Input Analyzer* (incluído no *software* de simulação utilizado). Este módulo possibilita a seleção de diferentes distribuições de probabilidade para teste, a estimativa dos valores dos parâmetros das distribuições, a avaliação da qualidade do

ajuste obtido através de testes de hipóteses (teste Qui-Quadrado e teste Kolmogorov-Smirnov) e a visualização gráfica da curva da distribuição selecionada sobre o histograma dos dados.

Na Figura 9 pode observar-se um exemplo do procedimento utilizado para o ajuste de distribuições teóricas de probabilidade aos dados recolhidos. Da esquerda para a direita pode ver-se, para uma determinada operação (*Sputtering Al*), a amostra inicial, o histograma e a distribuição ajustada (neste caso, uma distribuição Triangular de parâmetros 2.18, 2.73 e 4) e os respetivos testes de qualidade de ajuste.

Foram considerados 21 conjuntos de dados para as 21 operações do processo. Obtiveram-se, através do *Input Analyzer*, distribuições válidas para 20 desses conjuntos de dados uma vez que, para esses casos, os valores de prova obtidos nos testes de Kolmogorov-Smirnov foram sempre superiores a 15%. Para o conjunto de dados em que o grau de ajuste não foi satisfatório recorreu-se a uma distribuição empírica também gerada pelo módulo *Input Analyzer* (recurso forno de pré-aquecimento).

Houve, nesta fase do desenvolvimento conceptual, um especial cuidado em mapear, o mais fielmente possível, o modelo do sistema produtivo real, tarefa que foi facilitada pelo facto de uma das pessoas da equipa de desenvolvimento do modelo se encontrar a trabalhar na empresa. Esta situação facilitou o processo de recolha de dados e permitiu um melhor entendimento da forma de funcionamento da produção.

4.2 Modelação Operacional

Uma vez definido e discutido o modelo de dados de entrada e validado o modelo conceptual do sistema, efetuou-se a implementação do modelo de simulação tendo-se recorrido ao *software* de simulação Arena[®], desenvolvido pela Rockwell Automation Inc. para Microsoft Windows[®]. Este *software* combina a facilidade de utilização dos simuladores com a flexibilidade das linguagens de simulação e permite desenvolver modelos de simulação animados, para sistemas discretos e contínuos utilizando uma abordagem orientada a objetos (Kelton *et al.*, 2010). O Arena apresenta ao utilizador um conjunto de bibliotecas ou *templates* que incluem um conjunto de elementos

gráficos (módulos) que definem os componentes físicos e lógicos de um sistema e cuja combinação dá origem a um modelo de simulação operacional. É, também, possível a integração com outras tecnologias (por exemplo, bases de dados, folhas de cálculo, aplicações de desenho assistido por computador) o que possibilita a utilização combinada dos diferentes modelos de informação de uma organização.

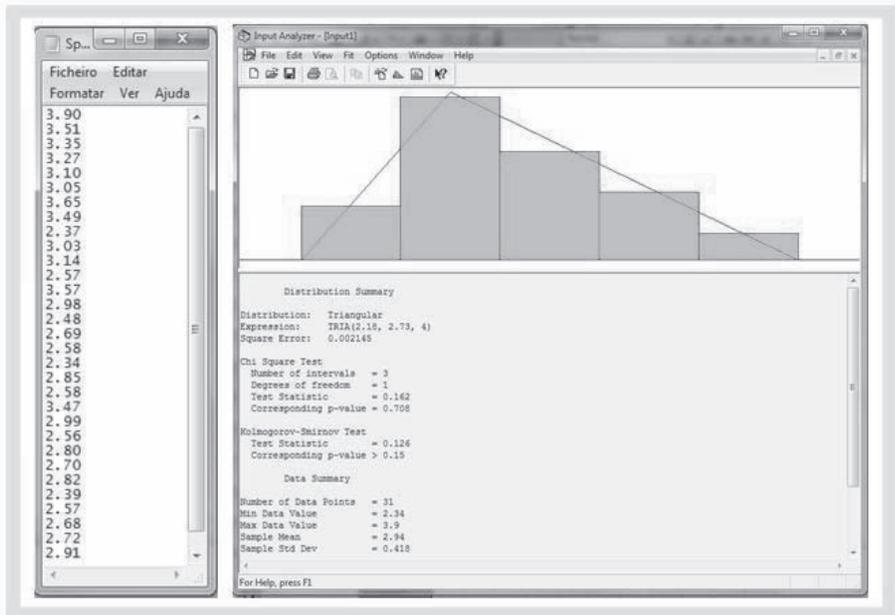


Figura 9: Procedimento para ajuste de distribuições teóricas de probabilidade via *Input Analyzer*

O Arena é uma ferramenta amplamente divulgada, de reconhecido mérito internacional e utilizada na Universidade de Aveiro (DEGEI) para ensino/investigação na área da simulação industrial. Por estas razões foi a ferramenta selecionada para desenvolver este projeto.

A fase de modelação operacional consumiu uma parte importante do tempo disponível devido à complexidade do modelo tendo, contudo, sido facilitada pelo trabalho prévio de conceptualização e modelação gráfica do sistema. O modelo desenvolvido tem cerca de 30 processos, 19 pontos de partida/chegada de carros de transporte, 75 distâncias entre sectores, 26 recursos diferentes (máquinas, operadores, carros de transporte), etc. A conjugação deste elevado

número de diferentes módulos possibilitou uma grande aproximação ao sistema real.

A Figura 10 apresenta duas imagens do modelo operacional do sistema de produção nomeadamente, o modelo lógico e o modelo de animação. O modelo lógico (desenvolvido pelo modelador e de pouca utilidade para os decisores da empresa) inclui um conjunto de módulos do *software* Arena (por exemplo, *Create*, *Process*, *Batch*, *Record*, *Variable*) conjugados de modo a recriar o funcionamento do sistema o mais fielmente possível. O modelo de animação constituiu uma mais-valia para os decisores uma vez que estes puderam identificar o sistema em análise e visualizar, de um modo “familiar”, os postos de trabalho e o comportamento da linha de produção. Esta identificação aumentou a credibilidade do estudo levado a cabo e contribuiu para a aceitação dos resultados provenientes dos modelos operacionais.

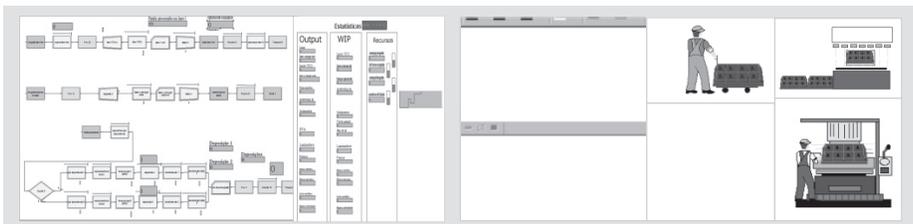


Figura 10: Excertos do modelo operacional desenvolvido no Arena: modelo lógico (esquerda) e modelo de animação (direita)

A verificação de um modelo de simulação (*debugging*) consiste em avaliar se o modelo operacional foi implementado corretamente e se funciona do modo pretendido. O processo de verificação tem como principal objetivo a deteção e remoção de erros não intencionais, relacionados com a lógica do modelo (Sargent, 2010).

As principais técnicas de verificação utilizadas neste estudo foram as seguintes:

- rastreabilidade do modelo: o *software* utilizado inclui o elemento *Trace* que permitiu examinar detalhadamente o movimento das entidades através do sistema e verificar as ações executadas em cada módulo por onde uma determinada entidade passou;

- controlador de execução interativo (*interactive run controller* ou *debugger*): o controlador de execução possibilitou a monitorização e o controlo da execução do modelo de um modo interativo ou seja, permitiu, através da utilização do teclado ou do rato, e de um modo iterativo, executar o modelo passo a passo, suspender a execução em pontos críticos, analisar dados e alterar o valor de determinadas variáveis para forçar a ocorrência de erros;
- teste de valores de entrada: a execução do modelo para um conjunto de valores dos parâmetros de entrada extremos ou inválidos permitiu avaliar o comportamento do modelo e a razoabilidade dos parâmetros de saída;
- visualização/animação: a visualização dinâmica do comportamento do modelo durante a sua execução, através da sua animação, possibilitou a deteção de determinados erros relacionados com a interação dos diferentes componentes do sistema.

Quando concluída a construção do modelo, foram efetuados diversos testes piloto. Estes testes permitiram, em diversas ocasiões, eliminar erros de código, perceber que algumas condições tinham sido erradamente introduzidas, ou não produziam o efeito desejado, e até mesmo identificar situações que existiam na realidade mas que não tinham sido transpostas para qualquer dos modelos (conceptual e operacional). Foram, também, para facilitar os testes, criados diversos indicadores (como contadores e gráficos) por forma a identificar problemas e permitir a sua resolução de uma forma expedita.

A validação de um modelo de simulação consiste em determinar se o modelo conceptual definido e implementado reflete adequadamente e com rigor o sistema modelado. Deste modo, é possível avaliar se as abstrações e simplificações definidas intencionalmente no modelo conceptual do sistema introduzem erros inaceitáveis nos resultados. A validade de um modelo está diretamente relacionada com os seus objetivos e, de um modo geral, são efetuados diversos testes e avaliações até que se obtenha um grau de confiança suficiente para considerar um modelo válido relativamente aos seus propósitos (Sargent, 2010).

As principais técnicas empregues para validar o modelo desenvolvido consistiram em:

- testes piloto e teste de Turing: a condução de testes piloto sobre o modelo verificado permitiu comparar certas estimativas obtidas para os valores de saída com os valores efetivamente observados para o sistema existente; a comparação entre os dois tipos de dados foi efetuada por especialistas no sistema simulado e através de métodos gráficos (por exemplo, diagramas de caixa);
- análise de sensibilidade: os valores de alguns parâmetros de entrada foram alterados, de modo a analisar a sensibilidade dos parâmetros de saída;
- discussão com a equipa de projeto: o modelo de simulação e o seu comportamento foram discutidos com os membros da organização pertencentes à equipa envolvida no projeto (especialistas na operação do sistema simulado) e foram avaliadas a sua razoabilidade e adequação;
- animação do sistema e gráficos operacionais: a observação, por parte dos especialistas, da animação do modelo e de gráficos que apresentam os valores de determinadas medidas de desempenho ao longo da duração da simulação permitiu, também, decidir sobre a validade do modelo desenvolvido.

A aplicação deste conjunto de técnicas permitiu confirmar a correta implementação do modelo de simulação no *software* utilizado e possibilitou a avaliação da sua adequação relativamente aos objetivos inicialmente definidos. Verificou-se que, comparando o *output* gerado pelo modelo com o que acontece na realidade, os resultados eram bastante semelhantes (produção de cerca de 480 painéis por dia, identificação de recursos com taxas de ocupação acima dos 90%, número de componentes em fila de espera nas diversas secções) e que o modelo era uma representação fidedigna da realidade.

Este processo de verificação e validação (V&V) permitiu ganhar confiança sobre os resultados e, deste modo, foi possível prosseguir o estudo e passar às fases de experimentação e análise de resultados.

5. Experimentação e Resultados

A experimentação de um modelo de simulação implica a definição do(s) cenário(s) a testar e a identificação de vários parâmetros relacionados com as condições de teste. Estes parâmetros estão relacionados com o tipo de simulação considerado (Nakayama, 2008): (i) simulação com uma condição natural de paragem (*terminating simulation*) em que o sistema possui uma condição inicial fixa e uma ocorrência que determina a duração da simulação; ou (ii) simulação sem uma condição natural de paragem (*nonterminating simulation*) em que o sistema não possui uma condição natural de paragem e a sua duração não é finita sendo o objetivo principal deste tipo de simulação a avaliação do comportamento estacionário do sistema e, como tal, os efeitos das condições iniciais devem ser removidos dos resultados da simulação.

A identificação do tipo de simulação mais apropriado para o sistema em causa condiciona a definição das condições experimentais do modelo de simulação. Tratando-se de uma simulação sem uma condição natural de paragem foram então definidas as seguintes condições de teste:

- a duração da simulação corresponde a uma semana de trabalho (7 dias de 24 horas) e reflete o período de planeamento típico do sistema;
- efetuaram-se vinte replicações independentes do modelo; o número de replicações foi definido através da execução de testes piloto e da avaliação da variabilidade associada aos resultados obtidos para as diferentes medidas de desempenho;
- as condições iniciais para cada replicação são caracterizadas pelo sistema parado e vazio;
- foi definido um período de aquecimento de 3 dias, que é o tempo suficiente para carregar toda a linha e, deste modo, iniciar a execução do modelo com todos os postos com material.

Quanto aos cenários analisados foram dois: (i) um cenário que representa a forma atual de funcionamento do sistema produtivo e que permitiu avaliar o desempenho do sistema e identificar pontos críticos e (ii) um cenário alternativo em que se procurou modelar as alterações ao sistema que a empresa pretendia implementar.

Quanto ao cenário inicial, toda a produção era feita em lotes que se movimentavam entre postos sendo os painéis processados de forma unitária em cada posto de trabalho. Os lotes eram lançados na deposição sendo constituídos por 48 painéis o que corresponde à capacidade da caixa de deposição (que entra dentro das câmaras). A empresa tinha capacidade, neste cenário atual, para produzir 12 lotes por dia, cada um constituído por 48 unidades.

Em relação ao cenário alternativo o que se pretendia avaliar era o impacto do aumento da capacidade das caixas de 48 para 64 painéis o que representaria o lançamento de mais 33% de produção. De notar que esta alteração implicaria o aumento da capacidade de processamento dos dois sistemas de *sputtering*, por via de alteração dos sistemas de bombagem o que significaria algum investimento para a empresa.

Uma outra alteração a introduzir era a troca dos equipamentos de laminação por um equipamento automático com uma taxa de saída de 6 unidades a cada 12 minutos o que também implicava um considerável investimento por parte da empresa.

Este cenário alternativo foi definido com base no facto de os decisores pretenderem saber qual o impacto, do ponto de vista das medidas de desempenho do sistema, dos investimentos planeados para, deste modo, avaliarem se eram ou não justificados perante o cenário inicial.

Tendo-se efetuado a definição das condições experimentais e dos cenários a estudar foram, então, realizadas as simulações e recolhidos os respetivos resultados. É necessário ter em consideração que, uma vez que os parâmetros de entrada de um modelo de simulação são, na sua maioria, variáveis estocásticas, os resultados obtidos através da sua experimentação são estimativas das medidas de desempenho do sistema.

Os resultados obtidos foram analisados de acordo com o método normalmente utilizado neste tipo de condições e que corresponde ao método das replicações independentes. Este método consiste em obter conjuntos de observações independentes, através de várias replicações (neste caso 20) inicializadas com diferentes sementes de números pseudoaleatórios (Law e

Kelton, 2000). A descrição do gerador de números aleatórios implementado no *software* Arena pode ser encontrada em Kelton *et al.* (2010).

Assim verificou-se, para o modelo que representa a situação inicial, que o volume total de produção da linha teve um valor médio de 3224 painéis em 7 dias, dando uma média de 461 painéis por dia (N). De realçar, também, que o número médio de lotes terminados por dia (R) corresponde a 10,7.

Um outro aspeto analisado foi a utilização dos equipamentos, também como forma de identificar quais os recursos mais críticos para o desempenho do sistema (gargalos). A Tabela 2 mostra os valores das taxas de ocupação dos recursos (Tr):

Pela tabela (e também pela análise das filas de espera associadas) pode concluir-se que os recursos que estão a limitar o sistema são o conjunto de laminadores com uma taxa de utilização a rondar os 100%, bem como ambos os *sputterings* (de alumínio e de zinco), com taxas de utilização também na ordem dos 100%.

Para o cenário alternativo introduziram-se no modelo as seguintes alterações:

- substituição das caixas existentes por outras de capacidade para 64 painéis, ao invés dos atuais 48;
- melhoria do sistema de bombagem das câmaras de entrada e saída dos *sputterings*, reduzindo assim, em cerca de 30%, o tempo de processamento dos painéis nestes postos de trabalho;
- substituição do atual sistema de laminação por um com a capacidade de processar 6 painéis a cada 12 minutos.

O volume de produção obtido com esta solução cifrou-se numa média diária (N) de 620 painéis embalados no fim da linha, verificando-se assim um aumento de cerca de 35%.

As novas taxas de ocupação dos recursos (Tr), encontram-se na Tabela 3.

Pode observar-se que os anteriores recursos gargalo deixaram de o ser apesar de manterem elevadas taxas de utilização. Com esta solução o posto que passaria a limitar o sistema produtivo seria o Laser 1, com uma taxa média de ocupação na ordem dos 97%.

Tabela 2: Taxas de ocupação dos recursos no cenário atual

Scheduled Utilization	Average	Minimum Average	Maximum Average
aplicacao eva posto1	0.4731	0.4709	0.4755
aplicacao eva posto2	0.00	0.00	0.00
arrefecimento maquina	0.9319	0.8866	0.9689
deposicao1 maquina	0.9543	0.9240	0.9858
deposicao2 maquina	0.9238	0.8457	0.9760
estacao electrica posto	0.3947	0.3907	0.3997
estacao mecanica maquina	0.5035	0.4988	0.5118
fita al1 posto	0.5010	0.4963	0.5048
fita al2 posto	0.00	0.00	0.00
forno envelhecimento 1 maquina	0.6540	0.5888	0.6974
forno envelhecimento 2 maquina	0.3968	0.3343	0.4716
forno pre aq1	0.9203	0.8847	0.9656
forno pre aq2	0.8829	0.7839	0.9483
isolamento maquina	0.6823	0.6761	0.6865
laminador	0.9991	0.9981	0.9997
laser 2	0.4771	0.4748	0.4794
laser 3	0.4771	0.4743	0.4800
laser1 maquina	0.8124	0.7733	0.8386
limpeza e embalagem posto	0.2505	0.2482	0.2534
maquina lavar	0.3556	0.3417	0.3670
maquina lixar	0.3437	0.3259	0.3564
Operador	0.7499	0.7410	0.7603
sputtering al maquina	0.9985	0.9941	1.0000
sputtering zn maquina	0.9979	0.9942	1.0000
teste modulos maquina	0.4947	0.4889	0.5006
teste painel maquina	0.5067	0.5015	0.5153

Observa-se, também, que o posto do Isolamento Periférico passa a ser utilizado a 87% da sua capacidade, o que é indicativo de que é um posto potencialmente limitador de futuros aumentos de cadência da linha. A taxa de ocupação dos operadores passa, também, para cerca dos 85% o que poderá colocar alguns problemas se se tiver em conta que não estão a ser contabilizados os tempos utilizados nas movimentações do produto.

Na Figura 11 são comparadas as taxas de ocupação dos recursos, para os dois cenários.

Tabela 3: Taxas de ocupação dos recursos no cenário alternativo

Scheduled Utilization	Average	Minimum Average	Maximum Average
aplicacao eva posto 1	0.6374	0.6181	0.6555
aplicacao eva posto2	0.00	0.00	0.00
arrefecimento maquina	0.8443	0.8090	0.8838
deposicao1 maquina	0.8174	0.7757	0.8648
deposicao2 maquina	0.8688	0.8275	0.8946
estacao electrica posto	0.5310	0.5083	0.5508
estacao mecanica maquina	0.6771	0.6489	0.7065
fita al1 posto	0.6418	0.6217	0.6616
fita al2 posto	0.00	0.00	0.00
forno envelhecimento 1 maquina	0.7390	0.7122	0.7829
forno envelhecimento 2 maquina	0.6843	0.6466	0.7339
forno pre aq1	0.6975	0.6530	0.7315
forno pre aq2	0.7569	0.7146	0.8089
isolamento maquina	0.8715	0.8437	0.8984
laminador	0.8608	0.8344	0.8850
laser 2	0.6101	0.5857	0.6293
laser 3	0.6099	0.5848	0.6305
laser 1 maquina	0.9732	0.9327	1.0000
limpeza e embalamento posto	0.3371	0.3204	0.3509
maquina lavar	0.4048	0.3906	0.4162
maquina lixar	0.4097	0.3924	0.4209
Operador	0.8501	0.8283	0.8646
sputtering al maquina	0.8927	0.8583	0.9212
sputtering zn maquina	0.8928	0.8544	0.9231
teste modulos maquina	0.6653	0.6325	0.6878
teste painel maquina	0.6510	0.6264	0.6719

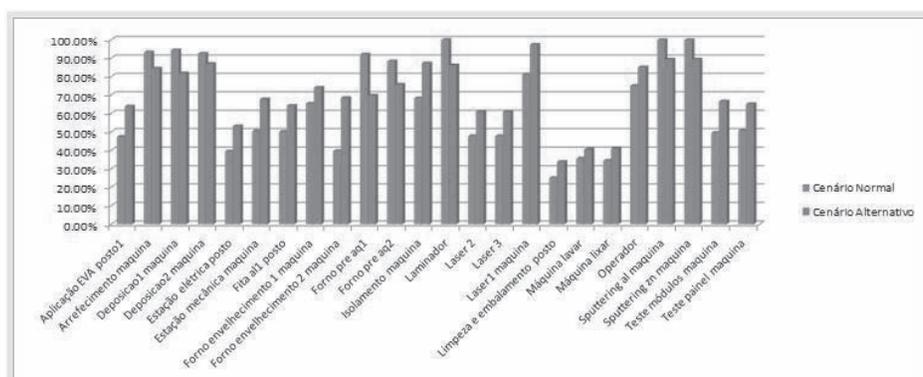


Figura 11: Comparação dos valores de taxas de ocupação

Nota-se, no novo cenário, uma subida da taxa de ocupação da maioria dos recursos, evidenciando uma melhor utilização da capacidade instalada. De realçar, também, a baixa taxa de aproveitamento de alguns postos de trabalho, como a colocação de fita de alumínio e aplicação de EVA, que são postos que trabalham em contínuo e dependem sempre da capacidade produtiva do posto imediatamente a seguir (a sua duplicação mostra-se desnecessária).

Genericamente: (i) quantificou-se o aumento de capacidade que potencialmente se conseguiria obter com a realização dos investimentos previstos, (ii) verificou-se que o sistema ficaria mais equilibrado em termos de utilização dos recursos e (iii) identificaram-se potenciais pontos gargalo do sistema. Com base nos resultados obtidos os decisores puderam fazer um estudo financeiro mais sustentado para justificar a realização dos investimentos.

6. Conclusões e Desenvolvimentos Futuros

O desenvolvimento de projetos de sucesso envolvendo parcerias Universidade/Empresa é, geralmente, difícil de alcançar. Se, por um lado, os académicos tendem a utilizar uma abordagem e uma linguagem “excessivamente” científicas desprovidas de sentido para os gestores e engenheiros, por outro, os elementos das empresas tendem a concentrar-se nas suas preocupações de curto prazo dedicando pouco esforço e atenção ao desenvolvimento do projeto.

Neste caso, em particular, estas dificuldades foram ultrapassadas devido ao facto de um elemento da equipa da Universidade trabalhar a tempo inteiro na empresa, durante toda a duração do projeto. Este aspeto possibilitou o estabelecimento de um canal de comunicação privilegiado entre a Universidade e a empresa e captou a atenção da gestão e dos colaboradores da empresa para o projeto. A utilização da modelação gráfica foi, também, um fator decisivo no sucesso do projeto pois permitiu a criação de um conjunto de elementos de trabalho simples e de fácil entendimento por parte de todos os membros da equipa (criou-se uma visão partilhada do sistema em análise).

Os objetivos da empresa foram alcançados uma vez que as abordagens escolhidas para a execução do projeto permitiram analisar, detalhada e dinamicamente, a situação atual e as alterações pretendidas. A empresa

identificou-se com os modelos gerados e com os resultados obtidos. Esses resultados estão a ser utilizados para fazer pequenos ajustes aos investimentos inicialmente previstos e para introduzir as modificações na linha de produção de painéis solares fotovoltaicos. Os objetivos da Universidade foram também alcançados uma vez que o projeto foi completado com sucesso e os resultados do mesmo estão a ser utilizados pela empresa para melhorar o seu desempenho.

De um modo resumido, a utilização da modelação gráfica e da simulação possibilitou: (i) a representação adequada da complexidade associada ao sistema em análise, (ii) a avaliação do comportamento dinâmico do sistema real e do cenário alternativo, contemplando um alargado conjunto de medidas de desempenho, (iii) a visualização dinâmica dos sistemas em análise captando aspetos operacionais relevantes, (iv) a criação de uma visão holística do sistema em estudo e (v) uma diminuição do risco associado à tomada de decisão.

Concluiu-se, também, que a simulação é uma ferramenta bastante poderosa na análise de cenários porque permitiu, num curto espaço de tempo, simular o comportamento de uma linha de produção sob diferentes condições operacionais. Este tipo de análise é praticamente inviável através de experimentação no sistema real, dado que é impraticável parar uma linha de produção para testar determinadas alterações ou encomendar equipamentos para verificar se produzem os efeitos desejados.

Em termos de desenvolvimentos futuros pode identificar-se um conjunto de ações, ao nível da modelação via simulação, que fariam sentido contemplar num horizonte de curto prazo tais como a atualização dos modelos de simulação (no que se refere ao conjunto de dados de entrada e aos pressupostos assumidos), o estudo de outras alternativas (por exemplo, em termos de novas formas da implantação), o estudo criterioso do sistema de transporte de materiais em utilização e a consideração de outras alternativas de manuseamento de materiais e a análise mais detalhada do que se pode chamar o “coração” da unidade fabril, o sistema de deposições (passando tanto pelo aumento da velocidade dos atuais equipamentos, com algumas alterações de *hardware*, bem como pela duplicação de alguns deles).

É de realçar que o investimento feito num estudo de simulação pode, a médio ou longo prazo, revelar-se bastante proveitoso, evitando investimentos desnecessários, promovendo um maior aproveitamento dos recursos já existentes, e facilitando o estabelecimento de um conjunto de prioridades para futuros investimentos.

Um desafio de mais longo alcance, e de âmbito mais geral, passa pela divulgação e promoção da simulação junto do tecido empresarial. A ferramenta reveste-se de uma enorme utilidade ao nível da melhoria do desempenho operacional das unidades industriais (e de serviços) e a sua utilização precisa de ser difundida e incentivada. A lecionação de unidades curriculares de Simulação nos ciclos de estudos de Engenharia e Gestão Industrial é, hoje, uma prática comum mas a sua aplicação em casos reais é ainda escassa. Os projetos que a equipa da Universidade de Aveiro tem levado a cabo evidenciam o grande potencial desta ferramenta de engenharia e o sucesso que a mesma tem obtido junto das empresas.

Agradecimentos

Este trabalho foi apoiado por fundos portugueses através do CIDMA (Centro de Investigação e Desenvolvimento em Matemática e Aplicações) e da FCT (Fundação para a Ciência e a Tecnologia) com o projeto PEst-OE/MAT/UI4106/2014.

REFERÊNCIAS

BUEDE, D., *The Engineering Design of Systems – Models and Methods*, 2nd ed New Jersey: John Wiley & Sons, Inc., 2009.

DAVID, P., Idasiak, V. e Kratz, F., "Reliability study of complex physical systems using SysML", *Reliability Engineering and System Safety*, volume 95, 2010, pp. 431 450.

FRIEDENTHAL, S., Moore, A. e Steiner, R., *A Practical Guide to SysML – the Systems Modeling Language*, Burlington: Morgan Kaufmann/OMG Press, Elsevier Inc., 2008.

KELTON, W., Sadowski, R. e Swets, N., *Simulation with Arena*, 5th ed. New York: McGraw-Hill Higher Education, 2010.

LAW, A., "How to build valid and credible simulation models", *Proceedings of the 2009 Winter Simulation Conference*, Austin, Texas, EUA, 2009.

LAW, A. e Kelton, W., *Simulation Modeling and Analysis*, 3rd ed. New York: McGraw-Hill Higher Education, 2000.

NAKAYAMA, M., "Statistical Analysis of Simulation Output", *Proceedings of the 2008 Winter Simulation Conference*, Miami, Florida, EUA, 2008.

RAMOS, A., "Engenharia de Sistemas Baseada em Modelos: Um Sistema para o Tráfego & Ambiente", *Tese de Doutoramento*, Universidade de Aveiro, 2011.

RUMBAUGH, J., Jacobson, I. e Booch, G., *The Unified Modeling Language Reference Manual*, 2nd ed. Massachusetts: Addison-Wesley, 1999.

SARGENT, R., "Verification and validation of simulation models", *Proceedings of the 2010 Winter Simulation Conference*, Baltimore, EUA, 2010.

STEVENSON, W., *Operations Management*, Nova Iorque: McGraw-Hill, 2007.

SUSSMAN, J., *Introduction to Transportation Systems*, Massachusetts: Artech House, Inc., 2000.

CASO 6

OTIMIZAÇÃO DO ESCALONAMENTO DA PRODUÇÃO NA INDÚSTRIA DE MOLDES: APLICAÇÃO AO CASO DA GECO

Bárbara Esperança Virgílio

GECO

esperanca.barbara@gmail.com

Marta Castilho Gomes

CESUR – Instituto Superior Técnico

Universidade de Lisboa

marta.gomes@tecnico.ulisboa.pt

Ana Barbosa Póvoa

CEG-IST – Instituto Superior Técnico

Universidade de Lisboa

apovoa@tecnico.ulisboa.pt

RESUMO

A indústria de moldes tem um papel de relevo no mundo atual, pois o fabrico de uma multiplicidade de objetos presentes na nossa vida diária depende da utilização de moldes. A produção destas ferramentas é um processo de alta tecnologia com grande implantação em Portugal, sendo o nosso País um dos maiores exportadores de moldes a nível mundial. A diminuição constante do ciclo de vida dos produtos e o aumento de competitividade do mercado tem obrigado as empresas do setor a melhorarem o seu desempenho, através do aumento da eficiência dos processos. Neste incremento, revela-se crítica a tarefa do planeamento e escalonamento da produção, a qual é ainda realizada por métodos tradicionais em muitas empresas de moldes.

Neste trabalho desenvolveu-se um modelo de programação linear inteira – OPTMESC – para resolver o problema do escalonamento da produção no setor dos moldes, que corresponde a uma indústria de produção discreta por encomenda do tipo *job shop*. O modelo foi implementado na linguagem de modelação GAMS e resolvido com um *solver* comercial para dados reais recolhidos na empresa GECO. Inclui-se um exemplo ilustrativo de aplicação do modelo (de reduzida dimensão) e um caso de estudo que consiste na obtenção de um plano de produção da empresa envolvendo peças de vários moldes. Para este último efetuou-se uma análise pós-otimização aos parâmetros mais relevantes do modelo. Os tempos computacionais obtidos são compatíveis com a frequência com que o problema é resolvido na fábrica. Assim, este estudo apresenta-se como um promissor primeiro passo no desenvolvimento de um sistema computacional de apoio ao escalonamento da produção na indústria de moldes.

PALAVRAS-CHAVE

Escalonamento na indústria de moldes, Produção por encomenda, Programação linear inteira, Problemas *job shop*.

1. Introdução

A produção orientada para o cliente é, cada vez mais, uma questão central em empresas de produção por encomenda. Esta orientação, e consequentemente a maior interação entre os sistemas de produção e o cliente, criam grandes desafios e necessidades de soluções inovadoras, quer ao nível da organização do processo produtivo e das tecnologias utilizadas, quer ao nível dos sistemas de planeamento e controlo da produção. O trabalho aqui descrito surge neste contexto, pretendendo estudar formas de melhoria do processo de escalonamento na GECO, uma empresa produtora de moldes.

O fabrico de moldes é uma indústria de alta precisão indispensável à produção de uma ampla diversidade de produtos do nosso quotidiano. Trata-se contudo de uma indústria praticamente desconhecida do grande público tal como refere Lopes (2000):

“A indústria de moldes está presente na caixa que suporta o ecrã da nossa televisão, na cadeira de jardim em que nos sentamos, no aspirador que temos em casa... Está lá, mas nós não a conseguimos ver, porque os moldes que deram origem a estes objetos continuam fechados nos pavilhões onde se processa o respetivo fabrico (...) É uma indústria que causa uma certa estranheza pelas suas dimensões. Quem imaginaria que um molde para um contentor de lixo pode pesar 40 toneladas ou que um molde para uma mala de viagem pode custar 50 mil contos?”

Em Portugal, o peso da indústria de moldes na economia é muito relevante, não só pelo volume de negócios que movimenta mas também por se tratar de um setor de excelência em que o nosso País se destaca a nível mundial. As empresas produtoras de moldes são tipicamente pequenas e médias empresas (PMEs) que atuam como parceiras das empresas de moldagem, geralmente grandes empresas que no contexto português se encontram essencialmente nos setores da indústria automóvel, de embalagem, eletrónica, de telecomunicações e de eletrodomésticos.

A indústria de moldes tipicamente trabalha em modo de produção por encomenda, onde o produto (molde) é concebido individualmente de acordo com as especificações do cliente (situação designada “one-of-a-kind production” na literatura). O intervalo de tempo para projetar e fabricar um molde tem

vindo a diminuir consideravelmente à medida que o ciclo de vida dos produtos obtidos por moldagem se torna mais curto. Desta forma, para melhorar a competitividade, muitas empresas de produção de moldes estão a desenvolver soluções para aumentar a eficiência dos seus processos, otimizando tempo e custos (Ni *et al.*, 2007), onde se inclui o desenvolvimento de ferramentas computacionais para apoiar o planeamento e escalonamento da produção.

O escalonamento é uma decisão de um processo produtivo que determina quando, onde e como produzir um conjunto de produtos obedecendo a um conjunto de requisitos num horizonte de tempo dado. O seu objetivo, em geral, corresponde à minimização do tempo total de processamento (makespan), à maximização do lucro ou à minimização de custos, no horizonte de tempo especificado (Floudas & Lin, 2004). Sendo uma função crítica para a melhoria do desempenho produtivo (Méndez *et al.*, 2006), o escalonamento da produção é um problema complexo que exige a aplicação de modelos de otimização, de técnicas heurísticas ou de simulação na sua resolução. Ao contrário de outras indústrias de produção discreta como sejam o fabrico de chips eletrónicos e de semicondutores, para as quais existe na literatura científica um vasto número de artigos abordando os respetivos problemas de planeamento e escalonamento da produção, verifica-se que para a indústria de moldes são escassos os trabalhos nesta área. De facto, Choy *et al.* (2011) afirmam que nesta indústria o escalonamento da produção é ainda normalmente efetuado por métodos tradicionais, e é essa a experiência das autoras no contato com empresas portuguesas do setor.

Assim, este trabalho teve como objetivo principal o desenvolvimento de um modelo de programação linear inteira para o apoio ao escalonamento da produção na indústria de moldes e a sua resolução com dados reais fornecidos por uma empresa (GECO). As decisões contempladas no modelo dizem respeito à afetação, ou atribuição, das operações às máquinas existentes no shop floor e a afetação no tempo (escalonamento) da produção de cada peça e dos moldes (conjuntos de peças) tendo em conta o cumprimento das datas de entrega acordadas com os clientes.

A organização deste texto é a seguinte: na secção 2 apresenta-se a contextualização do problema, efetuando-se na secção 3 uma revisão sumária

da literatura de escalonamento de *job shop*, com particular ênfase nos métodos de programação matemática. A secção 4 descreve em pormenor o modelo de programação linear inteira desenvolvido (OPTMESC) ilustrando-se a sua aplicação com um exemplo de reduzida dimensão. O modelo é resolvido na secção 5 com dados recolhidos na GECO, a que se segue uma análise pós-otimização aos parâmetros mais relevantes. A última secção (6) apresenta as conclusões do estudo e hipóteses de evolução futura do trabalho.

2. Contextualização do problema

Após uma perspetiva sumária da indústria de moldes em Portugal, descreve-se a constituição de um molde assim como o respetivo processo de fabrico. Segue-se uma breve apresentação da empresa GECO e uma descrição do processo de planeamento e escalonamento do fabrico de moldes atualmente em uso na mesma.

2.1 A indústria de moldes em Portugal

A indústria de moldes nasceu em Portugal no início do século XX, associada à indústria vidreira. Atualmente existem dois polos, nas regiões Centro e Norte do País - concelhos da Marinha Grande e Oliveira de Azeméis, respetivamente. Conta com cerca de 530 empresas, na sua maioria PMEs, e emprega aproximadamente 8250 trabalhadores. É considerada uma indústria metalomecânica pois incorpora todos os segmentos responsáveis pela transformação de metais nos produtos desejados.

O valor da exportação de moldes atingiu cerca de 364 milhões de euros em 2011, ano em que o valor total de produção foi cerca de 404 milhões de euros, ou seja, Portugal exportou mais de 90% da produção total. Os principais destinos dos moldes portugueses em 2011 foram a Alemanha (25%), França (18%), Espanha (17%), Brasil (4%), Polónia (3%), E.U.A. (3%) e México (3%). A dimensão internacional da indústria é bem patente no facto das vendas se terem destinado a 75 países distintos. O mercado de exportação de moldes, entre 2000 e 2011, registou uma diminuição da exportação que se deve em parte ao facto da indústria nacional, principalmente os plásticos, ter aumentado fruto do alargamento da cadeia de valor das empresas de moldes. De

mencionar ainda que o setor automóvel se tornou o principal dinamizador tanto a nível tecnológico como financeiro desta indústria, totalizando cerca de 14% do volume de produção total em 1991 e 72% em 2010. Finalmente, numa comparação com outros países, Portugal ocupava em 2004¹ a nona posição no *ranking* mundial de países exportadores de moldes (com uma quota de 3,6%) sendo o terceiro maior exportador a nível europeu (IAPMEI 2006).

Para colocar o setor dos moldes em Portugal numa nova senda de crescimento, com vista a torná-lo num líder mundial, a Sociedade Portuguesa de Inovação iniciou o desenvolvimento de um projeto estruturante a médio e longo prazo, envolvendo todas as partes interessadas (*stakeholders*): empresas, associações, centros tecnológicos, universidades e decisores políticos. A concretização deste projeto estruturante começou pelo planeamento estratégico do qual resultou o *Plano Estratégico para o Setor dos Moldes em Portugal* (2008).

Este estudo destaca o papel das universidades como fundamental para o aumento das capacidades e competências do setor dos moldes, quer ao nível da transferência de tecnologia quer ao nível da formação de quadros técnicos e de gestão. A indústria beneficiará ainda da interação existente entre as universidades e outros parceiros a nível internacional. Apesar da relação com a Academia tomar na atualidade diferentes formas, várias empresas do setor consideram existir insuficiente colaboração ao nível dos programas de ensino e no desenvolvimento de soluções técnicas para a engenharia de produto, produção ou o desenvolvimento de novos materiais. Surge assim a necessidade crescente de criação de um mecanismo de transferência de tecnologia e de conhecimento entre o setor empresarial e os centros de saber, que permita à indústria desenvolver produtos inovadores e ajustar-se aos ciclos de vida reduzidos dos produtos.

2.2 Descrição e processo de fabrico de um molde

Existem diferentes tipos de moldagem mas todos se podem resumir ao mesmo procedimento básico: o material que se pretende moldar é colocado, quente ou pré-aquecido, dentro do molde e através da compressão e/ou da

¹ Dados de contexto internacional mais recentes a que foi possível aceder.

ocorrência de reações químicas, endurece e adquire a forma pretendida. A figura 1 ilustra a moldagem por compressão de uma peça de plástico.

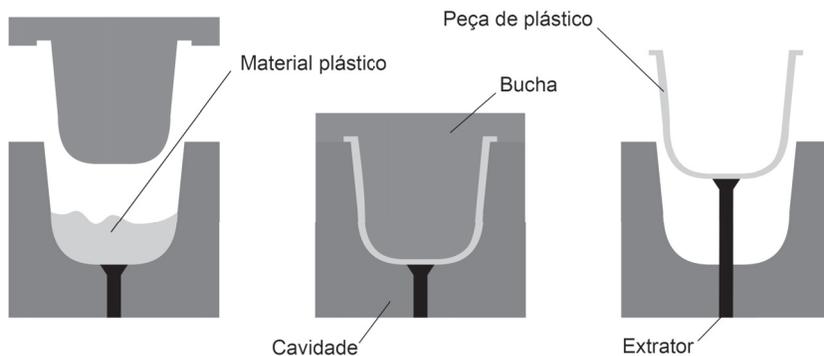


Figura 1: Funcionamento de um molde (moldagem por compressão).

Um molde é uma ferramenta constituída por diversas peças. A **estrutura** é a parte que serve de suporte ao molde. A **cavidade** é a parte côncava do molde, na qual é injetado, ou depositado, o material a moldar. A **bucha** é a parte convexa ou saliente do molde, por onde se efetua a extração da peça moldada. Os **elementos móveis** (localizados na maior parte dos casos na bucha) são elementos que se deslocam na horizontal através de uma guia (montada do lado da cavidade), durante a abertura do molde, para se poder extrair a peça moldada. Os **balancés** têm uma função semelhante aos elementos móveis mas deslocam-se na vertical (segundo um determinado ângulo) com a ajuda das chapas de extração permitindo a desmoldagem de peças com determinado tipo de orifícios, designados “zonas negativas”. Os **postigos** são colocados em zonas de difícil polimento, zonas para as quais o cliente pretende uma dureza diferente ou zonas a retificar ou a alterar posteriormente (do ponto de vista das dimensões). Por fim, os **acessórios** são peças que não terão contato com o material a moldar (plástico ou outro) mas que são montados na estrutura para um bom funcionamento do molde. A figura 2 apresenta diferentes perspectivas de um molde produzido na empresa GECO, onde se pode apreciar a complexidade do conjunto de peças que o constitui.

A figura 3 esquematiza as diferentes etapas do processo produtivo de um molde. Primeiramente decorre a fase de **projeto** do molde, em estreito contato com o cliente até que este aprove o desenho final. Dá-se então início ao fabrico propriamente dito, com a operação de **galgamento** que consiste no primeiro desbaste do bloco de aço, de modo a que as suas dimensões se aproximem das requeridas. O **desbaste** é uma operação destinada a remover material de modo a criar um esboço de geometria final da peça, deixando-a com uma espessura extra de 0,5 mm, para posterior acabamento. Na fase seguinte executam-se na peça as **furações** necessárias: de fixações, extração e águas. O **tratamento térmico** destina-se a conferir ao molde maior tenacidade e resistência ao uso; sendo um processo bastante dispendioso só é feito se o cliente exigir. Na operação de **retificação** anulam-se possíveis empenos (deformações) que a peça possa ter ganho no desbaste. No final desta operação a peça deve ter as cotas definidas no desenho (projeto). A **fresagem final**² destina-se a remover a sobreespessura de material deixada na fresagem de desbaste, de modo a deixar a peça com o acabamento e as dimensões pretendidas. Se não for possível definir uma determinada zona da peça por fresagem recorre-se à **erosão** (eletroerosão). No **controle** verificam-se as medidas finais de cada peça do molde, e se todas as operações foram completadas de acordo com o desenho do molde. Acabada a produção de todas as peças, procede-se à operação de **montagem**, podendo-se em seguida **testar** o molde. O **embarque** só tem lugar quando as amostras (peças moldadas) provenientes dos testes são do agrado do cliente.

² Esta poderá ser fresagem CNC (Comando Numérico Computorizado) ou DNC (Controlo Numérico Direto). A fresagem CNC é realizada por programação direta da máquina e só se usa em operações geometricamente simples. A fresagem DNC utiliza um *software* computacional que calcula cada cota necessária para a fabricação da peça através do seu desenho 3D e passa essa informação para a máquina automaticamente. Este é um processo mais atual e eficiente que veio substituir a fresagem CNC.

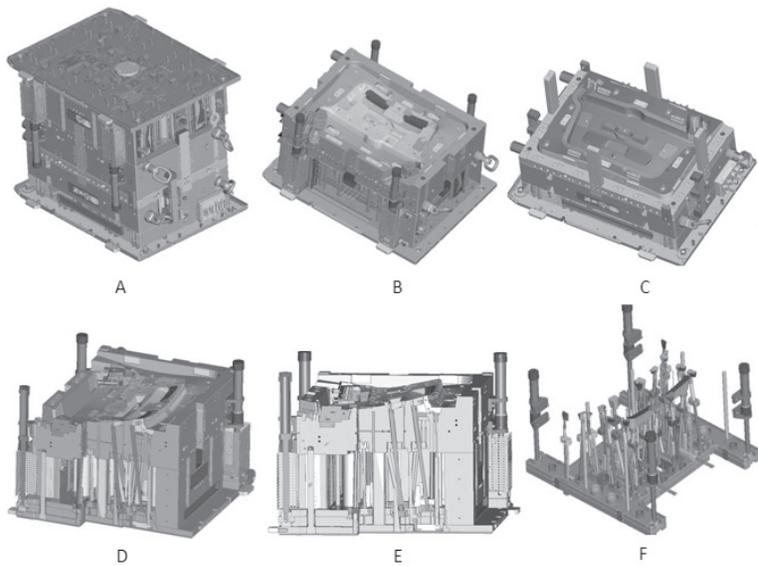


Figura 2: Diferentes perspetivas de um molde produzido na GECO.

A – Molde completo fechado; B – Bucha; C – Cavidade; D, E – Corte na bucha; F – Chapa de extração com os balancés



Figura 3: Processo de fabrico de um molde.

2.3 A empresa GECO

A empresa GECO iniciou a sua atividade em 1969 como Agente de Moldes e em 1975 tornou-se uma sociedade de Quotas de Responsabilidade Limitada com a denominação GECO (Gabinete Técnico e Controlo de Moldes em Fabricação, Lda.). Situada no Telheiro, freguesia de Maceira, distrito de Leiria, a GECO é uma sociedade de cariz familiar especializada no fabrico de moldes de injeção para a indústria de plásticos (termoplásticos, bi-componentes e injeção assistida a gás). A sua missão é:

“Transformar qualquer projeto numa ferramenta capaz de satisfazer as necessidades do cliente”.

O negócio cresceu e a empresa atualmente faz parte de um grupo industrial que integra fábricas autónomas e fábricas que apoiam a GECO na produção de moldes (unidades especializadas no fabrico de determinados tipos de peças ou na realização de certas operações). Ocupando uma área de 16.200 m², a empresa possui 316 funcionários dedicados à fabricação e comercialização de moldes. A produção destina-se à indústria automóvel (80% da produção total), eletrónica, de embalagem, farmacêutica e de equipamentos domésticos. Além das fábricas em Portugal a empresa possui unidades no Reino Unido, México e Brasil e escritórios na Alemanha, E.U.A. e Rússia. As fábricas situadas no estrangeiro providenciam apoio aos clientes, não se dedicando ao fabrico de moldes. Atualmente a empresa exporta praticamente toda a sua produção, tendo o volume de negócios sido de 14 milhões de euros em 2013.

2.4 Planeamento e escalonamento na GECO

Com o decorrer do tempo e com a evolução da tecnologia, os moldes produzidos na GECO foram tomando formas cada vez mais complexas, em resposta às necessidades dos clientes. Esta complexidade permitiu à empresa entrar em mercados novos aumentando substancialmente a produção total. Neste contexto, para cumprir as datas de entrega há que planear rigorosamente todo o processo produtivo dos moldes.

O planeamento e escalonamento inicia-se com a chegada da encomenda de um novo molde, pois só desta forma é que pode ser fornecido ao cliente um orçamento bem preparado. Os líderes de projeto (LP) averiguam as hipóteses

de produção do molde, tanto a nível de requisitos técnicos e de materiais, como de prazo, sendo a negociação com o cliente efetuada a partir destes dados. A cotação entregue ao cliente menciona o tipo de aço recomendado, os sistemas de injeção e de extração, acabamentos de superfícies e outras informações importantes referentes ao fabrico do molde. Após formalizada a encomenda, efetua-se o desenho preliminar do molde que o cliente deverá verificar, comentar e finalmente aprovar. Depois desta aprovação procede-se à encomenda do aço, iniciam-se os desenhos das peças e faz-se o escalonamento da produção.

Na GECO, o processo de planeamento e escalonamento da produção é efetuado pelo diretor de produção, designado líder de grupo (LG), de forma manual, recorrendo à folha de cálculo *Excel* e ao *software Microsoft Project* para armazenamento e visualização dos dados. As decisões do escalonamento são pois tomadas sem o recurso a um algoritmo de ordenação de tarefas implementado computacionalmente. Ao realizar esta tarefa o LG dá prioridade às peças dos moldes que constituem os fatores limitativos da entrega (bucha, cavidade e estrutura). Os LG também são responsáveis pelos vários setores de produção, tendo sempre conhecimento de como estão a decorrer as operações de fabrico. Semanalmente têm lugar reuniões entre os LP e os LG para apurar a situação em que se encontra a produção dos vários moldes e discutir modificações necessárias, de forma a minimizar erros e diminuir o tempo total de produção. No caso de não existir capacidade interna para realizar determinadas operações nos prazos estabelecidos, decide-se sobre o recurso a horas extraordinárias, ao apoio de outros complexos industriais do mesmo grupo empresarial ou à subcontratação de determinadas operações.

Durante o fabrico dos moldes é frequente surgirem erros que necessitam correção. Estes são muitas vezes detetados na fase de teste, quando a peça de plástico moldada não se apresenta como o cliente deseja. Os erros podem ter diversas origens: erros de projeto do molde, de programação das máquinas de fresagem DNC ou erros dos operadores das máquinas (por exemplo, troca no tamanho das ferramentas para a maquinação da peça ou engano no ângulo da furação da peça). Estes erros provocam atrasos no processo produtivo podendo comprometer as datas de entrega acordadas com os clientes. Do

ponto de vista do escalonamento da produção, este deve permitir acomodar operações adicionais de correção de erros quando necessário, de forma a observar os prazos.

A empresa GECO é bastante exigente no cumprimento das operações planeadas para as diversas fases de fabrico do molde, pois só assim poderá ir ao encontro da data de entrega com o máximo de rigor e qualidade. O tempo total de produção depende mais da complexidade do molde do que da sua dimensão, estando o cliente por vezes disposto a pagar um valor adicional para diminuir o prazo de entrega do molde.

Desta forma se entende como a função do planeamento e escalonamento da produção é crítica nesta indústria, não existindo atualmente na GECO qualquer sistema de apoio à decisão que a suporte.

3. Revisão da literatura

Na indústria podem-se identificar duas estratégias principais de produção: *flow shop* e *job shop*. A operação *flow shop* está normalmente associada à produção repetitiva ou contínua e é frequentemente caracterizada por elevados volumes de produção, grande uniformização, leques reduzidos de produtos e elevada taxa de automação. Já a operação *job shop* está tipicamente ligada à produção em descontínuo e fabrico por encomenda. Neste caso o leque de produtos é frequentemente diversificado, com volumes de produção reduzidos e a existência de diferentes sequências ou rotas de produção.

Os problemas de escalonamento em indústrias de produção discreta do tipo *job shop* têm atraído cada vez mais investigadores devido à sua generalidade e dificuldade inerente. O problema de escalonamento de *job shop* clássico, em que existe uma única máquina por operação, tem sido intensivamente estudado na literatura de Investigação Operacional. Todavia, como notam Tay & Ho (2008), na prática a configuração do *shop floor* consiste em diversas cópias das máquinas mais críticas de forma a reduzir os estrangulamentos devidos a operações longas ou à ocupação das máquinas. Além disso, em sistemas de produção complexos, uma tarefa (*job*) pode visitar uma máquina mais do que uma vez o que é designado recirculação ou *re-entrant process* na literatura.

Jain & Meeran (1999) e mais recentemente Potts & Strusevich (2009) apresentam revisões da literatura relativamente aos métodos de resolução que têm sido propostos para o problema do escalonamento de *job shop*. Da análise da literatura ressalta que, embora a maioria dos autores apresente formulações de programação matemática para o escalonamento de *job shop*, recorre a heurísticas para resolver o problema; o interesse por métodos de resolução baseados em programação matemática é bastante recente (Gomes, 2007).

Uma das principais vantagens deste tipo de métodos é a flexibilidade de modelação dos problemas de escalonamento, tornando relativamente simples o uso de várias funções objetivo e a incorporação de diferentes restrições na análise de um problema. Embora permitam obter a solução ótima, os modelos de programação matemática para o escalonamento de *job shop* possuem contudo a desvantagem dos tempos de computação necessários para a sua resolução crescerem exponencialmente com o aumento da dimensão dos problemas, o que se deve ao facto dos algoritmos aplicados, do tipo ramificação-e-limite³, serem algoritmos de enumeração implícita⁴. Assim, este é um método exato de resolução do problema, por oposição aos métodos aproximados (Gomes, 2007). Não obstante, a evolução dos recursos computacionais, tanto do ponto de vista do *hardware* como do *software* disponível, tem vindo a permitir resolver modelos de programação matemática de dimensão crescente.

Uma questão chave na modelação de problemas de escalonamento em programação matemática diz respeito à representação do tempo. Os modelos podem ser classificados em duas categorias principais (Méndez *et al.*, 2006): modelos de tempo discreto e modelos de tempo contínuo, dependendo da divisão do horizonte de escalonamento ser efetuada em intervalos de tempo uniformes ou em intervalos de diferente dimensão. Gomes *et al.* (2005), Pan & Chen (2005) e Ozguven *et al.* (2010) aplicaram a programação matemática na

³ *Branch-and-bound* em inglês.

⁴ Algoritmos que analisam o espaço de soluções (admissíveis) de um problema combinatorio de forma inteligente, evitando a enumeração explícita de todas as soluções. Em geral os *softwares* comerciais disponíveis combinam dois métodos de pesquisa do espaço de soluções: o *branch-and-bound* e a geração de planos de corte.

resolução de problemas de escalonamento de *job shop*: o primeiro trabalho apresenta modelos de tempo discreto enquanto os outros dois propõem formulações de tempo contínuo.

Por fim, a revisão da literatura efetuada mostrou ser reduzido o número de trabalhos que abordam especificamente o planeamento e escalonamento da produção na indústria de moldes. Ni *et al.* (2007) desenvolveram um modelo de informação de negócio para a integração de processos na indústria de moldes, de modo a racionalizar as atividades de negócio e maximizar o desempenho global. Choi & You (2006) apresentam uma extensa análise de desempenho de regras de despacho heurísticas para o escalonamento dinâmico destas indústrias, tendo recorrido para o efeito à simulação de eventos discretos. Choy *et al.* (2011) propuseram um modelo de apoio à decisão híbrido, combinando um algoritmo genético e um módulo de otimização.

4. Modelo de programação matemática para escalonamento na indústria de moldes

Esta secção descreve o modelo de programação linear inteira desenvolvido para o escalonamento da produção na empresa GECO, doravante designado OPTMESC⁵. Este baseia-se no modelo de tempo discreto descrito por Gomes *et al.* (2005) e Gomes (2007), que foi alterado e adaptado tendo em conta as características do caso de estudo em análise. O modelo mantém características genéricas que possibilitarão a sua utilização, com o mínimo de adaptações, no escalonamento de outras indústrias de produção discreta do tipo *job shop* por encomenda, nomeadamente indústrias que não sejam do tipo “*one-of-a-kind*”. A aplicação do modelo OPTMESC é ilustrada com um exemplo de reduzida dimensão.

4.1 Definição do problema

Previamente à apresentação do modelo, é necessário definir o problema em estudo. A caracterização que se segue incorpora já algumas especificidades do

⁵ Sigla derivada da expressão “**O**timização do **E**scalonamento de **M**oldes”, com rearranjo das letras.

modelo OPTMESC, nomeadamente a nível de nomenclatura e elementos que o constituem.

- Um molde (k) é constituído por várias peças (p), que diferem de molde para molde.
- Em geral, num molde as peças são únicas, mas podem existir algumas peças exatamente iguais, isto é, duas ou mais cópias da mesma peça⁶. Nesse caso, o modelo considera-as como duas ou mais unidades de uma mesma peça: são modeladas em conjunto, a nível de variáveis de decisão.
- A cada peça (p) corresponde desta forma um número (Q_p) de unidades idênticas a produzir (número inteiro: 1, 2, 3, etc.)
- Cada peça (p) segue uma sequência ou rota de produção (j). Peças diferentes podem seguir a mesma sequência de produção (j).
- Uma sequência de produção (j) é constituída por várias operações em sequência. São exemplos de operações o desbaste, furações, retificação e acabamento.
- Embora exista um fluxo genérico de operações, descrito em 2.2 e representado na figura 3, nem todas as operações estão presentes na sequência de produção de uma peça concreta. Para compreender este aspeto, é útil visualizar a figura 4 (na secção 4.3) onde se representam as sequências de produção para o conjunto de peças de um exemplo ilustrativo de aplicação do modelo.
- Cada operação (o) deve ser realizada por uma máquina (m).
- Existem diferentes máquinas (m) disponíveis para realizar uma dada operação (o). Além disso, uma máquina (m) pode executar diferentes operações. Por exemplo, uma máquina que realize o desbaste pode também executar o acabamento de estruturas.
- No modelo, cada operação (o) é precedida de um *buffer* ou fila de espera onde as peças aguardam para serem processadas. O *buffer* é comum a todas a máquinas que podem realizar essa operação.

⁶ Por exemplo, duas buchas e duas cavidades iguais, o que permitirá produzir duas peças moldadas em simultâneo de cada vez que o molde é usado.

- A descrição do modelo recorre à noção de *buffer* inicial (da primeira operação), *buffers* intermédios (das restantes operações) e *buffer* final (de peças acabadas) de cada sequência de produção. A figura 4 representa a alternância de *buffers* e operações nas sequências de produção do exemplo ilustrativo.
- As máquinas (m) caracterizam-se pela operação ou operações que executam, os tempos de produção e a capacidade (número total de unidades que podem processar em simultâneo).
- Os tempos de produção (T_{pom}) dependem da peça (p), da operação (o) e da máquina (m). Além disso, tomam valores inteiros uma vez que o modelo de escalonamento é de tempo discreto.
- Para cada molde (k) existe uma data de entrega (D_k).
- A produção das peças (p) deve ser escalonada para atender às datas de entrega (D_k) do conjunto de moldes (k) a que pertencem.

O modelo (OPTMESC) irá definir a afetação das máquinas na produção dos moldes de forma a cumprir as datas de entrega acordadas com os clientes, tanto quanto possível, num modo de operação *just-in-time*.

4.2 Formulação matemática

A formulação matemática proposta – OPTMESC – modela o problema descrito anteriormente: problema de escalonamento de *job shop* flexível em indústrias de produção discreta por encomenda. A formulação descrita nas secções 4.2.1 a 4.2.4 não incorpora explicitamente a operação de montagem das peças do molde; tal é feito na extensão do modelo apresentada em 4.2.5. Note-se que nenhuma das formulações considera as etapas de teste e embarque referidas na secção 2.2.

4.2.1 Conjuntos e índices

O modelo recorre à seguinte definição de conjuntos e índices:

K	moldes a escalonar	$k \in K$
P	peças a produzir	$p \in P$
J	sequências de produção	$j \in J$

O	operações	$o \in O$
M	máquinas	$m \in M$
P_j	peças que seguem a sequência de produção j	$P_j \subset P$
P_{kj}	peças pertencentes ao molde k j que seguem a sequência de produção	$P_{kj} \subset P_j$
J_k	seqüências de produção associadas ao molde k	$J_k \subset J$
O_j	operações na sequência de produção j	$O_j \subset O$
O_m	operações que podem ser realizadas na máquina m	$O_m \subset O$
M_o	máquinas que podem realizar a operação o	$M_o \subset M$
T	tempo	$t \in T$

O_j é um conjunto de elementos ordenados, sendo o aspeto da ordenação irrelevante no caso dos restantes conjuntos. Desta forma, para cada elemento (operação) de O_j estão definidas as operações antecessora e sucessora; excetuam-se a primeira operação da sequência, sem antecessora, e a última, sem sucessora. A ordenação de O_j torna-se necessária na escrita de certas restrições do modelo (restrições [7]).

Definem-se ainda os seguintes elementos particulares de cada sequência da produção:

- f_j primeira operação (“*first*”) na sequência de produção j
- l_j última operação (“*last*”) na sequência de produção j

4.2.2 Parâmetros

- H duração do horizonte de escalonamento, expressa em número de intervalos de tempo unitários
- Q_p quantidade (número de unidades idênticas) da peça p a produzir
- D_k data de entrega do molde k (corresponde a um instante do horizonte de escalonamento)⁷
- C_m capacidade da máquina m (número de unidades)
- S_o capacidade do *buffer* da operação o (número de unidades)

⁷ Nesta formulação do modelo corresponde ao instante em que se pretende que todas as peças do molde fiquem concluídas (prévio à montagem).

T_{pom} tempo de produção da peça p na máquina m , para a operação o (valor inteiro)

N_k penalidade por não terminar a produção de uma peça do molde k no horizonte de escalonamento definido

Z_k penalidade por uma peça do molde k se encontrar em espera num *buffer* intermédio (por unidade da peça e por unidade de tempo)

A_k penalidade pelo atraso na produção de uma peça do molde k , face à data de entrega (por unidade da peça e por unidade de tempo)

B_k penalidade por terminar a produção de uma peça do molde k antes da data de entrega (por unidade da peça e por unidade de tempo)

O horizonte de escalonamento é dividido em H intervalos de tempo de duração uniforme, visto que o modelo é de tempo discreto. Para a definição das restrições do modelo torna-se necessário definir um conjunto T de instantes no tempo mais abrangente:

$$T = \{ -t_{max}, -t_{max} + 1, -t_{max} + 2, \dots, -2, -1, 0, 1, 2, \dots, H, H+1 \}.$$

Nesta definição, t_{max} é o valor máximo dos tempos de produção (que assumem valores inteiros). O horizonte de escalonamento propriamente dito, ao qual se referem as decisões de carga das máquinas (ver secção 4.2.3), é o subconjunto de T :

$$\{ 0, 1, 2, \dots, H \}.$$

A necessidade de considerar instantes anteriores a 0 no conjunto T deve-se a certas restrições do modelo (questão discutida em 4.2.4) e a inclusão do instante $H+1$ deve-se à forma como são tidos em conta os limites dos intervalos de tempo unitários⁸ na definição das variáveis de decisão X_{pot} e $X_{fin_{pt}}$ (vd. secção 4.2.3).

Como pode ser produzida mais do que uma unidade de uma dada peça, os parâmetros Z_k , A_k e B_k são penalidades por unidade de uma peça do molde k e por unidade de tempo⁹. Dados A_k e B_k , define-se a seguinte penalidade para um instante t do horizonte de escalonamento:

⁸ Intervalos de tempo entre instantes consecutivos de T : por exemplo $[0,1]$, $[1,2]$, $[2,3]$.

⁹ Considerando uma unidade de tempo como um dos intervalos unitários em que o horizonte de escalonamento está dividido.

$$\psi_{kt} = \begin{cases} B_k(D_k - t) & t \leq D_k \\ A_k(t - D_k) & t > D_k \end{cases} \quad \forall k \in K, t = 0, \dots, H$$

É este parâmetro/penalidade que efetivamente surge na função objetivo, sendo obtido pela multiplicação do coeficiente por produzir a unidade da peça atrasada (A_k) ou adiantada (B_k) relativamente à data de entrega D_k pelo tempo compreendido entre o instante t e a data de entrega. Assim, a penalidade ψ_{kt} aumenta com o desfasamento entre o instante de conclusão da unidade da peça e a data de entrega.

4.2.3 Variáveis

- X_{pot} número de unidades da peça p no *buffer* da operação o entre os instantes $t-1$ e t
- $Xfin_{pt}$ número de unidades da peça p acabadas entre os instantes $t-1$ e t
- U_{pomt} número de unidades da peça p carregadas no instante t na máquina m , para executar a operação o

Todas as variáveis de decisão são inteiras (as peças são produzidas em unidades discretas). A variável X_{pot} é o número de unidades da peça p que se encontram no *buffer* que antecede a primeira operação (o) ou nos *buffers* intermédios que antecedem a operação o na sequência de produção j correspondente, no intervalo de tempo $[t-1, t]$ ¹⁰. Já as unidades da peça p acabadas no intervalo de tempo $[t-1, t]$, que se encontram no último *buffer*, são representadas pela variável $Xfin_{pt}$. O número de unidades terminadas no final do horizonte de escalonamento, ou seja, no intervalo $[H, H+1]$, é igual a $Xfin_{p(H+1)}$. A variável U_{pomt} contabiliza as unidades da peça p carregadas na máquina m , no instante t , para realizar a operação o .

4.2.4 Modelo OPTMESC

Tendo em conta as características do problema em estudo, o modelo OPTMESC apresenta a função objetivo e as restrições que se seguem:

¹⁰ Na definição das variáveis X_{pot} e $Xfin_{pt}$ os intervalos de tempo são fechados à esquerda e abertos à direita.

$$\begin{aligned}
& \text{Min} \sum_{k \in K} \sum_{j \in J_k} \sum_{p \in P_{kj}} \left[N_k \cdot (Q_p - Xfin_{p(H+1)}) \right. \\
& \left. + \sum_{t=0}^H \left(\psi_{kt} \cdot (Xfin_{p(t+1)} - Xfin_{pt}) + \sum_{o \in O_j | \{f_j\}} Z_k \cdot X_{pot} \right) \right] \quad [1]
\end{aligned}$$

Sujeito a:

$$U_{pomt} = 0 \quad \forall p \in P \quad \forall o \in O_j: p \in P_j \quad \forall m \in M_o, t = -t_{max}, \dots, -1 \quad [2]$$

$$X_{po0} = Q_p \quad \forall p \in P, o = f_j: p \in P_j \quad [3]$$

$$X_{po0} = 0 \quad \forall p \in P \quad \forall o \in O_j | \{f_j\}: p \in P_j \quad [4]$$

$$Xfin_{p0} = 0 \quad \forall p \in P \quad [5]$$

$$X_{po(t+1)} = X_{pot} - \sum_{m \in M_o} U_{pomt} \quad \forall p \in P, o = f_j: p \in P_j, t = 0, \dots, H \quad [6]$$

$$\begin{aligned}
X_{po(t+1)} &= X_{pot} - \sum_{m \in M_o} U_{pomt} + \sum_{m \in M_{o'}} U_{po'm'}(t - T_{po'm'}) \\
&\forall p \in P \quad \forall o \in O_j | \{f_j, l_j\}: p \in P_j, t = 0, \dots, H \quad [7]^{11}
\end{aligned}$$

$$\begin{aligned}
Xfin_{p(t+1)} &= Xfin_{pt} + \sum_{m \in M_o} U_{pom(t - T_{pom})} \\
&\forall p \in P, o = l_j: p \in P_j, t = 0, \dots, H \quad [8]
\end{aligned}$$

¹¹ o' é a operação que antecede a operação o na respetiva sequência de produção (relembre-se que O_j é um conjunto de elementos ordenados) e m' uma das máquinas que pode realizar a operação o' : $m' \in M_{o'}$.

$$\sum_{o \in O_m} \sum_{j \in J: o \in O_j} \sum_{p \in P_j} \sum_{\tau=t-T_{pom}+1}^t U_{pomt} \leq C_m \quad \forall m \in M, t = 0, \dots, H \quad [9]$$

$$\sum_{j \in J: o \in O_j} \sum_{p \in P_j} X_{pot} \leq S_o \quad \forall o \in O, t = 0, \dots, H + 1 \quad [10]$$

$$X_{pot}, Xfin_{pt}, U_{pomt} \in \mathbb{N}_0 \quad [11]$$

A função objetivo [1], a minimizar, é uma soma de penalizações: por existirem peças incompletas no final do horizonte de escalonamento (primeiro termo), pela finalização da produção de peças do molde k antes ou depois da respetiva data de entrega D_k (segundo termo) e pelo tempo de espera de peças inacabadas em *buffers* intermédios (terceiro termo). Relativamente ao segundo termo, a penalidade ψ_{kt} multiplica pelo número de unidades da peça p (do molde k) terminadas no instante t , que é igual à diferença das variáveis $Xfin_{p(t+1)}$ e $Xfin_{pt}$, isto é, do número de unidades no *buffer* final nos intervalos de tempo $[t, t+1[$ e $[t-1, t[$, respetivamente.

As restrições [2] a [5] traduzem as condições de inicialização. A inexistência de peças carregadas nas máquinas em instantes anteriores ao início do horizonte de escalonamento ($t=0$) é imposta por [2]¹². As restrições [3] garantem que no instante inicial, e no *buffer* inicial, estão as Q_p unidades (idênticas) de cada peça p que se pretende produzir. As restrições [4] e [5] garantem que os *buffers* intermédios e o *buffer* final estão vazios no início do horizonte de escalonamento, respetivamente.

As restrições [6], [7] e [8] correspondem aos “balanços de massa” do primeiro *buffer*, dos *buffers* intermédios e do *buffer* final para cada peça, que segue uma dada sequência de produção (j). As restrições [6] calculam o

¹² Tal justifica-se porque ao concretizar as restrições [7], [8] e [9] para alguns valores de t do horizonte de escalonamento, surgem variáveis U que se referem a instantes de tempo prévios ao instante inicial e é necessário definir o seu valor. No caso das máquinas estarem ocupadas no início do horizonte de escalonamento, em vez de serem igualadas a zero estas variáveis são igualadas a um valor positivo (ver secção 5.1).

número de unidades da peça p no *buffer* inicial entre os instantes $t-1$ e t subtraindo ao número de unidades existentes no *buffer* no intervalo de tempo anterior o número de unidades da peça p que iniciam a sequência de produção j no instante t . Estas peças são carregadas nas máquinas que podem realizar a primeira operação da sequência. Desta forma, o somatório das variáveis U estende-se a todas as máquinas de M_o , sendo o a primeira operação da sequência j : $o = f_j$. O *buffer* é portanto comum a todas as máquinas de M_o .

As restrições [7] determinam o número de unidades da peça p nos *buffers* intermédios somando ao número de unidades existentes no intervalo de tempo anterior o número de unidades que na operação anterior da sequência de produção (operação o') terminaram no instante t e, portanto, foram adicionadas ao *buffer* da operação o nesse mesmo instante, e subtraindo o número de unidades da peça p que iniciaram a operação o no instante t . Por último, as restrições [8] representam a variação do número de unidades da peça p acabadas ao longo do horizonte de escalonamento, que se encontram no *buffer* final da sequência de produção respetiva: este número é obtido somando às unidades terminadas no intervalo de tempo anterior aquelas que ficaram concluídas no instante t , e que portanto iniciaram a última operação no instante $t-T_{pom}$.

As restrições da capacidade das máquinas [9] são necessárias para contabilizar as peças (ou melhor, unidades de peças) em cada máquina m no instante t do intervalo de escalonamento (e que se mantém no intervalo $[t, t+1[$). Assim, a soma das variáveis U , estendida a todas as operações que podem utilizar a máquina m e às unidades de peças que iniciam aí a operação no instante t ou em instantes anteriores (estando ainda em processamento)¹³ tem de ser menor ou igual à respetiva capacidade, que é o número máximo de unidades que a máquina pode ter simultaneamente em processamento.

As restrições [10] referem-se à capacidade dos *buffers*. Todas as unidades das peças p têm de ser contabilizadas quando esperam no *buffer* de cada operação o entre os instantes $t-1$ e t . Então a soma das variáveis X , estendida a todas as sequências de produção contendo a operação o , é menor ou igual à

¹³ Pelo que nas restrições [9] o somatório no tempo (índice τ) se inicia no instante $t - T_{pom} + 1$ e termina no instante t .

capacidade do *buffer* desta operação. Estas restrições são escritas para todos os *buffers* intermédios das sequências (excetua-se o *buffer* inicial).

Finalmente, as restrições [11] estabelecem a integridade das variáveis de decisão.

4.2.5 Generalização do modelo: OPTMESC com montagem de peças

A implementação da operação de montagem no modelo OPTMESC pressupõe a inclusão de uma operação final de montagem em todas as sequências de produção das peças de um dado molde (a montar). A operação de montagem é específica de cada molde, sendo associada uma máquina virtual à sua realização. Esta não corresponde a nenhuma máquina existente no *shop floor* da fábrica, sendo criada apenas por exigência da formulação matemática, que associa máquinas a operações. Na generalização do modelo efetuada, contempla-se a hipótese de, no conjunto de moldes a escalonar, existirem moldes a montar e moldes sem operação de montagem.

Os novos conjuntos, elementos particulares, restrições e a modificação da função objetivo necessários são os seguintes:

Conjuntos:

K_1 moldes com operação de montagem

K_2 moldes sem operação de montagem

$K = K_1 \cup K_2$

P_k peças pertencentes ao molde k

$P_k \subset P$

Elementos particulares de conjuntos:

p_k^r peça representativa do molde k $p_k^r \in P_k$

$mont_k$ operação de montagem do molde k $mont_k \in O$

m_mont_k “máquina” onde é realizada a montagem do molde k

$M_o = \{ m_mont_k \}$ com $o = mont_k$

Restrições:

$$U_{p_k^r o m t} = U_{p o m t} \quad \forall k \in K_1 \quad \forall p \in P_k \setminus \{p_k^r\}, o = mont_k, \\ \forall m \in M_o, t = 0, \dots, H \quad [12]$$

Função objetivo:

$$Min \sum_{k \in K} \sum_{j \in J_k} \sum_{p \in P_{kj}} \left[N_k \cdot (Q_p - Xfin_{p(H+1)}) + \sum_{t=0}^H \sum_{o \in O_j \setminus \{f_j\}} Z_k \cdot X_{pot} \right] + \\ + \sum_{k \in K_1} \sum_{t=0}^H \psi_{kt} \cdot (Xfin_{p_k^r(t+1)} - Xfin_{p_k^r t}) + \\ + \sum_{k \in K_2} \sum_{j \in J_k} \sum_{p \in P_{kj}} \sum_{t=0}^H \psi_{kt} \cdot (Xfin_{p(t+1)} - Xfin_{pt}) \quad [1']$$

O conjunto dos moldes é dividido em dois conjuntos disjuntos, K_1 e K_2 , correspondentes aos moldes com e sem operação de montagem, respetivamente. As peças dos moldes com montagem são todas modeladas individualmente ($Q_p = 1$) e uma das peças do molde (qualquer) é tomada como a peça representativa do mesmo, necessária nas restrições [12] e na função objetivo; designa-se p_k^r .

Assim, as restrições [12] estabelecem a simultaneidade da operação de montagem para todas as peças dos moldes com montagem (aplicando-se apenas aos moldes $k \in K_1$). Por outro lado, nas restrições [9] de capacidade da “máquina” de montagem do molde k é necessário igualar a capacidade (C_m) ao número de peças do molde.

A função objetivo [1'] difere da anterior apenas no desdobramento da parcela relativa à data de entrega (penalização por finalização antes ou depois da data) para os conjuntos K_1 e K_2 . No caso de moldes sem montagem ($k \in K_2$) a penalidade é igual à de [1], afetando todas as peças do molde. Para moldes

com montagem ($k \in K_i$) a penalidade apenas se aplica à peça representativa do molde.

4.3 Exemplo ilustrativo

Para ilustrar e validar a aplicação do modelo OPTMESC, criou-se um exemplo baseado em dados reais recolhidos na GECO¹⁴. Este apresenta um conjunto de 8 operações, descritas na tabela 1, que inclui também uma designação abreviada de cada operação. A produção de uma peça pode incluir todas as operações ou apenas algumas, o que depende da sua complexidade. A cada operação (o) está associada uma ou mais máquinas (m) que a podem executar, igualmente indicadas na tabela 1 (última coluna). Consideram-se 7 máquinas, algumas das quais executam mais do que uma operação: por exemplo as máquinas $m1$ e $m2$ realizam o desbaste (*Desb*) e o acabamento de estruturas (*DNC_E*)¹⁵. Cada máquina pode apenas produzir uma peça de cada vez (a sua capacidade é unitária). Já à operação de montagem fez-se corresponder uma “máquina” $m8$ com capacidade igual a 2 (número de peças que se pretende montar).

Para cada operação existe um *buffer* ou fila de espera que a antecede tendo-se igualado a respetiva capacidade a 100 unidades. Escolheu-se um valor elevado para esta capacidade porque no escalonamento das operações na empresa não são tidas em conta restrições de espaço (este não é limitativo do escalonamento).

No exemplo existem três peças de um molde a produzir sendo a peça $p1$ uma bucha, a peça $p2$ uma cavidade e a peça $p3$ uma chapa da estrutura. Pretende-se montar as peças $p1$ e $p2$ e produzir a peça $p3$ em separado. A figura 4 ilustra as duas sequências de produção necessárias na implementação do modelo. As operações são representadas por retângulos e os *buffers* por círculos. Os círculos que se encontram entre as operações correspondem aos *buffers* intermédios, o primeiro círculo de ambas as sequências corresponde ao *buffer* inicial e o último círculo de cada sequência corresponde ao *buffer* final.

¹⁴ Na secção 5 inclui-se um comentário mais pormenorizado sobre o levantamento de dados efetuado na empresa.

¹⁵ Designação derivada das máquinas de fresagem DNC que executam a operação.

Existem operações comuns a ambas as sequências de produção e operações que só existem numa das sequências: por exemplo a operação *DNC_E* é específica da sequência de produção II. Está definida uma operação de montagem das peças *p1* e *p2* (na sequência de produção I), pelo que se utilizou o modelo OPTMESC com montagem de peças.

Tabela 1: Operações e máquinas associadas (exemplo ilustrativo).

Operações	Designação no modelo (<i>o</i>)	Máquinas (<i>m</i>)
Desbaste	Desb	<i>m1, m2</i>
Furar águas	Furag	<i>m3, m4</i>
Retificação	Rectif	<i>m5</i>
Acabamento de estruturas	DNC_E	<i>m1, m2</i>
Acabamento de peças moldantes	DNC_M	<i>m6</i>
Furar fixações e/ou extração	Furfix	<i>m3, m4</i>
Erosão	Erosao	<i>m7</i>
Montagem	Mont	<i>m8</i>

Os tempos de produção (em horas) são valores inteiros e as mesmas máquinas têm tempos de produção diferentes dependendo do tipo de peça e da sequência de produção. As tabelas 2 e 3 apresentam os valores considerados para estes parâmetros. Na última linha das tabelas figura o tempo total de produção mínimo de cada peça, correspondente à soma dos tempos de produção se forem escolhidas as máquinas mais rápidas (com tempos de produção menores) em operações em que há mais de uma máquina disponível. Os tempos de produção tanto da operação de furar águas como de furar fixações e/ou extração da peça *p2* são mais elevados do que os da peça *p1* devido ao facto da cavidade ser uma peça mais complexa do que a bucha. A cavidade é a peça por onde se efetua a injeção do material enquanto a bucha é a peça por onde se efetua a extração do molde, operação que requer um mecanismo mais simples do que a injeção.

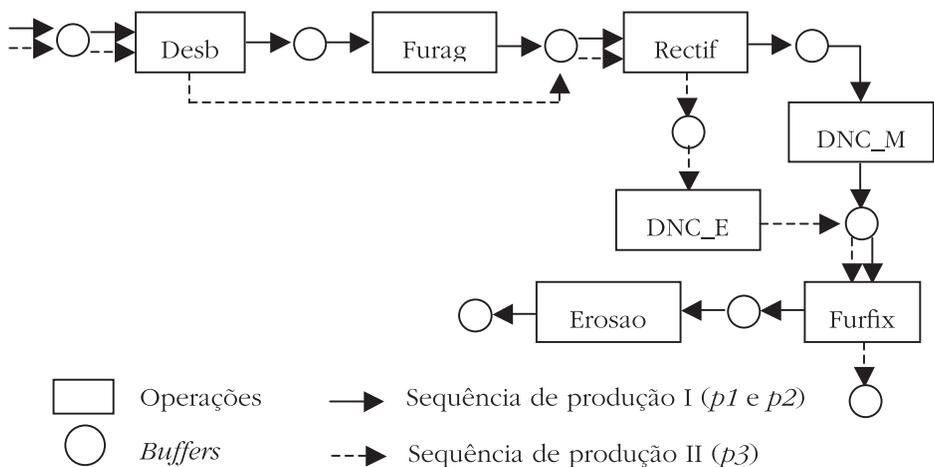


Figura 4: Seqüências de produção (exemplo ilustrativo).

Tabela 2: Tempos de produção da seqüência de produção I (exemplo ilustrativo).

Operação (o)	Máquinas (m)	Tempo de produção (horas)	
		$p1$	$p2$
Desb	$m1$	36	36
	$m2$	36	36
Furag	$m3$	74	86
	$m4$	72	84
Rectif	$m5$	2	2
DNC_M	$m6$	80	80
Furfix	$m3$	14	38
	$m4$	12	36
Erosao	$m7$	48	48
Mont	$m8$	15	15
Tempo total de produção mínimo (horas)		265	301

Tabela 3: Tempos de produção da sequência de produção II (exemplo ilustrativo).

Operação (<i>o</i>)	Máquinas (<i>m</i>)	Tempo de produção (horas)
		<i>p3</i>
Desb	<i>m1</i>	12
	<i>m2</i>	12
Rectif	<i>m5</i>	3
DNC_E	<i>m1</i>	36
	<i>m2</i>	36
Furfix	<i>m3</i>	18
	<i>m4</i>	16
Tempo total de produção mínimo (horas)		67

A escolha dos valores dos coeficientes na função objetivo teve como base a atribuição de Gomes (2007): $A_k = 20$, $B_k = 1$, $Z_k = 0,1$ e $N_k = 10^7$ (revendo: penalidade por produção após a data de entrega, antes desta, por tempo de espera nos *buffers* intermédios e pela existência de peças incompletas no final do horizonte de escalonamento). O último coeficiente é muito elevado quando comparado com os restantes para evitar a situação de peças incompletas. De referir que a empresa foi consultada sobre os valores recomendados para estas penalidades mas não foi possível obter uma contrapartida.

A data de entrega das peças *p1* e *p2* (após montagem) é de 357 horas e a da peça *p3* 342 horas. Será produzida uma unidade de cada peça, pelo que todas as variáveis de decisão no modelo apenas tomarão valores binários (0 ou 1). O horizonte de escalonamento utilizado foi igual à maior data de entrega: 357 horas dividido em intervalos de uma hora (este horizonte revelou-se suficiente para resolver o problema, cumprindo-se as datas de entrega).

Definidos todos os parâmetros, o modelo OPTMESC (com montagem de peças) foi implementado no sistema de modelação GAMS (versão 23.9.2) e resolvido com o *software* CPLEX versão 12.3 num computador *Intel Core* com um processador de 2,20 GHz, 6 GB de RAM e o sistema operativo *Windows 7*.

A tabela 4 sumariza os resultados obtidos mostrando as características do modelo (número de variáveis e de restrições), o número de iterações e o tempo computacional necessários para a resolução à otimalidade (a solução ótima é obtida impondo uma margem de erro - *relative gap* - de 0% ao correr o CPLEX) e o valor da função objetivo.

Tabela 4: Características numéricas do modelo e resultados (exemplo ilustrativo).

Nº de variáveis	Nº de restrições	Nº de iterações	Função objetivo	Tempo CPU (s)
19.954	16.039	3.166	5,4	4,6

Apresentam-se ainda os resultados detalhados da aplicação do modelo OPTMESC. As tabelas 5 e 6 indicam as sequências de produção para as três peças, com os instantes de início e fim de cada operação bem como a máquina a que esta foi afeta (atribuída). A tabela 7 mostra a ocupação de cada máquina, isto é, a sequência de operações na mesma.

A análise das tabelas 5 e 6 permite concluir que o intervalo de tempo entre inícios sucessivos de operações coincide com os tempos de produção indicados nas tabelas 2 e 3, pelo que não há necessidade das peças esperarem nos *buffers*. Excetua-se o caso da peça *p2* ao passar da operação de *Erosão (Erosao)* à operação de *Montagem (Mont)*. De facto, para esta peça mediaram 102 horas entre o início das duas operações enquanto o tempo de *Erosão* era apenas de 48 horas (tabela 2). Assim, após a operação de *Erosão* a peça esteve no *buffer* da operação de *Montagem* durante 54 horas, que se justifica pela necessidade da peça *p2* ter de esperar pela *p1* para passarem em simultâneo pela operação de *Montagem*. Em conformidade, a produção de ambas as peças terminou no mesmo instante.

Analisando a tabela 7 verifica-se que não há sobreposição de operações: não se encontra mais do que uma peça em cada máquina num determinado instante, tal como exigido pela restrição de capacidade das máquinas (dado que se considerou uma capacidade unitária). Também se pode verificar que há peças que visitam a mesma máquina mais de uma vez (recirculação de peças). É o caso da peça *p3* que é processada na máquina *m2* no *Desbaste (Desb)* e

volta a ser processada na operação de *Acabamento de estruturas (DNC_E)*; entre estas duas operações foi executada a operação de *Retificação (Rectif)* da peça na máquina *m5* (tabela 6). De notar que na solução devolvida pelo modelo não se fez uso da máquina *m1*.

Tabela 5: Sequência de produção com montagem das peças *p1* e *p2*.

Operação (o)	<i>p1</i>			<i>p2</i>		
	Máquina (m)	Instante de início (t)	Instante de fim (t)	Máquina (m)	Instante de início (t)	Instante de fim (t)
Desb	<i>m2</i>	90	126	<i>m2</i>	0	36
Furag	<i>m3</i>	126	200	<i>m4</i>	36	120
Rectif	<i>m5</i>	200	202	<i>m5</i>	120	122
DNC_M	<i>m6</i>	202	282	<i>m6</i>	122	202
Furfix	<i>m4</i>	282	294	<i>m3</i>	202	240
Erosao	<i>m7</i>	294	342	<i>m7</i>	240	288
Mont	<i>m8</i>	342	357	<i>m8</i>	342	357

Tabela 6: Sequência de produção da peça *p3*.

Operação (o)	<i>p3</i>		
	Máquina (m)	Instante de início (t)	Instante de fim (t)
Desb	<i>m2</i>	273	285
Rectif	<i>m5</i>	285	288
DNC_E	<i>m2</i>	288	324
Furfix	<i>m3</i>	324	342

Concluindo com a interpretação do valor ótimo da função objetivo (tabela 4), este resulta do facto da peça *p2* ter estado no *buffer* que antecede a operação de Montagem durante 54 horas, como já mencionado. Não existem peças incompletas no final do horizonte de escalonamento nem se aplica a penalidade por produzir peças antes ou depois da data de entrega (todas as

peças foram finalizadas exatamente na data de entrega). Sendo assim, o valor da função objetivo é 5,4 ($FO = 54 \times 0,1 = 5,4$).

Tabela 7: Ocupação das máquinas.

Máquina (<i>m</i>)	Operação (<i>o</i>)	Peça (<i>p</i>)	Instante de início	Instante de fim
<i>m2</i>	Desb	<i>p2</i>	0	36
	Desb	<i>p1</i>	90	126
	Desb	<i>p3</i>	273	285
	DNC_E	<i>p3</i>	288	324
<i>m3</i>	Furag	<i>p1</i>	126	200
	Furfix	<i>p2</i>	202	240
	Furfix	<i>p3</i>	324	342
<i>m4</i>	Furag	<i>p2</i>	36	120
	Furfix	<i>p1</i>	282	294
<i>m5</i>	Rectif	<i>p2</i>	120	122
	Rectif	<i>p1</i>	200	202
	Rectif	<i>p3</i>	285	288
<i>m6</i>	DNC_M	<i>p2</i>	122	202
	DNC_M	<i>p1</i>	202	282
<i>m7</i>	Erosao	<i>p2</i>	240	288
	Erosao	<i>p1</i>	294	342
<i>m8</i>	Mont	<i>p1</i>	342	357
	Mont	<i>p2</i>	342	357

5. Caso de estudo

Nesta secção descreve-se em primeiro lugar a aplicação do modelo OPTMESC à determinação de um plano de produção da empresa. Foram utilizados os dados de um plano real obtido pelo líder de grupo (LG), comparando-se a solução do modelo com este plano (solução manual). Para tal, foi necessário proceder a um levantamento das características do parque das máquinas da GECO (para cada máquina, as operações que podem ser

realizadas na mesma e as dimensões máximas nos eixos X, Y e Z das peças que suporta), bem como um levantamento das operações e respectivas durações (tempos de produção) de todas as peças envolvidas no plano de produção. A informação compilada está descrita na dissertação de mestrado de Virgílio (2011). Na segunda parte, apresenta-se uma análise pós-otimização aos coeficientes da função objetivo, às datas de entrega das encomendas e à duração do horizonte de escalonamento utilizado, os parâmetros mais relevantes do modelo.

5.1 Aplicação a um plano de produção da empresa

O plano de produção selecionado consiste em 42 peças pertencentes a 6 moldes diferentes e envolve a afetação de operações a 9 máquinas (de capacidade unitária). A tabela 8 descreve estas operações e a divisão que foi necessário efetuar pelo facto de certas peças só poderem ser carregadas em determinadas máquinas. São as operações resultantes da divisão referida que constituem o conjunto *O* de operações no modelo (vd. secção 4.2.1), indicando-se na tabela 8 as máquinas que lhes estão associadas. Por exemplo, a operação de desbaste foi dividida em três operações (*Desb1*, *Desb2*, *Desb3*) devido às máquinas disponíveis. A máquina *m1* pode efetuar qualquer destas operações, pois suporta peças de maior dimensão e portanto pode também desbastar peças de menor dimensão. Contudo, a máquina *m2* apenas pode efetuar a operação *Desb1* pois só suporta peças de dimensão reduzida.

Após a definição das operações de uma forma utilizável no modelo OPTMESC, definiram-se, com base na consulta do plano do LG, 12 seqüências de produção para as peças envolvendo no total 52 operações. Para uma mesma máquina os tempos de produção diferem consoante a operação e a peça. O conjunto das seqüências de produção e os tempos respetivos estão descritos em Virgílio (2011). De referir que, ao determinar o plano de produção, o LG recorre à sua experiência para definir as operações (que variam de acordo com o tipo de peça) e estimar os respetivos tempos de fabrico¹⁶. Tal

¹⁶ Os LG possuem por isso um conhecimento prático das várias etapas do processo indispensável a este “pré-processamento” das operações para a determinação do plano de produção.

como no exemplo ilustrativo (e pelas mesmas razões) optou-se por igualar a capacidade dos *buffers* das operações a 100 unidades.

Tabela 8: Operações e máquinas associadas (plano de produção).

Operação	Operação no modelo (<i>o</i>)	Máquinas (<i>m</i>)
Desbaste	Desb1	<i>m1, m2, m7, m8</i>
	Desb2	<i>m1, m7</i>
	Desb3	<i>m1</i>
Furar águas	Furag	<i>m9</i>
Acabamento de estruturas	DNC_E1	<i>m1, m7</i>
	DNC_E2	<i>m1</i>
Acabamento de peças moldantes	DNC_M1	<i>m3, m4, m5, m6</i>
	DNC_M2	<i>m3, m4</i>
	DNC_M3	<i>m4</i>
Furar fixações e/ou extração	Furfix	<i>m9</i>

Considerou-se a forma do modelo OPTMESC sem montagem de peças, pois esta operação não foi tida em conta pelo LG no plano de produção. Assim, todas as peças pertencentes a um dado molde surgem na função objetivo com a mesma data de entrega. As peças e datas de entrega figuram na tabela 9; cada uma das linhas corresponde a um molde (são 6 no total). É produzida uma unidade de cada peça ($Q_p = 1$) e as penalidades na função objetivo são idênticas às do exemplo ilustrativo.

O plano de produção do LG corresponde a um intervalo de tempo de 51 dias, isto é, 2 meses e 9 dias de laboração da fábrica (não contando fins de semana). Constatou-se que os tempos de produção expressos neste plano são múltiplos inteiros de 6 horas, período que corresponde a meio dia de laboração (um dia tem 12 horas de laboração). Assim, na grelha temporal utilizada no modelo OPTMESC a unidade de tempo corresponde a meio dia, estando as datas de entrega na tabela 9 expressas nessa unidade de tempo (meios-dias).

Verificou-se existirem no plano do LG várias máquinas ocupadas no início do horizonte de escalonamento (todas com exceção das máquinas $m1$ e $m3$). Esta pré-ocupação das máquinas foi considerada no modelo OPTMESC modificando as restrições de inicialização [2] de forma a garantir que as máquinas não estejam disponíveis no início do horizonte de escalonamento. As variáveis de decisão correspondem a operações que foram iniciadas e terminadas (no horizonte de escalonamento) no plano do LG.

Estudaram-se duas situações distintas na determinação do plano de produção, designadas plano A1 e plano B1. No primeiro são apenas tidas em conta as peças produzidas internamente na GECO no plano do LG. No segundo, além das peças do plano A1 incluíram-se peças que foram subcontratadas no plano do LG. Na verdade, desse plano constavam 56 peças das quais 42 foram produzidas na empresa (as peças que integraram o plano A1) e 14 foram subcontratadas por receio de que a capacidade das máquinas fosse insuficiente para o cumprimento das datas de entrega.

Para verificar se seria possível produzir mais peças internamente face ao plano do LG, introduziu-se uma peça de cada vez no conjunto de dados do plano A1 e resolveu-se de novo o modelo OPTMESC, repetindo-se o procedimento até deixar de ser possível finalizar todas as peças no horizonte de escalonamento estipulado. A estratégia utilizada na escolha das peças consistiu em dar prioridade às peças principais do molde (cavidades e buchas). Desta forma verificou-se ser viável produzir mais 7 das 14 peças subcontratadas, num total de 49 peças escalonadas. A introdução destas peças requer o acréscimo de novas sequências de produção para 5 peças, sendo utilizadas para as restantes 2 peças sequências de produção já definidas para o plano A1. De referir ainda que estas 7 peças pertencem aos 6 moldes do plano A1. As peças adicionais deste novo plano de produção (plano B1) estão discriminadas entre parêntesis na tabela 9.

Para a implementação e resolução do modelo OPTMESC utilizaram-se as mesmas ferramentas computacionais descritas em 4.3. O horizonte de escalonamento é de 105 meios-dias (o múltiplo de 5 imediatamente superior à maior data de entrega na tabela 9). As características dos modelos de programação linear inteira obtidos e os resultados da sua resolução são

apresentados nas tabelas 10A e 10B, para uma margem de erro de 0%. De referir que usando margens de erro de 5% e 10% se obteve a mesma solução (com os mesmos valores da função objetivo e das variáveis de decisão) em tempos computacionais semelhantes. Quanto às medidas de desempenho de ambas as soluções (planos de produção), além do valor da função objetivo indicam-se nas tabelas 10A e 10B as componentes desta função antes de multiplicadas pelos respetivos coeficientes: o tempo total de espera nos *buffers* intermédios e da produção de peças antes e após a data de entrega.

Tabela 9: Dados das peças para 6 moldes (plano A1 + peças adicionais consideradas no plano B1).

Peças (<i>p</i>)	Data de entrega (<i>D_i</i>)
<i>p1, p2, p3, p4, p5, p6, p7, p8</i>	26
<i>p9, p10, p11, p12, p13, p14, p15</i>	80
<i>p16, p17, p18, p19, p20, p21 + (p43, p44, p45, p47 – plano B1)</i>	76
<i>p22, p23, p24, p25, p26, p27, p28, p29</i>	86
<i>p30, p31, p32, p33, p34, p35, p36 + (p49 – plano B1)</i>	102
<i>p37, p38, p39, p40, p41, p42 + (p46, p48 – plano B1)</i>	102

Tabela 10A: Características numéricas do modelo e resultados para os planos de produção A1 e B1.

Plano	Nº de peças	Nº de variáveis	Nº de restrições	Nº de iterações	Nº de nós
A1	42	30.401	18.164	4.992	786
B1	49	34.302	20.490	16.983	6

Tabela 10B: Características numéricas do modelo e resultados para os planos de produção A1 e B1.

Plano	Função objetivo	Tempo CPU (s)	Tempo nos buffers	Tempo antes da data	Tempo após a data
A1	384,60	4,62	186	366	0
B1	485,40	3,29	234	462	0

Analisando os resultados, é de salientar que em ambos os casos o modelo OPTMESC foi resolvido à otimalidade em menos de 5 segundos. Não existem peças incompletas no final do horizonte de escalonamento nem peças terminadas depois da data de entrega em ambos os planos, sendo o valor da função objetivo superior para o plano B1 devido ao aumento do tempo total de peças à espera nos *buffers* e produzidas antes da data. De referir ainda que o tempo de CPU foi inferior no caso do plano B1, apesar da dimensão do modelo ser superior (maior número de variáveis e de restrições). Com efeito, o número de nós analisados na pesquisa *branch-and-bound* foi consideravelmente inferior para o plano B1, justificando um menor tempo de cálculo. Para comparar o desempenho do modelo com a solução manual, em termos do tempo de cálculo, inquiriu-se o LG: este, dada a experiência que acumulou, demora em média uma hora a planear um conjunto de peças semelhante. O uso do modelo OPTMESC poderá assim reduzir muito significativamente o tempo despendido na determinação de um plano de produção na GECO.

Na figura 5 efetua-se uma comparação entre os planos do LG e os planos A1 e B1 no que respeita às taxas de ocupação das máquinas. A taxa média de ocupação das máquinas no plano do LG e no plano A1 é de 71% mas o desvio padrão difere, sendo de 24% no primeiro caso e 19% no segundo. Consta-se portanto que o modelo OPTMESC produziu uma solução mais uniforme do ponto de vista da afetação das operações às máquinas. Quanto ao plano B1, apresenta uma taxa média de ocupação das máquinas superior (79%) devido à introdução de mais peças mas o desvio padrão, de 20%, mantém-se muito próximo do observado no plano A1.

Em anexo inclui-se a carta de *Gantt* correspondente ao plano B1, desenhada com recurso ao *software Microsoft Project*, com dois objetivos: permitir a visualização da solução produzida pelo modelo e ilustrar a representação dos planos de produção comumente usada na GECO.

5.2 Análise pós-otimização

Os coeficientes, ou penalidades, que figuram na função objetivo têm uma influência determinante sobre as características da solução obtida. Por esse

motivo averiguou-se junto do LG quais os valores recomendados para as penalidades, mas dada a falta de familiaridade deste com o modelo OPTMESC não foram sugeridos outros valores. Assim, torna-se importante uma análise dos coeficientes da função objetivo do modelo.

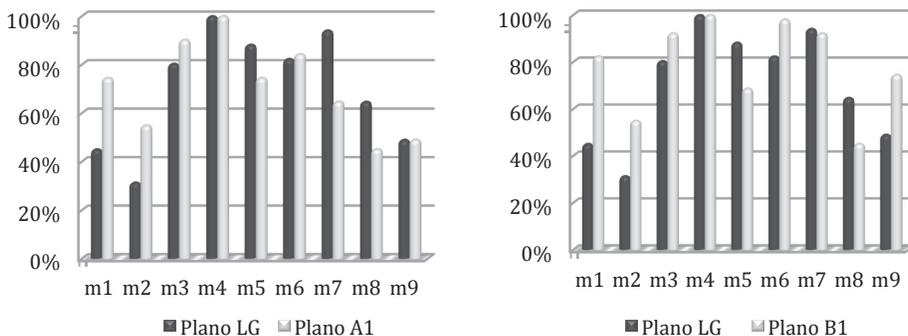


Figura 5: Ocupação das máquinas nos planos do LG, A1 e B1.

Partindo do plano com o maior número de peças (solução B1) variaram-se as penalidades do tempo de espera nos *buffers* (Z_k) e do tempo total de produção antes (B_k) e após a data de entrega (A_k), observando-se o impacto sobre a solução obtida. Não se variou a penalidade N_k pois esta só tem efeito na função objetivo quando não é possível finalizar todas as peças no horizonte de escalonamento, o que não é o caso para o horizonte de escalonamento utilizado (105 meios-dias). Os resultados figuram nas tabelas 11A e 11B. A solução B2 foi obtida com penalidades iguais para as três medidas de desempenho (e iguais a 1) enquanto as soluções B3 e B4 correspondem a dar maior importância a duas das componentes (penalidades iguais a 100) mantendo a outra penalidade igual a 1. Todas as soluções são ótimas (margem de erro de 0%).

De notar, em primeiro lugar, que os valores da função objetivo das quatro soluções não são comparáveis devido aos coeficientes serem diferentes; é relevante, sim, comparar os valores das medidas de desempenho tempo nos *buffers*, tempo de produção antes e após a data de entrega. Nas soluções B1 e B3 não são produzidas peças após as datas de entrega, ao contrário das soluções B2 e B4. A solução B3 é diferente da B1 sobretudo no tempo de

espera nos *buffers*, bastante reduzido (apenas 8 meios-dias no total) devido à penalidade respetiva ter aumentado face à usada em B1. A solução B1 é aquela com tempo de espera nos *buffers* mais elevado, B3 a que tem o pior valor de tempo de produção antes da data e B4 a que mais ultrapassa as datas de entrega, em consonância com uma menor importância relativa de cada uma destas medidas de desempenho na função objetivo (respetivamente, os valores de Z_k , B_k e A_k). Na solução B2, com penalidades iguais para as três medidas de desempenho, obtiveram-se valores intermédios das mesmas face às restantes soluções da tabela. Quanto ao tempo de computação, de referir que as soluções B1, B3 e B4 demoraram alguns segundos a ser calculadas enquanto o tempo requerido para a solução B2 foi uma ordem de grandeza superior (cerca de 30 segundos). De facto, o número de nós da árvore de procura (processo *branch-and-bound*) analisados foi muito superior para esta solução: cerca de 1000 contra um máximo de 25 na solução B3.

Partindo da solução B1 estudou-se ainda o impacto sobre o desempenho da solução ótima de aproximar (“compactar”) as datas de entrega. Não se variou a data de entrega do conjunto de peças do primeiro molde (26 meios-dias) mas as restantes foram diminuídas, em simultâneo, 10, 20, 30 e 40 meios-dias como se mostra na tabela 12. Manteve-se o horizonte de escalonamento de 105 meios-dias e a função objetivo da solução B1 (penalidades $A_k=20$, $B_k=1$ e $Z_k=0,1$), sendo os resultados desta análise apresentados na tabela 13.

A tabela 13 mostra que à medida que as datas de entrega são compactadas a função objetivo se degrada, aumentando de valor, e o tempo de CPU necessário para resolver o modelo à otimalidade cresce. Excetua-se a passagem da solução B1 à B5, em que o tempo de cálculo da segunda é ligeiramente inferior ao da primeira. O aumento do valor da função objetivo deve-se ao aumento do tempo de produção após a data, embora o tempo de espera nos *buffers* diminua tal como o tempo de produção antes da data.

As soluções B1 a B7 apresentam tempos de resolução que não ultrapassam 40 segundos enquanto para a solução B8 este foi de 656 segundos (cerca de 11 minutos). O número de nós analisados na árvore de procura de facto aproxima-se de 20.000, significativamente superior aos das soluções anteriores.

Para esta solução (B8) efetuou-se por último a experiência de encurtar o horizonte de escalonamento. Os resultados figuram nas tabelas 14A e 14B.

Tabela 11A: Análise às penalidades na função objetivo.

Solução	B1			B2		
Penalidades	$A_k = 20$	$B_k = 1$	$Z_k = 0,1$	$A_k = 1$	$B_k = 1$	$Z_k = 1$
Nº de variáveis	34.302					
Nº de restrições	20.490					
Nº de iterações	16.983			144.854		
Nº de nós	6			998		
Tempo CPU (s)	3,32			29,59		
Função objetivo	485,40			570		
Margem de erro	0			0		
Tempo nos buffers	234			57		
Tempo antes da data	462			472		
Tempo após a data	0			41		

Tabela 11B: Análise às penalidades na função objetivo.

Solução	B3			B4		
Penalidades	$A_k = 100$	$B_k = 1$	$Z_k = 100$	$A_k = 1$	$B_k = 100$	$Z_k = 100$
Nº de variáveis	34.302					
Nº de restrições	20.490					
Nº de iterações	15.110			21.153		
Nº de nós	25			3		
Tempo CPU (s)	4,77			5,51		
Função objetivo	1.422			50.106		
Margem de erro	0			0		
Tempo nos buffers	8			34		
Tempo antes da data	622			466		
Tempo após a data	0			106		

Tabela 12: Variação das datas de entrega face à solução B1.

Peças (p)	Variação das datas de entrega (D_k)				
	0	-10	-20	-30	-40
	Sol. B1	Sol. B5	Sol. B6	Sol. B7	Sol. B8
$p1, p2, p3, p4, p5, p6, p7, p8$	26	26	26	26	26
$p9, p10, p11, p12, p13, p14, p15$	80	70	60	50	40
$p16, p17, p18, p19, p20, p21 + (p43, p44, p45, p47 - \text{plano B1})$	76	66	56	46	36
$p22, p23, p24, p25, p26, p27, p28, p29$	86	76	66	56	46
$p30, p31, p32, p33, p34, p35, p36 + (p49 - \text{plano B1})$	102	92	82	72	62
$p37, p38, p39, p40, p41, p42 + (p46, p48 - \text{plano B1})$	102	92	82	72	62

Tabela 13: Análise da variação das datas de entrega.

Solução	B1	B5	B6	B7	B8
Variação das datas	0	- 10	- 20	- 30	- 40
Nº de variáveis	34.302				
Nº de restrições	20.490				
Nº de iterações	16.983	17.433	51.012	244.216	2.028.778
Nº de nós	6	0	110	2.148	19.123
Tempo CPU (s)	3,32	2,01	10,51	39,38	655,59
Função objetivo	485,40	566,20	1.832,20	3.980,60	7.193,00
Margem de erro	0	0	0	0	0
Tempo nos buffers	234	222	82	56	30
Tempo antes da data	462	504	464	375	290
Tempo após a data	0	2	68	180	345

Verifica-se que para horizontes de escalonamento de 104 a 94 meios-dias as soluções não variam, em termos do valor da função objetivo e das medidas de desempenho das três últimas colunas da tabela 14A, face à solução B8, obtida com um horizonte de escalonamento de 105 meios-dias.

Tabela 14A: Análise da variação do horizonte de escalonamento.

Solução	B8	B9	B10	B11
Horizonte temporal	105	104	94	93
Nº de variáveis	34.302	34.038	31.398	31.398
Nº de restrições	20.490	20.367	20.354	19.137
Nº de iterações	2.028.778	1.097.998	996.756	716.312
Nº de nós	19.123	13.458	13.352	6.571
Tempo CPU (s)	655,59	632,41	565,32	137,06
Função objetivo	7.193,00	7.193,00	7.193,00	7.745,40
Margem de erro	0	0	0	0
Tempo nos buffers	30	30	30	84
Tempo antes da data	290	290	290	277
Tempo após a data	345	345	345	373

Tabela 14B: Análise da variação do horizonte de escalonamento.

Solução	B12	B13	B14
Horizonte temporal	92	92	92
Nº de variáveis	30.870	30.870	30.870
Nº de restrições	18.891	18.891	18.891
Nº de iterações	3.680.171	36.155	36.155
Nº de nós	36.088	117	117
Tempo CPU (s)	1.272,33	303,67	303,63
Função objetivo	10.006.262,30	10.006.301,10	10.006.301,10
Margem de erro	0	5%	10%
Tempo nos buffers	23	61	61
Tempo antes da data	280	275	275
Tempo após a data	299	301	301

Indicam-se na tabela 14A as soluções para os dois valores do comprimento do horizonte de escalonamento referidos (104 e 94 meios-dias), designadas B9

e B10. Apenas para um horizonte de escalonamento de 93 meios-dias se obtém uma solução com um valor da função objetivo diferente (superior) ao da solução B8: trata-se da solução B11 na tabela. Comparando com a solução B8, observa-se que o aumento do valor ótimo da função objetivo resulta do aumento do tempo de produção após a data e do tempo de espera nos *buffers*, embora o tempo de produção antes da data tenha diminuído. De referir que esta nova solução (B11) foi calculada num tempo de CPU bastante mais curto (137 segundos contra 656, isto é, cerca de 2 minutos contra 11), em conformidade com um número de nós analisados inferior. Ao diminuir o horizonte de escalonamento para 92 meios-dias deixa de ser possível produzir todas as peças (solução B12, tabela 14B): há uma peça que não foi terminada, razão pela qual o valor da função objetivo aumentou muito ($N_k = 10^7$). O modelo foi agora mais difícil de resolver: o tempo de CPU para o cálculo da solução ótima (margem de erro nula) subiu para 1272 segundos (cerca de 21 minutos), tendo o número de nós analisados quase duplicado face a B8. De notar, todavia, que no caso deste horizonte de escalonamento (92 meios-dias), se impusermos uma margem de erro superior a zero ao correr o modelo é possível obter uma boa solução em tempos de CPU significativamente inferiores. De facto, a tabela 14B mostra que com margens de erro de 5% ou de 10%, o modelo devolve uma solução em 304 segundos (cerca de 5 minutos), idêntica nos dois casos em termos da função objetivo e medidas de desempenho (soluções B13 e B14, respetivamente).

Conclui-se assim que a influência do horizonte de escalonamento escolhido ao resolver o modelo é determinante, tanto do ponto de vista da qualidade da solução como do tempo de computação. A tolerância face à proximidade do ótimo (margem de erro) tem igualmente um impacto importante. Com o modelo OPTMESC o decisor poderá testar diferentes valores para estes parâmetros e escolher uma de entre as várias soluções geradas.

6. Conclusões e trabalho futuro

Este trabalho propõe um modelo de programação linear inteira, OPTMESC, para resolver o problema do escalonamento da produção no setor dos moldes. Este modelo, de tempo discreto, foi aplicado a dados reais disponibilizados

pela empresa GECO, determinando-se um plano (para um período de 2 meses e 9 dias) envolvendo a produção de várias peças de diferentes moldes. Com recurso aos modernos sistemas de implementação e resolução de modelos de programação matemática (linguagem GAMS e *solver* CPLEX, no caso presente) foram obtidas as soluções ótimas para duas instâncias do problema (planos A1 e B1) numa questão de segundos. Numa comparação com o plano do diretor de produção da empresa, mostrou-se que com a capacidade disponível das máquinas é possível escalonar um número significativo de peças adicionais cuja produção foi subcontratada no referido plano. Este é um resultado relevante pois a redução da subcontratação de peças reduz a incerteza associada à sua qualidade e ao cumprimento das datas de entrega dos moldes.

A análise pós-otimização realizada ilustra o tipo de testes que o decisor pode efetuar com o modelo OPTMESC de forma a escolher a solução (plano de produção) que julgue globalmente mais adequada, por exemplo atribuindo diferentes penalidades às medidas de desempenho na função objetivo. Mostrou-se ainda como o modelo pode ser utilizado para testar a possibilidade de tentar cumprir diferentes datas de entrega, dada a capacidade disponível das máquinas, e ilustrou-se o impacto de variações no horizonte de escalonamento e margem de erro sobre as soluções geradas.

A consulta do responsável de produção sobre os coeficientes (penalidades) utilizados no modelo OPTMESC merece uma reflexão. Não foi possível obter uma crítica ou uma contrapartida para os valores utilizados, tendo apenas sido referido que há cinco anos que a GECO não entrega uma encomenda atrasada e pretende manter essa imagem no mercado. Com efeito, um problema que surge frequentemente ao tentar aplicar a Investigação Operacional em organizações é a dificuldade em formular o objetivo a atingir, que num modelo de programação linear se consubstancia na função objetivo. Há portanto necessidade de trabalhar muito de perto com as pessoas responsáveis na empresa, integrando-as nas fases de estruturação do problema e definição dos objetivos a atingir. Este aspeto foi levado a cabo ao longo deste caso de estudo e apesar de não se terem conseguido todos os dados requeridos (nomeadamente os valores das penalidades na função objetivo), o resultado do trabalho contribuiu positivamente para uma análise mais estruturada do

problema de escalonamento, para além da informação positiva sobre uma forma de melhorar o planeamento da empresa.

Os tempos computacionais obtidos são compatíveis com uma periodicidade semanal de resolução do problema (obtenção de novos escalonamentos), que é a situação habitual na empresa. O modelo mostra-se portanto adequado à inclusão num sistema de apoio à decisão (SAD) para o escalonamento na indústria de moldes, na qual o problema é ainda comumente resolvido sem recurso a algoritmos implementados em computador. De facto, o diretor de produção da GECO viu com muito interesse o desenvolvimento deste projeto e os resultados obtidos. Porém, a operacionalização de tal ferramenta SAD vai requerer a criação de uma interface amigável, interativa e flexível com o utilizador, que permita nomeadamente visualizar o plano de carga das máquinas (carta de Gantt) diretamente a partir dos resultados do modelo. Conceber esta interface é uma das direções de trabalho futuro.

As outras direções de trabalho têm a ver com generalizações do modelo OPTMESC. Em primeiro lugar, surge a generalização que contemple a decisão sobre as peças a subcontratar em caso de falta de capacidade interna. Neste trabalho o procedimento usado na seleção de peças para obtenção do plano B1 foi uma heurística simples, e será importante estudar a otimização desta escolha através da sua inclusão no modelo de programação linear inteira. Um outro âmbito é a generalização do modelo ao problema do reescalonamento da produção (escalonar de novo para ter em conta alterações ocorridas), problema estudado em Gomes (2007) e Gomes *et al.* (2010). Nesta indústria é muito comum existir a necessidade de corrigir moldes já finalizados, tornando-se necessário inserir no plano de produção operações de correção das peças; atualmente esta inserção não é feita de forma sistemática pelo diretor de produção da empresa. Outra situação a estudar é a inserção de novas encomendas (novos moldes) num plano existente, fazendo uso da capacidade disponível das máquinas.

Em conclusão, este projeto revelou-se promissor no que respeita ao fortalecimento da ligação entre a universidade e a indústria de moldes, podendo vir a proporcionar oportunidades de colaboração com a empresa

GECO no desenvolvimento de uma ferramenta computacional de apoio ao planeamento e escalonamento da produção.

Agradecimentos

As autoras agradecem ao Sr. António Febra, diretor da GECO, a abertura e o incentivo à realização deste trabalho, recebendo a Bárbara Virgílio como estagiária na empresa, e ao Sr. Pedro Henriques, diretor de produção, e à sua equipa, todo o auxílio prestado no levantamento de dados bem como as orientações preciosas para a consecução do projeto. Foi com alegria e orgulho que as orientadoras viram a Bárbara Virgílio iniciar a sua vida profissional na GECO após a conclusão do Mestrado em Engenharia e Gestão Industrial do Instituto Superior Técnico, em novembro de 2011.

Agradecem ainda aos dois revisores anónimos pelas numerosas sugestões que permitiram melhorar este trabalho.

BIBLIOGRAFIA

A Indústria Portuguesa de Engineering and Tooling. (2010). Obtido em Outubro de 2010, do website da CEFAMOL: <http://www.cefamol.pt>.

CHOI, B. K. & You, N. K. (2006). Dispatching rules for dynamic scheduling of one-of-a-kind production. *International Journal of Computer Integrated Manufacturing*, 19, 383-392.

CHOY, K.L., Leung, Y.K., Chow, H.K.H., Poon, T.C., Kwong, C.K., Ho, G.T.S. & Kwok, S.K. (2011). A hybrid scheduling decision support model for minimizing job tardiness in a make-to-order based mould manufacturing environment. *Expert Systems with Applications*, 38, 1931-1941.

FLOUDAS, C. A. & Lin, X. (2004). Continuous-time versus discrete-time approaches for scheduling of chemical processes: a review. *Computers and Chemical Engineering*, 28, 2109-2129.

GOMES, M., Barbosa-Póvoa, A. & Novais, A.Q. (2005). Optimal scheduling for flexible job shop operation. *International Journal of Production Research*, 43, 2323-2353.

GOMES, C. (2007). *Reactive scheduling in make-to-order production systems: An optimization based approach*. Tese de Doutoramento em Engenharia de Sistemas, Instituto Superior Técnico, Universidade Técnica de Lisboa.

GOMES, M.C., Barbosa-Póvoa, A. & Novais, A.Q. (2010). A discrete time reactive scheduling model for new order insertion in job shop, make-to-order industries. *International Journal of Production Research*, 48, 7395-7422.

IAPMEI (2006). *A indústria portuguesa de moldes no contexto internacional - Contributos para o conhecimento da actividade*. Rede de gabinetes de empresa do IAPMEI.

JAIN, A. S & Meeran, S. (1999). Deterministic job shop scheduling: Past, present, future. *European Journal of Operational Research*, 113, 390-434

LOPES, F. (2000). *Retratos da indústria portuguesa de moldes para plásticos*. Edição da Universidade do Minho: Braga.

MÉNDEZ, C. A., Cerdá, J., Grossmann, I. E., Harjunkski, I. & Fahl, M. (2006). State-of-the-art review of optimization methods for short-term scheduling of batch processes. *Computers and Chemical Engineering*, 30, 913-946.

NI, Q., Feng, L. W., Yarlagadda, P. K. & Ming, X. (2007). Business information modeling for process integration in the mold making industry. *Robotics and Computer-Integrated Manufacturing*, 23, 195-207.

OZGUVEN, C., Ozbakir, L. & Yavuz, Y. (2010). Mathematical Models for Job Shop Scheduling Problems with Routing and Process Plan Flexibility. *Applied Mathematical Modelling*, 34, 1539-1548.

PAN, J.C.-H. & Chen, J.-S. (2005). Mixed binary integer programming formulations for the reentrant job shop scheduling problem. *Computers and Operations Research*, 32, 1197-1212.

POTTS, C. & Strusevich, V. (2009). Fifty years of scheduling: a survey of milestones. *Journal of the Operational Research Society*, 60, S41-S68.

Plano Estratégico para o Sector dos Moldes em Portugal. (2008). Obtido em Janeiro de 2011, do website da Sociedade Portuguesa de Inovação: <http://www.spi.pt/p.publicacoes> .

TAY, J.C. & Ho, N.B. (2008). Evolving dispatching rules using genetic programming for solving multi-objective flexible job-shop problems. *Computers and Industrial Engineering*, 54, 453-473.

VIRGÍLIO, Bárbara E.S. (2011), *Optimização do planeamento e escalonamento da produção na indústria de moldes: Aplicação ao caso da GEÇO*. Dissertação de Mestrado em Engenharia e Gestão Industrial, Instituto Superior Técnico, Universidade Técnica de Lisboa.

(Página deixada propositadamente em branco)

CASO 7

SISTEMA DE APOIO À DECISÃO DO ORÇAMENTO ANUAL DE PRODUÇÃO NA INDÚSTRIA DE BEBIDAS

Luís Guimarães

INESC TEC, Faculdade de Engenharia
Universidade do Porto
guimaraes.luis@fe.up.pt

Bernardo Almada-Lobo

INESC TEC, Faculdade de Engenharia
Universidade do Porto
almada.lobo@fe.up.pt

RESUMO

A criação do orçamento anual de produção na indústria das bebidas é uma tarefa complexa e exigente para os gestores. Consiste num plano de produção de longo prazo que define o escalonamento e dimensão dos lotes de produção numa série de linhas de enchimento localizadas em diferentes centros produtores. O objetivo do plano de produção a ser gerado é a satisfação da procura estimada minimizando os custos de produção, posse de inventário, transferência de produtos e horas extraordinárias.

O projeto foi realizado em colaboração com uma empresa do setor no âmbito da criação do seu orçamento anual de produção. Neste documento são apresentadas as várias etapas decorridas, desde a especificação do problema a resolver e metodologia de solução desenvolvida, até à sua validação e implementação. Atualmente a empresa utiliza a metodologia desenvolvida para a criação do orçamento anual de produção permitindo uma melhoria significativa da sua performance operacional, bem como a simulação de diferentes cenários para o ambiente produtivo, tendo já desempenhado um papel importante na reestruturação da sua cadeia de abastecimento.

PALAVRAS-CHAVE

Planeamento de longo prazo, Indústria de bebidas, Planeamento da produção e distribuição, Pesquisa em grandes vizinhanças, Programação inteira mista

1 Introdução

A indústria de bebidas fornece uma grande variedade de consumidores com produtos desde vinho até refrigerantes, passando pela cerveja, bebidas espirituosas e águas minerais. É uma das indústrias mais competitivas, muito fruto das diferenças culturais entre as populações, que criam o ambiente perfeito para o aparecimento de empresas especializadas em marcas reconhecidas a nível nacional ou regional. Nos últimos anos, a crescente concorrência no setor obrigou as empresas a expandir o seu leque de produtos. Aliado a este facto, o desenvolvimento de instalações produtivas mais automatizadas, despertou a necessidade latente de um planeamento da produção mais eficiente.

Este projeto foi inspirado num problema real de uma empresa que produz, distribui e comercializa cerveja e refrigerantes. O desafio prende-se com a criação de um plano de produção de longo prazo para as linhas de produção (enchimento) localizadas nos diferentes centros produtores, que se designa por orçamento de produção. É prática comum nesta indústria a existência de vários centros produtores para satisfazer uma procura geograficamente dispersa a um menor custo, uma vez que o transporte dos produtos finais pode representar uma percentagem significativa dos custos totais. Cada um destes centros produtores conta com múltiplas linhas de enchimento de onde saem os produtos finais para serem entregues aos clientes. Tipicamente, o orçamento é gerado para um horizonte rolante de 12 a 18 meses estando a escala temporal discretizada em períodos mensais. Decide-se, em cada período, as quantidades a produzir, a manter em inventário, a transferir entre os centros produtores e o número de horas extraordinárias necessárias, tendo por base a alocação de famílias de produtos às linhas de enchimento disponíveis. No contexto descrito, os centros produtores funcionam igualmente como centros distribuidores, uma vez que os armazéns de produto acabado localizados nas suas imediações são responsáveis por abastecer as encomendas geradas localmente. A coordenação dos centros produtores é essencial de forma a reduzir os custos de toda a cadeia de abastecimento.

O foco do projeto foi o desenvolvimento de um método de solução para a geração automática do orçamento de produção na indústria de bebidas que

sirva de apoio aos gestores na sua tomada de decisão. A complexidade inerente à resolução de instâncias reais de planeamento origina o aparecimento de problemas de grande escala. Neste contexto surge o compromisso entre eficácia e eficiência dos métodos de solução. Ao contrário dos métodos exatos que procuram identificar a solução ótima (provando a sua otimalidade), requerendo elevado esforço computacional (ou seja, a eficácia é colocada à frente da eficiência), os métodos aproximados, como as heurísticas e as metaheurísticas, tentam encontrar soluções de boa qualidade para o problema com um esforço computacional mais reduzido investindo sobretudo na eficiência pesquisa. Uma heurística é uma regra ou um conjunto simples de regras específicas para um problema e que, uma vez aplicadas, procuram gerar uma solução admissível ou melhorar uma solução incumbente. Por seu turno, as metaheurísticas são algoritmos genéricos utilizados na resolução de problemas de otimização combinatoria, guiando heurísticas na procura de soluções de maior qualidade. Entre estas heurísticas destacam-se as de pesquisa local, que iterativamente atualizam a solução incumbente, substituindo-a por uma outra melhor encontrada na sua vizinhança, até um ótimo local ser atingido. As diferentes metaheurísticas correspondem a diferentes técnicas para escapar ao aprisionamento em ótimos locais.

Uma vez que o objetivo deste trabalho consistia na criação de um sistema de apoio à decisão para ser implementado na empresa foi decidida a utilização de métodos aproximados para a obtenção dos planos. Em concreto, recorreram-se a metaheurísticas para a geração de planos de produção com elevada eficiência num curto período de tempo. O método de solução proposto utiliza os princípios do Variable Neighborhood Search (VNS) combinados com a formulação matemática desenvolvida, como forma de gerar rapidamente soluções de alta qualidade para o problema. O VNS assenta na mudança sistemática da estrutura de vizinhança utilizada para conduzir a exploração do espaço de soluções do problema (Hansen *et al*, 2008). No método desenvolvido, a pesquisa local é responsável por definir a alocação das famílias de produtos às linhas de enchimento que, no modelo matemático, corresponde às variáveis binárias de *setup*. Enquanto as restantes decisões que compõem a parte contínua da formulação, nomeadamente as de produção, de

inventário, de transferências e de horas extraordinárias, são determinadas com recurso à resolução do subsequente problema de programação linear (PL) obtido pela fixação das variáveis de *setup* no modelo. Uma vez que esta exploração das estruturas de vizinhança é computacionalmente exigente, a informação recolhida durante a resolução do PL é utilizada para, por um lado, guiar a pesquisa local e, por outro, identificar e rejeitar rapidamente soluções de qualidade inferior à atual.

A validação do método ocorreu em duas fases. Em primeiro lugar, a abordagem ao problema foi testada em instâncias aleatórias geradas de acordo com as características reais do problema. Após a validação na primeira fase, o método foi utilizado na geração do orçamento de produção da empresa para os anos de 2010 e 2011. A performance obtida, quer em termos do *output* gerado, quer dos tempos de execução, legitimou do ponto de vista prático o algoritmo, servindo de base para o desenvolvimento de um sistema de apoio à criação do orçamento. O sistema visa uma utilização e parametrização user-friendly do método por parte dos gestores da empresa.

A implementação na empresa do sistema de apoio à decisão (SAD) foi bem-sucedida. O SAD foi utilizado pelos gestores da empresa, em setembro de 2011, de forma independente na elaboração do orçamento de produção para o ano de 2012 e, na simulação de diferentes cenários de transformação do parque industrial da empresa.

O restante do documento está organizado seguindo as etapas decorridas no projeto. A Secção 2 descreve o processo produtivo dos tipos de produtos abordados (cerveja e refrigerantes). Na Secção 3 são apresentadas as principais características e condicionantes do planeamento de produção na indústria das bebidas, sendo também descrito o processo hierárquico de tomada de decisão. A descrição detalhada do problema e o enquadramento do projeto na empresa constituem a Secção 4. A metodologia de solução proposta é exposta na Secção 5 e a sua validação na Secção 6. Detalhes da implementação do sistema de apoio à decisão estão presentes na Secção 7. Por fim, a Secção 8 analisa o impacto do projeto na empresa e suas possíveis extensões.

2 Processo produtivo da indústria cervejeira e de refrigerantes

Os processos produtivos da cerveja e dos refrigerantes partilham algumas características a começar por ambos apresentarem dois estágios principais: a produção da bebida (estágio I) e o enchimento (estágio II).

O tipo de cerveja a fabricar determina a receita a ser seguida introduzindo algumas variações no processo do estágio I. As atividades a realizar podem ser agrupadas em três etapas: fabrico do mosto, fermentação e maturação, e filtração. A primeira etapa produz o mosto, uma mistura que contém os açúcares necessários para o processo de fermentação e que são extraídos da fonte de amido, no decurso desta etapa é adicionado o lúpulo. A etapa seguinte inicia-se com a fermentação que designa a transformação dos açúcares do mosto em etanol por ação de uma levedura. Segue-se a maturação que procede à eliminação de substâncias indesejáveis no plano sensorial. Esta etapa decorre em tanques cilindro-cónicos. Enquanto que a etapa de fabrico do mosto dura entre 6 e 8 horas, a etapa de fermentação e maturação pode arrastar-se, dependendo da cerveja a fabricar, de 4 dias a 3 semanas. A cerveja resultante das etapas anteriores apresenta-se turva, havendo, por isso, necessidade de se proceder a uma filtração que lhe confere o aspecto final desejado. É nesta etapa que se adicionam os concentrados no caso das cervejas com sabor, ou se remove o álcool no caso das cervejas não alcoólicas.

Por seu turno, os refrigerantes apresentam uma composição à base de água carbonatada, açúcar e xaropes de sabor concentrados, o que induz as diferenças no primeiro estágio do processo de fabrico. O estágio I da produção de refrigerantes inicia-se com clarificação da água à qual, seguidamente, são adicionados os concentrados e o açúcar em tanques de mistura próprios. O último processo do estágio I do fabrico de refrigerantes trata a carbonatação da bebida, que usualmente ocorre instantes antes do estágio seguinte.

O estágio II dos dois processos de fabrico ocorre em linhas de enchimento, que podem produzir tanto cerveja como refrigerantes. Estes equipamentos consistem numa série de correias transportadoras que fazem um percurso pré-definido através de máquinas responsáveis por lavar, encher, selar, rotular e embalar garrafas, latas ou barris. Inicialmente, os recipientes são lavados de forma a remover impurezas e a proceder à sua descontaminação, sendo, em

seguida, inspecionados. No passo seguinte os recipientes são cheios e selados, antes de passarem à fase de pasteurização que garante o período de validade dos produtos. Para o caso dos refrigerantes, a pasteurização pode ter lugar ainda no tanque de mistura. Na fase de rotulagem os recipientes são também inspecionados de forma a certificar que contêm o volume correto. A última operação na linha de enchimento é o embalamento dos recipientes em caixas de cartão, grades, packs ou outras unidades de venda e a sua paletização.

Uma vez que a sequência destas operações é independente do produto a produzir, doravante as máquinas e atividades a realizar na linha de enchimento serão tratadas como uma só, sendo todo o processo anteriormente descrito para o estágio II designado simplesmente por enchimento. A Figura 1 esquematiza o processo produtivo para ambos os tipos de produtos.

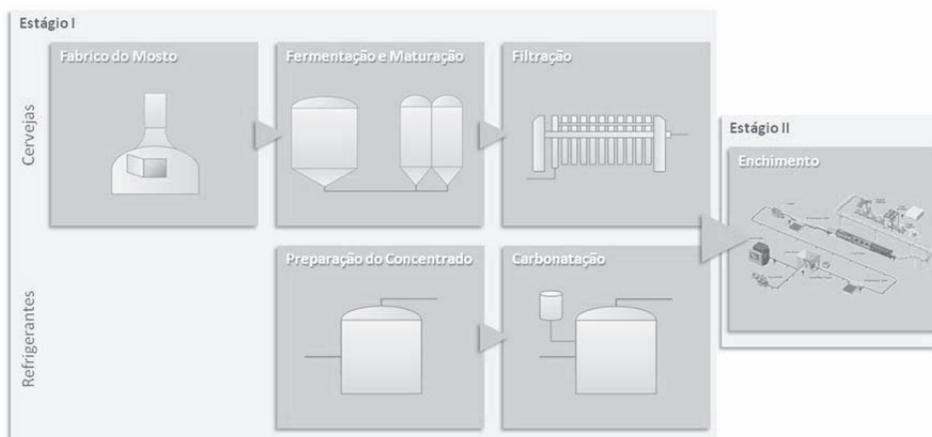


Figura 1: Principais etapas do processo produtivo

3 Planeamento da produção na indústria de bebidas

Ao longo dos últimos anos, o planeamento da produção na indústria de bebidas tem-se tornado progressivamente mais complexo, mas também desafiante para gestores e investigadores. Parte do desafio do planeamento de produção nesta indústria advém da automatização das várias etapas de fabrico do produto final. Tal facto, permite um maior controlo sobre o processo e potencia uma maior eficiência na utilização dos equipamentos produtivos. Por outro lado, a crescente competitividade no mercado forçou o alargamento do

portfólio de produtos das empresas, aumentando a dimensão e complexidade do problema. Este novo paradigma intensificou a urgência de melhores ferramentas de apoio à tomada de decisão. Os fatores condicionantes do planeamento da produção nesta indústria, revistos em seguida, tornam imperiosa a necessidade de um planeamento da produção criterioso, por forma a utilizar os recursos produtivos disponíveis eficientemente e, em último caso, assegurar a competitividade da empresa.

3.1 Perfil de vendas

Um dos principais condicionantes ao planeamento de produção na indústria de bebidas tem que ver com o perfil de vendas destes produtos. As vendas de cerveja e refrigerantes são altamente sazonais e variáveis, devido em parte aos padrões de consumo dos clientes finais. Durante as épocas festivas da Páscoa e do Natal, regista-se um aumento no consumo destes produtos, mas é na época de verão que o pico mais elevado nas vendas é atingido. Além destas variações anuais, as vendas são mais significativas na segunda metade de cada mês. Acresce a este comportamento a variabilidade introduzida por outras fontes como a gestão das marcas, a política comercial da empresa, ou ainda o facto de alguns dos principais clientes desta indústria serem grandes retalhistas com políticas de marketing bastante agressivas que requerem uma resposta quase imediata dos fornecedores.

Por seu turno, a capacidade instalada mantém-se praticamente inalterada durante todo o ano, podendo ser medida pelo número de horas disponíveis para produção. Embora o recurso à utilização de horas extraordinárias para fazer face a alguns destes picos seja uma possibilidade, esta capacidade extra não é suficiente para as alturas de maior consumo. Note-se que o aumento da capacidade instalada não é economicamente viável uma vez que a capacidade requerida no pico de consumo originaria baixas utilizações no resto do ano.

Estas características dos padrões de venda, em simultâneo com os *lead times* do processo de fabrico, condicionam fortemente a produção, e conduzem a indústria a trabalhar numa política de *make-to-stock*. Desta forma, a indústria garante uma resposta rápida e adequada ao mercado, mas também uma eficiente utilização dos seus recursos produtivos.

3.2 Cadeia de abastecimento

Tradicionalmente, as empresas que operam na indústria de bebidas contam com vários centros produtores localizados próximo dos principais pontos geográficos de consumo. O objetivo é controlar os custos de transporte que, de outra forma, representariam uma percentagem significativa do custo total do produto. Na maioria destes centros produtores existem áreas separadas dedicadas aos estágios I e II do processo produtivo. Ambos estão separados por unidades intermédias de armazenamento que possibilitam intervalos de espera entre a produção da bebida e o seu enchimento. Assim, dentro de um centro produtivo, é comum que o estágio I abasteça um conjunto de linhas de enchimento que operam em paralelo. O estágio do enchimento é considerado, na maioria das situações, o gargalo de todo o processo produtivo. No caso da produção de cerveja, a existência de *buffers* entre cada uma das etapas do estágio I torna-o mais flexível. Além disso, o elevado número de produtos finais a produzir no estágio II resume-se a um número substancialmente menor de diferentes tipos de cerveja (ou refrigerantes) a fabricar no estágio I. De facto, a maioria das diferenças entre os produtos finais encontram-se ao nível dos tipos de recipiente, rótulo ou de embalagem afetando apenas o estágio II.

As linhas de enchimento estão agrupadas segundo características tecnológicas, destacando-se o tipo de recipiente que são capazes de encher ou tipo de embalagem. Por exemplo, uma linha dedicada ao enchimento de barris é inadequada para o enchimento de garrafas ou latas. Importa distinguir de entre as linhas de enchimento para garrafas de vidro, as exclusivamente dedicadas a tara perdida e as de tara retornável, uma vez que as últimas necessitam de um processo extra de lavagem. Daí que, embora as linhas de enchimento de garrafas de tara retornável possam ser utilizadas no enchimento de garrafas de tara perdida, o contrário não se verifica. De qualquer forma, o leque de produtos compatível com uma determinada linha de enchimento é bastante alargado. Num mesmo centro produtor, um produto pode ser hipoteticamente alocado a várias linhas. Adicionalmente, os aspectos tecnológicos das linhas de enchimento condicionam também as taxas de enchimento (medida em litros por minuto (l/min) ou unidades de vendas por minuto (u.v./min)), que variam substancialmente de linha para linha.

Cada linha de enchimento pode apenas produzir um produto de cada vez, estando os equipamentos que a constituem configurados para um determinado tipo de líquido, recipiente e embalagem. Uma mudança entre lotes de produção de produtos diferentes despoleta ajustes nas configurações dos equipamentos, bem como possíveis operações de limpeza. Mudanças ao nível do líquido a encher envolvem, quase sempre, a limpeza dos equipamentos e o ajuste da configuração da máquina responsável pela pasteurização. Mudanças no recipiente, quer sejam em termos de tamanho (20 cl para 33 cl, por exemplo) quer de tipo (garrafa tara retornável para garrafa tara perdida), abrangem alterações mecânicas na maioria dos equipamentos. Por fim, as mudanças ao nível da embalagem são, em geral, as mais fáceis de efetuar em virtude de ser o último passo no processo de enchimento. Todas estas operações de ajuste e limpeza consomem tempo disponível para o enchimento, ou seja capacidade, e dependem da sequência de produção, podendo ainda gerar custos motivados pelo desperdício de matéria-prima ou produtos intermédios. Nestas condições, está-se perante *setups* dependentes da sequência: a mudança na configuração das máquinas do produto A para o produto B consome recursos diferentes (tempo e outros) do que a mudança do produto A para o C, por exemplo. Ora, o incremento do número de produtos verificado nos últimos anos reduziu significativamente o tempo operacional das linhas de enchimento, dado que mais *setups* são necessários. Para além deste facto, a constante pressão exercida pelos clientes no sentido de trabalhar com *stocks* mais baixos e entregas mais frequentes obriga à redução do tamanho médio dos lotes de produção, originando novamente mais *setups* e, conseqüentemente, maiores perdas de capacidade.

Como referido anteriormente, a capacidade das linhas de enchimento é medida pelo número de horas disponíveis para a produção. As linhas operam em regime de turnos, sendo que algumas se mantêm em laboração contínua. Desta forma, a utilização de horas extraordinárias nem sempre é uma opção, e os investimentos na capacidade instalada apenas produzem efeitos a longo prazo, permanecendo a capacidade constante no curto prazo.

A presença de mais de um centro produtor na cadeia de abastecimento permite a alocação de alguns produtos unicamente a um dos centros

produtores. Esta decisão leva a um aumento dos rendimentos operacionais das linhas de enchimento, e garante os critérios de qualidade desejados no estágio I. A presença de maiores lotes de produção no primeiro estágio, não só facilita o controlo do processo produtivo, como também reduz o seu custo. No entanto, para os produtos com maior rotação, a sua produção perto do local de consumo é sempre preferível devido aos já referidos custos de transporte.

3.3 Sistemas de planeamento de produção

O planeamento da produção na indústria de bebidas é um processo hierárquico, em que as decisões são tomadas em diferentes níveis da empresa, consoante a natureza das variáveis em jogo. Apesar da forte interligação entre as decisões a tomar, a criação de um único modelo para a tomada de decisão é impraticável. Não só porque tal modelo seria virtualmente impossível de manter, resolver e interpretar, como também os planos de produção nos diferentes níveis visam necessidades distintas. Aos diferentes níveis de decisão correspondem diferentes objetivos e horizontes de planeamento. O sistema hierárquico de planeamento da produção nesta indústria é constituído por três níveis de decisão (ver Figura 2): planeamento de longo prazo (com um carácter tático/estratégico), planeamento de médio prazo (tático/operacional) e planeamento de curto prazo (apenas operacional). O planeamento de longo prazo avalia a adequação da capacidade instalada face a uma previsão de vendas para um horizonte típico de 12 a 18 meses. Entre as decisões a estudar estão investimentos ou reduções da capacidade e a constituição de *stock* sazonal. Com um horizonte de entre 4 a 12 semanas, os planos de cariz tático são dirigidos à criação de planos para as operações, em especial produção e distribuição, tendo por objetivo a satisfação das previsões de vendas da forma mais económica. Por fim, o nível operacional destina-se à gestão detalhada das operações nos recursos produtivos para um horizonte de um dia a uma semana. Independentemente do nível da hierarquia, as decisões funcionam num esquema de horizonte temporal rolante. Na realidade, apenas o plano de alguns períodos iniciais do horizonte temporal é aplicado na prática, uma vez que o planeamento é realizado periodicamente e tem em conta o plano gerado no nível superior.

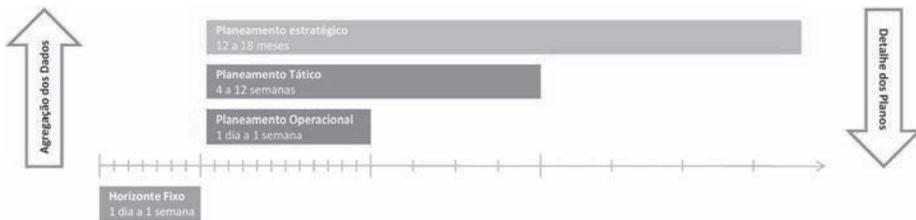


Figura 2: Horizontes de planeamento dos diferentes níveis de planeamento da produção

4 O caso de estudo

Este projeto é fruto da colaboração com uma empresa Portuguesa que compete no setor de bebidas. Com vendas distribuídas pelos vários continentes, a empresa é detentora de várias marcas de cerveja, refrigerante e águas lisas e com gás reconhecidas a nível nacional. A empresa possui 8 centros produtores com 20 linhas de enchimento, distribuídos geograficamente por todo o país. A produção de cerveja e de refrigerantes partilha algumas das linhas de enchimento, tendo o seu planeamento de ser executado em simultâneo. O objetivo deste projeto é o desenvolvimento de um método de solução do orçamento anual de produção (OrcP), que corresponde ao nível superior do sistema de planeamento da produção em vigor na empresa.

O OrcP faz parte do processo de orçamentação anual da empresa que se inicia no princípio de setembro e se prolonga até ao final de outubro, sendo uma ferramenta vital para transformar em objetivos a estratégia definida pela administração para os 12 meses seguintes. A primeira grande etapa deste processo é a criação do orçamento anual de vendas (OrcV). Da responsabilidade do departamento comercial, o OrcV é uma previsão de vendas mensal para todos os produtos que tenta capturar as principais tendências do mercado para o ano seguinte. Em simultâneo com a criação do OrcV, o departamento de produção elabora para cada linha de enchimento um calendário dos períodos de manutenção e estima a taxa de enchimento de cada produto. Estas estimativas são baseadas nas observações registadas nos anos anteriores e têm em consideração os ganhos ou perdas de eficiência

expectáveis. Como o objetivo do OrcP não passa por um plano detalhado para as linhas de enchimento, mas antes pela estimativa dos custos a incorrer no ano seguinte e pela validação da adequação da capacidade disponível para fazer face às previsões de vendas, o sequenciamento da produção é desconsiderado neste nível. A perda de capacidade resultante dos tempos de mudança é incorporada nas estimativas das taxas de enchimento, tendo novamente por base o histórico das observações.

Para além do OrcV, do calendário de manutenção e das estimativas das taxas de enchimento, o departamento de planeamento necessita da capacidade mensal disponível para conduzir a elaboração do OrcP. A capacidade estimada para cada linha de enchimento tem em consideração o calendário de manutenção definido e o número de dias de trabalho disponível em cada mês, estando dividida em três categorias: (1) normal, (2) sábados e, (3) domingos e feriados. Esta categorização distingue a laboração normal da laboração em regime de horas extraordinárias, sendo que a última decorre da utilização dos fins-de-semana e feriados.

A estimativa de vendas presente no OrcV é global, não estando discretizada por centro produtor. Para alocar as previsões aos centros produtores é utilizada a informação das vendas do ano anterior. Excepcionalmente, toda a procura pode ser alocada a apenas um centro produtor sempre que:

- os equipamentos adequados para a produção dos produtos em causa estejam apenas presentes num dos centros produtores, ou seja, por razões tecnológicas;
- para determinadas famílias de produtos, a previsão da procura desagregada pelos centros produtores é insuficiente para o cumprimento dos critérios de qualidade devido ao incumprimento de lotes mínimos e/ou frequência de produção;
- decisões estratégicas ditam a especialização dos centros produtores em determinadas famílias de produtos.

No OrcP apenas o estágio II (enchimento) é considerado, pois corresponde ao gargalo de todo o processo produtivo. O OrcP consiste num plano de produção de longo prazo que, em cada período do horizonte de planeamento, aloca os produtos às diferentes linhas de enchimento e determina os respetivos

lotes de produção. Através deste passo estão também implicitamente definidas as decisões de utilização de horas extraordinárias, inventário e transferências entre centros produtores. O objetivo do plano é a satisfação da previsão de vendas presente no OrcV da forma economicamente mais eficiente, isto é, minimizando os custos de produção, posse de inventário, alocação dos produtos às linhas, transferências entre centros produtores e utilização de horas extraordinárias.

Algoritmo 1: Método da empresa para a geração do OrcP

```
1: for all centros produtores do  
2:   for all necessidades identificadas do  
3:     atribuir necessidade à linha preferencial  
4:   end for  
5: end for  
6: for all linhas de enchimento do  
7:   for all períodos do  
8:     if produção > capacidade disponível then  
9:       realocar necessidades a outra linha no mesmo centro  
10:    end if  
11:    if produção > capacidade disponível then  
12:      realocar necessidades a períodos anteriores  
13:    end if  
14:    if produção > capacidade disponível then  
15:      realocar necessidades a uma linha noutra centro  
16:    end if  
17:  end for  
18: end for
```

Atualmente, a criação do OrcP destina-se ao planeamento de mais de 150 produtos, agrupados em 60 famílias, em 14 diferentes linhas de enchimento. Muito embora as restrições tecnológicas reduzam o tamanho do problema, a complexidade do planeamento exige tempo considerável por parte da equipa responsável. Por forma a simplificar o ato de planear, uma das estratégias seguidas na empresa é a definição de uma linha de enchimento preferencial. Esta regra determina automaticamente, para os produtos com mais de uma linha enchimento alternativa, a linha a alocar as necessidades identificadas. O Algoritmo 1 sistematiza o processo de geração do OrcP. Sempre que numa linha em determinado período ocorra um excesso de produção alocada, a

admissibilidade do plano é recuperada realocando os lotes de produção: a outras linhas de enchimento dentro do mesmo centro, a períodos anteriores na mesma linha, ou a linhas de outros centros produtores. Para linhas de enchimento que não operam em regime de laboração contínua, pode ser ainda considerada a utilização de horas extraordinárias. O OrcP é realizado uma vez por ano e os principais indicadores de performance são: utilização média das linhas de enchimento e quantidades totais de inventário, transferências e horas extraordinárias.

5 Metodologia de solução

Nesta secção é apresentada a metodologia seguida para a resolução do problema que surge na criação do OrcP. A abordagem seguida pode ser dividida em duas partes: a formulação do modelo matemático de base ao OrcP e o desenvolvimento do método de solução. O recurso a métodos exatos incorporados em *software* comercial foi colocado de parte pela necessidade de obtenção de boas soluções para o OrcP num limitado tempo de execução, e ainda pelo facto de a aquisição de licenças de software representar um investimento inadequado face às necessidades identificadas. Daí que se tenha desenvolvido um método de solução que, no entanto, se baseia no modelo matemático como se verá em seguida.

5.1 Modelo matemático

O modelo de programação inteira mista que aqui se apresenta, foi desenvolvido com o propósito de modelar o orçamento de produção na indústria de bebidas. Considere-se a criação de um plano de produção alocando produtos agregados em famílias a linhas de enchimento pertencentes a centros produtores para um horizonte temporal de períodos. Cada centro produtor detém uma procura e uma capacidade de armazenagem próprias. As transferências de produtos acabados são incluídas no modelo, possibilitando que um determinado centro produtor abasteça a procura alocada a outro centro incorrendo-se num custo extra de transporte. Uma vez que o enchimento é considerado o gargalo do processo produtivo, o planeamento é apenas realizado para este estágio, todavia a partilha de recursos produtivos

obriga a que as decisões sejam tomadas simultaneamente para cerveja e refrigerantes. Os produtos são agrupados em famílias sempre que possuam características semelhantes, nomeadamente os tipos de recipiente e de embalagem. É, também, equacionada a utilização de horas extraordinárias. O trabalho ao fim-de-semana e feriados é fundamental para fazer face aos picos de procura existentes ao longo do horizonte de planeamento, permitindo uma maior flexibilidade. Distingue-se o recurso a sábados e feriados (tipo I) do recurso a domingos (tipo II), devido à diferença dos custos com pessoal, mais elevados no tipo II do que no tipo I. A nomenclatura que se segue é utilizada para formular o problema:

Índices

i	produto: $i \in \mathbf{N} = \{1, \dots, N\}$
f	família: $f \in \mathbf{F} = \{1, \dots, F\}$
j, k	centro produtor: $j, k \in \mathbf{P} = \{1, \dots, P\}$
m	linha de enchimento: $m \in \mathbf{M} = \{1, \dots, M\}$
t	período: $t \in \mathbf{T} = \{1, \dots, T\}$

Conjuntos

M_j	conjunto das linhas de enchimento pertencentes ao centro produtor j
F_m	conjunto das famílias que podem ser produzidas na linha de enchimento m
N_m^{line}	conjunto dos produtos que podem ser produzidos na linha de enchimento m
N_f^{fam}	conjunto dos produtos pertencentes à família f

Parâmetros

cap_{mt}	capacidade disponível na linha de enchimento m no período t
j_m	centro produtor a que pertence a linha de enchimento m
f_i	família a que pertence o produto i
d_{ijt}	procura do produto i no centro produtor j no final do período t
h_{ijt}	custo de manter em stock uma unidade do produto i no centro produtor j no final do período t
r_{ijkt}	custo de transportar uma unidade do produto i desde o centro produtor j para o centro produtor k no período t

v_{imt}	custo de produção de uma unidade do produto i na linha de enchimento m no período t
c_{fmt}	custo de preparação (setup) da linha de enchimento m para produzir a família de produtos f no período t
P_{fmt}	taxa de enchimento da família f na linha de enchimento m no período t
b_{imt}	limite na quantidade a produzir do produto i na linha de enchimento m no período t
$co_{mt}^I (co_{mt}^{II})$	custo unitário de horas extraordinárias do tipo I (tipo II) na linha de enchimento m no período t
$mo_{mt}^I (mo_{mt}^{II})$	limite máximo de horas extraordinárias do tipo I (tipo II) na linha de enchimento m no período t

As seguintes variáveis de decisão capturam as decisões a serem tomadas:

X_{imt}	quantidade a produzir do produto i na linha de enchimento m no período t
I_{ijt}	quantidade em inventário do produto i no centro produtor j no final do período t
W_{ijkt}	quantidade a transferir do produto i do centro produtor j para o centro produtor k no período t
Y_{fmt}	(=1) se a linha de enchimento m é preparada para a família f no período t , (=0) caso contrário.
$O_{mt}^I (O_{mt}^{II})$	consumo de capacidade extra do tipo I (tipo II) na linha de enchimento m no período t

A formulação matemática do problema é apresentada em seguida:

$$\min \sum_{i \in \mathbf{N}} \sum_{j \in \mathbf{P}} \sum_{t \in \mathbf{T}} \left(h_{ijt} \cdot I_{ijt} + \sum_{k \in \mathbf{P}} r_{ijkt} \cdot W_{ijkt} \right) + \quad (1)$$

$$\sum_{m \in \mathbf{M}} \sum_{t \in \mathbf{T}} \left(\sum_{i \in \mathbf{N}} v_{imt} \cdot X_{imt} + \sum_{f \in \mathbf{F}_m} c_{fmt} \cdot Y_{fmt} + co_{mt}^I \cdot O_{mt}^I + co_{mt}^{II} \cdot O_{mt}^{II} \right) \\ I_{ij,t-1} + \sum_{m \in \mathbf{M}_j} X_{imt} + \sum_{k \in \mathbf{P} \setminus \{j\}} W_{ikjt} \\ = I_{ijt} + d_{ijt} + \sum_{k \in \mathbf{P} \setminus \{j\}} W_{ijkt}, \quad \forall i \in \mathbf{N}, j \in \mathbf{P}, t \in \mathbf{T} \quad (2)$$

$$\sum_{i \in \mathbf{N}_m^{line}} \frac{X_{imt}}{P_{f,mt}} \leq cap_{mt} + O_{mt}^I + O_{mt}^{II}, \quad \forall m \in \mathbf{M}, t \in \mathbf{T} \quad (3)$$

$$X_{imt} - b_{imt} \cdot Y_{f,mt} \leq 0, \quad \forall i \in \mathbf{N}_m^{line}, m \in \mathbf{M}, t \in \mathbf{T} \quad (4)$$

$$O_{mt}^I \leq mo_{mt}^I, O_{mt}^{II} \leq mo_{mt}^{II}, \quad \forall m \in \mathbf{M}, t \in \mathbf{T}. \quad (5)$$

$$(I_{ijt}, W_{ijkt}, X_{imt}, O_{mt}^I, O_{mt}^{II}) \geq 0, Y_{f,mt} \in \{0,1\} \quad (6)$$

A função objetivo (1) minimiza a soma dos custos de produção, posse de inventário, transferência, preparação das linhas de enchimento e utilização de horas extraordinárias. As restrições (2) asseguram a lógica do fluxo dos produtos em cada centro produtor. A procura do produto i no centro produtor j no período t só pode ser satisfeita através do inventário existente no início do período, da produção realizada nas linhas do centro ou das transferências de outros centros produtores. Não são permitidas encomendas em atraso ou perda de vendas, e as transferências entre os centros produtores ocorrem no mesmo período. As restrições (3) garantem que a produção em cada período não excede a capacidade normal disponível mais as horas extraordinárias. A produção de um determinado produto só pode ocorrer caso a linha de enchimento tenha sido preparada para a família à qual o produto pertence, como assegurado em (4), estando a quantidade máxima a produzir no período limitada a $b_{imt} = \min\{cap_{mt} \cdot p_{f,mt}, \sum_{j \in \mathbf{P}} \sum_{u=t}^T d_{iju}\}$, isto é o mínimo entre a quantidade que seria possível produzir se a linha de enchimento produzisse apenas o produto i e a procura total do produto ainda por satisfazer, ou seja, em todos nos centros neste e nos períodos seguintes. Em (5) são impostos os limites máximos de horas extraordinárias.

Exemplo 1: O seguinte exemplo ilustrativo considera o planeamento de 3 produtos durante um horizonte de 3 períodos em 2 centros produtores A e B . O centro A conta com apenas uma linha de enchimento - A_1 , enquanto que o centro B dispõe de duas linhas de enchimento - B_1 e B_2 . As capacidades das linhas são constantes ao longo do horizonte de planeamento e iguais a 250, 200 e 150, respectivamente para A_1 , B_1 e B_2 . O limite de utilização de horas extraordinárias para os tipos I e II é de 20, sendo que o custo do tipo I é de 5 e do tipo II de 7.5. O custo de posse de inventário h_{ikt} é também mantido constante e igual a 0.5 para todos os produtos em todos os centros. Os custos

variáveis de produção são desconsiderados. Os 3 produtos estão agrupados em 2 famílias, sendo a primeira família composta pelos produtos $i=1$ e $i=2$ e a segunda apenas pelo produto $i=3$. Os tempos de processamento e custos de *setup* das famílias não variam com o período nem com a linha de enchimento e são iguais a 1 e 100 respectivamente. Na Tabela 1 são apresentados os restantes dados em falta. A solução ótima é apresentada nas Tabelas 2 e 3 e foi obtida utilizando um solver comercial para resolver a formulação matemática do problema. É claramente notório que o planeamento em simultâneo dos dois centros produtores permite obter poupanças, em especial ao nível dos gastos em preparação das linhas e de horas extraordinárias. Este efeito é visível com a utilização da produção do centro *B* para abastecer a procura no centro *A* no período $t=1$ e o contrário no período $t=3$.

Tabela 1: Dados para o exemplo ilustrativo

	Centro produtor <i>A</i>				Centro produtor <i>B</i>			
	Procura (d_{ijt})			C. transf. para B (r_{ijkt})	Procura (d_{ijt})			C. transf. para A (r_{ijkt})
	$t=1$	$t=2$	$t=3$		$t=1$	$t=2$	$t=3$	
$i=1$	200	130	140	0.5	94	120	270	0.5
$i=2$	120	40	70	0.7	80	80	144	0.7
$i=3$	50	110	50	1	0	120	50	1

O modelo acima descrito pode ser visto como uma extensão do *single stage, multi-plant, multi-item and multi-period capacitated lot sizing problem* (MPCLSP) descrito em Sambasivan & Schimidt (2002), Sambasivan & Yahya, (2005) e Nascimento *et al* (2010). Note-se que o MPCLP corresponde a uma generalização do *capacitated lot sizing problem* (CLSP) para vários centros produtores, introduzindo o conceito de transferências entre centros produtores. Apesar de semelhante, o modelo descrito aqui difere do MPCLSP em alguns aspectos. Uma das principais diferenças é a possibilidade de existência de várias linhas de enchimento em cada centro produtor, ao contrário do MPCLSP que assume apenas uma linha. O motivo para esta extensão deve-se ao facto de, por razões de cariz tecnológico, as linhas de enchimento de um determinado centro produtor não poderem ser agregadas. Outra extensão é a utilização de horas extraordinárias a um custo adicional, não considerada anteriormente na literatura dedicada a este problema. A juntar a isto, o modelo

contabiliza somente os custos de *setup*, não considerando tempos de preparação das linhas de enchimento (contrariamente a Sambasivan & Yahya, 2005 e Nascimento *et al*, 2010), uma vez que a empresa opta por incluir na estimativa das taxas de enchimento o tempo médio dispendido nas trocas entre lotes de produção.

Tabela 2: Solução para o exemplo ilustrativo

	<i>i</i> = 1			<i>i</i> = 2			<i>i</i> = 3		
	<i>t</i> = 1	<i>t</i> = 2	<i>t</i> = 3	<i>t</i> = 1	<i>t</i> = 2	<i>t</i> = 3	<i>t</i> = 1	<i>t</i> = 2	<i>t</i> = 3
Centro produtor A									
<i>Procura</i> (d_{ijt})	200	130	140	120	40	70	50	110	50
<i>Inventário Inicial</i> ($I_{ij,t-1}$)	0	0	0	0	0	0	0	80	50
<i>Produção na linha de enchimento A₁</i> (X_{imt})	154	130	180	120	40	70	0	80	0
<i>Saídas para o centro B</i> (W_{ijkt})	0	0	40	0	0	0	0	0	0
<i>Entradas com origem no centro B</i> (W_{ikjt})	46	0	0	0	0	0	130	0	0
<i>Inventário Final</i> (I_{ijt})	0	0	0	0	0	0	80	50	0
Centro produtor B									
<i>Procura</i> (d_{ijt})	94	120	270	80	80	144	0	120	50
<i>Inventário Inicial</i> ($I_{ij,t-1}$)	0	0	0	0	0	0	0	20	0
<i>Produção na linha de enchimento B₁</i> (X_{imt})	140	120	60	80	80	144	0	0	0
<i>Produção na linha de enchimento B₂</i> (X_{imt})	0	0	170	0	0	0	150	150	0
<i>Saídas para o centro A</i> (W_{ijkt})	46	0	0	0	0	0	130	0	0
<i>Entradas com origem no centro A</i> (W_{ikjt})	0	0	40	0	0	0	0	0	0
<i>Inventário Final</i> (I_{ijt})	0	0	0	0	0	0	20	50	0

Por fim, os custos de *setup* são incluídos tendo por base a definição de famílias de produtos e não individualmente por produto como originalmente no MPCLSP.

A opção de inclusão dos custos de preparação prende-se com o objetivo de minimizar o número de linhas a produzir a mesma família, uma estimativa de futuras perdas de capacidade motivadas pelos tempos de *setup* dependentes da sequência de produção. Ao minimizar o número de famílias a produzir em cada linha num nível estratégico estamos, potencialmente, a minimizar futuras perdas de capacidade, uma vez que os produtos pertencentes à mesma família originam menores tempos de *setup* nas mudanças entre si do que os tempos verificados nas mudanças entre produtos de famílias diferentes.

Tabela 3: Análise da utilização das linhas de enchimento na solução ótima do exemplo ilustrativo

	Linha de enchimento A_1			Linha de enchimento B_1			Linha de enchimento B_2		
	$t=1$	$t=2$	$t=3$	$t=1$	$t=2$	$t=3$	$t=1$	$t=2$	$t=3$
<i>Tempo total de produção</i> ($\sum_{i \in N_m^{line}} \frac{X_{imt}}{p_{fimt}}$)	274	250	250	220	200	204	150	150	170
<i>Capacidade (cap_{mt})</i>	250	250	250	200	200	200	150	150	150
<i>Horas extraordinárias Tipo I</i> (O_{mt}^I)	20	0	0	20	0	4	0	0	20
<i>Horas extraordinárias Tipo II</i> (O_{mt}^{II})	4	0	0	0	0	0	0	0	0
<i>Utilização da capacidade</i>	100%	100%	100%	100%	100%	100%	100%	100%	100%
<i>Capacidade máxima com horas extraordinárias</i>	290	290	290	240	240	240	190	190	190
<i>Utilização da capacidade máxima</i>	94%	86%	86%	92%	83%	85%	79%	79%	89%

5.2 Método de solução

A dificuldade inerente à resolução do problema acima descrito (NP-hard), bem como a dimensão das instâncias presentes no caso real motivaram o desenvolvimento de um eficiente método de solução. Para estas instâncias não conseguimos obter soluções ótimas em tempo útil. A heurística proposta é

inspirada em elementos fundamentais do VNS e da programação linear. O VNS é uma metaheurística que usa a exploração sistemática de várias estruturas de vizinhança simultaneamente, como o intuito de encontrar e escapar de ótimos locais. Uma solução diz-se vizinha de outra se puder ser obtida através de uma (normalmente pequena) alteração da inicial. A estas pequenas alterações dá-se o nome de movimentos. O conjunto de todas as soluções vizinhas de uma incumbente x obtidas aplicando um determinado tipo de movimento constitui a estrutura de vizinhança $N(x)$.

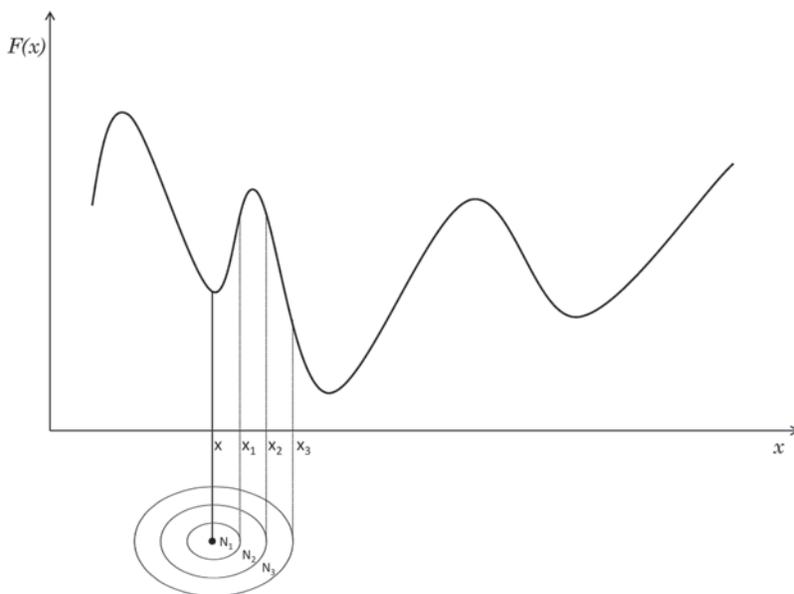


Figura 3: Representação esquemática do VNS

O VNS assenta na exploração destas definições tendo por base três princípios: (1) um ótimo local para uma certa estrutura de vizinhança pode não ser um ótimo local de uma outra estrutura de vizinhança, (2) um ótimo global é ótimo local para todas as estruturas de vizinhança e, finalmente, (3) para muitos problemas, os ótimos locais encontram-se relativamente perto do ótimo global. Começando com uma solução inicial x , o VNS baseia-se na utilização de heurísticas de pesquisa local para encontrar melhores soluções recorrendo a um conjunto ordenado de estruturas de vizinhança $N_k (k = 1, \dots, k_{max})$, sendo

que a solução inicial para o problema pode ser obtida através de uma qualquer heurística de construção. Na Figura 3 estão esquematizadas a solução incumbente x , um ótimo local, e três soluções vizinhas obtidas aplicando diferentes estruturas de vizinhança. As três etapas descritas em seguida sistematizam o esquema básico do VNS (ver Mladenovic & Hansen, 1997) e são repetidas até que o critério de paragem definido seja atingido.

<i>Perturbação</i>	Representa a componente estocástica do VNS e consiste na seleção aleatória de uma solução x' na estrutura de vizinhança $N_k(x)$ com o intuito de escapar a ótimos locais.
<i>Pesquisa Local</i>	Aplicação de uma heurística de melhoria à solução x' , considerando a estrutura de vizinhança N_k até que um ótimo local x'' seja atingindo.
<i>Movimentos e Estruturas de Vizinhança</i>	Se o ótimo local x'' encontrado durante a pesquisa local for melhor que a solução incumbente x , então x'' é aceite e substituí x como solução incumbente, regressando-se à estrutura de vizinhança inicial $k=1$. Caso contrário, o algoritmo segue para a exploração da próxima estrutura de vizinhança $k=k+1$ (se $k > k_{max}$, então $k=1$).

Aplicando estas regras ao exemplo na Figura 3 a próxima solução incumbente seria x_3 , uma vez que a exploração das estruturas de vizinhança N_1 e N_2 não permitiram encontrar soluções melhores do que a atual.

O algoritmo desenvolvido e seguidamente descrito resulta da adaptação de uma das variantes do VNS mais aconselhada para problemas de larga escala, ou sempre que a pesquisa local se mostrar computacionalmente dispendiosa. Nestes casos para não comprometer a eficiência o Reduced VNS (RVNS) pode representar uma alternativa interessante, uma vez que a pesquisa local é eliminada. O RVNS é um método completamente estocástico de geração de soluções vizinhas na estrutura de vizinhança atual (N_k), sendo a solução atualizada em caso de melhoria. Naturalmente que a eficácia obtida no RVNS é inferior à do VNS, mas os ganhos de eficiência são bastante consideráveis na maioria dos casos.

5.2.1 Representação da solução

O método proposto baseia-se numa representação incompleta da solução, contemplando apenas as variáveis de *setup* das famílias nas linhas de enchimento Y_{fmt} . Esta representação identifica implicitamente uma única solução do problema, uma vez que introduzindo estas variáveis no modelo matemático apresentado, as restantes decisões podem ser obtidas resolvendo o subsequente problema de PL. Ou seja, as restantes decisões são exclusivamente dependentes do padrão de *setup* definido (Y'_{fmt}). Isto permite que as estruturas de vizinhança sejam definidas por alterações em Y'_{fmt} .

5.2.2 Solução inicial

Obter uma solução inicial para o problema em causa é difícil, pois passa pela criação de um padrão de *setup* que obedeça simultaneamente às restrições de capacidade das linhas de enchimento e procura dos produtos. O problema da inadmissibilidade de um qualquer padrão de *setup* pode ser resolvido introduzindo no modelo matemático as variáveis de decisão I_{ij}^0 , que simbolizam o *stock* inicial no centro produtor j do produto i , e penalizando na função objetivo o seu uso através de um custo bastante elevado h_{ij}^0 . Deste modo, qualquer padrão de *setup* é desta forma uma solução admissível para o problema, aumentando a flexibilidade da pesquisa conduzida pelo VNS. O termo $\sum_{i \in N} \sum_{j \in P} h_{ij}^0 \cdot I_{ij}^0$, introduzido na função objetivo pode ser visto como uma medida da inadmissibilidade da solução. Sempre que se verifique em instâncias reais o recurso a *stock* inicial, o custo extra é apenas incorrido para quantidades superior à disponível.

Observe-se que qualquer padrão de *setup* determina uma solução inicial válida, desde um padrão sem qualquer *setup* até um padrão onde todos os *setups* possíveis são realizados. A solução inicial implementada é inspirada na política de produção lote por lote. Este procedimento analisa período a período a procura total de cada família em cada centro produtor: $R_{jft} = \sum_{i \in N_f^{fam}} d_{ijt}$. Caso $R_{jft} > 0$, um *setup* é alocado na linha de enchimento pertencente ao centro produtor ($m \in M_j$) cuja taxa de enchimento (p_{fmt}) seja

mais elevada. As restantes decisões são obtidas resolvendo o subseqüente problema de PL que resulta de fixar o padrão de *setup* no modelo matemático.

5.2.3 Estruturas de vizinhança

Os vizinhos de uma solução incumbente x são encontrados introduzindo pequenas alterações no padrão de *setup* no modelo e re-otimizando o subseqüente PL. Para o problema em estudo foram definidos três tipos de movimento que exploram diferentes ideias para a melhoria da solução incumbente:

- 1) **inserir**(f, m, t) consiste na mudança do estado do *setup* da família f na linha de enchimento m no período t de 0 para 1, assim a estrutura de vizinhança $N_I(x)$ inclui todas as possíveis mudanças em Y'_{fmt} de 0 para 1. O objetivo é encontrar um novo padrão de *setup* que, apesar de um incremento nos custos de preparação das linhas de enchimento, resulte em ganhos em termos de custos de produção, *stock*, transferência ou horas extraordinárias.
- 2) **remover**(f, m, t) é o inverso do movimento inserir, sendo que a estrutura de vizinhança $N_R(x)$ contempla todas as possíveis mudanças em Y'_{fmt} de 1 to 0. Também o seu objetivo é o inverso do movimento inserir, procurando com um aumento nos outros custos reduzir os custos de preparação das linhas de enchimento.
- 3) **transferir**(f, m_o, t_o, m_d, t_d) realoca a um lote de produção da família f atualmente alocado à linha de enchimento m_o no período t_o para uma nova posição na linha de enchimento m_d no período t_d . Deste modo a estrutura de vizinhança $N_T(x)$ é composta por todos os movimentos em que $Y'_{f m_o t_o} = 1$ e $Y'_{f m_d t_d} = 0$. Sendo o movimento de interpretação mais complexa, a sua análise pode dividir-se em:
 - a) *transferências dentro da própria linha* - o objetivo passa por antecipar ou atrasar lotes de produção com o objetivo de eliminar ou introduzir inventário. A constituição de inventário reduz, possivelmente, os gastos com horas extraordinárias e/ou transferências.
 - b) *transferências para outras linhas* - no caso de transferências dentro de apenas um período, o intuito é ponderar os custos de produção,

preparação das linhas, transferência e horas extraordinárias. Sendo que a transferência para outras linhas em diferentes períodos acrescenta os custos de inventário à equação, uma vez que neste caso pode dar-se o caso de criação ou eliminação de *stock*. Independentemente do tipo de transferência, estes movimentos são também particularmente úteis no equilíbrio do consumo de capacidade em casos de elevada utilização.

Para qualquer dos movimentos, aquando da avaliação de uma solução vizinha, o PL é resolvido para todo o horizonte de planeamento, ou seja todas as quantidades de produção, inventário, transferência e horas extraordinárias são re-otimizadas. Como consequência, alterações num período podem ter repercussões ao longo de todo o horizonte, sendo as descrições fornecidas em cima uma simplificação do seu real impacto. Aliás, este facto motivou a escolha de uma representação incompleta da solução, pois permite para um determinado padrão de *setup* obter as restantes decisões otimizadas. Caso estas fossem definidas heurísticamente, poder-se-iam erradamente rejeitar padrões de *setup* interessantes.

5.2.4 Desenho do algoritmo

Em teoria, maiores estruturas de vizinhança conduzem a ótimos locais de melhor qualidade resultando num método de solução mais eficaz. No entanto, quanto maior a vizinhança a explorar maior também o tempo necessário para a sua exploração. O esquema original do VNS requer a exploração exaustiva das estruturas de vizinhanças durante a fase de pesquisa local, o que para instâncias de grande dimensão pode conduzir a tempos computacionais impraticáveis. No caso prático em estudo, as decisões a tomar têm um cariz tático/estratégico podendo em teoria ser tomadas num período relativamente alargado de tempo. Porém, isto é apenas válido se o objetivo for a criação de apenas um plano, o que não se verifica na prática. Frequentemente, estas decisões resultam da análise de vários cenários alternativos variando os dados de entrada. Para cada um dos cenários testados tem de ser gerada uma solução, o que reduz significativamente o tempo disponível. Para que o método de solução seja capaz de responder às exigências da sua inerente aplicação prática

é necessário um importante equilíbrio entre a sua eficácia e eficiência. De seguida, são apresentadas algumas componentes do método desenvolvido que permitem um aumento de eficiência substancial sobre o esquema original do VNS.

A avaliação das soluções vizinhas no método desenvolvido envolve a resolução de um PL, representando um esforço computacional considerável, uma vez que são necessárias múltiplas repetições do processo. Tendo em consideração que soluções vizinhas apresentam pequenas alterações no padrão de *setup* face à solução incumbente, a informação da resolução do PL da incumbente pode ser preciosa na melhoria da eficiência da pesquisa, acelerando o processo. Desde logo, a solução incumbente pode constituir a base inicial para o novo PL evitando que este seja resolvido desde início e poupando algum do tempo necessário para obter a solução ótima. As soluções vizinhas também são descartadas durante o seu processo de re-otimização caso o seu custo ultrapassar o custo da melhor solução conhecida até ao momento.

Não obstante as regras definidas durante a avaliação das soluções vizinhas possibilitem uma poupança significativa do tempo dispendido no processo de avaliação, para grandes instâncias do problema o tempo necessário para uma avaliação exaustiva de toda a vizinhança seria proibitivo. Tradicionalmente, em casos onde a pesquisa local é demasiado dispendiosa em termos computacionais, é comum optar-se pela utilização do RVNS como forma de aumentar a eficiência na exploração do espaço de solução. Contudo, a aleatoriedade na sugestão das soluções vizinhas potencia a geração frequente de soluções de qualidade inferior, comprometendo a eficácia da pesquisa e limitando os ganhos em eficiência. Com efeito, foram criadas novas regras para a seleção das soluções vizinhas a explorar de forma a melhorar o esquema básico do RVNS. Tal como sugerido em Hung *et al* (2003), a ideia é aproveitar a informação disponível no final da resolução do PL para conduzir a pesquisa local. A cada solução vizinha é associada uma probabilidade de seleção de acordo com a melhoria potencial da solução incumbente. Para tal os preços sombra das restrições (3) e (4) e a capacidade por utilizar na linha de enchimento m no período t na solução incumbente, são utilizados para avaliar a melhoria potencial de um vizinho. O princípio por detrás das regras

definidas assenta no facto dos preços sombra poderem ser vistos como a utilidade marginal dos recursos. Neste contexto, um movimento do tipo $\text{inserir}(f, m, t)$ estima a utilidade de um *setup* adicional da família f na linha de enchimento m o período t , assumindo que este é válido para o máximo entre a capacidade por utilizar e a procura da família. No caso de um movimento do tipo $\text{remover}(f, m, t)$ é estimada a utilidade associada à “libertação” de capacidade na linha de enchimento no período em análise. A avaliação de um movimento do tipo $\text{transferir}(f, m_o, t_o, m_d, t_d)$ é a mais difícil, uma vez que este movimento introduz várias alterações na solução, sendo por isso menos fiável a informação retirada da análise de sensibilidade do PL. No entanto, este pode ser visto como a combinação de um movimento do tipo $\text{inserir}(f, m_d, t_d)$ e outro do tipo $\text{remover}(f, m_o, t_o)$.

Tendo em conta que as melhorias potenciais calculadas constituem apenas estimativas da melhoria real verificada na função objetivo, estas não devem limitar em demasiado a pesquisa das soluções vizinhas. Deste modo, após calculada a melhoria potencial para todas as soluções vizinhas da estrutura de vizinhança em análise ($r \in N_k(x)$), estas são ordenadas por ordem decrescente de melhoria. Seja $\sigma(r)$ a classificação da solução vizinha r . A probabilidade $\mu(r)$ de uma determinada solução r ser escolhida como próxima solução a ser explorada é dada por:

$$\mu(r) = \frac{\text{bias}(r)}{\sum_{r' \in N_k(x)} \text{bias}(r')},$$

sendo $\text{bias}(r)$ a função identidade. Dado que o objetivo de avaliar e ordenar as soluções vizinhas é dar prioridade às soluções melhor colocadas a função identidade é dada por $\text{bias}(r) = 1/\sigma(r)$. Para selecionar a solução vizinha a explorar, um número aleatório é gerado recorrendo a uma distribuição uniforme e, em seguida, comparado com as probabilidades associadas a cada uma das soluções.

6 Validação do método desenvolvido

O método desenvolvido foi validado sequencialmente, primeiro de um ponto de vista científico e, posteriormente, de um ponto de vista da aplicação prática simulando a geração do OrcP para os anos de 2010 e 2011 (o ano corrente à data da realização do projeto).

O primeiro conjunto de testes validou o algoritmo desenvolvido, sendo que este foi testado num vasto conjunto de instâncias geradas de forma a reproduzir as características reais do problema. Detalhes dos testes realizados podem ser encontrados em Guimarães *et al* (2012).

A segunda fase de testes visava aferir a aplicabilidade do algoritmo na resolução de problemas reais. Com efeito, foi simulada a criação do OrcP para os anos de 2010 e 2011, tendo em vista a comparação dos planos obtidos com os gerados pela empresa. Em ambos os casos o planeamento é realizado para os centros produtores dedicados à produção de cerveja e refrigerantes. Os dados recolhidos para a instância de 2010 do OrcP referem-se a um total de 3 centros produtores, com um número de linhas de enchimento disponíveis em cada centro variável entre 1 e 5, num total de 10 linhas. O OrcV de 2010 contém as previsões de vendas para um total de 125 produtos, posteriormente agregados em 62 famílias. Apesar das restrições tecnológicas presentes, o total de possíveis *setups* (alocações famílias-linhas) é de cerca de 100 em cada mês. O OrcP de 2011 debruça-se sobre os mesmos 3 centros produtores, mas alterações tecnológicas verificadas nas linhas de enchimentos implicam agora o planeamento simultâneo de 14 linhas, variando o número de linhas de enchimento em cada centro entre 4 e 5 linhas. O número de produtos sofreu um aumento para 160, com 68 famílias. Em consequência, o número de possíveis *setups* por período subiu para 120. Para ambos os anos as taxas de enchimento, custos de produção, posse de inventário, transferência e horas extraordinárias estão disponíveis e são baseados em estimativas da empresa. Três cenários foram comparados:

- C1 OrcP criado pela empresa,

- OrcP obtido através da utilização das alocações família-linha (variáveis Y'_{fml}) definidas em C1 no modelo matemático da Secção 5 e resolvendo o subsequente problema de PL para
- C2 determinar as quantidade de produção, inventário, transferência e horas extraordinárias, correspondendo ao melhor plano que seria possível de criar com as alocações das famílias às linhas de enchimento definidas pela empresa,
- C3 Utilização do algoritmo para a criação do OrcP.

Até à data deste projeto, a criação do OrcP na empresa era um processo realizado manualmente com o auxílio de folhas de cálculo. A experiência da equipa de planeamento responsável pelo processo constituía um ponto-chave, uma vez que a tomada de decisão seguia o conhecimento implícito dos custos associados. Assim, a comparação direta entre os planos gerados através da aplicação do algoritmo desenvolvido com os planos definidos na empresa pode não traduzir ganhos diretos. As razões que podem motivar erros na comparação advêm do facto de planos gerados manualmente nem sempre obedecerem rigorosamente a todas as restrições definidas para o problema e/ou de a tomada de decisão responder a critérios não traduzidos devidamente nos dados do problema. A criação do cenário C2 visa colmatar esta dificuldade e permite, simultaneamente, perceber as desvantagens da utilização de linhas preferencias para a alocação das famílias às linhas de enchimento.

Os resultados obtidos para os três cenários nos dois anos simulados são apresentados na Tabela 4. Todos os custos são medidos em termos de unidades monetárias (u. m.) e foram adulterados por questões de confidencialidade. Ambos os cenários testados apresentam ganhos substanciais face ao OrcP da empresa. O cenário C2 obtém poupanças no custo total do plano estimadas em 24% e 36% para os anos de 2010 e 2011, respectivamente. Grande parte dos ganhos é obtida através da redução dos custos de horas extraordinárias, mesmo que isso implique um aumento nos restantes custos. Este facto é bastante interessante, dado que a equipa de planeamento coloca um grande enfoque no controlo dos custos de posse de inventário. Para ambos os anos, o algoritmo gera os planos mais competitivos do ponto de vista

económico, mostrando a vantagem de um sistema de produção flexível onde a mesma família pode ser produzida em mais de uma linha de enchimento e de um algoritmo que tire partido desse facto. No total a poupança no cenário C3 ascende a 35% em 2010 e 39% em 2011.

Tabela 4: Comparação dos resultados obtidos na simulação dos OrcP de 2010 e 2011

	C1	C2	Poupança (C2)	C3	Poupança (C3)
2010					
<i>Custo total</i>	1.962.720	1.509.115	453.605 (23%)	1.275.769	686.951 (35%)
<i>Custo de posse de inventário</i>	108.589	185.748	-77.159 (-71%)	134.426	-25.837 (-24%)
<i>Custo de transferência</i>	204.129	162.247	41.883 (21%)	148.772	55.357 (27%)
<i>Custo de setup</i>	42.550	42.550	0 (0%)	41.650	900 (2%)
<i>Número de setups</i>	851	851	0 (0%)	833	18 (2%)
<i>Custo de horas extraordinárias</i>	1.607.451	1.118.570	488.881 (30%)	950.921	656.530 (41%)
<i>Custo total (s/ custo setup)</i>	1.920.170	1.466.565	453.605 (24%)	1.234.119	686.051 (36%)
2011					
<i>Custo total</i>	3.259.777	2.163.237	1.096.540 (34%)	1.976.865	1.282.912 (39%)
<i>Custo de posse de inventário</i>	450.926	171.429	279.496 (62%)	163.060	287.866 (64%)
<i>Custo de transferência</i>	317.965	312.157	5.808 (2%)	365.503	-47.538 (-15%)
<i>Custo de setup</i>	48.350	48.350	0 (0%)	46.050	2.300 (5%)
<i>Número de setups</i>	967	967	0 (0%)	921	46 (5%)
<i>Custo de horas extraordinárias</i>	2.442.536	1.630.334	812.202 (33%)	1.402.252	1.040.283 (43%)
<i>Custo total (s/ custo setup)</i>	3.211.427	2.114.887	1.096.540 (34%)	1.930.815	1.280.612 (40%)

O *trade-off* entre os diferentes custos fica bem patente nos resultados obtidos. Por exemplo, para o ano de 2010, tanto os cenários C2 como C3, reduzem o custo total face ao plano da empresa através de um aumento dos custos de inventário e, para o ano de 2011, a solução mais eficiente (cenário C3) é a que apresenta os custos de transferência mais elevados, reduzindo significativamente a utilização de horas extraordinárias. Em ambos os anos, C3 apresenta o menor número de *setups* sugerindo futuros planos a um nível operacional mais eficientes em virtude da necessidade de menos mudanças nas linhas de enchimento.

As Figuras 4 e 5 apresentam as quantidades totais mensais de produção, inventário, transferência e horas extraordinárias para cada um dos cenários testados nos anos de 2010 e 2011, respectivamente, ajudando a perceber com maior detalhe as diferenças entre as soluções geradas para o OrcP.

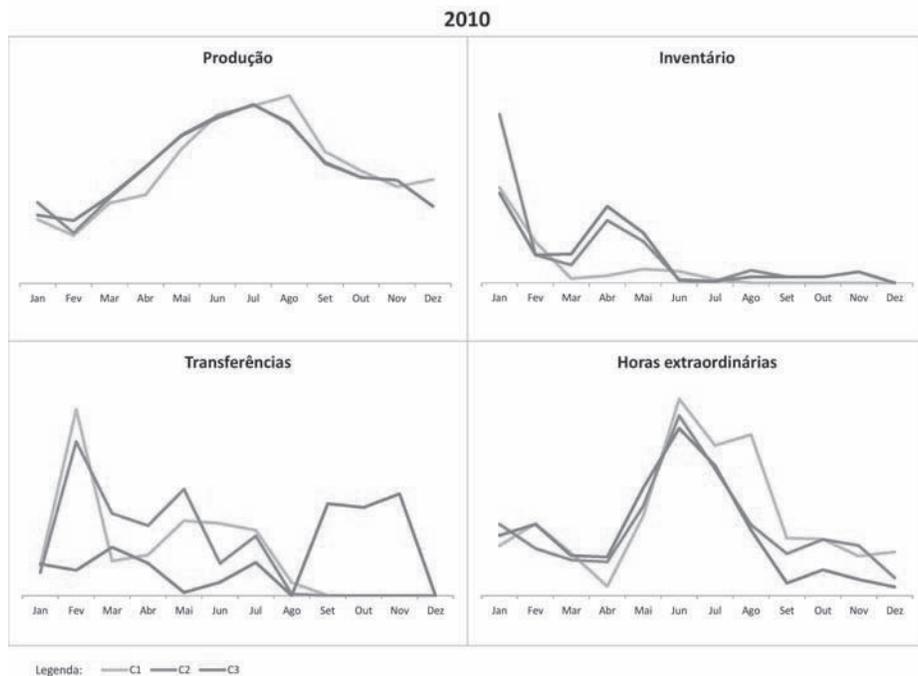


Figura 4: Comparação das quantidades totais mensais de produção, inventário, transferência e horas extraordinárias para os diferentes cenários do OrcP 2010

Atente-se nas diferenças verificadas para a instância relativa a 2010. As quantidades mensais produzidas nos cenários C2 e C3 são bastante semelhantes entre si, mas substancialmente diferentes do cenário C1 para o período compreendido entre os meses de março a setembro. A consequência no cenário C1 é o uso de um maior número de horas extraordinárias nos meses de pico de procura o que em grande parte explica a diferença de custos. Note-se que, ao contrário do cenário C1, os cenários “otimizados” constituem *stock* nos meses anteriores ao pico da procura (março-maio). Finalmente, a diferença entre os cenários C2 e C3 encontra-se principalmente na definição das transferências entre os centros produtores, possibilitada por uma diferente

alocação das famílias às linhas de enchimento. Estas são evidentemente mais baixas durante os meses de janeiro a agosto no plano C3 face a C2, sendo que o aumento verificado no final do ano possibilita a redução do número de horas extraordinárias nesse período.

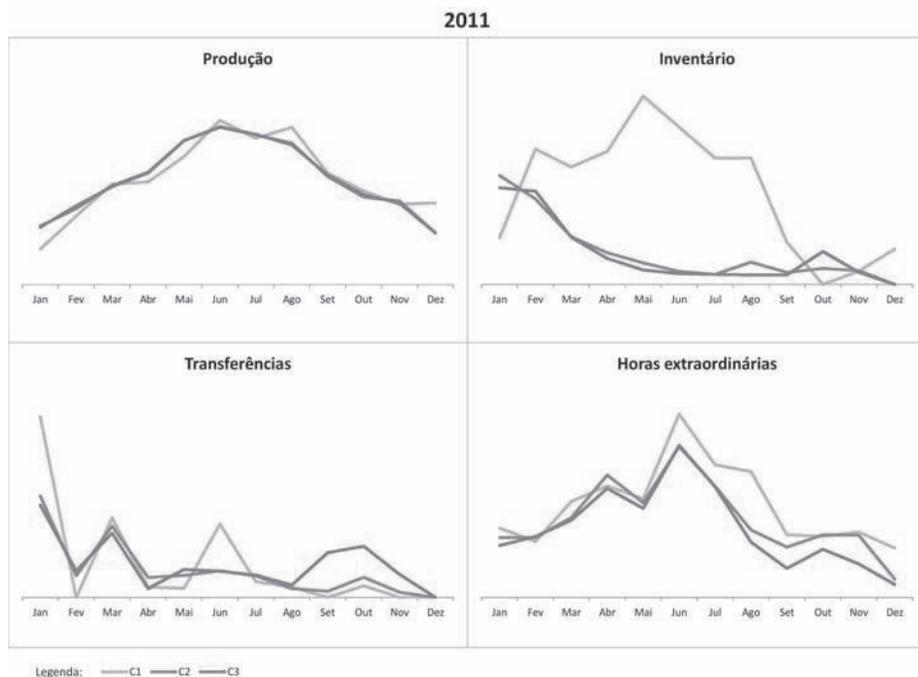


Figura 5: Comparação das quantidades totais mensais de produção, inventário, transferência e horas extraordinárias para os diferentes cenários do OrcP 2011

No que diz respeito à instância de 2011, as diferenças nos custos face ao cenário da empresa (C1) são também, implicadas pelas quantidades totais mensais produzidas. Tal como em 2010, as produções totais estão bastante alinhadas nos cenários C2 e C3, mas desfasadas face ao cenário C1. Isto implica a necessidade no plano da empresa de maiores investimentos em custos de posse de inventário e horas extraordinárias, especialmente visível nos meses de pico. Por seu turno, o cenário C3 tira partido de uma diferente alocação das famílias às linhas de forma a reduzir os custos de horas extraordinárias face a

C2, apesar do incremento registado no final do ano, ao nível das quantidades transferidas.

Os planos do cenário C2 foram obtidos em 12 e 14 segundos para as instâncias de 2010 e 2011, respetivamente, correspondendo ao tempo necessário para resolver até à otimalidade o problema de PL. O algoritmo necessitou de 100 segundos para gerar o OrcP 2010 e de 180 segundos para o ano de 2011. O tempo de execução contempla a obtenção de uma solução inicial e a sua melhoria, tendo sido utilizado o critério de paragem do algoritmo de 1000 iterações sem se registar melhoria na solução incumbente.

7 Implementação na empresa

Os resultados alcançados durante a validação do método desenvolvido motivaram a concepção de um SAD, a ser implementado na empresa, e que permitisse à equipa de planeamento a utilização do método na criação do OrcP. O objetivo do sistema é, com o menor esforço possível, gerir a entrada e saída de dados e efetuar toda a parametrização necessária para a execução do algoritmo nele embebido, sem que para isso sejam necessários conhecimentos profundos na área da Investigação Operacional. O SAD foi apelidado de LTP e o interface desenvolvido encontra-se ilustrado na Figura 6.

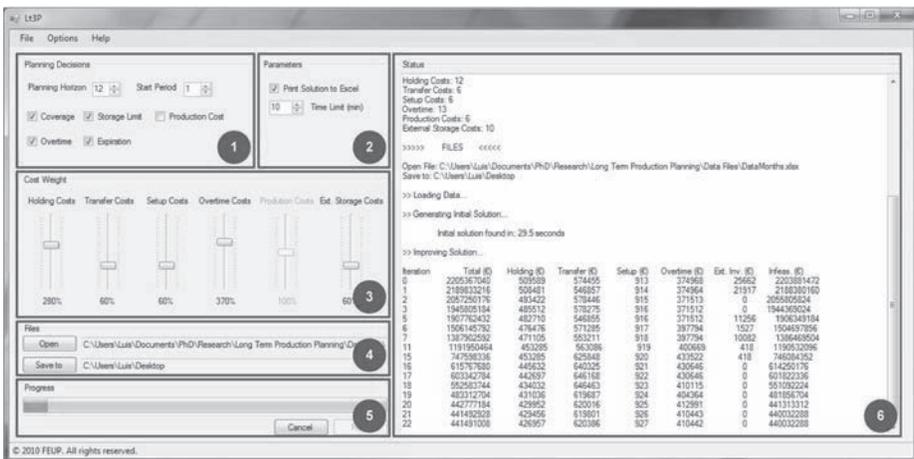


Figura 6: Interface do sistema de apoio à decisão LTP

Na Figura 6 estão identificadas e numeradas as 6 áreas que compõem o interface do SAD e que se descrevem em seguida.

1 - Opções de planeamento

Nesta área o utilizador pode definir o horizonte de planeamento a considerar, assim como o período inicial para o plano a ser gerado, facilitando a utilização de ficheiros de dados de planos anteriores para o planeamento rolante. Também nesta área são ativadas/desativadas algumas características a incluir no plano, designadamente o *stock* de segurança (“Coverage”), as horas extraordinárias (“Overtime”), o limite à capacidade de armazenagem (“Storage Limit”), o prazo de validade dos produtos (“Expiration”) e os custos de produção (“Production Cost”). Ao seleccionar/desseleccionar uma opção o interface ajusta dinamicamente o modelo apresentado na Secção 5, que constitui a base do algoritmo.

2 - Opções do algoritmo

Nesta área o utilizador define o tempo limite para a execução do algoritmo. Desta forma, o algoritmo fica totalmente parametrizado, uma vez que o único elemento em falta é o critério de paragem.

3 - Ponderação dos custos

A finalidade desta área é permitir ao utilizador alterar conscientemente a relação entre os diferentes custos envolvido na avaliação dos planos, visando a criação de cenários alternativos. Inicialmente, o peso relativo de todos os custos está nivelado nos 100% representando o custo real estimado pela empresa. Deslocando a barra de controlo associada a cada um dos indicadores é possível definir o factor pelo qual os custos reais vão ser multiplicados na avaliação conduzida pelo algoritmo das diferentes soluções, aumentando ou reduzindo a sua importância relativa.

4 - Ficheiros de *input* e *output*

Desde o início do desenvolvimento do LTP que um dos objetivos claros era assegurar uma transição suave entre o método utilizado anteriormente e a nova ferramenta. Uma das formas para acautelar esta transição foi o uso de folhas de cálculo como ficheiros de *input* de dados e de *output* da solução gerada. Foram definidos templates para os ficheiros de entrada e saída de dados de forma a, por um lado, modificar rapidamente os dados do problema e, por

outro, facilitar a consulta da solução obtida e dos seus principais indicadores de performance.

5 - Estado da execução

A barra de progresso informa o utilizador da percentagem do tempo total decorrido, sendo possível interromper o processo a qualquer altura.

6 - Informação sobre o estado da execução

Esta área é dedicada a comunicar ao utilizador o estado corrente da execução do algoritmo. Sempre que a solução incumbente é atualizada, é impressa a informação relativa à iteração do algoritmo, bem como os custos (ou outra unidade alternativa definida pelo utilizador) associados a esta solução para cada um dos conjuntos de decisões a tomar (por exemplo, custo de produção, posse inventário, transferência, horas extraordinárias, etc.).

Após o seu desenvolvimento, o LTP foi intensamente testado no período compreendido entre o final de 2010 e setembro de 2011. Sendo que, em outubro de 2011, a empresa estava confiante na utilização do sistema, tendo em vista a elaboração do OrcP para 2012. De facto, a empresa baseou o seu processo de criação do OrcP 2012 no LTP e de uma forma completamente autónoma.

8 Impacto na empresa e futuros desenvolvimentos

Após as fases de concepção, desenvolvimento, teste e implementação o L3tP é hoje uma ferramenta capaz de planear de uma forma automática, otimizada e integrada as operações das diferentes fábricas da empresa, possibilitando a tomada de melhores decisões, suportadas por métodos quantitativos e considerando a sensibilidade dos agentes de decisão através dos vários *inputs* recolhidos.

Durante este processo levantaram-se algumas dificuldades. Uma das mais importantes prendeu-se com os dados de *input* necessários para alimentar o algoritmo. Dado que o problema abordado procura gerar planos para um futuro mais distante, uma parte importante da informação ainda não se encontrava disponível nos sistemas de informação da empresa. Assim, existe um trabalho importante de recolha desta informação e seu tratamento que não

pode ser desprezado e, sobretudo, existe a necessidade de validação dos mesmos por forma a minimizar o impacto de possíveis erros.

Adicionalmente, em especial na fase de testes, observou-se uma tendência por parte dos agentes de decisão da empresa em querer que o algoritmo replicasse de forma exata as regras heurísticas seguidas pela equipa de planeamento. A utilização do LTP obrigou a uma mudança de paradigma na forma de planear, muito em parte motivada pela adoção da avaliação dos planos em termos de custo explícito. Esta mudança gerou uma desconfiança inicial face aos *outputs* da ferramenta. De forma a ultrapassar a desconfiança inicial e estabelecer uma confiança plena no LTP por parte da equipa de planeamento, o modelo que lhe serve de base foi sendo refinado durante a fase de testes para que representasse irrepreensivelmente a realidade da empresa. O modelo apresentado na Secção 5 é hoje uma versão simplificada do modelo em funcionamento na empresa.

Ultrapassadas as dificuldades iniciais, as vantagens da utilização do LTP são largamente recompensadoras. O LTP despoletou a sistematização de todo o processo de elaboração do orçamento de produção e a definição de quais os dados necessários e os principais indicadores de desempenho a considerar. A comparação resultante entre o processo anterior e a nova metodologia, permitiu perceber as limitações associadas às regras definidas anteriormente. A avaliação dos planos através de custos reais é sem dúvida uma mais-valia, possibilitando uma melhor comparação entre cenários alternativos e favorecendo o processo de tomada de decisão.

Um impacto importante do projeto está relacionado com a poupança de custos que um melhor planeamento potencia e que ficou demonstrada nos testes de validação conduzidos. Todavia, um dos principais benefícios que pode ser atribuído a este projeto reside na mudança profunda da tarefa da equipa de planeamento durante a criação do orçamento. Antes da introdução do L3tP, a tarefa de gerar um plano válido para o OrcP consumia quase todo o tempo da equipa (note-se que a elaboração do orçamento pode demorar 3 semanas), sendo que alterações dos pressupostos assumidos ou nos dados de entrada, obrigavam a que a equipa se concentrasse na obtenção de uma nova solução admissível para o problema. Com a introdução do LTP, a situação

mudou radicalmente, uma vez que os planos são agora gerados em poucos minutos. Deste modo, o foco da equipa de planeamento foi redirecionado da criação de um único OrcP para a análise de cenários alternativos para o OrcP. Uma tarefa de pouco valor acrescentado e bastante dispendiosa em termos do tempo exigido, foi transformada no estudo de planos alternativos, potenciando ainda mais a redução dos custos operacionais. A utilização do LTP permite de uma forma simples antecipar e analisar possíveis fontes de distúrbios nos planos gerados, tais como a insuficiência de capacidade produtiva ou de armazenagem, a escassez de vasilhame e as necessidades excessivas de antecipação de *stock*. Simultaneamente, o LTP possibilita de uma forma expedita simular vários cenários extremos, obtendo informação sobre os seus efeitos ao longo do ano e possíveis repostas para estes casos. Entre estes cenários estão os aumentos ou quebras na procura, os ganhos ou perdas de produtividade nas linhas de enchimento, as necessidades adicionais de manutenção nas linhas de enchimento, ou ainda outras fontes de instabilidade que afetem os pressupostos assumidos.

O sucesso da implementação do LTP motivou o desejo por parte da empresa em incorporar decisões adicionais, tendo em vista a extensão do sistema para além do planeamento da produção. Durante o ano de 2012 o LTP foi estendido para a inclusão de decisões de distribuição e utilizado com esta nova variante para a criação do orçamento para o ano de 2013. Conhecidas as localizações e conseqüentemente os custos de transporte entre os centros produtores e os centros distribuidores, bem como a procura de cada produto associada a cada centro distribuidor e sua capacidade de armazenagem, a ferramenta sugere também um plano de entregas dos centros produtores para os centros distribuidores, satisfazendo as procuras identificadas da forma mais eficiente possível (minimizando os custos de produção e de distribuição).

Este contexto potencia uma das principais vantagens do LTP que é o estudo da abertura e/ou fecho de novas unidades de produção, ou linhas de enchimento, e agora centros de distribuição, avaliando estes cenários através de ganhos e perdas. Desta forma, para além de otimizar os planos considerando a cadeia de abastecimento atual a ferramenta possibilita a procura de melhores configurações para mesma. Uma vez que todos os

parâmetros no modelo são definidos em função do período de tempo, este permite também estudar a calendarização destas novas aberturas/fecho de instalações.

Durante o ano de 2012 o LTP desempenhou um papel importante na reestruturação da cadeia de abastecimento da empresa, permitindo avaliar e validar do ponto de vista operacional as mudanças estratégicas a serem implementadas. Também o orçamento para o ano de 2013 beneficiou (e muito) das novas funcionalidades do LTP.

Agradecimentos

Os autores agradecem à Unicer - Bebidas, S.A. pela colaboração neste projeto, nomeadamente ao Eng. Carlos César (COO - Chief Operating Officer), ao Eng. Pedro Ribeiro (Diretor do Planeamento Operacional) e à sua equipa Ana Ferreira, Benjamin Herbin, Jorge Gomes e Maria João Pissarra. O primeiro autor agradece à Fundação da Ciência e Tecnologia a atribuição da bolsa de doutoramento com a referência SFRH/BD/62010/2009.

REFERÊNCIAS

GUIMARÃES, L., Klabjan, D. e Almada-Lobo, B. (2012). Annual production budget in the beverage industry. *Engineering Applications of Artificial Intelligence*, 25(2):229 – 241.

HANSEN, P., Mladenovic, N., e Pérez, J.M. (2008). Variable neighbourhood search: methods and applications. *4OR: A Quarterly Journal of Operations Research*, 6(4):319–360.

HUNG, Y., Chen, C., Shih, C. e Hung, M. (2003). Using tabu search with ranking candidate list to solve production planning problems with setups. *Computers & Industrial Engineering*, 45(4):615 – 634.

MLADENOVIC, N. e Hansen, P. (1997). Variable neighborhood search. *Computers & Operations Research*, 24(11):1097 – 1100.

NASCIMENTO, M., Resende, M. e Toledo, F. (2010). Grasp heuristic with path-relinking for the multi-plant capacitated lot sizing problem. *European Journal of Operational Research*, 200(3):747 – 754.

SAMBASIVAN, M. e Schmidt, C.P. (2002). A heuristic procedure for solving multi-plant, multi-item, multi-period capacitated lot-sizing problems. *Asia Pacific Journal of Operational Research*, 19:87–105.

SAMBASIVAN, M. e Yahya, S. (2005). A lagrangean-based heuristic for multi-plant, multi-item, multi-period capacitated lot-sizing problems with inter-plant transfers. *Computers and Operations Research*, 32(3):537–555.

(Página deixada propositadamente em branco)

CASO 8

OTIMIZAÇÃO DA SEQUÊNCIA DE VEÍCULOS NA MONTAGEM FINAL DE UMA EMPRESA DA INDÚSTRIA AUTOMÓVEL: ESTUDO DE UM CASO

Cláudia Sofia Rodrigues Duarte

MIT Portugal, EDAM area, Departamento de Produção e Sistemas
Escola de Engenharia, Universidade do Minho
claudia.s.r.duarte@gmail.com

José Valério de Carvalho

Departamento de Produção e Sistemas, Escola de Engenharia
Universidade do Minho
vc@dps.uminho.pt

Ana Paula Barbosa Póvoa

Centro de Estudos de Gestão do IST, Instituto Superior Técnico
Universidade de Lisboa
apovoa@tecnico.ulisboa.pt

RESUMO

Atualmente, o aumento da concorrência nos mercados faz com que o ciclo de vida dos produtos diminua e a variedade dos produtos aumente para satisfazer as necessidades dos clientes. Por estas razões, as empresas foram obrigadas a introduzir flexibilidade nos sistemas produtivos a fim de conseguirem satisfazer a procura. Esta flexibilidade trouxe desafios nomeadamente na configuração dos layouts e no planeamento da produção.

Neste contexto surge o presente trabalho que se propõe desenvolver um modelo que ajude a responder aos desafios que surgem na indústria automóvel, em que, para lidar com as flutuações da procura ao longo do ano usando eficazmente a capacidade disponível, se adotou a estratégia de montar veículos diferentes na mesma linha de produção. Esta estratégia dá origem a problemas novos que é necessário abordar.

Para o efeito, desenvolvemos um modelo de programação inteira para decidir a sequenciação de veículos na linha de produção que considera a capacidade de produção limitada, que os carros para mercados especiais devem ser produzidos no início do turno, e que os carros da mesma cor devem ser agrupados.

Este trabalho foi desenvolvido no âmbito de um projeto de investigação do Programa MIT-Portugal, e teve como referência a linha de montagem de um parceiro industrial deste projeto, cuja identidade está protegida por um acordo de confidencialidade. Os resultados da investigação indicam que é possível desenvolver uma ferramenta de apoio à decisão que permite fazer o planeamento diário da sequenciação, obtendo um melhor planeamento de trabalhadores extra quando as restrições de capacidade não podem ser cumpridas, para além de permitir uma melhor sincronização entre a pintura e a montagem final e de apoiar uma melhor logística com os fornecedores de peças pintadas.

PALAVRAS-CHAVE

Sequenciação de veículos, Planeamento da Produção, Otimização, Programação Inteira, Heurísticas

Introdução

Recentemente o nosso parceiro industrial implementou o conceito de linha única permitindo a produção de três modelos (A, B e C) em simultâneo, numa única linha de montagem final. O planeamento da produção e mais concretamente a sequenciação destes três modelos na linha única revela-se de extrema importância para evitar paragens de linha ou subaproveitamentos da linha de produção. Neste sentido, o principal objetivo deste projeto é o desenvolvimento de um modelo matemático e computacional que permita determinar a sequência dos carros na montagem final, em cada turno.

Na literatura, existem algumas propostas como as de Drexl and Kimms (2001), Prandstetter and Raidl (2008) e Fliedner and Boysen (2008) que utilizam abordagens exatas baseadas em modelos matemáticos de programação inteira, para problemas semelhantes. O problema é NP-difícil (Kis 2004), e Estellon, Gardi e Nouioua (2008) observaram que este tipo de abordagens atinge o seu limite quando se consideramos aproximadamente cem veículos com poucas opções. Assim, de forma a conseguir obter soluções aceitáveis para o problema, na literatura aparecem várias heurísticas e meta-heurísticas como forma de resolução do problema (Christine Solnon, Cung, Nguyen, & Artigues, 2008). Exemplos destas heurísticas e meta-heurísticas são Ant Colony Optimization, Algoritmos Progressivos e Algoritmos de Pesquisa Local (Gottlieb, Puchta, & Solnon, 2003), (Ribeiro, Aloise, Noronha, Rocha, & Urrutia, 2008), (C Solnon, 2008) Algoritmos Genéticos (Warwick & Tsang, 1995), Simulated Annealing e Tabu Search (Cordeau, Laporte, & Pasin, 2008), (Reis, 2007). O software de programação matemática tem melhorado nos últimos anos. Em experiências computacionais preliminares, observámos que é possível resolver problemas com uma dúzia de opções e cerca de 300 carros em tempos relativamente curtos. Por isso, decidiu-se dar mais um passo, e explorar o conceito de procurar sequências de montagem em que os carros da mesma cor estivessem agrupados.

O conceito de agrupar cores é diferente da tradicional minimização do número de mudanças de cor na sequência de montagem, apresentado, por exemplo, em Prandstetter e Raidl (2008), que considera o número de vezes em que dois carros consecutivos na sequência de montagem são de cores

diferentes. Minimizar o número de mudanças de cor não conduz necessariamente a que os carros da mesma cor apareçam agrupados na sequência. Neste artigo, defendemos que a implementação deste conceito pode melhorar o desempenho do sistema.

Na próxima secção apresentamos o sistema de produção genérico do nosso parceiro industrial, de forma a enquadrar o problema. Na secção 4 descrevemos o problema enquanto na secção 4 apresentamos o modelo desenvolvido utilizando o IBM ILOG 12.2. Os testes que verificaram a robustez da nossa abordagem são apresentados na secção 5. Na secção 6 apresentamos as conclusões finais do trabalho. Importa referir que devido a razões de confidencialidade os valores reais foram alterados, no entanto, o caso de estudo em análise descreve a tendência observada na realidade da produção em análise.

Sistema de Produção

O Sistema de Produção em estudo engloba três áreas principais: carroçarias, pintura e montagem final (Figura 1)



Figura 1: Principais áreas produtivas numa empresa de montagem final de veículos.

Na área das carroçarias a estrutura do carro é montada através de processos de soldagem que unem peças de metal estampadas e prensadas. Nesta área, os tempos de setup das prensas são elevados e a melhor sequência é a que minimiza os tempos de setup.

Na área da pintura, a estrutura do carro é tratada, preparada e pintada. Um objetivo importante nesta área é minimizar a quantidade de solvente utilizado para limpar os bicos das pistolas de pintura. Estes bicos têm de ser limpos quando a cor é alterada ou por razões de manutenção e de qualidade. Por estes motivos, é definido um número máximo de carros com a mesma cor que podem ser pintados seguidos, sem recorrer à limpeza dos bicos. Este uso de

solventes para limpar os bicos também representa um tempo de setup que deve ser minimizado, para além do custo associado que interessa minimizar.

Na área da montagem final, os componentes mecânicos, elétricos e outros são adicionados ao carro. Alguns destes componentes podem exigir operações com maior duração que o tempo de ciclo (tempo disponível para montagem em cada estação de trabalho). Os carros que necessitam destes componentes devem estar afastados entre si na sequência para nivelar a carga de trabalho nas estações em causa. O objetivo é garantir o balanceamento da carga de trabalho e o fornecimento de componentes para minimizar o tempo necessário para construir a sequência na montagem final.

Como 70% do valor do carro é adicionado na montagem final, os problemas de sequenciação focam-se em encontrar uma boa sequência para a montagem final, utilizando restrições de capacidade que asseguram o balanceamento da carga de trabalho e o fornecimento de componentes (Gravel, Gagné, & Price, 2005). Por este motivo, na próxima secção, iremos detalhar as características da Montagem Final.

Características da Linha de Montagem Final

A Montagem Final é constituída por uma linha única composta por um grupo de estações de trabalho. Estas estações de trabalho estão ligadas por um sistema de transporte que assegura o movimento dos carros desde a primeira estação até à última. Em cada estação de trabalho, enquanto o carro se movimenta pela estação, são executadas diferentes operações que dependem do modelo do veículo. Estas estações têm um comprimento correspondente ao tempo de ciclo de produção. Quando as operações correspondentes ao modelo do veículo que está a atravessar a estação estiverem terminadas, o operador retorna ao início da estação ou para no início da estação, à espera do próximo carro. A representação gráfica das estações de trabalho encontra-se na Figura 2.



Figura 2: Representação da linha de produção e das estações de trabalho.

Na generalidade, quando, numa determinada estação de trabalho, existem operações para o modelo de carro com uma duração significativamente maior do que o tempo de ciclo, ou quando, por alguma razão os operadores não conseguem terminar as operações no tempo disponível, há duas estratégias possíveis para ultrapassar os atrasos, seguidas em diferentes contextos de produção. Nos Estados Unidos, os operadores não estão autorizados a passar as fronteiras das estações de trabalho, pelo que trabalhadores extra são necessários para terminar as operações. No Japão os operadores param a linha, para poderem terminar o trabalho. O nosso parceiro industrial pretende implementar a estratégia utilizada nos Estados Unidos, recorrendo a trabalhadores extra. No entanto, neste momento, quando há problemas de qualidade ou atrasos, a linha é parada para garantir a correção dos problemas.

Uma das questões mais relevantes para a definição de uma sequência de veículos de modelos distintos é o tempo de montagem de determinados modelos em estações de trabalho específicas. O tempo disponível em cada estação é constante, igual ao tempo de ciclo, e está relacionado com a velocidade do transportador. Se existe um modelo com um tempo de montagem ligeiramente superior ao tempo de ciclo, numa determinada estação, não é necessário proceder a uma paragem da linha, desde que o modelo seguinte compense este excesso de tempo, de modo a nivelar o tempo de operação em cada estação de trabalho. Se não houver esta compensação, é

necessário recorrer a trabalhadores extra ou então poderá haver uma paragem de linha.

Definição do Problema

O problema a tratar pode ser definido da seguinte forma:

- Dados os seguintes parâmetros:
 - Opções ($o \in O$)
 - Cores ($c \in C$)
 - Variantes ($v \in V$). Uma variante v pode ter várias opções o , mas apenas uma cor c .
 - A matriz 0-1 $\hat{O}=[o_{o,v}]$ representa as opções de cada variante. Se a variante v inclui a opção o , o valor de $o_{o,v}$ é 1; no caso contrário, o valor é 0.
 - A matriz 0-1 $\hat{C}=[c_{c,v}]$ representa a cor de cada variante. Se a variante v tem a cor c , o valor de $c_{c,v}$ é 1; no caso contrário, o valor é 0.
 - Cores dos carros especiais $dc \subseteq C$. As primeiras cores da matriz são as cores dos carros especiais dc .
 - Procura de cada variante (D_v)
 - Número de carros especiais dt ($dt = \sum_v^V D_v * c_{c,v}, \forall c \in dc$)
 - Posições na sequência ($t \in T$), de acordo com a capacidade máxima da linha, por turno.
 - Tempo extra necessário para a montagem da opção o (w_o) – foi considerado um valor igual para todos os casos, e igual a 1, na atual versão simplificada do modelo.
 - As restrições de capacidade ($H_o:N_o$) – que significa que, em qualquer parte da sequência com um comprimento de N_o veículos, apenas podem estar H_o veículos que requeiram excesso de tempo.
 - valor dos pesos e – o peso e é atribuído à parte da função objetivo que avalia as violações de restrições de capacidade. O peso e é atribuído à parte da função

objetivo que avalia se os carros especiais estão nas primeiras dt posições da sequência. A escolha destes pesos será discutida no final da secção 4.2.

- Pretende-se determinar a sequência de produção, de forma a minimizar o número de violações de restrições de capacidade, a produzir os carros especiais nas primeiras posições da sequência e a agrupar os carros por cor. Estes objetivos devem ser considerados pela ordem de importância com que foram mencionados.

Para resolver este problema e tendo em conta as características da linha de montagem, descritas na secção 2, houve um conjunto de hipóteses que tiveram de ser assumidas à partida:

- Não há stock entre as estações de trabalho;
- O problema é estático, ou seja, o mix dos modelos não pode ser alterado;
- Vários modelos são produzidos na linha da Montagem Final. Estes modelos têm diferentes componentes e requerem diferentes tarefas e diferentes tempos de processamento;
- Há regras para a sequenciação dos carros. Estas regras traduzem-se em restrições de capacidade ($H_o: N_o$ com $o \in O$).

Desta forma, estamos em condições para apresentar, na próxima secção o modelo de sequenciação de veículos desenvolvido.

Modelo de Sequenciação de Veículos

Um modelo matemático foi desenvolvido para o problema de sequenciação de veículos. O modelo inicial, que considera apenas restrições de capacidade baseia-se no Modelo OPL incluído no IBM ILOG 12.2 CPLEX Optimization Studio e no modelo de Drexl and Kimms (2001). O nosso modelo considera que:

- os carros são produzidos numa linha de montagem;
- os carros se movem ao longo das estações de trabalho, onde são instaladas as opções dos veículos, como o ar condicionado ou os rádios;

- a linha de montagem pode ser vista como um conjunto de posições, em número igual à procura, onde cada carro vai ser alocado a uma das posições;
- os carros não podem ser alocados arbitrariamente porque existe capacidade limitada;
- a cada opção que limite a capacidade de produção, porque a sua montagem exige mais do que o tempo de ciclo, está associada uma restrição de capacidade;
- os carros para mercados especiais devem ficar nos primeiros lugares da sequência, por motivos de qualidade e de restrições de envio;
- os carros com a mesma cor devem estar juntos na sequência.
- a nossa função objetivo considera três características:
 - O número de vezes que uma restrição de capacidade é violada, através do método introduzido no *ROADEF Challenge 2005* (Prandstetter, 2005). Cada restrição de capacidade tem diferentes níveis de prioridade tendo em conta o tempo extra necessário para montar essa opção.
 - Os carros especiais devem ser sequenciados nas primeiras posições;
 - A dispersão de carros da mesma cor deve ser o menor possível, num compromisso com os restantes objetivos do sistema.

De seguida, iremos apresentar o modelo matemático e explicá-lo em detalhe.

Novo modelo de programação matemática para sequenciação de veículos

A nossa formulação inclui um conjunto de parâmetros constituídos pelas opções ($o \in O$), por um conjunto de cores ($c \in C$), pelas cores dos carros especiais ($dc \subseteq C$) e por um conjunto de variantes ($v \in V$) requeridas pelos clientes. Cada variante (v) é constituída por um conjunto de opções (o) e contém apenas uma cor (c). A matriz binária \hat{O} considera as opções de cada variante. Se a opção (o) está incluída na variante (v), o valor é 1, se não está incluída o valor é 0. A matriz binária \hat{C} considera as cores de cada variante. Se a

cor (c) pertence à variante (v), o valor é 1, se não pertence o valor é 0. As primeiras cores na matriz das cores são as cores dos carros especiais (dc).

O número de carros especiais (di) é igual a $\sum_v^V D_v * c_{c,v}, \forall c \in dc$. A procura total é igual a $\sum_v^V D_v$, sendo D_v a procura de cada variante. A procura total corresponde ao número total de posições na sequência (T). A cada opção corresponde uma restrição de capacidade ($H_o:N_o$) que significa que apenas podem ser sequenciados H_o carros numa sequência de N_o carros que requerem a opção $o \in O$. Cada restrição de capacidade tem diferentes níveis de prioridade de acordo com o tempo extra necessário para a montagem da opção (w_o).

A função objetivo considera três medidas. Todas têm um peso associado. O peso é atribuído à parte da função objetivo que avalia as violações das restrições de capacidade. O peso é atribuído à parte da função objetivo que avalia se os carros especiais se encontram nos primeiros lugares da sequência. O peso 1 é atribuído à parte da função objetivo que avalia a dispersão de cor dos carros.

As variáveis de decisão que constam deste modelo são as seguintes:

- $custo$ - valor da função objetivo
- $pen_{o,t}$ - penalidade associada à violação de restrições de capacidade
- $\hat{X} = [x_{v,t}]$ - matriz 0-1 cujo valor de $x_{v,t}$ é 1 se a variante v for colocada na posição t . Caso contrário, o valor é zero.
- $\hat{S} = [start_{c,t}]$ - matriz 0-1 cujo valor de $start_{c,t}$ é 1 entre a primeira e a última posição da sequência. Nos restantes casos o valor é zero.
- $\hat{E} = [end_{c,t}]$ matriz 0-1 cujo valor de $end_{c,t}$ é 1 depois da última posição do carro com a cor c . Nos restantes casos o valor é zero.

O modelo matemático desenvolvido pode então ser descrito da seguinte forma:

$$\min[custo] \tag{1}$$

$$custo = \alpha * \left(\sum_{o=1}^O \sum_{t=1}^T pen_{o,t} * w_o \right) \tag{2}$$

$$\begin{aligned}
& + \beta * \left(\sum_{v=1}^V \sum_{c=1}^{dc} \sum_{t=dt+1+I}^T c_{c,t} * x_{v,t} * (t - dt) \right) \\
& + \sum_{c=1}^C \sum_{n=1}^T (start_{c,n} - end_{c,n})
\end{aligned}$$

s.a

$$\sum_{v=1}^V x_{v,t} = 1, \forall t \in T \quad (3)$$

$$\sum_{t=1}^T x_{v,t} = D_v, \forall v \in V \quad (4)$$

$$\sum_{v=1}^V \sum_{j=k}^{k+N_o-1} (x_{v,j} * o_{o,v}) - pen_{o,j} \leq H_o, \forall o \in O, \forall k \in \{1 \dots T - N_o + 1\} \quad (5)$$

$$\sum_{v=1}^V (x_{v,t} * c_{c,v}) \leq start_{c,t} - end_{c,t}, \forall c \in C, \forall t \in T \quad (6)$$

$$start_{c,s} \leq start_{c,s+1}, \forall c \in C, \forall s \in \{1 \dots T - 1\} \quad (7)$$

$$end_{c,s} \leq end_{c,s+1}, \forall c \in C, \forall s \in \{1 \dots T - 1\} \quad (8)$$

$$pen_{o,t} \geq 0, \forall o \in O, \forall t \in T \quad (9)$$

$$x_{v,t} \in \{0,1\}, \forall v \in V, \forall t \in T \quad (10)$$

$$start_{c,t} \in \{0,1\}, \forall c \in C, \forall t \in T \quad (11)$$

$$end_{c,t} \in \{0,1\}, \forall c \in C, \forall t \in T \quad (12)$$

A função objetivo (equação 1) considera: o número de violações de restrições de capacidade, usando o método introduzido no *ROADEF Challenge 2005*, expresso na primeira linha da equação 2, o posicionamento dos carros especiais, expresso na segunda linha da equação 2, e a soma da dispersão dos carros da mesma cor, expresso na terceira linha da equação 2. As equações 3 a 12 descrevem as restrições do modelo. A equação 3 garante que apenas um carro pode ser produzido na posição t . Para assegurar que o número de carros de cada variante corresponde exatamente à procura dos clientes, a restrição 4 foi implementada. As restrições 5 e 9 garantem que cada violação de restrição de capacidade é contada corretamente usando o método aplicado no *ROADEF Challenge 2005* e a variável $pen_{o,t}$. As restrições 6, 7, 8, que o valor $start_{c,t}$ da

matriz \hat{E} é 1 entre o primeiro e último carro da sequência e zero nos outros casos, e que o valor $end_{c,t}$ da matriz \hat{E} é 1 depois da última posição do carro com a cor c e zero nos outros casos. As restrições 10, 11 e 12 garantem que as matrizes \hat{S} , \hat{E} e \hat{X} são binárias.

Nas secções seguintes, vamos explicar com detalhe como foram considerados os carros especiais e a dispersão de cor.

Os carros especiais nas sequências

O nosso parceiro industrial definiu que é importante proceder à montagem dos carros especiais nos primeiros lugares da sequência. Isto porque estes carros têm de ser entregues no início do dia, e, por isso, é importante produzi-los primeiro. Sendo a programação matemática uma ferramenta poderosa para modelar características especiais decidiu introduzir-se uma penalização na função objetivo do modelo de programação matemática, como se pode verificar na figura seguinte.



Figura 3: Representação da função objetivo para 30 carros especiais e 40 posições, considerando que os carros especiais devem ficar em primeiro lugar na sequência.

Como exemplo, suponhamos que 30 carros especiais devem ser produzidos num determinado turno. Se eles não forem sequenciados nos primeiros 30 lugares da sequência, uma penalidade é adicionada na função objetivo tal como está representada na Figura 3. Carros especiais até a posição 30, inclusive, não terão penalidades. Por outro lado, se um carro especial é sequenciado após a posição 30, o valor da penalidade corresponde à diferença

entre à posição do carro e o número de carros especiais, multiplicada por uma constante, que neste exemplo é 10.

Observe-se que pode não ser possível colocar os 30 carros nas primeiras 30 posições da sequência sem violar as restrições de capacidade. Portanto, como estes objetivos podem ser conflituosos, foram então dados pesos diferentes na função objetivo. Com o auxílio do nosso parceiro industrial considerou-se que a questão mais importante na sequência final de montagem de automóveis é obedecer às restrições de capacidade. Por este motivo, a penalidade associada às restrições de capacidade é a mais elevada ($\alpha = 50000$) e a penalidade associada à não sequenciação dos carros especiais nos primeiros lugares da sequência tem um valor intermédio ($\beta = 10$). Outra penalidade menor (peso = 1) é considerada para a dispersão de cor. Este valor é menor, porque, de acordo com o nosso parceiro industrial, obedecer às restrições de capacidade e sequenciar os veículos especiais nos primeiros lugares da sequência, é mais importante do que medir a dispersão de cor, e entendeu-se que $\beta = 10$ seria adequado. No que diz respeito à escolha de $\alpha = 50000$, esta destina-se a assegurar que uma solução com um número maior de violações de restrições de capacidade nunca irá ocorrer se existir alguma solução com um número menor de violações de restrições de capacidade. Este número é suficientemente grande para garantir que a solução com menor número de violações de restrições de capacidade é escolhida, mesmo que tenha maior dispersão de cor, por exemplo; Por outro lado, não é demasiado grande para causar problemas de estabilidade numérica do modelo.

Dispersão de carros da mesma cor

Na literatura, os modelos desenvolvidos anteriormente, como o de Prandstetter and Raidl (2008), consideram o número de mudanças de cor. Estes modelos assumem que há uma mudança de cor na sequência se dois carros adjacentes têm cores diferentes. Nestes modelos o número de mudanças de cor é o mesmo quando dois carros estão afastados duas posições (dois carros com outra cor no meio) ou quando estão separados 100 posições. Por esta razão, introduzimos o conceito de uma sequência que minimiza a soma dos valores

de dispersão de carros com a mesma cor, tentando colocar os carros de uma determinada cor o mais próximo possível na sequência de montagem.

A solução fornecida pelo nosso modelo encontra-se representada na Figura 4. A primeira linha representa as cores dos carros e a segunda representa a opção de cada carro, considerando apenas o modelo do carro (A, B e C). Na segunda linha, os carros amarelos têm uma restrição de capacidade de 1:2 e os carros cor de laranja tem uma restrição de capacidade de 2:3. Como já foi explicado, anteriormente, a restrição 1:2 significa que apenas é permitido 1 carro em cada 2 carros sequenciados, o mesmo para a restrição 2:3 que significa que apenas são permitidos 2 carros em cada 3 carros sequenciados. Este exemplo considera 187 carros, 32 variantes, 10 opções, 11 cores, 74 carros especiais, $\beta = 10$ e $\alpha = 50000$.



Figura 4: Solução fornecida pela formulação de programação inteira.

Considerando as cores dos carros e, por isso, a primeira linha, neste exemplo, todos os carros especiais, que têm cor preta, são colocados nas primeiras posições da sequência, e a dispersão de cor correspondente é igual a zero. Considerando os carros vermelhos da primeira linha, verificamos que para evitar a colocação de dois carros com a opção amarela da segunda linha juntos, e, por isso, evitar a violação de restrições de capacidade, há dois carros de outra cor colocados entre os carros vermelhos. Portanto, a dispersão da cor vermelha é igual a dois. Globalmente esta solução, não ótima, mostra uma dispersão total de 23 e foi obtida em, aproximadamente, uma hora e meia. Esta solução não otimiza a função objetivo, uma vez que não foi dado tempo suficiente para o *Cplex* encontrar essa solução.

Soluções com uma dispersão de cores reduzida permitem definir lotes de carros com a mesma cor para formar uma sequência na área de pintura. Para fins ilustrativos, na solução apresentada na Fig. 4, os carros podem ser pintados utilizando a sequência: preto, azul esverdeado, vermelho, branco, etc. Isto

permite uma redução do consumo de solventes na área de pintura, porque as cores serão alteradas vezes menos.

A estratégia consiste em utilizar a sequência para determinar a ordem dos carros na pintura. Após os carros serem pintados, eles podem entrar na linha de montagem final utilizando a sequência determinada pelo modelo. Mudanças de posição são necessárias, mas eles são permitidas, porque a empresa quer manter sempre um número de carros no *buffer* intermédio entre a pintura e montagem final.

Esta estratégia pode também proporcionar uma melhor logística com os fornecedores, por exemplo, no caso dos para-choques, que têm atributo cor e são fornecidos numa sequência que segue a sequência da montagem final. A entrega de para-choques quando as cores estão agrupadas pode permitir ao fornecedor uma produção em lotes e uma diminuição das operações necessárias para organizar a sequência, reduzindo os custos, e até mesmo evitar eventuais erros de abastecimento de sequência.

Este novo modelo que conta a dispersão de cada cor tem uma complexidade maior, face ao modelo que considera apenas restrições de capacidade e carros especiais. Por sua vez apresenta, também, maiores tempos computacionais para obter uma solução. Por esta razão, foi estudado o desenvolvimento de uma heurística que será descrita na próxima secção.

Nova heurística integrada com o modelo de programação inteira

Resultados preliminares mostraram que o modelo que considera apenas restrições de capacidade e carros especiais é resolvido, em média, em segundos para 300 carros, no IBM ILOG 12.2. No entanto, quando se insere no modelo a dispersão de cor o problema torna-se mais difícil de resolver. Por este motivo, uma heurística foi desenvolvida, mantendo a abordagem original do problema.

A ideia básica por trás da heurística é uma medida aproximada do valor de dispersão. Em vez de medir a dispersão considerando o número total de posições ocupadas pelos carros de uma determinada cor, a heurística mede a dispersão considerando o número total de intervalos com, por exemplo, 5, 10 ou 25 posições ocupadas pelos carros e pelas respectivas cores. Assim, a

heurística simplifica o problema dividindo a sequência em intervalos, reduzindo assim o número de variáveis e, conseqüentemente, a complexidade do problema.

Para reduzir o número de variáveis da matriz das cores, o conceito de número de intervalos foi criado. Cada intervalo (intervalo) representa um grupo de carros de acordo com a seguinte equação:

$$\text{número de intervalos} = \frac{\text{número de posições}}{\text{largura dos intervalos}}$$

Para cada cor, considera-se um intervalo de carros, constituído por um determinado número de carros, como a Figura 5 explica.

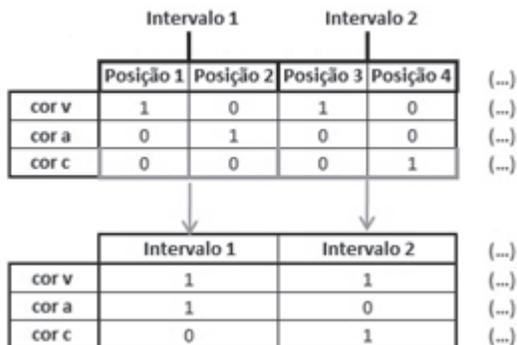


Figura 5: Aglomerar de 4 carros considerando três cores, um número de intervalos igual a 2.

A Figura 5 permite explicar melhor o conceito de aglomerar carros em intervalos. Por exemplo, a Posição 1 e 2, vão passar a ser representadas pelo intervalo 1. Esse intervalo adquire assim a cor dos carros da Posição 1 e 2. Portanto, cada intervalo pode ter mais do que uma cor, pois a sua cor corresponde à cor das posições aglomeradas.

O número de intervalos adequado, para cada problema de otimização, deve ser avaliado testando diversas larguras de intervalo e tendo em conta os resultados computacionais desses testes. Isto porque o número de intervalos deve ser o mais próximo possível do número de posições da sequência, para alcançar o resultado mais próximo possível da solução ótima do problema real. Estima-se que esta aproximação esteja de acordo com a seguinte equação.

dispersão do problema real

*= dispersão da heurística * largura dos intervalos*

Isto significa que a dispersão do problema real é, no máximo, igual à largura do intervalo multiplicado pelo valor da dispersão do resultado da heurística. Esta questão não será explorada neste artigo, mas é desenvolvida em Duarte (2014).

Resumindo, a heurística simplifica o problema dividindo a sequência em pequenos intervalos, conseguindo-se desta forma a redução do número de variáveis. Por outro lado, em vez de uma solução ótima global, à saída obteremos uma solução ótima de um problema aproximado.

A formulação deste novo modelo inclui uma função objetivo similar (equação 1 da secção 4) e as restrições 3, 4, 5, 9 e 10 do modelo de programação apresentado na secção 4. No entanto, as variáveis $start_{c,t}$ e $end_{c,t}$ terão uma formulação diferente devido ao conceito do número de intervalos ($nint$). Assim, foram criados os parâmetros $nint$ e $nlarge$. O parâmetro $nlarge$ representa a largura do intervalo e é obtida considerando a seguinte equação.

$$nlarge = \begin{cases} \frac{T}{nint} + 1, & t \% nint \neq 0 \\ \frac{T}{nint}, & otherwise \end{cases}$$

Uma nova variável de decisão foi criada e as variáveis $start_{c,t}$ e $end_{c,t}$ foram modificadas da seguinte forma:

- $\hat{S} = [start_{c,s}]$ - Matriz binária. Valor de $start_{c,s}$ igual a 1 quando o primeiro intervalo de carros, s , com $s \in \{1 \dots nint\}$, com a cor c é sequenciado. Mantém-se igual a 1 até ao último intervalo.
- $\hat{E} = [end_{c,s}]$ - Matriz binária. Valor de $end_{c,s}$ igual a 1 depois do último intervalo de carros, s , com $s \in \{1 \dots nint\}$, com a cor c ser sequenciado. Mantém-se igual a 1 até ao último intervalo.
- $\hat{B} = [bint_{c,s}]$ - Matriz binária. Valor igual a 1 se a cor c é sequenciada no intervalo s , com $s \in \{1 \dots nint\}$.

A nova formulação que inclui a heurística é definida pelas seguintes equações.

$$\min[\text{custo}] \quad (1)$$

$$\text{custo} = \alpha * \left(\sum_{o=1}^O \sum_{t=1}^T \text{pen}_{o,t} * w_o \right) + \beta * \left(\sum_{v=1}^V \sum_{c=1}^{dc} \sum_{t=dt+1}^T c_{c,t} * x_{v,t} * (t - dt) \right) \quad (2a)$$

$$+ \sum_{c=1}^C \sum_{n=1}^{nint} (\text{start}_{c,n} - \text{end}_{c,n})$$

$$\sum_{v=1}^V x_{v,t} = 1, \forall t \in T \quad (3)$$

$$\sum_{t=1}^T x_{v,t} = D_v, \forall v \in V \quad (4)$$

$$\sum_{v=1}^V \sum_{j=k}^{k+N_o-1} (x_{v,j} * o_{o,v}) - \text{pen}_{o,j} \leq H_o, \forall o \in O, \forall k \in \{1 \dots T - N_o + 1\} \quad (5)$$

$$\sum_{v=1}^V (x_{v,t} * c_{c,v}) \leq \text{bint}_{c,[(T-1)\backslash nlarge]+1}, \forall c \in C, \forall t \in T \quad (13)$$

$$\text{bint}_{c,s} \leq \text{start}_{c,s} - \text{end}_{c,s}, \forall c \in C, \forall s \in \{1 \dots nint\} \quad (6a)$$

$$\text{start}_{c,s} \leq \text{start}_{c,s+1}, \forall c \in C, \forall s \in \{1 \dots nint - 1\} \quad (7a)$$

$$\text{end}_{c,s} \leq \text{end}_{c,s+1}, \forall c \in C, \forall s \in \{1 \dots nint - 1\} \quad (8a)$$

$$\text{pen}_{o,t} \geq 0, \forall o \in O, \forall t \in T \quad (9)$$

$$x_{v,t} \in \{0,1\}, \forall v \in V, \forall t \in T \quad (10)$$

$$\text{start}_{c,s} \in \{0,1\}, \forall c \in C, \forall s \in \{1 \dots nint - 1\} \quad (11a)$$

$$\text{end}_{c,s} \in \{0,1\}, \forall c \in C, \forall s \in \{1 \dots nint - 1\} \quad (12a)$$

$$\text{bint}_{c,s} \in \{0,1\}, \forall c \in C, \forall s \in \{1 \dots nint - 1\} \quad (13)$$

Na secção seguinte vamos apresentar os resultados do novo modelo que considera a heurística.

Resultados Preliminares

O modelo proposto foi testado considerando os dados da procura do nosso parceiro industrial. Estes dados incluem 42 semanas de 2010 e 2011. Com estes dados foi criado um gerador aleatório, no Excel, para obtermos cenários semelhantes aos que ocorreram durante estas 42 semanas.

Nestes resultados testamos 6 instâncias aleatórias, que foram testadas considerando 5 intervalos para a heurística e um limite de tempo de execução de 30 minutos. Consideramos que este é um tempo razoável, uma vez que o planejamento é para ser efetuado antes do início do turno. Isto porque pode ser necessário fazer reajustamentos à sequência devido, por exemplo, a problemas de qualidade ou problemas com fornecedores. O número de intervalos é igual a 5, porque testes anteriores que avaliaram a sensibilidade deste parâmetro, revelaram que este valor traz bons resultados, em 30 minutos.

As instâncias criadas pelo gerador aleatório têm as características descritas na Tabela 1.

Na Tabela 1 existem opções que não têm procura porque as 6 instâncias são geradas considerando a procura de 42 semanas dos anos de 2010 e 2011, como já foi referido no início desta secção. Como a procura das opções com estas restrições de capacidade é baixa, as instâncias geradas não contiveram carros com estas opções.

Os resultados das diferentes instâncias foram testados considerando os três problemas seguintes:

1. Problema restrições de capacidade (considerando apenas a primeira linha da equação 2 na secção 4) – minimiza o número de violações das restrições de capacidade;
2. Problema restrições de capacidade + Carros especiais (considerando a primeira e segunda linhas da equação 2 na secção 4) – minimiza o número de violações das restrições de capacidade e tenta colocar os carros especiais nas primeiras posições da sequência;
3. Problema restrições de capacidade + Carros especiais + Dispersão (considerando o problema global equação 2 na secção 4) – minimiza o número de violações das restrições de capacidade, tenta colocar os carros especiais nas primeiras posições da sequência e minimiza a dispersão dos carros com a mesma cor.

Tabela 1: Características das Instâncias.

Características das Instâncias	Instância 1	Instância 2	Instância 3	Instância 4	Instância 5	Instância 6	
Número Carros	300						
Mix (número carros modelos A, B e C – A:B:C)	37:121:142	74:51:175	48:99:153	54:83:163	52:105:143	48:103:149	
Número Opções	14	14	14	15	14	15	
Número Variantes	115	103	113	121	102	109	
Número Cores	16	19	17	19	18	18	
Número Carros Especiais	79	52	69	65	53	80	
Taxa de Utilização	Opção 1 Restrição capacidade - 1:2	0,95	1,17	1,02	1,09	0,95	0,99
	Opção 2 Restrição capacidade - 1:2	0,37	0,74	0,48	0,54	0,52	0,48
	Opção 3 Restrição capacidade - 1:3	0,17	0,09	0,21	0,12	0,18	0,15
	Opção 4 Restrição capacidade - 1:3	0,28	0,23	0,20	0,39	0,19	0,23
	Opção 5 Restrição capacidade - 1:4	0,23	0,19	0,15	0,20	0,15	0,20
	Opção 6 Restrição capacidade - 1:4	0,43	0,17	0,35	0,17	0,17	0,09
	Opção 7 Restrição capacidade - 1:26	0,24	0,16	0,12	0,24	0,16	0,24
	Opção 8 Restrição capacidade - 1:6	0,08	0,14	0,06	0,10	0,00	0,10
	Opção 9 Restrição capacidade - 1:6	0,12	0,05	0,12	0,07	0,10	0,08
	Opção 10 Restrição capacidade - 1:3	0,16	0,11	0,13	0,16	0,04	0,19
	Opção 11 Restrição capacidade - 1:4	0,00	0,00	0,00	0,00	0,00	0,00
	Opção 12 Restrição capacidade - 1:16	0,08	0,02	0,11	0,05	0,05	0,08
	Opção 13 Restrição capacidade - 1:3	0,00	0,00	0,00	0,04	0,00	0,02
	Opção 14 Restrição capacidade - 1:6	0,00	0,00	0,00	0,00	0,01	0,00
	Opção 15 Restrição capacidade - 1:2	0,25	0,21	0,25	0,27	0,23	0,33
	Opção 16 Restrição capacidade - 1:2	0,06	0,05	0,05	0,05	0,04	0,03
Média	0,26	0,26	0,25	0,25	0,21	0,23	

O segundo e o terceiro problemas são hierárquicos e respeitam a seguinte figura.

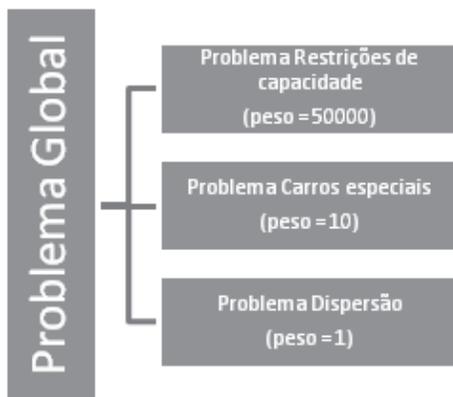


Figura 6: Abordagem hierárquica multiobjetivo.

Analisando a Figura 6 podemos verificar que a prioridade é respeitar as restrições de capacidade. A segunda prioridade é colocar os carros especiais nas primeiras posições da sequência e, a última, é minimizar a dispersão dos carros com a mesma cor. A escolha dos pesos a dar a cada sub-problema deve ser feita para assegurar que uma solução com mais violações de restrições de capacidade nunca vai ocorrer se for possível obter uma solução com menos violações, mesmo que essa solução seja pior em termos de dispersão de cor. O conjunto de valores indicado na Figura 6 assegura esse objetivo.

A hierarquização do modelo foi considerada tendo em conta os objetivos da empresa e da indústria automóvel em geral, tendo em conta que este modelo foi desenvolvido para, em primeiro lugar, responder às necessidades da montagem final.

Nas tabelas com os resultados finais (Tabela 2 e Tabela 3), apresentámos os resultados em termos de tempo, número de violações de restrições de capacidade, posição do último carro especial comparando com o número de carros especiais (valor entre parêntesis) e dispersão. Os valores de dispersão são comparáveis para qualquer problema que estejamos a analisar, porque são obtidos da mesma forma em todos os problemas, somando dispersão de cada cor. O valor do GAP, mencionado na Tabela 3, é calculado da seguinte forma.

$$GAP = \frac{\text{Valor actual da função objectivo} - \text{limite inferior função objectivo}}{\text{Valor actual da função objectivo}}$$

Note-se que o valor da função objetivo é diferente do valor da dispersão, e o GAP diz respeito ao valor da função objetivo.

De seguida vamos apresentar os resultados do problema das restrições de capacidade e do problema das restrições de capacidade e dos carros especiais na Tabela 2. Os resultados do problema global encontram-se na Tabela 3.

Tabela 2: Solução ótima para o problema das restrições de capacidade e para o problema das restrições de capacidade e dos carros especiais.

		Problema Restrições de capacidade			Problema Restrições de capacidade + Carros especiais			
		Solução Óptima			Solução Óptima			
		Tempo (s)	Violações de Capacidade	Dispersão	Tempo (s)	Violações de Capacidade	Posição último carro especial	Dispersão
Instância 1	Sem Heurística	9.25	0	3203	20.93	0	79 (79)	2336
	Com Heurística <i>nint</i> = 5							
Instância 2	Sem Heurística	9.56	49	4012	185.33	49	55 (52)	3317
	Com Heurística <i>nint</i> = 5							
Instância 3	Sem Heurística	18.35	5	3485	22.98	5	78 (69)	833
	Com Heurística <i>nint</i> = 5							
Instância 4	Sem Heurística	11.81	25	3759	115.43	25	65 (65)	2832
	Com Heurística <i>nint</i> = 5							
Instância 5	Sem Heurística	6.58	0	3997	88.42	0	53 (53)	3123
	Com Heurística <i>nint</i> = 5							
Instância 6	Sem Heurística	8.31	0	3544	3.68	0	85 (80)	2863
	Com Heurística <i>nint</i> = 5							

A Tabela 3 apresenta apenas um resultado tanto para o problema “Sem heurística” como para o problema “Com Heurística $nint=5$ ” porque para o problema das restrições de capacidade e para o problema das restrições de capacidade e dos carros especiais a heurística não é considerada. A heurística é apenas utilizada quando é considerado o problema de dispersão de cor.

As soluções encontradas e descritas na Tabela 3 para as Instâncias são soluções ótimas para o problema de minimização da função objetivo. Isto porque o tempo em segundos é sempre menor que 1800 segundos, o tempo máximo que demoraria a simulação. Assim, por exemplo, para a Instância 2 o número mínimo de violações de restrições de capacidade é 49.

Analisando os resultados das tabelas 2 e 3, podemos verificar que para a capacidade considerada de 300 carros por turno:

- O modelo resolve o problema das restrições de capacidade e o problema das restrições de capacidade e dos carros especiais em menos de 60 segundos, em média;
- O modelo das restrições de capacidade origina maiores valores de dispersão do que os restantes modelos testados como se pode verificar pela análise da Instância 3 onde o resultado ótimo do problema “Restrições de capacidade + carros especiais” é inferior ao resultado ótimo do problema “Restrições de capacidade + Carros especiais + Dispersão”. Isto justifica-se pelo facto de este resultado ótimo ser aproximado uma vez que foi considerada a Heurística com $nint=5$;
- A complexidade do problema aumenta quando introduzimos a dispersão na função objetivo porque o número de restrições e de variáveis aumenta. Isto é demonstrado pelo facto dos tempos para alcançar uma solução ótima aumentarem;
- Considerando o problema global, podemos afirmar que a abordagem heurística origina melhores resultados em termos de dispersão, em 30 segundos.

Tabela 3: Soluções para o problema das restrições de capacidade, dos carros especiais e da dispersão dos carros.

		Problema Restrições de capacidade + Carros especiais + Dispersão															
		Melhor solução aos 5 minutos				Melhor solução aos 20 minutos				Melhor solução aos 30 minutos				Solução Ótima			
Instância		Violações de Capacidade	Posição último carro especial	Dispersão	GAP	Violações de Capacidade	Posição último carro especial	Dispersão	GAP	Violações de Capacidade	Posição último carro especial	Dispersão	GAP	Tempo (s)	Violações de Capacidade	Posição último carro especial	Dispersão
1	Sem Heurística	0	79 (79)	3242	100%	0	79 (79)	3242	100%	0	79 (79)	3241	100%	Not found	Not found	Not found	Not found
	Com Heurística <i>nint</i> = 5	0	79 (79)	1291	10%	0	79 (79)	804	14%	0	79 (79)	802	10%	Not found	Not found	Not found	Not found
2	Sem Heurística	49	55 (52)	3665	77%	49	55 (52)	3665	77%	49	55 (52)	3665	77%	Not found	Not found	Not found	Not found
	Com Heurística <i>nint</i> = 5	49	55 (52)	2315	76%	49	55 (52)	2315	76%	49	55 (52)	2315	76%	Not found	Not found	Not found	Not found
3	Sem Heurística	5	78 (69)	Not found	Not found	5	78 (69)	3872	97%	5	78 (69)	3872	97%	Not found	Not found	Not found	Not found
	Com Heurística <i>nint</i> = 5	Optimal solution found before	Optimal solution found before	Optimal solution found before	Optimal solution found before	Optimal solution found before	Optimal solution found before	Optimal solution found before	Optimal solution found before	Optimal solution found before	Optimal solution found before	Optimal solution found before	Optimal solution found before	57.64	5	78 (69)	191
4	Sem Heurística	25	65 (65)	Not found	Not found	25	65 (65)	4468	87%	25	65 (65)	4468	87%	1118.17	25	65 (65)	165
	Com Heurística <i>nint</i> = 5	25	65 (65)	2548	85%	Optimal solution found before	1118.17	25	65 (65)	3							
5	Sem Heurística	0	53 (53)	4032	100%	0	53 (53)	4032	100%	0	53 (53)	4032	100%	Not found	Not found	Not found	Not found
	Com Heurística <i>nint</i> = 5	0	53 (53)	1120	27%	0	53 (53)	1120	27%	0	53 (53)	1120	27%	Not found	Not found	Not found	Not found
6	Sem Heurística	0	85 (80)	3923	100%	0	85 (80)	3923	100%	0	85 (80)	3923	100%	Not found	Not found	Not found	Not found
	Com Heurística <i>nint</i> = 5	0	85 (80)	2062	100%	0	85 (80)	975	3%	0	85 (80)	975	3%	Not found	Not found	Not found	Not found

- Analisando ao pormenor, verificamos que as instâncias 3 e 4 conseguem alcançar uma solução ótima em menos de 30 minutos. Estas instâncias têm em comum o facto de, devido à procura da opção 1, não ser possível encontrar uma solução ótima com zero violações de restrições de capacidade e com todos os carros especiais seguidos nos primeiros lugares da sequência. Desta forma, podemos concluir que a existência de violações de restrições de capacidade na solução que minimiza o número de violações de restrições de capacidade, e o facto dos carros especiais não poderem ser sequenciados seguidos nas primeiras posições da sequência, não torna o problema mais complexo. No entanto, mais testes são necessários para confirmar este facto.

Tendo como referência os resultados da primeira instância descritos na Tabela 3, a dispersão das soluções ao fim de 30 minutos para a abordagem exata e para a abordagem heurística, está representada na Figura 7 e na Figura 8.



Figura 7: Representação da dispersão – abordagem exacta, instância 1.



Figura 8: Representação da dispersão – abordagem heurística, instância 1.

Analisando a Figura 7 e a Figura 8 podemos afirmar que a abordagem heurística nos dá melhores soluções do que a abordagem exata. Na Figura 8, podem distinguir-se grupos maiores de cores. Maiores grupos de cor, significam que os carros da mesma cor estão mais próximos. Neste sentido, as soluções são melhores em termos de dispersão. Os carros pretos, cinzentos e brancos, por exemplo, estão mais próximos nesta figura.

Apesar de serem necessários mais testes, acreditamos que esta heurística nos permite poupar tempo e que é possível desenvolver uma ferramenta, baseada nestes modelos, que possa ser usada diariamente para sequenciação de veículos, no nosso parceiro industrial.

Conclusões

Através da análise dos resultados preliminares verificamos que estes modelos têm potencial para poderem ser utilizados para obtenção das seqüências diárias de produção, no nosso parceiro industrial. Estes modelos minimizam as violações das restrições de capacidade, a dispersão de cor e tentam sequenciar os carros especiais nos primeiros lugares da seqüência.

Estes resultados mostram, também, que soluções que minimizem as violações de restrições de capacidade e que tentam sequenciar os carros especiais nos primeiros lugares da seqüência podem ser obtidos em poucos segundos. Também podemos afirmar, que a heurística apresenta melhores resultados quando comparada com o modelo inicial.

Atualmente, o parceiro industrial deste projeto, considera apenas as restrições de capacidade no modelo de planeamento. No entanto, quando por razões de procura elevada, alguma restrição de capacidade tem de ser violada, o que a empresa faz é relaxar as restrições de capacidade. Por exemplo, se existirem 10 carros e 6 deles tiverem uma opção com a restrição de capacidade 1:2, isto significa, que teremos de violar esta restrição uma vez. O nosso parceiro industrial, neste momento, transforma a restrição de capacidade 1:2 em 1:3, o nosso modelo minimiza a violação de restrições de capacidade.

Em termos de outros impactos no nosso parceiro industrial acreditamos que este modelo nos permitirá alcançar um melhor planeamento do número de trabalhadores extra, quando são violadas restrições de capacidade. Isto porque esta é uma das respostas do modelo que nos informa do número de minutos extra necessários para completar a seqüência. Também acreditamos que este modelo pode melhorar a sincronização entre a pintura e a montagem final, reduzir o custo com solventes na área de pintura se a seqüência da montagem final for considerada para a criação de lotes de cor na pintura, e pode melhorar a logística e reduzir os custos com os fornecedores da montagem final,

permitindo a entrega de peças por cores e reduzindo os custos do rearranjo da sequência.

A redução do custo com solventes na área de pintura, já era tido em conta nos modelos desenvolvidos pelo Prandstetter (2005), no entanto, o conceito de dispersão de cor irá, também, permitir a redução do tempo que os carros passam no stock entre a pintura e a montagem final (WIP), uma vez que desta forma tentamos aproximar todos os carros da mesma cor.

Sendo assim, o uso da heurística juntamente com o modelo inicial, parece ser uma boa abordagem para sequenciar veículos, não só no nosso parceiro industrial, mas também, noutras indústrias de montagem de veículos.

Este artigo apenas contém resultados preliminares pelo que será necessário desenvolver mais testes para provar que efetivamente o modelo pode ser aplicado no nosso parceiro industrial. É necessário também um estudo aprofundado e uma análise de sensibilidade dos pesos a dar na função multiobjectivo. Uma análise mais aprofundada também é necessária para definir a melhor escolha quanto ao número de intervalos a utilizar na heurística.

REFERÊNCIAS

Cordeau, J.-F., Laporte, G., & Pasin, F. (2008). Iterated tabu search for the car sequencing problem. *European Journal of Operational Research*, 191, 945-956.

Drexler, A., & Kimms, A. (2001). Sequencing JIT Mixed-Model Assembly Lines Under Station-Load and Part-Usage Constraints. *Management Science*, 47(3), 480-491.

Duarte, C. S. R. (2014). *Single Line for Assembly Just-in-Sequence Multiple Models*. (PhD), University of Minho.

Estellon, B., Gardi, F., & Nouioua, K. (2008). Two local search approaches for solving real-life car sequencing problems. *European Journal of Operational Research*, 191, 928-944.

Fliedner, M., & Boysen, N. (2008). Solving the car sequencing problem via Branch & Bound. *European Journal of Operational Research*, 191, 1023-1042.

Gottlieb, J., Puchta, M., & Solnon, C. (2003). A study of greedy, local search and ant colony optimization approaches for car sequencing problems *Applications of Evolutionary Computing* (Vol. 2611/2003): Springer Berlin / Heidelberg.

Gravel, M., Gagné, C., & Price, W. L. (2005). Review and comparison of three methods for the solution of the car sequencing problem. *Journal of the Operational Research Society*, 56, 1287-1295.

Prandstetter, M. (2005). *Exact and heuristic methods for solving the Car Sequencing Problem*. (Master), Vienna University, Vienna.

Prandstetter, M., & Raidl, G. R. (2008). An integer linear programming approach and a hybrid variable neighborhood search for the car sequencing problem. *European Journal of Operational Research*, 191, 1004–1022.

Reis, R. J. d. O. (2007). *Solving the Car Sequencing Problem from a Multiobjective Perspective*. (Master), Instituto Superior Técnico, Lisbon.

Ribeiro, C. C., Aloise, D., Noronha, T. F., Rocha, C., & Urrutia, S. (2008). *European Journal of Operational Research*(191), 981-992.

Solnon, C. (2008). Combining two pheromone structures for solving the car sequencing problem with ant colony optimization. *European Journal of Operational Research*(191), 1043-1055.

Solnon, C., Cung, V. D., Nguyen, A., & Artigues, C. (2008). The car sequencing problem: Overview of state-of-the-art methods and industrial case-study of the ROADEF'2005 challenge problem. *European Journal of Operational Research*, 191, 912-927.

Warwick, T., & Tsang, E. P. K. (1995). Tackling Car Sequencing Problems Using a Generic Genetic Algorithm. *Evolutionary Computation*, 3(3), 267-298.

CASO 9

INTEGRAÇÃO E OTIMIZAÇÃO DE PROCESSOS NA INDÚSTRIA DE PRENSADOS DE MADEIRA

João Miranda

Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Portalegre
CERENA - Centro de Recursos Naturais e Ambiente, Instituto Superior Técnico
jlmiranda@estgp.pt, joaoluismiranda@ist.utl.pt

Ana Seita

Centro de Processos Químicos, Instituto Superior Técnico
ana.seita@ist.utl.pt

Cristina Fernandes

CERENA - Centro de Recursos Naturais e Ambiente
Instituto Superior Técnico, Universidade de Lisboa
mc.fernandes@ist.utl.pt

Henrique A. Matos

CERENA - Centro de Recursos Naturais e Ambiente
Instituto Superior Técnico, Universidade de Lisboa
henrimatos@ist.utl.pt

Clemente Pedro Nunes

CERENA - Centro de Recursos Naturais e Ambiente
Instituto Superior Técnico, Universidade de Lisboa
c.pedronunes@ist.utl.pt

RESUMO

A integração energética e a produção de utilidades são de vital importância nos processos químicos, em diferentes setores industriais, desde a petroquímica, aos têxteis, à cerâmica, entre outros. Os processos devem estar energeticamente balanceados, e a viabilidade econômica da unidade industrial depende da otimização simultânea dos fluxos energéticos e dos fluxos de produção. Este caso de estudo visa melhorar o desempenho econômico de unidades industriais de prensados de madeira, situadas em Tomar e na Nazaré, através da análise de diferentes configurações para produzir as utilidades quentes necessárias. Num enquadramento de sustentabilidade energética, as necessidades térmicas devem ser geradas numa fornalha de biomassa que recorre aos resíduos produzidos em diversas operações do processo (casca, aparas de placas e pó). As unidades industriais utilizavam o fuelóleo, como um dos recursos energéticos, em particular durante os meses de inverno. Neste estudo pretende-se avaliar diferentes configurações de um novo sistema energético das unidades industriais que recorram somente à queima de biomassa. Nos processos industriais as utilidades podem ser geradas de forma direta ou indireta, respetivamente, a partir dos gases de combustão ou através de um termofluido intermédio. Assim, os modelos de investigação operacional podem ajudar a decidir qual a melhor configuração para o sistema energético que permita satisfazer as necessidades de utilidades quentes que o processo requer. As alternativas foram comparadas atendendo aos equipamentos selecionados e seus atributos: a potência requerida para a fornalha e o consumo de biomassa, o número de permutadores/geradores de vapor e área total de transferência de calor, assim como a eficiência global do sistema energético. Para apoio à decisão, foi desenvolvido um modelo de Programação Não-Linear Inteira Mista (MINLP) que considera os detalhes operacionais. A introdução de variáveis binárias permitiu modelar a combinatória das diversas alternativas. Foram ainda, consideradas restrições de caráter prático e/ou baseadas na experiência real de unidades industriais desta natureza.

PALAVRAS-CHAVE

Otimização; Integração de processos; Eficiência energética, Biomassa, Geração de vapor.

1. Introdução

A economia portuguesa possui atualmente uma elevada intensidade energética e uma forte dependência da importação de energia primária. Assim, nos últimos anos, com a subida do preço do petróleo verificou-se uma perda de competitividade das empresas portuguesas. De acordo com um relatório interno da ADENE/DGEG em fevereiro de 2008, a evolução da intensidade energética em Portugal, expressa em termos do consumo de energia final (tep) por unidade de Produto Interno Bruto (em milhão de euros, a preços constantes de 2000) entre 1997 e 2007, divergiu significativamente da média europeia. A intensidade energética nacional, com valores acima dos 140, teve um crescimento desde 1997 a 2005, verificando-se somente um ligeiro decréscimo nos últimos anos. No entanto, na EU27 a intensidade energética teve durante este período um comportamento sempre decrescente e com valores sempre inferiores a 130 (Magueijo e outros, 2010).

Uma estratégia de utilização eficiente de energia, associada ao controlo e redução do consumo de energia para a mesma riqueza criada, é essencial para a competitividade económica de um país, como indica o *Livro Verde para a Eficiência Energética* (Direcção-Geral da Energia e dos Transportes / Comissão Europeia, 2005). O aumento da eficiência energética requer desenvolvimentos tecnológicos, bem como medidas que regulamentem o consumo energético e as emissões de gases com efeito de estufa e que estimulem em simultâneo a competitividade económica global das empresas portuguesas. Para além disso, os países que assinaram o protocolo de Quioto têm de criar estratégias para mitigar as emissões destes gases, através do uso de energias renováveis. A biomassa é um recurso renovável e neutro em carbono, e tem sido recentemente utilizado para produzir energia e biocombustíveis (Gassner e Marechal, 2009).

De entre as medidas de eficiência energética aplicáveis a um processo industrial, a Integração de Processos (IP) é uma das mais promissoras, podendo contribuir para uma redução de cerca de 1,8% no consumo energético até 2020 (Magueijo e outros, 2010). Esta metodologia é uma abordagem sistemática, utilizada na conceção e/ou otimização de processos produtivos, com vista a uma adequada articulação intra e interprocessual que

permita reduzir significativamente os custos de produção, satisfazendo políticas de utilização racional de energia e de matérias-primas, bem como, entre outros, a redução de efluentes e subprodutos e a minimização de emissões gasosas. O aumento do custo do petróleo, bem como as taxas associadas às emissões de CO₂, têm motivado estudos de novas alternativas para otimizar os sistemas energéticos na indústria.

Note-se que será necessário proceder a uma constante reavaliação dos diversos fatores económicos, para que um estudo de otimização e integração energética possa ser utilizado para decisões específicas. Numa primeira etapa é necessário identificar e quantificar o consumo de utilidades quentes em cada unidade do processo, bem como as fontes de energia disponíveis. Numa segunda etapa constrói-se uma superestrutura que representa as várias alternativas possíveis de produzir as utilidades quentes desejadas a partir de diferentes recursos energéticos, usando ou não um fluido intermédio para a transferência de calor. A função objetivo é definida em termos dos custos de equipamento, da área total de transferência e do número de equipamentos.

Soršak e outros (2002) dedicaram-se ao estudo e otimização de redes de permutadores de calor de diferentes tipos. Esse artigo descreve a síntese simultânea de redes de permutadores de calor integrados compreendendo diferente tipos de equipamentos de permuta. A superestrutura geral para integração energética de Yee e Grossmann (1990) foi utilizada e adaptada a vários tipos de permutadores de calor.

Drobez e outros (2010) estudaram a integração energética num processo de produção de biogás através de resíduos animais, com base num modelo MINLP. A solução apresentada indica que se o processo e as instalações auxiliares forem integrados energeticamente é possível poupar a quase totalidade das utilidades quentes e cerca de um terço das utilidades frias. Assim, a maior parte da eletricidade e do calor produzido na cogeração pode ser vendido como excedente para as redes de distribuição. Estas propostas podem melhorar o desempenho económico de uma empresa e reduzir significativamente o impacto ambiental, convertendo resíduos prejudiciais ao meio ambiente, orgânicos e de origem animal, em produtos de valor.

Mais recentemente, Damours e outros (2011) apresentaram um modelo de *Programação Linear Inteira Mista* (MILP) visando a integração da cadeia de aprovisionamento de produtos florestais e respetivas biorefinarias. Avaliaram também diferentes estratégias de integração (Flisberg e outros, 2011) numa empresa florestal do Chile. Este tipo de modelos surge como ferramenta de apoio à decisão, sendo naquele estudo claramente explicitados os atributos de validade e oportunidade que são cruciais na tomada de decisão.

Revisite-se também o estudo que visava satisfazer a procura crescente de fertilizantes químicos na Índia (Erlenkotter e Manne, 1968), instalando unidades de produção que suprissem tais necessidades. Ora, entre os decisores não se encontraram resistências quanto a diversas aproximações, tendo o estudo decorrido como previsto. Porém, das duas localizações potenciais que foram selecionadas, apenas uma se concretizou e num faseamento mais conservador do que o indicado no estudo.

Atendendo à mais-valia da modelação dos sistemas processuais, apresentam-se, na secção 2, os enquadramentos da integração e otimização de sistemas de energia e duma aplicação na indústria de prensados de madeira. Na secção 3, descrevem-se as abordagens sequencial e de otimização simultânea à IP. Na secção 4 descreve-se a modelação matemática da superestrutura de integração energética para o caso de estudo. Na secção 5, apresentam-se considerações de carácter tecnológico, aspetos computacionais e comparam-se resultados. Na secção 6, referem-se as principais conclusões e os desenvolvimentos futuros.

2. Integração de Processos numa unidade de prensados de madeira

2.1 A Integração Energética de Processos em Portugal

A *Integração Energética de Processos* pode a um nível mais elevado recorrer a componentes processuais que ajudem a aumentar a eficiência global do Processo, fazendo a integração global com todo o sistema de utilidades. A produção de energia a partir de fontes renováveis adquiriu especial importância no quadro da política energética da UE em virtude do agravamento da situação de forte dependência energética em relação ao petróleo pelo recente aumento do seu preço. O uso de biomassa florestal para

produção de energia elétrica constitui uma alternativa interessante à utilização de combustíveis de origem fóssil, de que Portugal é importador, e deve portanto ser explorada de forma sustentada e sempre que possível integrada numa estratégia adequada de cogeração, onde o aproveitamento da energia térmica permite valores de eficiência superiores a 85 %.

A aplicação das Metodologias de Integração de Processos tem um enquadramento português através do *Grupo Nacional para a Integração de Processos* (GNIP, <http://gnip.ist.utl.pt/>). Este grupo é um consórcio de instituições, empresas e universidades constituído em 7 de abril de 1995 com o objetivo genérico de divulgar e promover as metodologias de Integração de Processos em Portugal. Este consórcio manteve o apoio necessário à participação portuguesa no *Acordo Internacional para a Integração de Processos* no âmbito da AIE-Agência Internacional Energia (1995-2005). O GNIP participou ainda num Anexo sob o título «*Energy Systems Integration Between Society and Industry Including Cogeneration Systems and Power Plants*», que integrou ainda a Finlândia, Canadá e Suécia.

Atualmente o GNIP participa no Acordo da AIE denominado *IETS-Industrial Energy-related Technology & Systems* (IETS). Este acordo engloba vários anexos: Sistemas de Separação Energeticamente Eficientes (Anexo IX); Secagem Energeticamente Eficiente (Anexo X); Biorefinarias baseadas na Indústria (Anexo XI); Tecnologia de Membranas para a Separação Energeticamente Eficiente de Hidrocarbonetos (Anexo XII); Bombas de Calor Industriais (Anexo XIII) entre outros. Outra informação adicional pode ser obtida em <http://www.iea-industry.org/>. O GNIP tem nos últimos anos participado em alguns projetos de divulgação e aplicação das metodologias de IP. Diferentes empresas portuguesas foram já alvo de aplicação das metodologias de IP, podendo ser apontados os estudos no setor da Química de Base, Adubos, Têxtil, Celulose, Petroquímica e Madeira.

Nas últimas décadas, além dos trabalhos de investigação desenvolvidos pelas universidades, também os trabalhos de aplicação em unidades industriais ilustram a importância da obtenção de processos integrados e otimizados. O objetivo destes estudos é, não só definir a própria configuração/estrutura

processual, como também calcular os valores das variáveis importantes do processo.

Qualquer alteração de um equipamento, ou de características de uma corrente de processo (ex: temperatura de vapor, caudal, composição), propaga-se ao longo de todo o sistema em termos de operação e rentabilidade. Assim, a análise e compreensão global de todo o processo deverá ser um requisito mínimo para uma atuação de sucesso no incremento da competitividade na Indústria de Processos.

2.2 O sistema energético de uma unidade de prensados de madeira

Num processo de produção de prensados de madeira, as fibras são obtidas por processos termo-mecânicos, a partir de toros de madeira destrocados que são previamente aquecidos com vapor antes do desfibramento. Algumas unidades requerem uma atmosfera húmida para a produção dos painéis, que posteriormente são sujeitos a uma prensa e enviados para as câmaras de estabilização para absorção da humidade. Durante a compressão do painel, produz-se um licor negro que é enviado para uma unidade de evaporação. Noutras unidades, caso dos painéis de MDF, *Median Density Fiberboard*, após o desfibrador são adicionadas resinas e corantes. A fibra é seca, em contra corrente com ar quente, passando posteriormente por um *pendistor*, e por uma formadora de manta. Os painéis são produzidos a partir da compressão a frio e a quente da manta de fibras, tal como se apresenta no diagrama da Figura 1.

Assim, tanto no caso dos prensados de madeira como no caso da produção de MDF, verifica-se que estas fábricas têm um elevado consumo de utilidades quentes, maioritariamente vapor, produzido a partir da biomassa e do fuelóleo. Durante o inverno quando escasseia a biomassa e esta é mais húmida, o fuelóleo era usado para garantir o caudal de vapor necessário para satisfazer as necessidades do processo. Consequentemente, o custo final destes derivados da madeira é amplamente dependente da eficiência e dos combustíveis utilizados para produção de energia térmica.

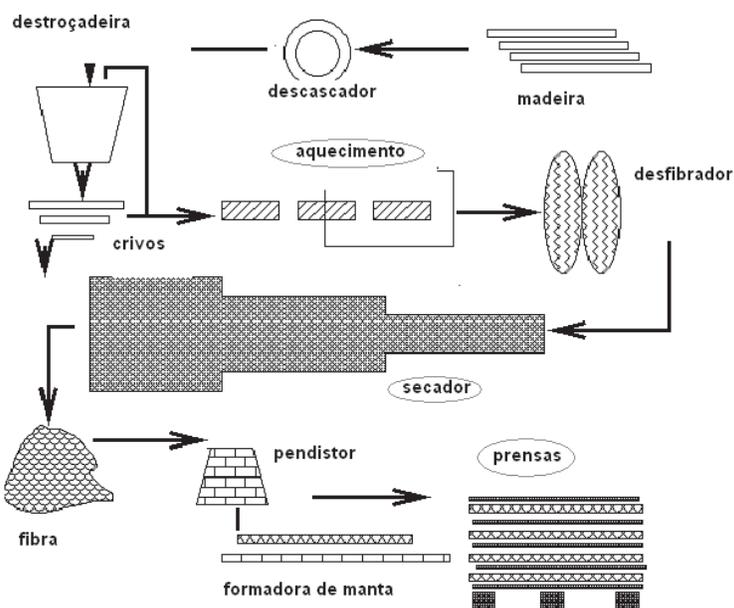


Figura 1 – Diagrama simplificado do processo de formação de MDF

Na Figura 2 apresenta-se o diagrama que descreve as necessidades de utilidades quentes de uma unidade de cartões prensados de madeira. Neste diagrama, o vapor é gerado em duas caldeiras que utilizam os gases provenientes da queima de biomassa e/ou fuelóleo numa fornalha de grelha. O forno queima serradura, aparas de cartão duro, casca e resíduos produzidos no processo. A caldeira principal gera vapor de 21 bar, que é usado nas prensas e nas câmaras térmicas, enquanto a outra caldeira produz vapor a 11 bar para o desfibramento da madeira destrocada, para as câmaras de humidificação e para a unidade de evaporação (a 2 bar). Note-se que embora os processos industriais em Tomar e Nazaré sejam diferentes, os sistemas energéticos eram inicialmente similares e a Figura 2 constitui uma boa aproximação a ambos os sistemas.

Neste estudo, pretende-se substituir os sistemas de produção de utilidades quentes da unidade de prensados e da unidade de MDF, definindo-se o tipo de equipamento, número de unidades de permuta e respetiva dimensão (área de

transferência), bem como selecionar a melhor configuração estrutural para a recuperação energética dos gases de combustão, de forma a garantir que:

- se utilize somente biomassa;
- em todas as alternativas propostas, apenas é necessário uma fornalha para gerar os gases de combustão que permitem posteriormente produzir as utilidades necessárias ao processo. Esta nova fornalha inclui uma grelha móvel capaz de queimar exclusivamente biomassa com humidade elevada bem como o pó gerado no processo.
- entre as várias alternativas o vapor pode ser gerado diretamente a partir dos gases de combustão ou através de fluido intermédio (termofluido).
- o balanço deve considerar aspetos económicos, como os custos de investimento em equipamentos de permuta energética, custos de utilidades quentes e frias.

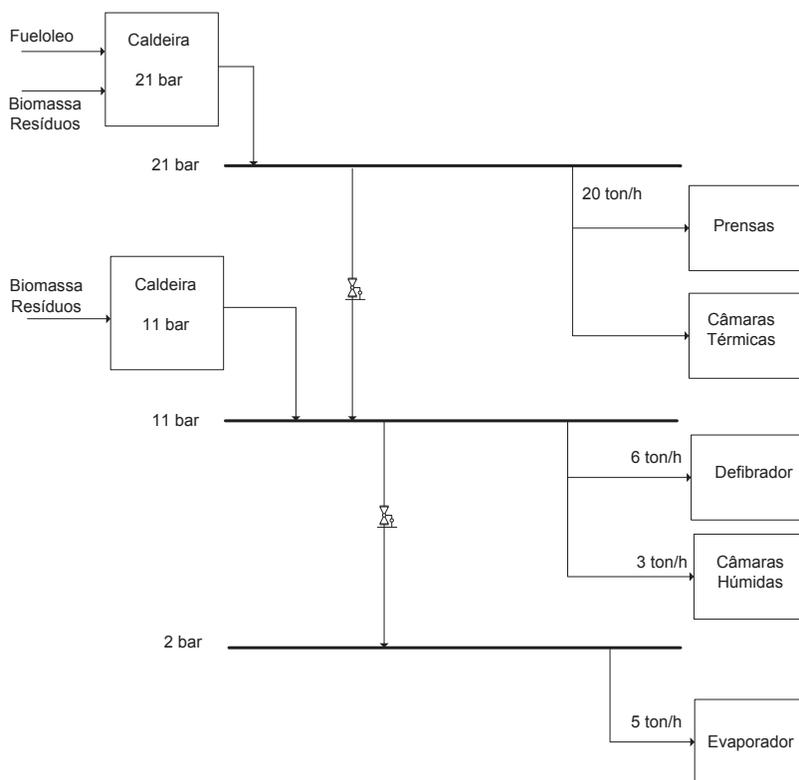


Figura 2 – Rede inicial das utilidades quentes (vapor) numa unidade de cartões prensados de madeira.

Na secção seguinte apresentam-se as abordagens que permitem a escolha da melhor alternativa económica, sendo o caso em estudo um problema típico de IP onde se pretende configurar a melhor rede de permutadores de calor.

3 Abordagens à modelação para a Integração de Processos

Nesta secção, visam-se as abordagens à modelação para a IP de forma a melhorar a sua contextualização, bem como para evidenciar as fortes ligações entre os métodos da Engenharia de Sistemas de Processos (*Process Systems Engineering*) e da Investigação Operacional. Nomeadamente, apresenta-se a analogia com o problema de Transporte e o seu caso especial de Transexpedição, ilustrando-se com um exemplo numérico simples. Note-se que os intervalos de temperatura e conteúdos energéticos da referida analogia podem ser utilizados na construção da superestrutura da rede de permutadores, tal como se descreve posteriormente na secção 4.

3.1 A abordagem sequencial e por otimização simultânea

A síntese de redes de permutadores de calor (*heat exchanger network synthesis*, HENS) é um problema bem conhecido da Engenharia de Sistemas de Processos (*Process Systems Engineering*), estando bem disseminadas as abordagens sequenciais que se baseiam no método do ponto de estrangulamento (*pinch point method*), como mostraram Linnhof e outros (1982). Nesta abordagem sequencial, minimiza-se o consumo de utilidades energéticas, em seguida minimiza-se o número de permutadores requeridos, e finalmente minimiza-se o custo total de equipamentos e energia.

Através deste tipo de abordagem sequencial, Seita e outros (2009) desenvolveram um estudo de integração energética numa unidade de produção de prensados de madeira de MDF. Após uma análise global da unidade industrial, aplicou-se o método do ponto de estrangulamento para estabelecer a rede de recuperação máxima (*Maximum Energy Recovery*, MER).

O método *pinch* funciona como uma primeira abordagem para a análise de integração energética entre as várias correntes de um processo. Esta metodologia permite determinar entre que correntes poderá ser realizada a integração, mas não permite a discussão e a avaliação económica dos vários

cenários estruturais alternativos, nem a discussão da sua viabilidade técnica-económica. Assim, a abordagem sequencial com base no método *pinch* pode conduzir a desvios face ao ótimo global uma vez que a qualidade da solução em cada fase depende da qualidade da solução obtida numa fase anterior. Então, a otimização simultânea do objetivo económico, balanceando o custo de equipamentos e o custo de utilidades, revela-se necessária. Deste modo, podem conjugar-se os vários aspetos do problema HENS, integrar-se de forma efetiva as interações entre os diversos subproblemas, formular custos fixos e custos variáveis, bem como promover a robustez do modelo e da solução.

Miranda e Casquilho (2007) apresentaram um modelo robusto de HENS, com variáveis binárias associadas à seleção de equipamentos e com cenários probabilísticos para emular a incerteza. O enquadramento em Programação Estocástica em Duas Fases (*Two-stage Stochastic Programming*) visava: i) a seleção dos equipamentos na primeira fase, por especificação das variáveis binárias $\{0, 1\}$; ii) após concretização dos cenários, a minimização dos custos de recurso em segunda fase visava a definição das variáveis contínuas das utilidades energéticas. No entanto, a resolução de tal modelo revelou grandes dificuldades, inerentes à complexidade do problema em questão.

A classificação de complexidade computacional de vários problemas de HENS foi estabelecida por Furman e Sahinidis (2001): enquanto os modelos de Programação Linear correspondiam a problemas da classe P, os modelos MILP por eles estudados correspondem a problemas NP-difíceis em sentido forte (*NP-hard in strong sense*). Tais estudos de complexidade computacional permitem também perspetivar esquemas de resolução que possam ser úteis perante a dificuldade de grandes instâncias destes problemas. A Tabela 1 apresenta sinoticamente algumas referências de interesse e, quando possível, a classificação de complexidade computacional dos problemas tratados.

3.2 A analogia com o problema de Transporte

Ora, na formulação de modelos são usuais as abordagens construtiva e por «template». A abordagem construtiva é mais fundamental, sendo comuns os procedimentos de carácter geral que visam a definição do objetivo a otimizar, a definição das variáveis de decisão e a especificação das restrições funcionais. A

abordagem por «template» utiliza modelos existentes que possam ser ajustados de forma suficiente à situação real. Além disso, a situação real pode exigir a combinação de formulações de diferentes tipos de problemas, quando não é possível que um só modelo represente os aspetos importantes da situação.

Tabela 1 - Sinopse de referências no âmbito da Integração de Processos

Ano	Autores	Assunto
1982	Linnhoff <i>e outros</i>	Análise termodinâmica a redes de integração energética
1983	Cerda e Westerberg	Modelação dos fluxos caloríficos em rede de transportes (<i>transportation</i>); problema NP-difícil em sentido forte para modelo MILP de permutas (<i>matches</i>)
1983	Papoulias e Grossmann	Modelação dos fluxos em rede de transexpedição (<i>transshipment</i>); problema P
1986	Saboo <i>e outros</i>	Implementação computacional dos modelos de <i>transshipment</i>
1990	Yee <i>e outros</i>	Otimização simultânea do sistema processual e da rede energética; problema NP-difícil em sentido forte
1994	Daichendt e Grossmann	Avaliação preliminar de alternativas utilizando modelos agregados
2001	Furman e Sahinidis	Estudo de complexidade computacional de problemas de HENS
2008	Relvas <i>e outros</i>	Aplicação do método <i>pinch</i> a redes de transferência de massa
2011	Martín e Grossmann	Otimização energética na produção de hidrogénio a partir de biomassa lenhocecelulósica

Esta abordagem por «template» permite agilizar a modelação do problema, existindo uma tipificação de problemas de otimização como sejam o conhecido problema de Transporte (*Transportation*) e os seus casos especiais (*Transshipment, Production Allocation, Assignment*). Por exemplo, Cerda e Westerberger (1983) recorrem a formulações de Transporte para minimização dos consumos energéticos em redes de permutadores, considerando: as correntes quentes como origens e os respetivos fluxos energéticos como disponibilidades; as correntes frias como destinos, com as correspondentes necessidades energéticas.

Tabela 2 Dados para analogia de Transporte

Corrente	FCp (MW/°C)	Tin(°C)	Tout(°C)	Q= - FCp (Tout-Tin)
H1	1,5	390	190	300
H2	1	290	190	100
C1	2	25	125	-200
C2	1	125	325	-200

Exemplifica-se a analogia com duas correntes quentes (H1 e H2) e duas correntes frias (C1 e C2). Na Tabela 2 apresentam-se os respeitantes conteúdos energéticos (FCp), temperaturas de entrada (Tin) e de saída (Tout), bem como as quantidades energéticas em permuta (Q). Então, obtém-se o usual grafo bipartido do problema de Transporte (Figura 3).

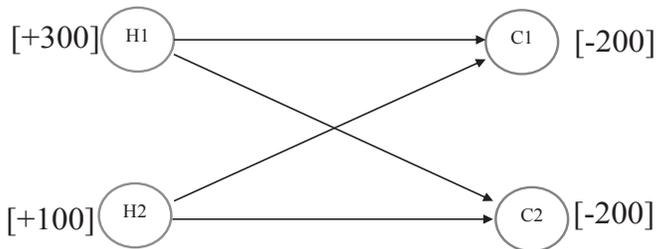


Figura 3 – Ilustração da analogia com o problema de Transporte

A generalização ao problema de Transexpedição (*Transshipment*) desenvolve-se considerando os intervalos de temperatura como entrepostos.

Tabela 3 - Conteúdos energéticos e intervalos de temperatura

Intervalos (em °C)	H1	H2	C1	C2
I - de 390 a 325	97,5			
II - de 325 a 290	52,5			35
III -de 290 a 190	150	100		100
IV -de 190 a 125				65
V - de 25 a 125			200	
SOMA	300	100	200	200

Na Tabela 3 apresentam-se os valores dos conteúdos energéticos em cada intervalo de temperaturas, para os valores do exemplo anterior. Por razões de simplicidade mas que não afetam o objeto deste exemplo ilustrativo, note-se que não foram consideradas as diferenças de temperatura entre as correntes quentes e frias. Obviamente que sem tal força motriz não haveria transferência de energia térmica!

Os valores em excesso em cada intervalo de temperatura são transferidos para o intervalo de temperatura imediatamente inferior, definindo assim os fluxos residuais de energia. Tais fluxos residuais são apresentados entre parêntesis na Figura 4, ilustrando a designada «cascata de calor» e que corresponde à estrutura de fluxos térmicos segundo a formulação em Transexpedição.

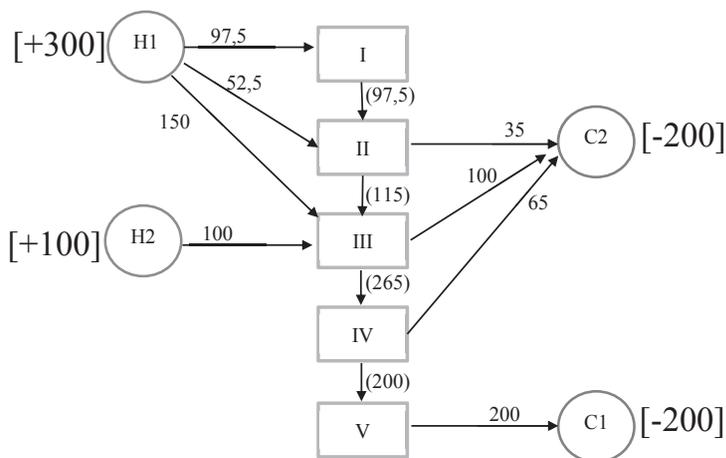


Figura 4 – Ilustração de superestrutura por analogia com o problema de Transexpedição

Este simples exemplo ilustra um caso em que as correntes são satisfeitas de forma interna sem necessidade de consumo externo de utilidades. Atente-se também que as temperaturas e os conteúdos energéticos das correntes são aqui parâmetros do problema, e não variáveis a otimizar.

4. Modelação do caso de estudo

Apresenta-se na Figura 5 o esquema da superestrutura desenvolvida para o caso de estudo industrial onde se considera queima de biomassa como

combustível e a respetiva produção de gases de combustão (a encarnado). Estes gases serão posteriormente aproveitados para produzir as utilidades quentes necessárias ao processo (vapor de alta, média e baixa pressão; a negro), diretamente e/ou através de um termofluido intermédio (a verde).

Nesta superestrutura conjugam-se as múltiplas e diversas alternativas a considerar em termos de equipamentos, permutadores, correntes diversas (gases de combustão, termofluidos quentes e frios, vapores a diferentes pressões), visando-se o aproveitamento energético dos gases de combustão produzidos na fornalha. A seleção da melhor configuração recorre ao desenvolvimento de um modelo de programação matemática que considere simultaneamente os fatores tecnológicos e os económicos.

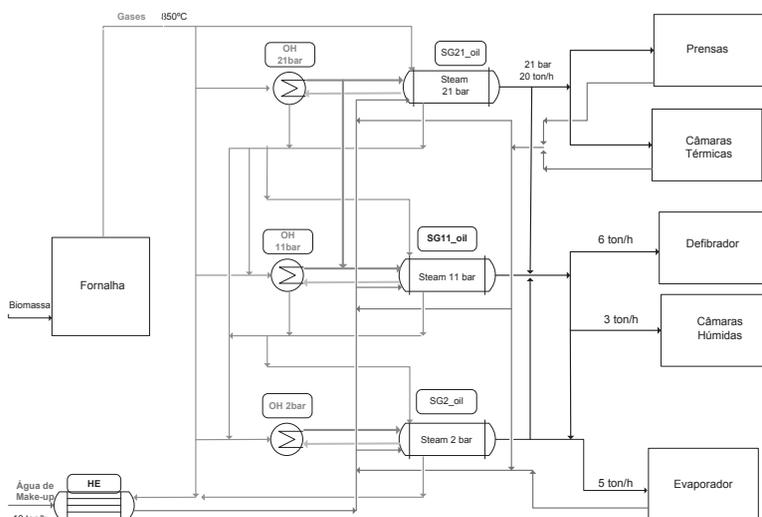


Figura 5 – Superestrutura para aproveitamento energético de gases de combustão

O modelo robusto de Miranda e Casquilho (2007) foi ajustado ao contexto deste caso de estudo. Tal modelo usa como «template» a superestrutura geral para integração energética de Yee e outros (1990), a qual considera simultaneamente os fatores de projeto: consumo de utilidades, número de permutadores, área total para transferência de energia térmica. Tal «template»

encontra-se disponível no ambiente GAMS (*General Algebraic Modeling System*) com a designação «SYNHEAT».

A superestrutura geral para integração energética (Yee e outros, 1990) apresenta as diversas possibilidades de permuta entre as correntes, mas não exige a definição do ponto de estrangulamento ou a partição em subredes. Considera-se o seguinte procedimento:

1. Definir o número de estágios («entrepostos»), selecionando-se o valor máximo entre o número de correntes frias e de correntes quentes, pois o número ótimo de permutadores não é usualmente superior ao número de correntes; numa abordagem mais rigorosa, deve considerar-se igual número de estágios e de intervalos de temperatura (Daichendt e Grossmann, 1994);
2. Em cada estágio, a corrente é dividida e direcionada para um possível permutador entre cada corrente quente e cada corrente fria; as correntes de saída dos permutadores são misturadas isotermicamente, definindo as correntes de entrada para o estágio seguinte;
3. As temperaturas à saída de cada estágio são tratadas como variáveis a otimizar, não se requerendo a pré-definição dos intervalos de temperatura.

Essa superestrutura geral contém algumas limitações. Não supõe a ocorrência de correntes de passagem (*by-pass*), exceto para situações específicas, como por exemplo:

- quando interessar manter uma força motriz elevada, ou
- onde seja necessário garantir a observância do ΔT_{\min} entre as correntes em permuta, para o que seria necessário diminuir o fluxo mássico da corrente fria através de um *by-pass* (geralmente, a adição de uma corrente deste tipo poderá beneficiar a rede em termos de números de unidades, mas com um aumento das áreas de transferência).

A superestrutura geral para integração energética é incorporada frequentemente em problemas mais complexos, conduzindo a uma formulação não linear (NLP) ou mesmo a uma formulação MINLP se se introduzirem variáveis binárias, como se apresenta em seguida.

O balanço energético global a cada corrente quente i considera as permutas com as correntes frias j ao longo dos diversos estágios k , Q_{ijk} , bem como a necessidade de se recorrer a uma corrente para arrefecimento, onde se permuta a quantidade, Q_{cu_i} :

$$\mathbf{h.1} \quad FCP_i(TIN_i - TOUT_i) = \sum_{k \in K} \sum_{j \in C} Q_{ijk} + Q_{cu_i} \quad , \quad \forall i$$

No primeiro membro define-se o conteúdo entálpico da corrente quente i , FCP_i , que no âmbito da otimização simultânea é tratado como variável, tal como as temperaturas de entrada TIN e saída $TOUT$ das correntes. Assim, obtém-se um grupo de equações não lineares, mas caso sejam definidos ou estimados os valores daqueles fluxos, as equações descritas nestas restrições reduzem-se a relações lineares.

De igual modo, tem-se o balanço energético para cada uma das correntes frias onde poderá ter de se recorrer a utilidades para aquecimento, Q_{hu_j} :

$$\mathbf{h.2} \quad FCP_j(TOUT_j - TIN_j) = \sum_{k \in K} \sum_{i \in H} Q_{ijk} + Q_{hu_j} \quad , \quad \forall j$$

O balanço energético em cada estágio k , tanto para as correntes quentes como para as correntes frias, origina também restrições não-lineares no primeiro membro:

$$\mathbf{h.3} \quad \begin{cases} FCP_i(T_{ik} - T_{i,k+1}) = \sum_{j \in C} Q_{ijk} \quad , \quad \forall i, k \\ FCP_j(T_{jk} - T_{j,k+1}) = \sum_{i \in H} Q_{ijk} \quad , \quad \forall j, k \end{cases}$$

As temperaturas de entrada das correntes quentes e frias, respetivamente, definem as temperaturas no primeiro e último estágio na superestrutura:

$$\mathbf{h.4} \quad \begin{cases} TIN_i = T_{i1} \quad , \quad \forall i \\ TIN_j = T_{j,K+1} \quad , \quad \forall j \end{cases}$$

Presume-se um perfil monotonamente decrescente das temperaturas ao longo dos diversos estágios:

$$\mathbf{h.5} \quad \begin{cases} T_{ik} \geq T_{i,k+1} \quad , \quad \forall i, k \\ T_{jk} \geq T_{j,k+1} \quad , \quad \forall j, k \end{cases}$$

Assume-se também a existência de folgas nas temperaturas à saída da superestrutura, as quais servem posteriormente para definir os requisitos de utilidades:

$$\mathbf{h.6} \quad \begin{cases} TOUT_i \leq T_{i,K+1}, \quad \forall i \\ TOUT_j \geq T_{j1}, \quad \forall j \end{cases}$$

O consumo de utilidades, quente Q_{hu_j} e fria Q_{cu_i} , é assim determinado pela diferença entre as temperaturas à saída da superestrutura e as temperaturas requeridas pelas correntes de processo:

$$\mathbf{h.7} \quad \begin{cases} FCp_i (T_{i,K+1} - TOUT_i) = Q_{cu_i}, \quad \forall i \\ FCp_j (TOUT_j - T_{j1}) = Q_{hu_j}, \quad \forall j \end{cases}$$

Repare-se que no exemplo ilustrativo da secção 3.2 os consumos de utilidades são nulos ($Q_{hu_j} = Q_{cu_i} = 0$): o problema típico de Transporte requer como condição necessária que a soma das disponibilidades nas origens iguale a soma das procuras nos destinos (e a soma de todos os fluxos). Então, as permutas energéticas, Q_{ijk} , entre correntes quentes i e frias j ao longo dos estágios k (intervalos de temperatura) podem ser lidas diretamente da Tabela 3.

Assim, a corrente quente $H1$ será a origem de 300 MW que serão distribuídos pelos três primeiros estágios (entrepósitos), concretamente, $Q_{1j1} = 97,5$ MW para o primeiro estágio, $Q_{1j2} = 52,5$ MW para o segundo e os restantes $Q_{1j3} = 150$ MW para o terceiro. E os valores das respetivas permutas para o quarto e quinto estágio são nulos ($Q_{1j4} = Q_{1j5} = 0$).

As permutas energéticas referentes às restantes correntes, $H2$, $C1$ e $C2$, podem obter-se de forma similar. Atente-se que para favorecer a leitura, na definição dos estágios deste exemplo ilustrativo não foram consideradas as diferenças mínimas de temperatura (ΔT_{min}) que devem existir para promover a transferência de energia térmica entre correntes quentes e frias!

A área de cada permutador é definida através da relação habitual, envolvendo a carga energética em permuta Q_{ij} , o coeficiente de transferência global U_{ij} , bem como a diferença de temperaturas logarítmica média (*log mean temperature difference*) $LMTD_{ij}$:

h.8

$$A_{ij} = \frac{Q_{ij}}{U_{ij}LMTD_{ij}}$$

onde

h.9

$$LMTD_{ij} = \frac{dT|_2 - dT|_1}{\ln(dT|_2) - \ln(dT|_1)}$$

Considerando a otimização simultânea do fluxograma e da rede de integração energética, a introdução de variáveis binárias possibilita a especificação das unidades de permuta, donde se pode estabelecer um conjunto de restrições de caráter lógico sobre as demais variáveis de decisão.

As trocas caloríficas entre as correntes quentes e frias, Q_{ijk} , e as utilidades de interesse, Q_{cu_i} e Q_{hu_j} , são tratadas como variáveis semicontínuas (ou se anulam, ou encontram-se delimitadas superiormente por Q^{Upp}), conforme as restrições formuladas com as variáveis binárias y_{ijk} , ycu_i , yhu_j :

h.10

$$\begin{cases} Q_{ijk} \leq Q^{Upp} \cdot y_{ijk} & , \quad \forall i, j, k \\ Q_{cu_i} \leq Q^{Upp} \cdot ycu_i & , \quad \forall i \\ Q_{hu_j} \leq Q^{Upp} \cdot yhu_j & , \quad \forall j \end{cases}$$

Para cada um dos diferentes tipos de permuta energética, a determinação da força motriz dT para a transferência energética também utiliza relações de índole lógica, estabelecendo-se (até por necessidade numérica!) uma abordagem de diferença mínima de temperaturas, ΔT_{min} :

h.11

$$\begin{cases} dT_{ijk} \leq T_{ik} - T_{jk} + dT^{Upp}(1 - y_{ijk}) \\ dT_{ij,k+1} \leq T_{i,k+1} - T_{j,k+1} + dT^{Upp}(1 - y_{ijk}) \end{cases} , \quad \forall i, j, k$$

h.12

$$\begin{cases} dTcu_i|_1 \leq T_{i,K+1} - TOUT_{cu} + dT^{Upp}(1 - ycu_i) \\ dTcu_i|_2 \leq TOUT_i - TIN_{cu} + dT^{Upp}(1 - ycu_i) \end{cases} , \quad \forall i$$

h.13

$$\begin{cases} dThu_j|_1 \leq TOUT_{hu} - T_{j1} + dT^{Upp}(1 - yhu_j) \\ dThu_j|_2 \leq TIN_{hu} - TOUT_j + dT^{Upp}(1 - yhu_j) \end{cases} , \quad \forall j$$

h.14

$$dT_{ijk}, dTcu_i, dThu_j \geq \Delta T_{min}, \quad \forall i, j, k$$

As temperaturas das correntes de saída das utilidades são parâmetros do problema, mas as diferenças de temperatura dT são variáveis de decisão. No caso específico da aplicação às utilidades, os índices 1 e 2 introduzidos referem as diferenças de temperaturas em cada uma das extremidades dos permutadores, ou seja, à entrada (1) e à saída (2) dos permutadores supostamente em contracorrente.

A função objetivo prevê a minimização dos custos das utilidades e dos custos para investimento, privilegiando também economias de escala através do expoente fracionário que valoriza as áreas de transferência. Lembra-se que as áreas de transferência dependem das quantidades energéticas permutadas, dos coeficientes de transferência globais, da diferença logarítmica média das temperaturas (*LMTD*), em conformidade com as equações h.8 e h.9. Com a introdução de variáveis binárias, também se consideram os custos fixos dos equipamentos. Assim, a função objetivo é da forma:

$$\begin{aligned}
 \text{h.15} \quad [\min] z = & Ccu \sum_{i \in H} Qcu_i + Chu \sum_{j \in C} Qhu_j \\
 & + CV_{ij} \sum_{i \in H} \sum_{j \in C} \sum_{k \in K} [A_{ijk}]^{\beta_{ij}} + CV_{cu,i} \sum_{i \in H} [Acu_i]^{\beta_{cu,i}} + CV_{hu,j} \sum_{j \in C} [Ahu_j]^{\beta_{hu,j}} \\
 & + \sum_{i \in H} \sum_{j \in C} \sum_{k \in K} CF_{ij} y_{ijk} + \sum_{i \in H} CF_{cu,i} ycu_i + \sum_{j \in C} CF_{hu,j} yhu_j
 \end{aligned}$$

Então, agrupando as relações h.1—h.15 obtém-se no grupo de relações h.16 (a-d) um modelo não linear com variáveis inteiras (binárias), o modelo *minlp_simsup*. Tal modelo avalia os custos de utilidades e os custos de investimento, de natureza fixa ou variável, e em simultâneo define a rede ótima de permutadores de calor a partir da superestrutura:

Modelo *minlp_simsup*:

$$\begin{aligned}
 \text{h.16-a} \quad [\min] z = & Ccu \sum_{i \in H} Qcu_i + Chu \sum_{j \in C} Qhu_j \\
 & + CV_{ij} \sum_{i \in H} \sum_{j \in C} \sum_{k \in K} [A_{ijk}]^{\beta_{ij}} + CV_{cu,i} \sum_{i \in H} [Acu_i]^{\beta_{cu,i}} + CV_{hu,j} \sum_{j \in C} [Ahu_j]^{\beta_{hu,j}} \\
 & + \sum_{i \in H} \sum_{j \in C} \sum_{k \in K} CF_{ij} y_{ijk} + \sum_{i \in H} CF_{cu,i} ycu_i + \sum_{j \in C} CF_{hu,j} yhu_j
 \end{aligned}$$

sujeito a,

Relações *h.1 — h.14*

$$\mathbf{h.16-b} \quad Q_{cu_i}, Q_{hu_j}, Q_{ijk}, A_{cu_i}, A_{hu_j}, A_{ijk}, T_{ik}, T_{jk} \geq 0, \quad \forall i, j, k$$

$$\mathbf{h.16-c} \quad FC_{p_i}, FC_{p_j}, TIN_i, TOUT_i, TIN_j, TOUT_j \geq 0, \quad \forall i, j$$

$$\mathbf{h.16-d} \quad y_{cu_i}, y_{hu_j}, y_{ijk} \in \{0,1\}, \quad \forall i, j, k$$

Embora os fluxos e temperaturas das correntes de processo sejam considerados parâmetros nas analogias de Transporte e Transexpedição, tais fluxos e temperaturas são variáveis a determinar no âmbito da otimização simultânea de todo o sistema energético, pelo que o modelo se torna não-linear. Repare-se que:

- Caso não se incorporem na função objetivo h.15 os custos exponenciais referentes às áreas de transferência, obtém-se um submodelo MILP.
- Caso não se considerem as parcelas referentes às variáveis binárias em h.15, bem como se especifiquem as relações de índole lógica h.10—h.14, obter-se-ia uma formulação NLP. Esse submodelo de NLP seria não-convexo, pois várias restrições apresentam termos bilineares. A configuração ótima da rede de permutadores viria implicitamente da leitura dos valores não-nulos para as áreas, bem como dos valores para as transferências energéticas.
- Se for definido um nível específico de recuperação energética, através da fixação do nível de utilidades requeridas pela rede, o problema reduz-se à minimização da área de transferência.
- Caso ocorra a fixação dos valores dos fluxos FC_p , obtém-se um modelo onde ocorrem apenas restrições lineares, revelando-se as não-linearidades somente na função objetivo. A resolução é mais simples, pois não só o espaço de procura das soluções possíveis será convexo, como a dimensionalidade do problema será menor.

Estas alterações, generalizações ou simplificações, são importantes para a resolução de problemas MINLP de grande dimensão, pois permitem obter resultados parciais e configurações de boa qualidade para a rede de integração energética. Embora tais alterações possam conduzir a soluções subótimas, promovem a análise, a comparação e delimitam a procura de configurações melhores. Caso tal procura seja sistemática e satisfaça os procedimentos de

decomposição hierárquica de Daichendt e Grossmann (1997), para além de se evitar a resolução de um problema MINLP de grande dimensão, também a otimalidade dum configuração pode ser assegurada através dum esquema com características similares ao branch-and-bound.

Considerando o problema detalhado para a síntese do subsistema de integração energética, observa-se a não-convexidade do modelo MINLP proposto, o que agrava ainda mais a dificuldade inerente aos modelos MINLP de grande dimensão. Têm sido desenvolvidos esforços para melhorar a eficiência dos esquemas de resolução, como sejam a introdução de modelos agregados e convexos, ou a incorporação de aspetos de índole termodinâmica, visando reduzir as dificuldades na fase inicial do cálculo, mas tendo o cuidado de preservar a otimalidade da solução.

5 Aspetos de Otimização e análise de resultados

5.1 Considerações de carácter tecnológico exigidas na implementação em ambiente real

Na formulação do problema foi necessário ter em consideração algumas aproximações e garantir alguns parâmetros tecnológicos de funcionamento:

- 1) A diferença de temperatura mínima (ΔT_{min}) entre as correntes quentes e as correntes frias é de 30°C, exceto no primeiro permutador de calor que usa os gases provenientes da fornalha onde o ΔT_{min} é de 90°C.
- 2) A variação de temperatura do termofluido nos permutadores é de 40°C.
- 3) O termofluido para produção de vapor a 21 bar funciona de 280°C a 240°C, enquanto que o termofluido para produção de vapor a 11 bar funciona entre 250°C e 210°C.
- 4) Considerou-se o poder calorífico inferior da biomassa como sendo 14374 kJ/kg, nas condições médias de humidade admitidas à fornalha.
- 5) Os custos dos permutadores de calor foram estimados admitindo permutadores de caixa e tubo e a função da área total foi estimada através dos coeficientes globais de transferência de calor listados na Tabela 4.

Tabela 4 -Coeficientes globais de transferência de calor

	U, Wm ⁻² K ⁻¹
Vapor-óleo	300
Óleo-gases	100
Gases-água	120

5.2 A implementação computacional

Dada a dificuldade de resolução dos modelos MINLP de grande dimensão, associada com a ocorrência de variáveis binárias e de funções não-convexas, visou-se diminuir o espaço de soluções possíveis, ou seja, reduzir-se o espaço de procura para o problema MINLP original. Obtém-se assim o problema MINLP-reduzido por introdução de restrições pertinentes, valores de corte (*upper bounds*), eliminação de configurações não-possíveis, fixação ou delimitação de variáveis binárias, como se descreve em seguida.

Consideraram-se vários aspetos de interesse (Daichendt e Grossmann, 1994) nomeadamente os relacionados com:

- O carácter de transexpedição (*transshipment*) dos fluxos caloríficos— potencia a determinação do número total K de estágios a considerar na superestrutura. Para tal, consideram-se os sucessivos intervalos entre as temperaturas de entrada das correntes de processo, quentes e frias, bem como o valor desejado para ΔT_{min} ; ou seja,

$$\text{h.17} \quad \begin{cases} Th_k \in \{TIN_i, TIN_j + \Delta T_{min}\} \\ Tc_k \in \{TIN_j, TIN_i - \Delta T_{min}\}, \quad \forall k \\ Th_k - Tc_k = \Delta T_{min} \end{cases}$$

- A localização dos estágios (entrepósitos) onde é possível considerar a ocorrência de permutas energéticas— para tal assinalando os intervalos de temperatura onde cada par (i,j) de correntes quente e fria se encontram presentes.

As correntes de utilidades são tratadas como correntes de processo, respetivamente, quentes ou frias, generalizando o seu uso ao longo de todos os estágios onde sejam necessárias. Através de uma análise preliminar às trocas

entre correntes, eliminam-se permutas impossíveis, impõem-se permutas necessárias ou conjugam-se as diversas alternativas. O procedimento para avaliação preliminar das alternativas (*preliminar screening procedure*) envolve:

- 1) A formulação de um modelo agregado, desenvolvido a partir do modelo MINLP original que é simplificado de forma a focalizar um dos seus subproblemas; por exemplo, evitando não-convexidades e obtendo um modelo MILP; ou ainda, fixando variáveis binárias e tratando um modelo NLP ou LP.
- 2) A determinação de uma solução inicial possível, desenvolvendo uma configuração-base para o subsistema energético, que permita o estabelecimento de um valor de corte (*upper bound*) para o subproblema de minimização dos custos.
- 3) A análise preliminar, utilizando o modelo agregado para delimitar o subconjunto de configurações que apresentem um valor-objetivo não superior ao custo daquela configuração-base.
- 4) A eliminação de configurações da superestrutura do problema MINLP original, obtendo-se uma superestrutura reduzida, à qual corresponde um espaço de procura também reduzido.
- 5) A resolução do problema MINLP com espaço de procura reduzido.

Quanto a aspetos do cálculo computacional, a igualdade h.9 padece de dificuldades numéricas quando as diferenças de temperatura dT nas extremidades inicial e final do permutador em contracorrente forem similares, originando valores próximos de zero no denominador da fração em h.9. Nesse caso, aplicam-se as aproximações de Chen [1987] e Paterson [1984], respetivamente:

h.18

$$LMTD_{ij} \approx \sqrt[3]{dT|_2 \cdot dT|_1 \cdot \left(\frac{dT|_2 + dT|_1}{2} \right)}$$

$$\approx \frac{2}{3} \sqrt{dT|_2 \cdot dT|_1} + \frac{1}{3} \left(\frac{dT|_2 + dT|_1}{2} \right)$$

Os aspetos específicos da rede são concretizados através da fixação ou delimitação de variáveis, ou mesmo adicionando novas restrições. Nomeadamente:

- As permutas proibidas ou requeridas configuram-se diretamente, forçando as variáveis binárias para os valores zero ou um, respetivamente.
- As restrições de projeto são diretamente incorporadas no modelo da superestrutura: por exemplo, uma permuta proibida (*forbidden*) é formulada anulando os valores das variáveis associadas, das quantidades energéticas permutadas e das áreas de transferência respetivas, ou mesmo igualando as temperaturas de entrada e saída das correntes nos permutadores proibidos.
- O número total de unidades de permuta (*exchangers*) é fixado ou delimitado superiormente através da soma das variáveis binárias,

h.19

$$\sum_{i \in H} \sum_{j \in C} \sum_{k \in K} y_{ijk} + \sum_{i \in H} ycu_i + \sum_{j \in C} yhu_j \leq Nexch^{Upp}$$

- O número de partições das correntes é controlado delimitando o número de permutas em cada estágio; também pode não ser permitida qualquer partição, caso se requeiram permutas únicas para cada corrente em cada estágio:

h.20

$$\begin{cases} \sum_{i \in H} y_{ijk} \leq 1, & \forall j, k \\ \sum_{j \in C} y_{ijk} \leq 1, & \forall i, k \end{cases}$$

Note-se que a introdução de constrangimentos, embora providencie uma aproximação mais direta à situação real, origina uma perda de qualidade no objetivo económico. Deste modo, o modelo agregado ou o modelo reduzido conduzem sempre a configurações subótimas, as quais são úteis para promover a procura da melhor configuração, bem como para eliminar todas as configurações inferiores, num esquema sistemático e análogo ao branch-and-bound.

Por outro lado, a consideração de um maior número de estágios na superestrutura permite a duplicação de permutas entre correntes e a utilização de um número maior de permutadores. Porém, o número de variáveis e

restrições aumenta, o que se traduz num crescimento significativo do esforço computacional.

5.3 Análise de resultados

Usando as abordagens descritas analisou-se o sistema energético das duas unidades fabris localizadas em Tomar e na Nazaré. A abordagem sequencial baseada no método *pinch* foi aplicada em ambas as unidades. A abordagem de otimização simultânea foi utilizada apenas na análise do sistema energético de Tomar, não sendo aplicada na unidade da Nazaré.

A análise realizada à unidade da Nazaré, permitiu a implementação industrial de um novo sistema energético que conduziu a decréscimos no consumo energético específico e no custo de operação específico de 24% e 29%, respetivamente, conforme reportaram Seita e outros (2009), obtendo-se também um menor impacto ambiental dos gases de combustão. Tal sistema é apresentado no diagrama da Figura 6 e corresponde à vista apresentada na Figura 7.

Nesta instalação da Nazaré, a configuração energética escolhida e implementada após o estudo de Integração de Processos é constituída por uma fornalha que produz os gases de combustão a partir apenas da biomassa e de alguns resíduos do processo.

Estes efluentes quentes vão aquecer um termofluido num permutador de calor. Os gases de combustão após esta troca de calor vão ainda ser usados no secador para redução do teor da humidade das fibras. O termofluido quente vai gerar posteriormente vapor de média pressão e água quente sob pressão.

Esta alternativa exigiu, para além da fornalha, três equipamentos de transferência de calor (aquecedor de óleo por serpentina, gerador de vapor de alta pressão e permutador de calor para água quente). Na Figura 7, apresenta-se uma vista da câmara de combustão (fornalha) e da unidade de recuperação de energia da fábrica de prensados de madeira da Nazaré, onde são aproveitados resíduos de madeira e biomassa para aquecimento do termofluido intermédio utilizado no aproveitamento térmico.

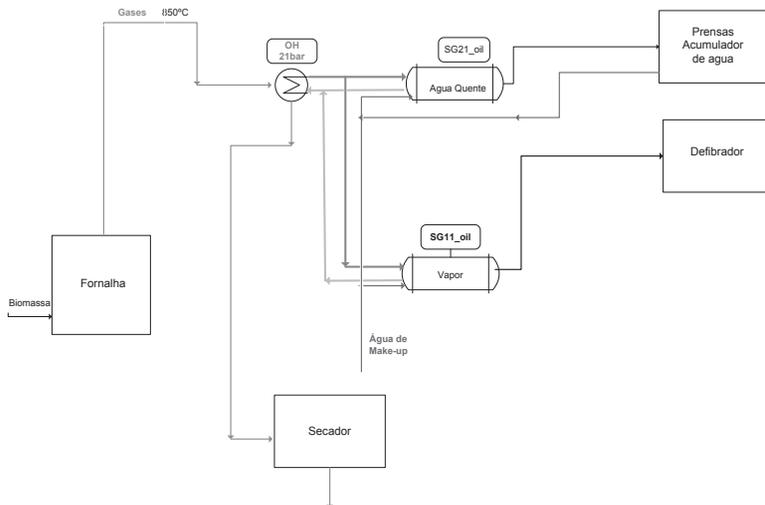


Figura 6 – Aproveitamento energético na instalação da Nazaré



Figura 7 – Vista da câmara de combustão na unidade industrial de prensados de madeira da Nazaré (fonte: Seita e outros, 2009)

Para a unidade de Tomar, apresenta-se na Figura 8 o diagrama correspondente à melhor configuração obtida no âmbito da avaliação preliminar das alternativas (*preliminar screening procedure*). Neste cenário o vapor de mais alta pressão é produzido indiretamente através do termofluido, enquanto as restantes utilidades quentes são obtidas diretamente dos gases. Esta alternativa requer quatro equipamentos de transferência de calor com uma

área estimada de cerca de 1700 m². Os gases na saída da chaminé encontram-se a 118°C e a eficiência energética global é de 94%, determinando assim um consumo de biomassa da ordem das 5 ton/h.

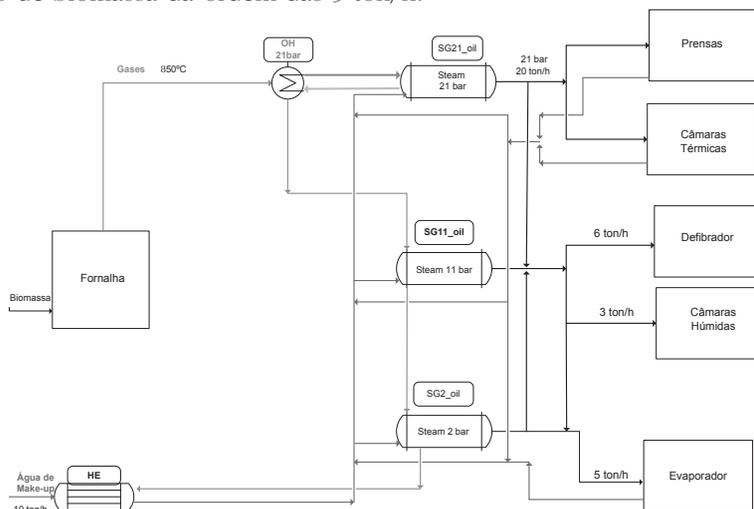


Figura 8 – Alternativa proposta para o estudo de Tomar com base na otimização simultânea.

Utilizou-se o ambiente GAMS, tendo sido seleccionado o solver DICOPT, específico para MINLP. No âmbito da avaliação preliminar de alternativas, atendendo a restrições adicionais e à fixação de variáveis que permitem reduzir o espaço de procura, as versões do modelo *minlp_simsup* apresentam tempos de execução diminutos (inferiores a 10 segundos, em computador portátil ASUS-F3JC, com processador Intel Core2 T5500 a 1.55GHz e 2GB de RAM).

Em termos de comparação, como refere a Tabela 5, a abordagem sequencial baseada no método *pinch* e a avaliação preliminar da abordagem simultânea aplicadas ao sistema energético de Tomar conduzem à mesma qualidade de configurações. No entanto é bem conhecido que a avaliação preliminar de alternativas é apenas uma fase inicial, servindo de referência para melhorias subsequentes e visando soluções e configurações de superior qualidade.

Tabela 5 – Matriz comparativa de abordagens.

	Nazaré	Tomar
Abordagem sequencial	Conduziu a um novo sistema de reaproveitamento energético!	Qualidade inferior, comparável a uma avaliação preliminar por MINLP.
Abordagem simultânea	Abordagem não aplicada.	Qualidade superior, mas desenvolvimentos descontinuados.

A abordagem através da modelação MINLP considera simultaneamente o consumo energético, bem como o custo de investimento em equipamentos (número e dimensão em área de transferência). Deste modo, perspetivam-se melhorias significativas relativamente à abordagem sequencial, onde a qualidade da solução para a escolha do equipamento depende da qualidade da solução da fase anterior relativa à recuperação de energia, e a dinâmica dos fatores económicos que lhes estão associados não é tratada. Ora, atente-se que a decisão de implementar uma unidade industrial depende não só da validade técnica como da oportunidade económica, pelo que é necessário proceder a uma constante apreciação dos diversos fatores económicos.

6 Conclusões

No âmbito da otimização e integração de processos, enquadrrou-se um estudo na indústria de prensados de madeira. Este estudo constitui um elemento de apoio à decisão, a partir do qual serão detalhadas os projetos de integração energética, e cuja funcionalidade não depende diretamente da sofisticação matemática e computacional.

De facto, as duas abordagens à síntese de redes de permutadores de calor conduziram a situações distintas na Nazaré e em Tomar, embora os resultados numéricos sejam similares na parte em que podem ser comparáveis. No caso da Nazaré aplicou-se apenas a abordagem sequencial e implantou-se um novo sistema de aproveitamento energético, enquanto que em Tomar se aplicaram ambas as abordagens mas não ocorreram desenvolvimentos subsequentes. Há ainda que referir que:

- A abordagem sequencial é baseada no método pinch, o qual se encontra bem estabelecido e reconhecido no meio industrial. A concretização de um maior aproveitamento energético dos gases de combustão na unidade da Nazaré possibilitou a redução de custos de energia primária, a redução de custos de operação, além de um menor impacto ambiental.
- A abordagem de otimização simultânea considera o consumo de utilidades e o custo de investimento, isto é, estabelece-se o compromisso entre o número de permutadores e a área total de transferência de energia em função da da recuperação de energia e da minimização das utilidades. Esta abordagem conduz a um modelo MINLP de resolução difícil, com aspetos computacionais exigentes, mas os respetivos resultados não foram concretizados em Tomar.
- Embora os processos industriais sejam diferentes nas duas unidades de prensados de madeira, o sistema energético em Tomar era similar ao da Nazaré. Para a unidade de Tomar, as configurações propostas em ambas as abordagens também foram algo semelhantes, pelo que os resultados de ambas as abordagens vieram a coincidir.

Os resultados obtidos servem de premissas para um projeto específico relativamente aos tipos de equipamento, dimensões e faseamento. Os resultados do estudo não foram implementados na unidade de Tomar pois, independentemente da qualidade das soluções e do modelo de programação matemática, não podemos negligenciar as dificuldades inerentes à aplicação de recursos num período económico em que tais recursos estão deveras limitados.

Na Nazaré, a utilização da fornalha encontra-se atualmente a cerca de metade do seu valor nominal, o que poderia ter sido evitado caso se tivesse utilizado um modelo de otimização simultânea com penalização do excesso de capacidades. Uma alternativa possível para uma utilização superior da fornalha considera a implantação de uma unidade para a produção de energia elétrica. Atendendo à validade e oportunidade de tal alternativa, a existência de um modelo que considere a otimização dos investimentos, em Programação Estocástica a Duas Fases ou até multifase, pode contribuir para a robustez da decisão e constitui um novo desafio.

Nomenclatura

Índices e conjuntos

- C – Conjunto de correntes frias, j (*Cold*);
- H – Conjunto de correntes quentes, i (*Hot*);
- i – Corrente quente, i ;
- j – Corrente fria, j ;
- k – Intervalo de temperatura (estágio da superestrutura), k ;
- K – Conjunto dos intervalos de temperatura (estágios), k ;

Parâmetros

- ΔT_{min} – Diferença mínima de temperaturas;
- $Ccu(j)$ – Custo unitário da utilidade fria, j ;
- CF – Custos de investimento de natureza fixa, referentes a permutadores entre correntes quentes e frias (ij), ou equipamentos de utilidades quentes (hu) ou frias (cu);
- $Chu(i)$ – Custo unitário da utilidade quente, i ;
- CV – Custos de investimento de natureza variável, para a permutadores entre correntes quentes e frias (ij), equipamentos de utilidades quentes (hu) ou frias (cu);
- dT^{upp} – Limite superior (*upper bound*) para a diferença de temperatura entre correntes quentes e frias;
- FCp – Conteúdos caloríficos da correntes, quentes i ou frias j ;
- NC – Número de correntes frias j (*Cold*);
- $NExch^{upp}$ – Limite superior (*upper bound*) para o número de unidades de permuta;
- NH – Número de correntes quentes i (*Hot*);
- NK – Número de intervalos de temperatura (estágios da superestrutura), k ;
- Q^{upp} – Limite superior (*upper bound*) para a energia permutada entre correntes quentes e frias;
- TIN – Temperaturas de entrada das correntes, quentes i ou frias j ;
- $TOUT$ – Temperaturas de saída das correntes, quentes i ou frias j ;
- U_{ij} – Coeficiente global de transferência entre a corrente quente i e fria j ;

Variáveis

- Acu_i – Área de transferência para o arrefecimento da corrente quente i pelas utilidades frias;
- Ahu_j – Área de transferência para o aquecimento da corrente fria j pelas utilidades quentes;
- A_{ij} – Área de transferência para a permuta entre a corrente quente i e a corrente fria j ;
- dT – Diferenças de temperatura entre correntes quentes e frias;
- $dTcu_i$ – Diferença de temperatura no arrefecimento da corrente quente i pelas utilidades frias;
- $dThu_j$ – Diferença de temperatura no aquecimento da corrente fria j pelas utilidades quentes;
- dT_{ij} – Diferença de temperatura na permuta entre a corrente quente i e a corrente fria j ;
- $LMTD_{ij}$ – Diferença de temperatura logarítmica média (*log mean temperature difference*) na permuta entre as correntes i e j ;

- Q_{cu_i} – Consumo de utilidades frias no arrefecimento da corrente quente i ;
- Q_{hu_j} – Consumo de utilidades quentes no aquecimento da corrente fria j ;
- Q_{ij} – Energia permutada entre a corrente quente i e a corrente fria j ;
- Q_{ijk} – Energia permutada entre cada corrente quente i e as correntes frias j ao longo dos diversos estágios k ;
- T – Temperaturas das correntes quentes i , correntes frias j , dos diversos estágios k ;
- y_{cu_i} – Variáveis binárias referentes ao arrefecimento da corrente quente i por utilidades frias;
- y_{hu_j} – Variáveis binárias referentes ao aquecimento da corrente fria j por utilidades quentes;
- y_{ijk} – Variáveis binárias para a permuta entre a corrente quente i e a correntes fria j no estágio k ;

Agradecimentos

Os autores agradecem os dados fornecidos pelas empresas do Grupo Investwood situadas em Tomar e Famalicão da Nazaré, nomeadamente ao Engenheiro Costa Cabral e Engenheira Alexandra Gouveia pelo apoio técnico fornecido na recolha de dados. Os autores também agradecem aos revisores anónimos, cujas sugestões muito contribuíram para valorizar este artigo.

REFERÊNCIAS

- ADENE/DGEG, *Relatório Interno*, Fevereiro de 2008
- Cerda, J., Westerberg, A., "Synthesizing Heat Exchanger Networks Having Restricted Stream/Stream Matches Using Transportation Problem Formulations", *Chemical Engineering Science*, Vol. 38, 1983, 1723-1740
- Daichendt, M., Grossmann, I.E., "Preliminary Screening Procedure for the MINLP Synthesis of Process Systems – II.Heat Exchanger Networks", *Computers and Chemical Engineering*, Vol. 18, 1994, 679-709
- Daichendt, M., Grossmann, I.E., "Integration of hierarchical decomposition and mathematical programming for the synthesis of process flowsheets", *Computers and Chemical Engineering*, Vol. 22, 1997, 147-175
- Damours, S., Feng, Y, LeBel, L., Nourelfath, M., "Integrated Bio-refinery and Forest Products Supply Chain Network Design", *Proceedings of IFORS2011 — 19th Conference for the International Federation of Operational Research Societies*, Melbourne, Australia, July-2011, 15
- Direcção-Geral da Energia e dos Transportes / Comissão Europeia, *Fazer mais com menos – Livro Verde sobre a eficiência energética*, Serviço das Publicações Oficiais das Comunidades Europeias, 2005
- Drobež, R., Pintarič, Z.N., Pahor, B., Kravanja, Z., "Simultaneous heat integration and the synthesis of biogas processes from animal waste", *Asia-Pacific Journal of Chemical Engineering*, Vol. 6, 2011, 734–749

- Erlenkotter, D., Manne, A.S., "Capacity Expansion For India`s Nitrogenous Fertilizer Industry", *Management Science*, Vol. 14, n.º 10, June-1968, B553-B572
- Flisberg, P., Damours, S., Ronnqvist, M., Troncoso, J.J., Weintraub, A., "Integrated Value Chain Planning", *Proceedings of IFORS2011 — 19th Conference for the International Federation of Operational Research Societies*, Melbourne, Australia, July-2011, 20
- Furman, K.C., Sahinidis, N.V., "Computational Complexity of Heat Exchanger Network Synthesis", *Computers and Chemical Engineering*, Vol. 25, 2001, 1371–1390
- Gassner, M., Marechal, F., "Methodology for the optimal thermo-economic, multi-objective design of thermochemical fuel production from biomass", *Computers and Chemical Engineering*, Vol. 33, 2009, 769–781
- Linnhoff, B., Townsend, D.W., Boland, D., Hewitt, G.F., Thomas, B.E.A., Guy, A.R., Marsland, R.H., *A User Guide On Process Integration For The Efficient Use Of Energy*, The Institution of Chemical Engineers, Rugby, 1982
- Magueijo, V., Fernandes, C., Matos, H., Pedro Nunes, C., Calau, J.P., Oliveira, F., Carneiro, J., *Medidas de eficiência Energética aplicáveis à indústria Portuguesa: um enquadramento tecnológico sucinto*, ADENE-Agência para a Energia, 2010
- Martín, M., Grossmann, I.E., "Energy optimization of hydrogen production from lignocellulosic biomass", *Computers and Chemical Engineering*, Vol. 35, 2011, 1798– 1806
- Miranda, J.L., Casquilho, M.A.S., "Robust optimization in Heat Exchanger Network", *Proceedings of ICEEM2007—2nd International Conference of Energy, Environment and Energy Management*, University of Extremadura, Badajoz, Spain, June-2007, 4pp
- Relvas, S., Matos, H., Fernandes, M.C., Castro, P., Pedro Nunes, C., "Aquo Min: A Software tool for Mass-Exchange Networks targeting and design", *Computers and Chemical Engineering*, Vol. 32, 2008, 1085-1105
- Saboo, A.K., Morari, M., Colberg, R.D., "RESHEX: an interactive software package for the synthesis and analysis of resilient heat-exchanger networks - I.Program description and application", *Computers and Chemical Engineering*, Vol. 10, 1986, 577-589
- Seita, A., Matos, H., Fernandes, M.C., "Process Integration in a Biomass based Industrial Plant", *CibBB2009-Conference on Industry-based Bioenergy and Biorefinery*, INETI, Lisboa, 19-20 Novembro, 2009
- Soršak, A., Kravanja, Z. "Simultaneous MINLP synthesis of heat exchanger networks comprising different exchanger types", *Computers and Chemical Engineering*, Vol. 26, 2002, 477-775
- Yee, T.F., Grossmann, I.E., "Simultaneous Optimization Models for Heat Integration – II.Heat Exchanger Network Synthesis", *Computers and Chemical Engineering*, Vol. 14, 1990, 1165-1184

(Página deixada propositadamente em branco)

CASO 10

OTIMIZAÇÃO DE PROCESSOS NA PRODUÇÃO DE CAPAS PARA ESTOFOS DE AUTOMÓVEIS: O CASO DA COINDU

Cláudio Alves

Centro de Investigação Algoritmi da Universidade do Minho
Escola de Engenharia, Universidade do Minho
claudio@dps.uminho.pt

Pedro Brás

Centro de Investigação Algoritmi da Universidade do Minho
Escola de Engenharia, Universidade do Minho
pedro.bras@dps.uminho.pt

José Valério de Carvalho

Centro de Investigação Algoritmi da Universidade do Minho
Escola de Engenharia, Universidade do Minho
vc@dps.uminho.pt

Telmo Pinto

Centro de Investigação Algoritmi da Universidade do Minho
Escola de Engenharia, Universidade do Minho
telmo@dps.uminho.pt

RESUMO

Este artigo descreve a aplicação de ferramentas de otimização no quadro da atividade da empresa Coindu sediada no Norte de Portugal. A empresa produz capas para estofos de automóveis a partir de diferentes matérias-primas incluindo tecido e couro natural. Em 2011, produziu mais de 700000 capas para produtores de estofos nacionais e internacionais que fornecem por sua vez grandes marcas do setor automóvel. Neste artigo, analisamos em particular os processos de corte de peles de couro da Coindu. As peles de couro são usadas para produzir os modelos de gama mais alta. As peças dos estofos são cortadas diretamente a partir das peles. Dado que as peças e as peles são figuras geométricas com contornos irregulares, o processo de corte gera inevitavelmente desperdícios. O custo das peles de couro e os elevados requisitos de qualidade impostos pelos clientes explicam a intenção da empresa em automatizar estes processos para melhorar assim o aproveitamento dessas matérias-primas.

Diferentes abordagens de otimização foram desenvolvidas para maximizar o aproveitamento das peles. O problema de otimização associado é um problema de corte e empacotamento a duas dimensões que envolve figuras geométricas fortemente irregulares. Dada a complexidade do problema, optou-se pelo desenvolvimento de métodos heurísticos. Os detalhes dessas abordagens e os resultados obtidos são descritos nas secções seguintes.

PALAVRAS-CHAVE

Otimização de processos; Indústria automóvel; Heurísticas; Meta-heurísticas.

1. Introdução

A Coindu é uma empresa multinacional do ramo têxtil que produz essencialmente capas de estofos para a indústria automóvel. Atualmente, a empresa é composta por três unidades de produção situadas em Mogege (Vila Nova de Famalicão), Arcos de Valdevez e Curtici (Roménia), que empregam mais de 1700 colaboradores. A principal unidade situada em Mogege emprega praticamente metade de toda a mão-de-obra da empresa. No total, são cerca de 840 colaboradores que garantem a operacionalidade de doze linhas de produção e do setor de corte de tecido e couro. As outras unidades nasceram no quadro de sucessivas expansões levadas a cabo pela empresa para responder ao aumento da procura. A nível nacional, a Coindu é líder no setor dos têxteis para a indústria automóvel. A empresa alcançou uma posição de destaque a nível internacional, e é hoje fornecedora de capas em tecido e couro natural para as principais empresas mundiais de produção de estofos e bancos de automóveis. Os clientes finais dessas empresas são marcas bem conhecidas do setor automóvel.

As capas e outros acessórios fornecidos pela Coindu são compostos por diversos elementos e materiais. A superfície das capas é produzida a partir de couro natural, PVC e tecido. Por baixo dessa superfície, podem estar telas, espumas, linhas, etiquetas, entre outros elementos, que fazem parte da estrutura da capa e que entram por isso no processo de produção. A produção das capas envolve diversas operações entre as quais se destacam as operações de corte das matérias-primas. Neste artigo, focamo-nos num destes processos e abordamos em particular aquele que envolve o corte do couro.

O corte do couro é, por vários motivos, um dos processos produtivos mais críticos da empresa. O couro é usado naturalmente para produzir as capas de estofos dos segmentos de automóveis mais altos. Os requisitos de qualidade impostos pelos clientes são altos e condicionam em parte estes processos. A Coindu compra peles de couro tratadas a partir das quais corta as diferentes peças dos estofos. Atualmente, grande parte do corte das peles de couro é executado de forma manual. As peças são representadas através de estruturas metálicas com a forma das peças dos estofos que são designadas por

«cortantes». Os cortantes são colocados na pele por operadores manuais de modo a aproveitar ao máximo a superfície evitando os defeitos da pele e cumprindo os requisitos de qualidade definidos para cada peça. Uma vez preenchida, a pele passa por uma prensa que irá provocar o corte das peças dos estofos. Todos os espaços à superfície da pele que não são aproveitados por serem demasiado pequenos acabam como desperdício. Dado o valor de uma pele de couro e as quantidades que são processadas pela empresa, o desperdício que é gerado neste processo de corte representa um custo significativo.

Apesar das políticas de melhoria implementadas pela empresa ao nível dos processos de corte de couro (formação dos operadores, atribuição de prémios de produtividade, por exemplo), a dependência destes processos do fator humano levanta invariavelmente várias dificuldades. Um operador necessita de algum tempo para alcançar a experiência que garanta níveis competitivos de aproveitamento das peles. Por outro lado, a dificuldade em encontrar profissionais motivados e interessados em prosseguir com o ofício é real. Finalmente, aspetos como o cansaço e o empenho dos operadores afetam de forma significativa os resultados alcançados.

Por todos estes motivos, a empresa procurou avaliar o potencial de sistemas de corte automático comparando-os com o desempenho dos operadores atuais. O principal desafio consiste em desenvolver módulos automáticos para o cálculo dos planos de corte. Como veremos mais adiante, enquanto os problemas de corte têm sido em geral muito estudados, existem muito poucas soluções para o problema específico de corte de couro da Coindu. O principal objetivo deste trabalho consistiu assim em desenvolver, implementar e testar algoritmos adaptados ao caso específico da empresa.

O processo de corte das peles de couro pode ser representado através de um problema de otimização que envolve figuras geométricas a duas dimensões (as peles e as peças dos estofos), e onde o objetivo consiste em determinar a posição das peças à superfície das peles de forma a evitar que uma peça se sobreponha a outra e a garantir que ela esteja totalmente dentro da superfície da pele. O facto de a pele ser um produto natural com defeitos

(incluindo buracos) e zonas de qualidade distinta levanta outras dificuldades. As peças devem evitar os defeitos e devem ser colocadas de forma que os requisitos de qualidade especificados pelos clientes sejam rigorosamente satisfeitos. O critério de otimização depende do aproveitamento total da pele que é conseguido.

Este problema de otimização pertence à família dos problemas de corte e empacotamento. Na literatura, o problema é designado por problema de corte em peles de couro (*leather nesting problem*). Os principais elementos deste problema (as peças e as peles de couro) são figuras geométricas fortemente irregulares. Os contornos das figuras não seguem nenhum padrão específico e a própria superfície das peças e das peles não são espaços homogêneos. As superfícies podem ter buracos (e outros defeitos no caso das peles), e podem ser divididas em zonas de qualidade segundo a natureza do produto no caso das peles ou segundo as especificações dos clientes no caso das peças. Todos estes aspetos tornam este problema num dos mais complexos no domínio dos problemas de corte e empacotamento.

Neste artigo, descrevemos os detalhes do problema de corte de peles de couro da Coindu, e discutimos as abordagens que foram desenvolvidas para a sua resolução. A atividade da empresa, os seus produtos e as principais características do processo de corte de couro são apresentados na Secção 2. Os elementos que caracterizam o problema de otimização associado são descritos na Secção 3. Na Secção 5, revemos de forma sumária as contribuições descritas na literatura para este problema. Duas classes de algoritmos que foram desenvolvidos e implementados são apresentados nas secções 6 e 7. A primeira abordagem consiste num conjunto de heurísticas construtivas baseadas em regras de agrupamento e seleção de peças, de seleção de regiões a explorar à superfície da pele e em funções de avaliação das diferentes posições de colocação. Essas heurísticas são descritas na Secção 6. Na Secção 7, descrevemos uma meta-heurística baseada em pesquisa local que consiste num algoritmo de pesquisa de vizinhança variável (*variable neighborhood search*). Os resultados obtidos com estes algoritmos são discutidos na Secção 8.

2. A Coindu

O diagrama da Figura 1 ilustra os principais elementos e processos relacionados com a atividade de produção da Coindu.

O processo de armazenagem consiste na receção das matérias-primas provenientes dos fornecedores e está sujeito a normas rigorosas de controlo e documentação. A avaliação da qualidade dos materiais é um processo exigente, nomeadamente no que diz respeito ao couro. Ao receber um lote de peles, é feita uma inspeção geral ao couro para certificar o nível de qualidade declarado pelo fornecedor. Se for detetado algum desvio relativamente ao nível de qualidade declarado, é desencadeado um processo de renegociação dos termos do contrato de compra e venda referentes ao lote de peles em causa. Estes desvios são relativamente comuns dado que o couro é um produto natural que apresenta características variáveis e que está sujeito a condições que o poderão afetar tais como a temperatura e a humidade.

O processo de produção é despoletado por uma ordem de fabrico que envolve a especificação das matérias-primas necessárias à produção do artigo em causa. A maioria das matérias-primas está sujeita a um processo de corte. De acordo com as especificações das ordens de fabrico, a secção de corte faz a preparação e o corte das matérias-primas necessárias à produção. Os objetivos desta secção são o corte atempado de todos os materiais solicitados com o melhor aproveitamento possível. O nível de aproveitamento das matérias-primas é um dos fatores com maior impacto na competitividade global da empresa.

O processo de corte da Coindu divide-se em três etapas:

- o corte de couro;
- o corte de outros materiais;
- o controlo de qualidade dos materiais cortados.

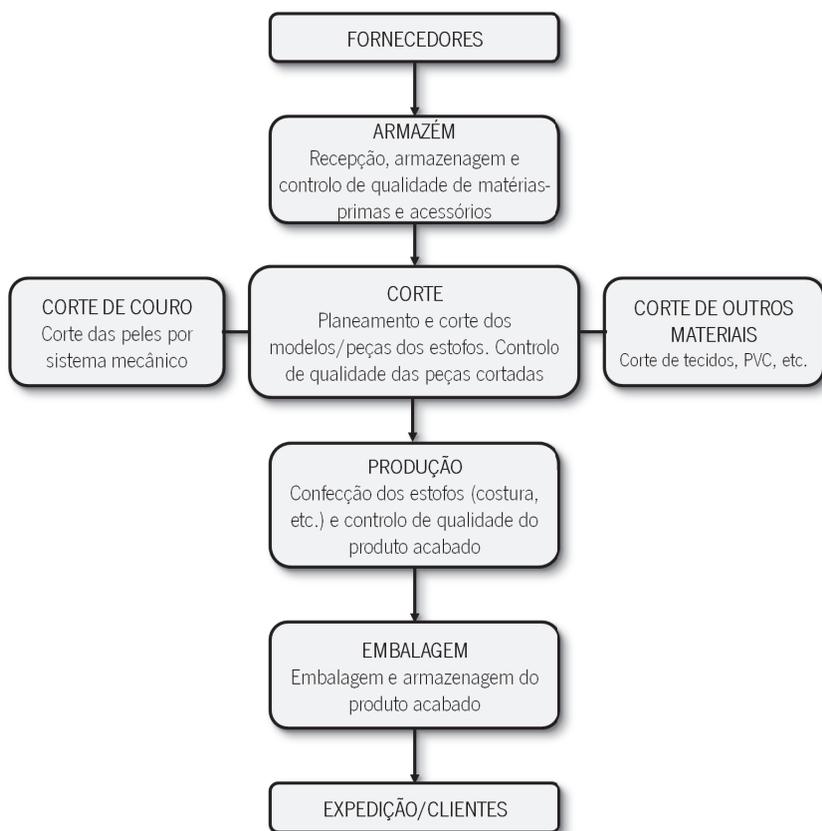


Figura 1: Visão geral das atividades de produção da Coindu

Atualmente, o corte de couro na Coindu é feito de forma manual. Apesar de muitas outras empresas já terem abdicado desta forma de corte e terem adotado sistemas automáticos, a Coindu mantém este setor a operar de forma manual devidos aos elevados níveis de eficiência que outros métodos ainda não permitem alcançar. O processo de corte de couro da Coindu depende assim de operadores especializados e com muita experiência.

Para atingir níveis elevados de aproveitamento do couro, a Coindu traçou uma filosofia para o setor do corte que assenta na formação de novos cortadores, no estabelecimento de objetivos e no estímulo à competição. A empresa aposta fortemente na formação de operadores de corte. Essa

formação consiste essencialmente num processo de aprendizagem prático e diário que junta os cortadores mais experientes e eficazes aos cortadores em formação. Com este método, a experiência e técnicas de corte são legadas de forma progressiva e natural. A Coindu estabelece também objetivos de aproveitamento do couro relativos a cada modelo a ser cortado e atribui prémios a quem cumprir ou ultrapassar essas metas. Desta forma, a empresa recompensa os cortadores mais empenhados pelos níveis de eficiência alcançados.

Os processos de corte manual de couro têm também as suas desvantagens relacionadas inevitavelmente com o fator humano. O processo de formação é moroso, consome recursos e afeta, de alguma forma, a eficácia do setor do corte. O retorno só se faz verdadeiramente sentir quando o operador já conquistou experiência e autonomia, o que depende evidentemente da disposição deste em prosseguir com o ofício. A dificuldade em encontrar novos profissionais com motivação para esta atividade é também um entrave. Outras desvantagens relacionam-se com outras variáveis relativas à natureza humana, tais como o cansaço, a motivação, o ânimo e o empenho.

O corte efetuado através de cortantes pelos operadores manuais consiste nos passos seguintes:

1. identificação e requisição do lote de couro a ser cortado de acordo com a ordem de fabrico;
2. seleção dos cortantes associados às peças que constam da ordem de fabrico;
3. seleção da pele do lote requisitado para a execução do corte;
4. colocação da pele na mesa de corte;
5. revista da pele de forma a realçar com um giz ou um marcador próprio os defeitos e as diferentes zonas de qualidade da pele;
6. colocação dos cortantes sobre a pele de forma a criar um plano de corte compacto e eficiente, tendo em conta as zonas de qualidade, os defeitos e as quantidades de cada peça requerida na ordem de fabrico;

7. passagem da mesa de corte para a prensa para a execução do corte;
8. recolhimento dos cortantes, peças cortadas e desperdício da mesa de corte;
9. colocação das peças cortadas na mesa de inspeção.

A revista e o corte das peles de couro são realizados por dois operadores em cada mesa de corte. A coordenação dessas atividades é normalmente definida por estes operadores. Uma boa comunicação e interação entre os cortadores são essenciais para a eficácia dessas operações.

Após o corte, as peças são submetidas a um rigoroso controlo de qualidade cujo objetivo é identificar e recolher as peças defeituosas evitando assim que estas avancem no processo produtivo. Os defeitos devem-se a falhas no processo de revista do couro que resultam em peças cortadas em zonas com defeito ou a anomalias no processo de corte (arrastamento dos cortantes pela prensa, peças deformadas). A inspeção das peças é realizada por um revistador que opera em mesa própria (com forma e iluminação adequada), e termina com a separação das peças cortadas por tipo e modelo.

O processo de corte de outros materiais engloba o corte de tecidos (de vários tipos), PVC, telas, espumas, alcatifas, alcântara e TNT (Tecido Não Tecido). Esses materiais são fornecidos sob a forma de rolos com várias dezenas de metros de comprimento. O corte é feito de forma automática através de lâminas. Para executar este processo, são estendidas sobre a máquina de corte várias camadas do material a cortar, formadas por dezenas de folhas de tecido que são fixadas à mesa de corte por meio de vácuo (ao conjunto destas camadas chama-se o colchão). Sobre o colchão movimentam-se as cabeças de corte que possuem lâminas, vazadores (ferramenta que permite fazer cortes circulares de diâmetro fixo) e outras ferramentas de corte. Essas cabeças executam o corte de um plano de corte pré-definido que foi elaborado por meio de uma aplicação informática própria. Evidentemente, o plano é gerado de forma a originar o máximo aproveitamento da matéria-prima e a respeitar as quantidades impostas pela ordem de fabrico. Neste tipo

de corte, e ao contrário do que se passa no corte de couro, nem todas as peças são revistadas após terem sido cortadas dado que o número de defeitos é tipicamente baixo e as elevadas quantidades de peças cortadas tornam financeiramente inviável uma revista exaustiva. Os defeitos que poderão surgir dependem exclusivamente do processo de corte e são rastreados através da amostra de peças que são recolhidas.

Em qualquer um dos processos de corte descritos (couro e outros materiais), as peças cortadas são identificadas através de timbragem. Em cada peça, é impresso ou carimbado um conjunto de informações que inclui o número de referência do cliente, a identificação sumária da peça, o número da ordem de fabrico e a data. Essa identificação permite a rastreabilidade das peças ao longo do processo de fabrico até ao cliente final, e possibilita que, no caso de uma reclamação de defeito de fabrico num estofo, seja possível identificar a fábrica que o produziu, a data de produção e a equipa responsável pela produção e aprovação do mesmo. Após essa identificação, as peças validadas seguem para a área de produção da empresa.

A produção consiste em operações de costura que visam agregar as peças cortadas e outros materiais para dar origem ao produto final (as capas). É o setor da Coindu que utiliza o maior número de recursos humanos. A costura é feita de forma manual por costureiras em máquinas de costura industriais. A área de produção divide-se em linhas, cada uma delas especializada num determinado tipo de modelo de estofo. Essa especialização é indispensável já que cada modelo de estofo envolve um conjunto de operações próprias. Os materiais necessários ao processo são as peças cortadas provenientes do setor de corte, linhas, elásticos, e perfis, entre outros. No final, os produtos passam por um controlo de qualidade que consiste na inspeção geral e meticulosa de toda a capa de forma a garantir que estas estão livres de defeitos e em conformidade com todos os requisitos de qualidade. As capas defeituosas são encaminhadas para análise de forma a apurar as causas que estiveram na origem do defeito. Os produtos acabados e sem defeitos seguem para a embalagem.

O setor da embalagem completa o ciclo produtivo da Coindu. Consiste no acondicionamento dos produtos acabados em embalagens para que estes cheguem em perfeitas condições ao cliente. O tipo de caixas a usar, a disposição do produto dentro destas e o tipo de etiquetagem são requisitos definidos pelo cliente e que a empresa segue escrupulosamente.

3. Produtos e matérias-primas

A Coindu produz essencialmente capas para estofos de automóveis. Os clientes finais destes produtos são marcas bem conhecidas da indústria automóvel entre as quais se destacam a Audi, BMW, Peugeot, Citroën e Volvo. A empresa produz vários modelos para cada uma destas marcas, cada um com as suas características próprias ao nível do *design*, das funcionalidades, dos tipos de materiais, e da configuração e quantidade de peças.

Os materiais usados para a produção das capas são variados (couro, PVC – vinil – , licra, alcântara, espuma, tecidos, telas e linhas). O couro distingue-se dos outros materiais pelas suas propriedades e características únicas. O couro é um produto de origem natural que é extraído a partir da pele que cobre o corpo de um animal. No seu estado original, essas peles estão cobertas por pêlos ou lã e são compostas por várias camadas: epiderme, derme e tecido subcutâneo. O couro usado pela Coindu é constituído apenas pela derme da pele proveniente de bovinos. A empresa compra peles de couro prontas a serem usadas que foram já tratadas pelos fornecedores através dos processos tradicionais de preparação desses materiais (curtimento, tingimento, amaciamento e acabamento). Essa preparação confere às peles de couro o aspeto que é conhecido, mas não elimina todas as imperfeições. As características do couro variam de pele para pele em termos de espessura, rigidez, flexibilidade, ao nível da textura estampada à superfície e do aspeto superficial (poros dos pelos, rugas da barriga e pernas). Uma pele de couro poderá ter também outros tipos de marcas de origem não natural tais como picadas de insetos, cicatrizes, arranhões, e marcas ou defeitos associados ao processo de curtimento.

Em contexto industrial, todas estas características das peles de couro traduzem-se em zonas diferenciadas de qualidade e em defeitos que são considerados como áreas não aproveitáveis. A Coindu usa uma escala de classificação constituída por cinco zonas de qualidade designadas por A, B, C, D e «lixo» (em ordem decrescente de qualidade, sendo a zona A a que representa o melhor nível de qualidade e a zona de «lixo» a que caracteriza as regiões do couro que não podem ser aproveitadas). A aplicação dessa escala às peles de couro depende de um conjunto de orientações e normas definidas previamente por cada cliente. Assim, uma zona que é considerada como sendo de um determinado nível de qualidade para um cliente poderá ser avaliada de forma diferente por outro. Para cada cliente e para cada modelo de carro, existe um catálogo de couro que contém toda a informação de qualificação das zonas e defeitos do couro.

Em geral, o grau de exposição que uma peça tem num assento de um automóvel determina o nível de qualidade associado. As zonas das peças que, após montagem num assento, estiverem mais expostas, deverão ser de qualidade superior, e o oposto sucede com as zonas mais escondidas de um assento. Seguindo o mesmo princípio, é atribuído o nível de qualidade mais elevado às peças nas zonas do banco que sofrem maior uso. É o caso, por exemplo, das zonas laterais dos assentos onde o desgaste é maior devido à entrada e saída dos ocupantes do automóvel.

O setor de corte do couro da Coindu utiliza um sistema de cores associado às diferentes zonas de qualidade. Este sistema é aplicado aos cortantes para identificar as zonas de qualidade das peças. Esse método visual permite aos cortadores identificar as restrições de posicionamento de um dado cortante de forma rápida. Neste artigo, usaremos a cor preta para representar as zonas A das peles. As zonas B, C e D serão representadas a vermelho, verde e azul, respetivamente. As zonas de «lixo» serão representadas a cinzento.

As peles de couro podem ser divididas em várias partes às quais estão tipicamente associados determinados níveis de qualidade. Por exemplo, as zonas da pele localizadas perto das articulações têm características que são

distintas das zonas mais próximas do dorso do animal. Uma pele de couro pode ser decomposta em seis regiões distintas:

1. parte de trás do dorso: esta é a zona mais espessa e resistente das peles donde provém o couro de melhor qualidade (nível A);
2. pescoço: nessa região, a pele é grossa e resistente, mas tem também um grande número de defeitos (marcas de crescimento, vincos e cicatrizes);
3. dorso: juntamente com a primeira região, é a zona nobre da pele que fornece o couro de melhor qualidade;
4. barriga: esta é uma zona fina da pele onde o corte de peças é delicado já que o operador tem de evitar o couro com fraca elasticidade, farpado no avesso ou com marcas profundas derivadas de vincos naturais na pele;
5. patas: região da pele que é fina com vincos ao nível das articulações;
6. zona da cauda: região da pele com uma grande variedade de defeitos incluindo picadas de insetos e cicatrizes.

A Figura 2 ilustra a localização dessas regiões numa pele de couro.

Um automóvel possui sempre dois assentos situados à frente, e poderá ter uma ou mais filas de bancos traseiros. Cada um dos lugares poderá também incluir encostos de cabeça, apoios de braço, apoio para copos, apoios para cadeira de bebé, airbags ou LCD (para os encostos de cabeça). Na Coindu, o estofos de um automóvel, constituído pelos bancos da frente e de trás, é designado por carro completo.

Na generalidade dos modelos, é necessário distinguir o lado esquerdo do lado direito dos estofos de um automóvel por várias razões (ausência de simetria, funcionalidade dos bancos, existência de airbags ou detalhes estéticos). As partes que constituem o estofos de um carro completo são constituídas por diferentes peças que envolvem um conjunto variado de

materiais. A composição e especificação de cada uma dessas partes e os seus componentes são descritos numa ficha técnica (Figura 3). Cada modelo de carro possui um conjunto destes documentos que compila informações como as codificações internas e do cliente, a lista de materiais necessários, a nomenclatura de todas as peças e pequenos materiais, o esquema de montagem, a descrição de procedimentos técnicos de corte, costura e operações especiais, a gama de controlo de defeitos, e as instruções de acondicionamento das peças na embalagem.

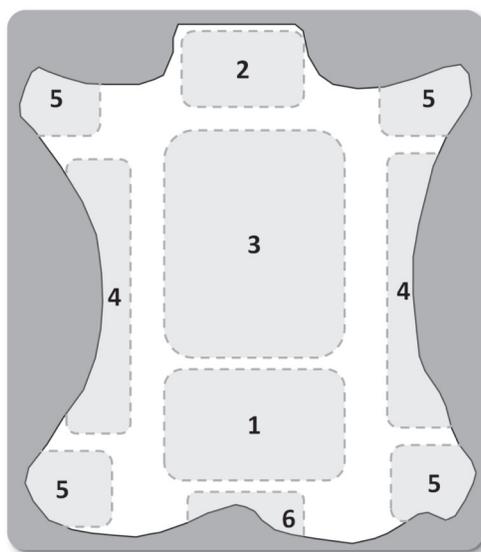


Figura 2: Distribuição das diferentes regiões de uma pele de couro

Atualmente, a Coindu produz mais de uma dezena de modelos de automóveis que têm até dez variantes distintas em alguns casos. O estofa de cada um destes modelos possui características, materiais e peças que poderão diferir consideravelmente em relação a outro modelo. A caracterização geométrica das peças que compõem o estofa de um modelo genérico fabricado na Coindu é feita com base nos seguintes elementos:

- dimensão: a área das peças pode variar consideravelmente desde 50 cm² até 1500 cm²;

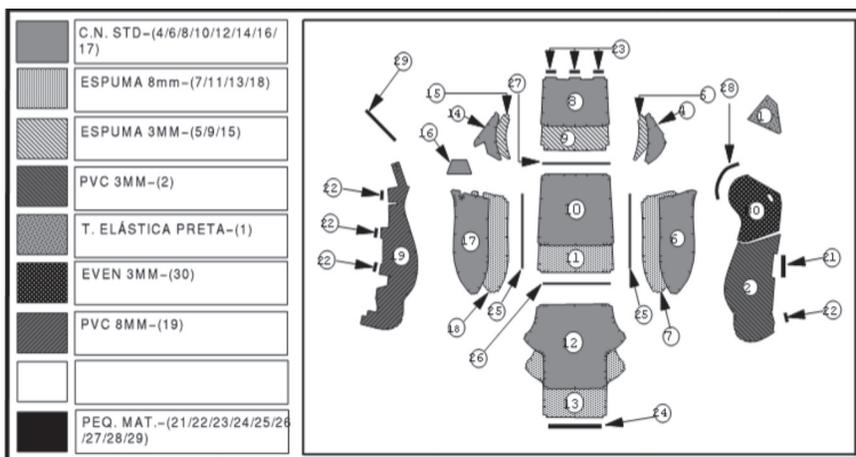


Figura 3: Parte de uma ficha técnica (assento da frente direito)

- zonas de qualidade: uma peça é constituída por uma ou várias zonas de qualidade (até um máximo de quatro zonas distintas). Na fase de criação das peças, em *software* CAD, estas zonas são concebidas como camadas sobrepostas ao traçado do contorno da peça, sendo assim consideradas como formas geometricamente independentes. Essas zonas definem requisitos mínimos de qualidade do couro a partir do qual a peça será cortada;

- buracos: a presença de buracos no interior das peças é uma característica pouco comum que se verifica em peças onde se encontram encostos de cabeça, airbags, cintos de segurança ou cadeiras de bebé;

- irregularidade: dependendo do modelo de automóvel e da parte do estofado em causa, algumas peças poderão ter contornos regulares (algumas quase de forma retangular) ou fortemente irregulares (com várias concavidades e contornos com detalhes sinuosos);

- número de vértices: as peças podem ser representadas através de um conjunto de segmentos consecutivos (polígonos). O número de vértices de uma peça depende da sua dimensão, irregularidade e das suas zonas de qualidade. No seu conjunto (contorno e zonas de qualidade), uma peça

pequena poderá ser representada através de um polígono com poucas dezenas de vértices, enquanto que uma peça de grandes dimensões e várias zonas de qualidade poderá envolver centenas de pontos distintos.

4. O problema de corte em peles de couro

O processo de corte do couro da Coindu pode ser formulado como um problema de otimização no qual se pretende posicionar um conjunto de pequenas figuras irregulares (as peças dos estofos) em superfícies irregulares de maiores dimensões que representam as peles de couro, com o objetivo de maximizar a sua ocupação. O problema pertence à família dos problemas de corte e empacotamento que têm merecido uma atenção crescente por parte da comunidade académica. Entre 1995 e 2004, foram referenciadas mais de 400 publicações nessa área em (Wäscher, Haußner, & Schumann, 2007). Na literatura, o problema é designado por *leather nesting problem* (problema de corte em peles de couro). Apesar da sua relevância prática, existem poucos resultados para a variante do problema associado ao caso da Coindu. A maior parte das aplicações descritas na literatura referem-se a outras indústrias como a do calçado onde se consideram restrições adicionais que não têm correspondência na indústria automóvel.

No caso da Coindu, as ordens de fabrico são compostas por listas de peças que ocupam normalmente mais do que uma pele. O objetivo do problema de otimização consiste assim em determinar os planos de corte com os melhores aproveitamentos para um conjunto de peles de couro.

Conforme foi referido acima, as peças dos estofos são compostas por zonas de qualidade e poderão ter eventualmente buracos no seu interior. Uma zona de qualidade representa o nível mínimo de qualidade para essa região da peça. Como tal, nunca poderá ser colocada numa região da pele de qualidade inferior. Ao longo deste artigo, as zonas de qualidade A das peças serão representadas a magenta, enquanto que as outras zonas de qualidade serão representadas com as cores utilizadas para as zonas de qualidade das peles de couro. As figuras 4 e 5 representam respetivamente uma pele de couro e um conjunto de peças de modelos produzidos pela Coindu. As

caraterísticas das peças dependem do tipo de indústria que é considerado. Na indústria do calçado, por exemplo, as peças são de pequenas dimensões enquanto na indústria automóvel existem peças com uma grande variedade de tamanhos.

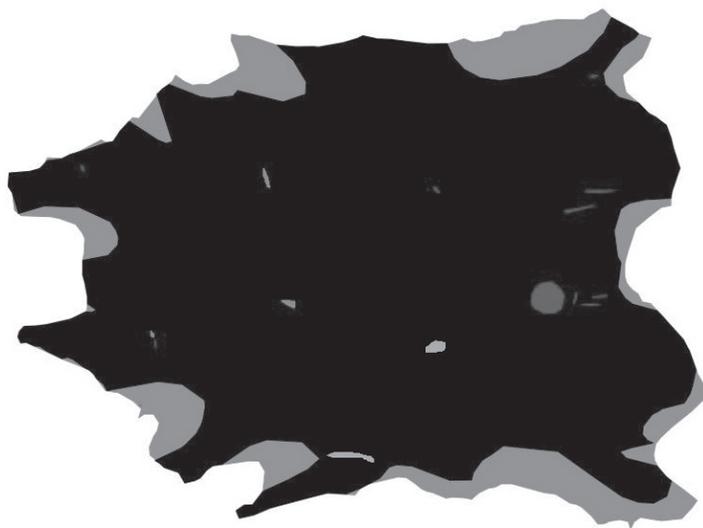


Figura 4: Pele de couro



Figura 5: Peças de estofos de automóveis

Um plano de corte consiste num conjunto de peças posicionado à superfície das peles de couro. Um plano é considerado válido se todas as peças estiverem integralmente em cima da pele, se não houver sobreposições entre peças e se nenhuma peça for colocada em cima de um defeito, de um buraco ou de uma zona de qualidade inferior. A eficiência de um conjunto de planos de corte associados a uma ordem de fabrico é calculada com base na razão entre a área total de todas as peças cortadas e a área total das peles envolvidas subtraindo a área considerada como utilizável da última pele.

Um dos aspetos importantes a ter em conta no desenvolvimento de abordagens para o problema de corte em peles de couro tem a ver com a forma como são representadas as figuras. No nosso caso, optou-se por uma representação baseada em polígonos que são gerados a partir de uma ferramenta de CAD. O número de vértices que são necessários para representar uma peça varia entre 150 e 300. Para as peles, esse valor ultrapassa frequentemente os 1000 vértices. O número de vértices dos polígonos tem um impacto significativo no desempenho dos algoritmos. Como tal, na aplicação que foi desenvolvida, os polígonos usados para representar as peças e peles do problema foram sujeitos a um conjunto de operações de simplificação baseadas em aproximações por defeito ou por excesso.

No caso das peles, foram aplicadas aproximações por defeito para a área de qualidade A que representa a base da pele (assumimos que todas as outras zonas de qualidade se sobrepõem a essa base). As outras zonas de qualidade inferior são representadas através de aproximações por excesso. Se duas ou mais zonas de qualidade se intersectarem após estas simplificações, as restrições referentes às zonas de qualidade inferiores irão prevalecer, o que garante a validade dos planos de corte finais. A título de exemplo, se dois polígonos associados a zonas de qualidade B e D se intersectarem, nunca será colocada na zona de interseção uma peça cuja qualidade seja superior a D. Aos polígonos que representam as peças, são aplicadas aproximações por excesso. Esse procedimento garante que os planos continuam a ser válidos mesmo quando as peças simplificadas são substituídas pelas peças originais. Esse procedimento de simplificação dos polígonos permitiu reduzir o número médio de vértices das peles para 300. No caso das peças, o número médio de vértices baixou para 80.

Outra dificuldade associada aos problemas de corte em peles de couro está relacionada com a forma como são detetadas interseções. Todos os polígonos envolvidos são figuras com contornos irregulares e, em muitos casos, com concavidades. A existência de zonas de qualidade nas peças e nas peles torna o problema ainda mais complexo. Com efeito, os polígonos

usados para representar as zonas de qualidade de uma peça e da pele poderão eventualmente ficar sobrepostos se os níveis de qualidade associados forem compatíveis. Bennell e Oliveira (2008) descrevem várias abordagens para o cálculo das posições relativas entre dois polígonos. Na nossa implementação, usámos uma ferramenta designada por *invólucro de posicionamento* (*no-fit polygon* - NFP) que determina o conjunto de pontos onde pode ser colocado um polígono de modo a não interseccionar com outro. O *invólucro de posicionamento interior* (*inner-fit polygon* - IFP) é um instrumento similar que permite determinar se um polígono se encontra totalmente no interior de outro polígono.

Dados dois polígonos A e B, o invólucro de posicionamento de A relativamente a B (NFP_{AB}) é um polígono cujo contorno é definido pelo percurso de um ponto de referência do polígono A quando este percorre o contorno de B. Ao longo desse percurso, assume-se que A e B se mantêm sempre em contacto sem nunca se sobrepor. Uma vez calculado NFP_{AB} , a posição de A relativamente a B é avaliada com base na posição do ponto de referência de A relativamente a NFP_{AB} . Assim, se o ponto de referência se encontrar dentro de NFP_{AB} , então os polígonos A e B interseccionar-se-ão. Se o ponto de referência estiver na fronteira de NFP_{AB} , então os polígonos A e B estarão em contacto sem contudo se interseccionar. Finalmente, se o ponto de referência de A estiver fora de NFP_{AB} , os polígonos A e B não estarão nem em contacto nem sobrepostos. A Figura 6 ilustra um invólucro de posicionamento entre dois polígonos irregulares com concavidades.

O invólucro de posicionamento interior entre dois polígonos A e B (IFP_{AB}) é obtido fazendo deslizar o ponto de referência do polígono A ao longo do contorno de B, mas de forma a que o polígono A se mantenha sempre no interior do polígono B. Se o ponto de referência do polígono A estiver dentro do invólucro IFP_{AB} , então A estará completamente no interior de B sem tocar na sua fronteira. Se o ponto de referência estiver fora do polígono IFP_{AB} , então o polígono A não estará inteiramente dentro de B. Finalmente, se o ponto de referência estiver na fronteira de IFP_{AB} , então A estará dentro de B e em contacto com a sua fronteira.

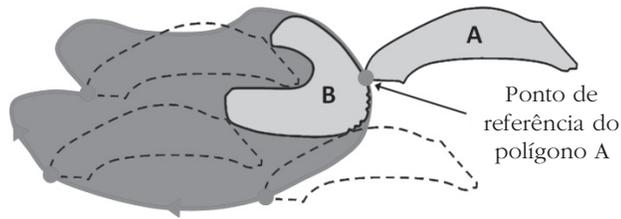


Figura 6: Invólucro de posicionamento entre dois polígonos

Para garantir a validade dos planos de corte, é necessário calcular os NFP entre todos os pares de peças e todos os IFP entre o contorno das peças e a área útil das peles. Além disso, é necessário também calcular os NFP entre as zonas de qualidade das peças e todas as zonas de qualidade inferior das peles.

5. Abordagens da literatura

O problema genérico de corte com figuras irregulares foi abordado por diversos autores. O problema é designado na literatura anglo-saxónica por *nesting problem*. Devido à sua complexidade, a maior parte dos métodos propostos assentam em procedimentos heurísticos (Babu & Babu, 2001; Bennell & Dowsland, 1999; Heckmann & Lengauer, 1995). O problema de corte em peles de couro é uma das variantes desse problema, e é seguramente um dos mais complexos no grupo dos problemas a duas dimensões devido às características que foram descritas acima.

Existem poucas abordagens descritas na literatura para o problema de corte em peles de couro, e a maior parte delas são aplicadas a problemas com restrições adicionais que não existem no problema de corte do couro da Coindu. É o caso das restrições ao nível da orientação das peças nas peles que ocorrem na indústria do calçado. Por esse motivo, passaremos a referir-nos ao problema da Coindu como o problema genérico de corte em peles de couro.

O primeiro algoritmo para a resolução do problema genérico de corte em peles de couro foi proposto por Heistermann e Lengauer (1995). Estes autores

desenvolveram uma heurística construtiva que pretende reproduzir os passos seguidos por operadores humanos. Em cada iteração, uma peça é colocada numa determinada zona (restrita) da pele que é designada por *foco*. A localização desse foco na pele pode ser fixa se se mantiver aproximadamente constante, ou variável se for alterada a cada iteração do algoritmo. Numa primeira fase, são selecionadas as peças cujos contornos melhor se ajustam ao contorno desse foco. Para facilitar essa escolha, as peças são previamente agrupadas em classes de acordo com a forma do seu contorno. O posicionamento das peças na pele é feito tendo em conta a área da peça e a distância entre o bordo da peça e os bordos da pele e do plano de corte construído até ao momento.

Heistermann e Lengauer (1995) descrevem algumas estratégias para melhorar a qualidade das soluções. Propõem a criação de figuras agregadas compostas por duas ou mais peças originais. O objetivo é tirar proveito de encaixes previamente estudados que poderão conduzir a uma diminuição do desperdício de couro que ocorre naturalmente quando se tenta juntar várias peças. Os autores usam também um procedimento para compactar as peças e eliminar assim os espaços vazios entre elas. Apresentam resultados computacionais para instâncias em que as peças correspondem a partes de um sofá. O algoritmo demora entre 60 a 140 segundos a construir planos de corte válidos com níveis de aproveitamento compreendidos entre os 56% e 67% (valores médios aproximados).

Yuping, Shouwei e Chunli (2005) descrevem um algoritmo de arrefecimento simulado para um problema de corte em peles de couro sem zonas de qualidade. Para gerar posicionamentos válidos, são efetuados movimentos de translação horizontal ou vertical e rotações das peças à superfície das peles. Os autores compararam os resultados da sua abordagem com os resultados obtidos por operadores manuais. Constataram que os operadores demoravam em média entre 15 a 30 minutos para preencher completamente uma pele. Nesse espaço de tempo, o algoritmo de arrefecimento simulado produz soluções com níveis de aproveitamento comparáveis aos dos operadores manuais.

Crispin *et al.* (2005) propuseram dois algoritmos genéticos para um problema de corte em peles de couro na indústria do calçado. Nesse caso, as peças devem ser colocadas na pele segundo uma determinada direção o que reduz de forma significativa os graus de liberdade do problema. O primeiro algoritmo consiste em colocar as peças em zonas restritas da pele de modo a maximizar o aproveitamento local nessa zona. A segunda abordagem é baseada em grafos de conectividade em que os arcos representam as peças que poderão estar em contacto num plano de corte válido. À semelhança de Yuping, Shouwei & Chunli (2005), os autores defendem que essas abordagens permitem aproximar os resultados obtidos pelos operadores humanos sem contudo os melhorar.

Yuping e Caijun (2009) descrevem um procedimento baseado em algoritmos genéticos e pesquisa por arrefecimento simulado para o problema de corte em peles de couro sem zonas de qualidade. As peças são colocadas inicialmente na pele usando uma estratégia do tipo *bottom-left*. Para melhorar a qualidade do plano de corte, recorrem a uma combinação entre um algoritmo genético e um procedimento de pesquisa por arrefecimento simulado. O esquema de arrefecimento controla a taxa de mutação do algoritmo genético, o que permite melhorar a convergência do algoritmo. Os resultados computacionais descritos nesse artigo mostram que o algoritmo é capaz de fornecer em uma hora soluções com níveis de aproveitamento da ordem dos 70%.

Entre todas estas contribuições, apenas o artigo de Heistermann e Lengauer (1995) trata de um problema similar ao problema de corte da Coindu. Contudo, só apresenta resultados para instâncias da indústria do mobiliário onde as peças são tipicamente grandes, o que favorece a construção de planos de corte com elevados níveis de aproveitamento. Nas secções seguintes, apresentamos duas classes de métodos que foram desenvolvidos especificamente para o problema de corte de couro da Coindu. O primeiro consiste num conjunto de regras que permitem definir diferentes heurísticas construtivas. O segundo foi desenvolvido para melhorar as soluções obtidas a partir dessas heurísticas construtivas, e baseia-se em

pesquisa de vizinhança variável. Os dois algoritmos que são descritos neste artigo foram apresentados originalmente em (Alves *et al.*, 2012a) e (Alves *et al.*, 2012b).

6. Heurísticas construtivas

O primeiro grupo de algoritmos que foi desenvolvido para o problema de corte em peles de couro da Coindu assenta num conjunto de estratégias que se confundem com os passos de uma heurística construtiva e que incidem nos seguintes aspetos: o agrupamento de peças, a seleção da próxima peça a posicionar na pele, a seleção da região de posicionamento admissível, e a avaliação dos possíveis posicionamentos.

As peças começam por ser agrupadas de acordo com o valor de um determinado atributo (área, concavidades, por exemplo). Os grupos reúnem as peças cujo valor do atributo escolhido se encontra dentro de um determinado intervalo. Em cada iteração do algoritmo, é selecionada a próxima peça a ser colocada na pele. A seleção pode ser feita considerando todas as peças ainda por colocar, ou apenas aquelas que pertencem a um determinado grupo. O passo seguinte consiste em identificar uma região de posicionamento admissível na pele para a peça escolhida. Uma vez escolhida essa região, são avaliados os pontos de posicionamento válido no interior dessa região. Foram considerados vários critérios de avaliação. Os mais simples baseiam-se, por exemplo, na distância entre a peça e o bordo da pele, ou entre a peça e o centro da pele. Os mais sofisticados assentam, por exemplo, na qualidade do ajuste entre a peça e o bordo da pele e/ou do plano de corte construído até ao momento. Nas próximas secções, descrevemos os detalhes das diferentes estratégias.

6.1 Estratégias para o agrupamento de peças

As estratégias de agrupamento de peças são usadas para garantir um tratamento indiferenciado às peças do mesmo grupo que apresentam características muito próximas. As características das peças são representadas pelos atributos seguintes:

- (G1) área;
- (G2) grau de irregularidade;
- (G3) grau de concavidade;
- (G4) razão entre altura e largura do retângulo envolvente;
- (G5) valor das peças baseado nas suas zonas de qualidade e respectivas áreas;
- (G6) homogeneidade das zonas de qualidade.

Iremos designar respetivamente por A_j , I_j , C_j , R_j , V_j e H_j os valores dos atributos (G1) a (G6) associados a uma peça j . Além destes atributos, considerámos ainda um sétimo (G7) cujo valor corresponde a uma combinação linear dos outros atributos (G1) a (G6).

A separação das peças em grupos definidos a partir destes atributos permite evitar que seja dada uma prioridade maior a uma peça que é apenas ligeiramente diferente de outra. O número de peças num grupo depende da instância e também do número máximo de grupos que são considerados. Seja n o número máximo de grupos admitidos, e sejam min e max os valores mínimo e máximo dos atributos de todas as peças de uma dada instância, respetivamente. O grupo m , com $m=1, \dots, n-1$, inclui as peças cujos valores dos atributos estão no seguinte intervalo:

$$\left[min + \frac{max - min}{n} \times (m - 1); min + \frac{max - min}{n} \times m \right]$$

No caso do n -ésimo grupo, o intervalo é definido da seguinte forma:

$$\left[max - \frac{max - min}{n}; max \right]$$

Os grupos são ordenados por ordem crescente do valor do atributo. Assim, o valor de uma peça no grupo $i+1$ será sempre maior que o valor do atributo de uma peça do grupo i . O valor de n tem um impacto no número

de peças em cada grupo. Quando o valor de n é pequeno, o número de peças por grupo tende a ser elevado. À medida que n aumenta, o número de peças por grupo tende a diminuir.

O grau de irregularidade I_j de uma peça j é calculado com base no número de pontos do polígono que representa o contorno da peça. O grau de concavidade C_j consiste na diferença entre a área do invólucro convexo do polígono que representa o contorno da peça e a área total da peça.

A razão R_j entre a altura e a largura do retângulo envolvente da peça j permite distinguir as peças compridas das peças estreitas. Assumimos que a orientação original das peças é tal que o respetivo retângulo envolvente tem a sua maior aresta paralela ao eixo dos x ou ao dos y num sistema de coordenadas Cartesiano. As peças para as quais o valor R_j se aproxima da unidade estão associadas a retângulos envolventes que se aproximam de um quadrado.

Os atributos anteriores não consideram a existência de zonas de qualidade nas peças. No entanto, na fase de posicionamento, poderão existir vantagens em dar primazia às peças com uma maior distribuição de zonas de qualidade, por exemplo. Para lidar com estes casos, definimos o valor V_j do atributo (G5) da peça j como sendo o quociente entre a soma das áreas de cada uma das zonas de qualidade, multiplicada por um determinado fator, e a área da peça. Na nossa implementação, usámos um fator v_k para uma zona de qualidade k com valor igual respetivamente a 4, 3, 2 e 1 para as zonas de qualidade A, B, C e D. Seja a_j^k a área da zona de qualidade k de uma peça j . Temos que:

$$V_j = \frac{\sum_{k=1}^q (v_k a_j^k)}{A_j},$$

sendo q o número total de zonas de qualidade da peça j .

Outra característica de uma peça com potencial impacto no seu posicionamento é o seu grau de homogeneidade. Por exemplo, peças com muitas zonas de qualidade poderão dificultar o seu posicionamento, uma vez que é necessário fazer corresponder as zonas de qualidade da peça com as da pele. Usar o número de zonas de qualidade de uma peça como medida da

sua homogeneidade não é adequado, uma vez que uma peça com uma grande zona de qualidade de um tipo e várias outras zonas de pequenas dimensões pode ser considerada erradamente como sendo fortemente heterogénea. Para medir a homogeneidade de uma peça j (H_j), usámos a relação entre a área total das zonas de qualidade da peça (ponderadas através de um fator h_k que decresce com a área de cada zona) e a área total da peça. Assumimos que as zonas de qualidade das peças estão ordenadas por ordem decrescente da sua área. Para uma zona de qualidade k , h_k toma o valor $\frac{1}{i}$, sendo i o índice da i -ésima zona de qualidade da peça. O valor H_j de uma peça j é dado por:

$$\frac{\sum_{k=1}^q (h_k a_j^k)}{A_j}.$$

Para peças com uma única zona de qualidade, esse valor é igual a 1.

Por fim, considerámos um agrupamento baseado na combinação de todos os atributos anteriores. Para uma dada instância do problema com p peças, sejam A^{max} , I^{max} , C^{max} , R^{max} , V^{max} e H^{max} os valores máximos dos atributos (G1) a (G6), respetivamente. Uma peça j será agregada a um grupo de acordo com o valor do atributo dado pela seguinte expressão:

$$\alpha_1 \times \frac{A_j}{A^{max}} + \alpha_2 \times \frac{I_j}{I^{max}} + \alpha_3 \times \frac{C_j}{C^{max}} + \alpha_4 \times \frac{R_j}{R^{max}} + \alpha_5 \times \frac{V_j}{V^{max}} + \alpha_6 \times \frac{H_j}{H^{max}},$$

em que α_i , $i = 1, \dots, 6$ representa o peso atribuído a cada atributo.

6.2 Estratégias para a seleção da próxima peça

As estratégias de seleção da próxima peça a colocar na pele dividem-se em dois grupos. A primeira é também a mais simples do ponto de vista computacional. Consiste em escolher uma peça por ordem crescente ou decrescente do valor de um dos seus atributos (G1) a (G7). Essa estratégia é designada abaixo através da sigla (S1.I/D). A segunda estratégia restringe a escolha da próxima peça a um dos grupos de peças definidos na secção anterior. Trata-se de uma estratégia genérica designada abaixo através da sigla (S2.I/D). O grupo de peças são escolhidas por ordem crescente (S2.I) ou

decrecente (S2.D) do valor do seu índice. A título de exemplo, se as peças tiverem sido agrupadas com base na sua área, e se for escolhida uma estratégia do tipo (S2.D), as primeiras peças a serem posicionadas na pele serão as peças de maior área.

A seleção de uma peça dentro do grupo que foi previamente escolhido pode ser feita de duas formas distintas: com base nas características dos IFP das peças relativamente à pele (estratégias (S2.I/D.1)), ou com base no valor da função usada para avaliar a qualidade dos pontos válidos de posicionamento (estratégias (S2.I/D.2)). No primeiro caso, foram exploradas diferentes alternativas que são resumidas no final desta secção. Na estratégia (S2.D.1.a) e assumindo que as peças foram agrupadas com base no valor da sua área, por exemplo, o grupo das peças maiores é escolhido em primeiro lugar. Cada uma das peças desse grupo possui um conjunto de IFP que determinam os pontos válidos de posicionamento dentro da pele. Nessa estratégia, as peças desse grupo são escolhidas por ordem crescente da área do menor dos seus IFP. O objetivo é tentar encontrar uma peça cuja qualidade de encaixe numa determinada zona da pele é (previsivelmente) o melhor possível. No caso das estratégias (S2.I/D.2), o critério usado para escolher uma peça do grupo é o valor que é dado pela função que avalia a qualidade dos pontos de posicionamento. Essas funções são introduzidas na Secção 6.4.

As estratégias de seleção da próxima peça podem ser resumidas da forma seguinte:

(S1.I/D) seleção de uma peça a partir do conjunto completo de peças ordenadas por ordem crescente (S1.I) ou decrescente (S1.D) do valor de um determinado atributo;

(S2.I/D) seleção de um grupo de peças por ordem crescente (S2.I) ou decrescente (S2.D) do valor do seu índice, e seleção de uma peça desse grupo com base:

(S2.I/D.1) nas características dos IFP;

(S2.I/D.2) no valor da função utilizada para avaliar os pontos de posicionamento.

As características dos IFP foram escolhidas como critério de seleção da peça porque, em alguns casos, os IFP podem efetivamente fornecer informação sobre a qualidade do ajuste de uma peça numa determinada zona da pele. Como já referimos, se o IFP de uma peça de área reduzida tiver ele também uma área pequena, é previsível que essa peça se ajuste bem nessa região da pele. Com base nesse princípio, foram exploradas as seguintes estratégias para selecionar uma peça dentro do grupo de peças que foi previamente escolhido:

- (S2.I/D.1.a) seleção da peça que possui o IFP com menor área;
- (S2.I/D.1.b) seleção da peça que possui o IFP com maior área;
- (S2.I/D.1.c) seleção da peça com o maior ou o menor IFP em função do índice do grupo de peças selecionado.

É importante notar que ao aplicar uma destas estratégias, a região da pele de posicionamento admissível é selecionada ao mesmo tempo que a peça.

A aplicação da estratégia (S2.I/D.2) implica avaliar os pontos de posicionamento para cada uma das peças do grupo usando uma das funções de avaliação que serão discutidas mais abaixo. No grupo de peças que foi escolhido, o par peça/ponto que produz o melhor resultado é aquele que é escolhido. Essa estratégia implica a seleção simultânea da próxima peça a colocar e do ponto exato na pele onde essa peça será colocada. Os pontos que são efetivamente avaliados devem pertencer a uma determinada região da pele que é previamente escolhida. As estratégias usadas para selecionar essas regiões são descritas na secção seguinte.

6.3 Estratégias para a seleção da região de posicionamento admissível

A área da região da pele na qual se procura uma posição válida para colocar a próxima peça tem um impacto evidente no desempenho dos

algoritmos. Para que seja mais eficiente, a procura é limitada a uma região da pele. A seleção dessa região permite também controlar o preenchimento da pele dando prioridade, por exemplo, a regiões com determinadas características.

Algumas estratégias para a seleção da próxima peça a colocar na pele obrigam à seleção da região de posicionamento. É o caso das estratégias que dependem das características dos IFP. Para os outros casos, as regiões selecionadas são as seguintes:

- (P1) todas as zonas livres da pele;
- (P2) níveis verticais da direita para a esquerda da pele;
- (P3) níveis verticais da esquerda para a direita da pele;
- (P4) menor IFP da peça;
- (P5) maior IFP da peça;
- (P6) maior ou menor IFP da peça, dependendo do grupo da peça selecionada;
- (P7) zona livre mais pequena da pele;
- (P8) maior zona livre da pele;
- (P9) zona livre da pele com o menor índice de qualidade;
- (P10) zona livre da pele com o maior índice de qualidade;
- (P11) zona livre da pele com o contorno menos irregular;
- (P12) zona livre da pele com o contorno mais irregular.

A estratégia (P1) corresponde ao caso extremo em que a região escolhida é na realidade a totalidade da pele.

As estratégias (P2) e (P3) inspiram-se nas estratégias de corte do couro usados pelos operadores da Coindu. Consistem em preencher a pele colocando as peças disponíveis de forma a ocupar o maior espaço possível em cada um dos níveis. Os operadores só começam a preencher o nível seguinte quando não existem mais peças que possam ser colocadas no nível atual.

As estratégias (P4) a (P6) restringem o posicionamento do ponto de referência de uma peça a um determinado IFP dessa peça com a pele. A título de exemplo, se for usada a estratégia de agrupamento baseada na área das peças (G1) em conjunto com a estratégia (S1.D) e (P4), a primeira peça a ser colocada será a maior das peças, e será colocada numa zona da pele que é delimitada pelo IFP mais pequeno dessa peça com a pele. A posição exata dessa peça nessa região será determinada através das funções de avaliação de pontos de posicionamento que são descritas na próxima secção.

As zonas livres da pele referidas nas estratégias (P7) a (P12) referem-se a espaços vazios e disjuntos na pele onde possa ser colocada a peça que foi escolhida nos passos anteriores. Explorámos também a possibilidade de escolher uma zona livre da pele usando um critério baseado na sua qualidade. A qualidade de uma zona livre é medida com base na percentagem da sua superfície coberta por zonas de qualidade A. Se nenhuma zona livre é de qualidade A, o índice de qualidade é medido a partir das zonas de qualidade B e assim sucessivamente. Finalmente, a irregularidade de uma zona livre usada como critério nas estratégias (P11) e (P12) é medida a partir do número de pontos do polígono que definem o contorno dessa zona.

6.4 Estratégias de avaliação dos pontos de posicionamento

Após a seleção de uma peça e de uma região de posicionamento admissível, a fase seguinte consiste em determinar a posição exata da peça na pele. Essa escolha é feita com base no valor de uma função de avaliação dos pontos onde a peça poderá ser colocada. Várias funções foram consideradas. Algumas baseiam-se em critérios simples como a distância da peça ao bordo da pele, enquanto outras, mais complexas, procuram medir a qualidade do ajuste da peça ao plano de corte atual, ao bordo da pele ou às zonas de qualidade da pele. Alguns critérios baseiam-se na área da interseção de um *offset* da peça com a região exterior da pele e com o plano de corte atual. A Figura 7 ilustra esse conceito.

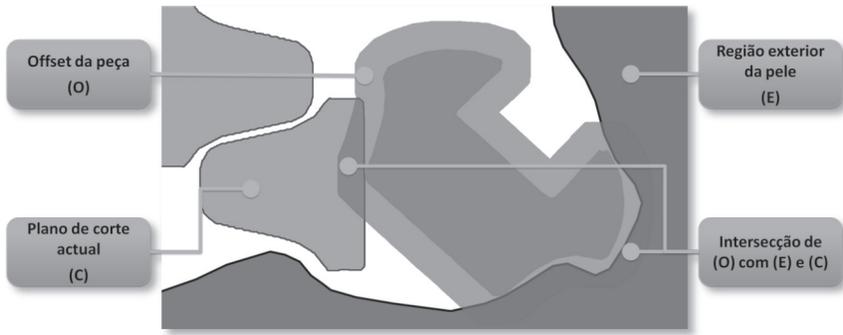


Figura 7: Interseção entre o *offset* de uma peça, o exterior da pele e o plano de corte atual

Seja I o conjunto dos polígonos que resultam da interseção do *offset* O de uma peça p com o plano de corte atual e com a zona exterior da pele, e seja $area(I_i)$ a área do i -ésimo polígono de I . Foram consideradas as seguintes funções para avaliar o posicionamento de uma dada peça p :

(E1) $\sum_{i=1}^{|I|} area(I_i)$: área total da interseção de O com o plano de corte atual e com a zona exterior da pele;

(E2) $\sum_{i=1}^{|I|} area(I_i)/(area(O) - area(p))$: área relativa da interseção de O com o plano de corte atual e com a zona exterior da pele;

(E3) $\max_{i=1, \dots, |I|} area(I_i)$: área do maior polígono resultante da interseção de O com o plano de corte atual e com a zona exterior da pele;

(E4) $\max_{i=1, \dots, |I|} area(I_i)/(area(O) - area(p))$: área relativa do maior polígono resultante da interseção de O com o plano de corte atual e com a zona exterior da pele;

(E5) área da interseção de O com o plano de corte atual e com a zona exterior da pele mais a área das zonas de qualidade de p que estão sobre zonas de qualidade da pele equivalente;

(E6) área relativa da interseção de O com o plano de corte atual e com a zona exterior da pele mais a área das zonas de qualidade de p que estão sobre zonas de qualidade da pele equivalente;

(E7) área da interseção de O com o plano de corte atual e com a zona exterior da pele mais a área das zonas de qualidade de p que estão sobre zonas de qualidade B, C ou D da pele;

(E8) área relativa da interseção de O com o plano de corte atual e com a zona exterior da pele mais a área das zonas de qualidade de p que estão sobre zonas de qualidade B, C ou D da pele;

(E9) número de espaços vazios gerados ao colocar a peça p ;

(E10) área total do desperdício gerado ao colocar a peça p ;

(E11) distância de p ao bordo da pele;

(E12) distância de p ao centro da pele;

(E13) distância de p à região com menor nível de qualidade da pele;

(E14) distância de p à região com maior nível de qualidade da pele.

Na Figura 7, a interseção do *offset* da peça com a região exterior da pele e com as peças do plano de corte atual dá origem a dois polígonos disjuntos. A função (E1) corresponde à soma de todas as áreas dos polígonos resultantes dessa interseção. A função (E3) procura privilegiar as posições em que grande parte da peça está em contacto com o contorno da pele ou com o plano de corte atual. As funções (E5) a (E8) privilegiam as posições em que a peça tem ao mesmo tempo um bom ajuste ao contorno da pele e ao plano atual, e um bom aproveitamento das zonas de qualidade da pele. As funções (E5) e (E6) favorecem as posições em que há uma correspondência perfeita entre as zonas de qualidade da peça e as da pele. Por seu lado, as funções (E7) e (E8) valorizam as posições em que se consegue um melhor aproveitamento das zonas de qualidade inferior.

A função (E9) contabiliza o número dos espaços vazios que seriam gerados se a peça fosse colocada no ponto de posicionamento em análise. Nessa função, não são considerados os espaços vazios em que não seja possível colocar qualquer uma das peças disponíveis. Esses espaços são considerados como desperdício na função (E10).

A região com menor ou maior nível de qualidade da pele usada nas funções (E13) e (E14), respetivamente, é determinada através da divisão da

pele em secções quadrangulares. O nível de qualidade dessas secções é avaliado pela estratégia de agrupamento de peças (G5).

Em alguns casos, é possível optar por um ponto de posicionamento que minimize ou maximize uma determinada função de avaliação, como são os casos de (E12) a (E14). No caso de (E12), por exemplo, designaremos por (E12.MIN) e (E12.MAX), respetivamente, os casos em que se procura um ponto de posicionamento que minimiza ou maximiza essa função de avaliação. Para as outras funções apenas faz sentido procurar o ponto que maximiza o valor da função (funções (E1) a (E8)), ou que minimiza o seu valor (funções (E9) a (E11)).

6.5 Síntese e avaliação da qualidade das soluções

A partir destes quatro grupos de estratégias, é possível definir três grandes variantes para o algoritmo construtivo. As principais diferenças entre essas variantes resultam da estratégia que é usada para selecionar a próxima peça a colocar na pele:

- se for utilizada a estratégia (S1.I/D), a próxima peça a colocar na pele, a região de posicionamento admissível e a posição da peça nessa região são selecionadas sequencialmente e de forma independente;
- se for utilizada a estratégia (S2.I/D.1), a próxima peça é selecionada com base nas características dos seus IFP. A escolha da peça implica a seleção de uma região de posicionamento admissível. Para as variantes que usam a estratégia (S2.I/D.1), não é necessário especificar uma estratégia de seleção das regiões de posicionamento admissível na pele;
- se for utilizada a estratégia (S2.I/D.2), a peça que é selecionada é aquela que resultar no maior valor da função de avaliação do posicionamento num dos pontos da região de posicionamento admissível. A próxima peça a ser colocada na pele, a região de posicionamento admissível e a posição final da peça são determinadas simultaneamente.

A eficiência dos planos de corte criados para uma determinada ordem de fabrico é avaliada com base em dois critérios. Para todas as peles exceto a última a ser cortada, essa eficiência é medida com base na percentagem de área da pele que é coberta pelas peças do plano de corte (nível de aproveitamento). Na última pele, a qualidade do plano de corte é medido com base na área não utilizada que ainda pode ser aproveitada para cortar peças de outras ordens de fabrico. Para medir o valor dessa área, a Coindu usa um procedimento que consiste em colocar um quadrado com uma determinada dimensão nas zonas livres da última pele. Os espaços onde esse quadrado não entra são considerados desperdícios. Ao longo desse procedimento, o quadrado mantém a sua orientação e não é feita qualquer distinção entre zonas de qualidade.

7. Um algoritmo de pesquisa por vizinhança variável

Além das heurísticas construtivas descritas na secção anterior, foi desenvolvido um algoritmo de melhoria baseado num procedimento de pesquisa por vizinhança variável. Foram exploradas novas funções de vizinhança a partir de uma representação das soluções baseada essencialmente em sequências de peças.

O plano de corte inicial é obtido a partir da heurística construtiva que resulta da aplicação das estratégias (G1), (S2.D.1.c), (P6) e (E11). A heurística determina a sequência pela qual as peças são colocadas e a posição exata de cada peça na pele. Um plano de corte pode ser completamente definido por um conjunto de peças, as suas orientações e correspondentes posições na pele. Em alternativa, um plano pode ser representado através de uma sequência de peças e por um procedimento de colocação dessas peças baseado na aplicação das estratégias de seleção da região de posicionamento admissível e de avaliação dos pontos de posicionamento apresentados na secção anterior (as peças são seleccionadas e colocadas na pele uma após a outra segundo a ordem definida pela sequência, e o processo é repetido até não haver mais peças na sequência). No algoritmo de pesquisa por

vizinhança variável que foi desenvolvido, consideramos esta última representação.

Foram definidos quatro tipos de movimentos a partir dessa representação:

1. substituição de uma peça p da sequência por outra que não está na sequência, eliminação da subsequência de peças desde a peça p até ao final da sequência original, e aplicação da heurística construtiva para preencher o resto da pele;
2. substituição de uma peça p da sequência por uma peça p' que não está na sequência, colocação das peças desde p' até ao final da sequência original aplicando as estratégias (P6) e (E11) da heurística construtiva, e aplicação da heurística construtiva completa para preencher o resto da pele;
3. permutação de duas peças p e p' da sequência, se a colocação da peça p' resultar num valor da função de avaliação dos pontos de posicionamento superior ao da peça p , colocação das restantes peças da sequência aplicando as estratégias (P6) e (E11) da heurística construtiva, e aplicação da heurística construtiva completa para preencher o resto da pele;
4. eliminação de uma peça p da sequência, colocação das restantes peças da sequência aplicando as estratégias (P6) e (E11) da heurística construtiva, e aplicação da heurística construtiva completa para preencher o resto da pele.

A peça p é selecionada com base na qualidade do seu encaixe na pele que é medido a partir da área relativa de interseção do *offset* da peça com o plano de corte atual e com a zona exterior da pele (estratégia (E2)). Seja P o conjunto de peças candidatas a partir do qual é selecionada a peça p . O conjunto de peças que são candidatas à substituição de peça p na sequência é designado por P' , enquanto p' representa a peça que é selecionada nesse conjunto. A sequência de peças associada ao atual plano de corte é designada por S .

As peças do conjunto P são escolhidas a partir de uma determinada parte da sequência associada ao plano de corte atual. Essa subsequência de peças é escolhida com base na percentagem de aproveitamento da pele que se verifica imediatamente após ter colocado as respectivas peças. O conjunto P pode ser definido, por exemplo, como um subconjunto de peças da sequência original cuja colocação gera aproveitamentos da pele compreendidos entre 50% e 70%. Apenas serão escolhidas as peças que, após terem sido colocadas na pele, gerarem aproveitamentos nesse intervalo. Seja U o vetor composto pelos valores dos aproveitamentos obtidos após colocar uma peça da sequência S :

$$U = (u_1, u_2, \dots, u_{|S|})$$

com $u_k < u_{k+1}, k = 1, \dots, |S| - 1$. O vetor U representa a evolução do aproveitamento de uma pele à medida que as peças de S vão sendo colocadas. Para identificar uma subsequência da sequência inicial de peças, usam-se os valores u_{min} e u_{max} que designam respetivamente o limite inferior e superior para o aproveitamento da pele. Para o exemplo anterior, teríamos $u_{min} = 50\%$ e $u_{max} = 70\%$.

A definição das vizinhanças depende dos movimentos que foram descritos acima e dos seguintes parâmetros:

- q : número de peças candidatas a serem substituídas, permutadas com outra peça ou eliminadas da sequência ($q=|P|$);
- r : número de peças candidatas à substituição da peça p ($r=|P'|$);
- subsequência (s_i, \dots, s_j) da sequência de peças inicial, a partir da qual são selecionadas as peças de P , com

$$i = \arg \min_{l \in \{1, \dots, |S|\}} \{u_l : u_l \geq u_{min}, u_l \in U\},$$

e

$$j = \arg \max_{l \in \{1, \dots, |S|\}} \{u_l : u_l \leq u_{max}, u_l \in U\}.$$

Quando $q = 1$, a peça p que é substituída, permutada ou eliminada é a peça da subsequência que resulta no menor valor da função de avaliação dos pontos de posicionamento de uma dada sequência. Por seu lado, quando $r = 1$, a peça p é substituída por uma peça p' que resulta no maior valor da função de avaliação dos pontos de posicionamento. Variando os valores desses parâmetros, é possível definir diferentes vizinhanças.

A exploração destas classes de vizinhanças é feita através de um algoritmo baseado na versão básica do método de pesquisa por vizinhança variável descrito em (Hansen, Mladenovic & Pérez, 2010). A pesquisa é efetuada em vizinhanças específicas definidas a partir dos parâmetros apresentados acima. A cada iteração do algoritmo, a pesquisa de uma solução numa vizinhança diferente pode permitir melhorar a qualidade do plano de corte atual uma vez que o ótimo local numa vizinhança não é necessariamente o mesmo que o ótimo local da vizinhança que foi previamente explorada.

8. Experiências computacionais

Para avaliar o desempenho dos algoritmos, foram conduzidas uma série de experiências computacionais realizadas a partir de instâncias reais que correspondem a modelos de automóveis em produção na Coindu. Os algoritmos foram implementados em C++, e integrados na aplicação *COINDUNester* desenvolvida para o efeito e que possui funcionalidades adicionais de tratamento de peles, peças e planos de corte. Para implementar algumas funções geométricas da aplicação, recorreremos à versão 3.7 da biblioteca *CGAL (Computational Geometry Algorithms Library)*. Os testes foram executados num PC com um processador *Intel Core i3* com 2.27 GHz e 4GB de RAM.

Na Coindu, o tempo médio necessário para construir um plano de corte numa pele de couro é de 600 segundos por pele. Essa operação é efetuada por dois operadores manuais. Esse valor foi usado como referência para avaliar o desempenho dos algoritmos e também como um parâmetro para limitar o tempo máximo de execução.

Para avaliar o desempenho das heurísticas construtivas, foram realizados dois conjuntos de experiências. No primeiro, comparou-se o desempenho das diferentes estratégias descritas na Seção 6. No segundo conjunto de experiências, usámos as estratégias que foram identificadas como sendo as mais rápidas e as mais promissoras em termos de qualidade dos planos de corte gerados, e aplicámos essas estratégias a um conjunto de várias instâncias diferentes.

Os testes foram realizados a partir de dois conjuntos de dados diferentes que correspondem cada um a um modelo de automóvel atualmente em produção na Coindu. Cada conjunto corresponde a um dos estofos de um carro completo (assentos e encostos da frente e de trás) dos referidos modelos. A instância associada ao primeiro modelo é composta por 23 peças diferentes. A do segundo modelo tem 22 peças. Esses conjuntos serão designados por modelo 1 e modelo 2, respetivamente.

A diversidade de peças entre os dois modelos é grande. No modelo 1, a peça de maior dimensão é cerca de 28 vezes maior que a peça mais pequena (10 vezes no caso do modelo 2). Uma análise comparativa entre as áreas (A), perímetros (P) e áreas dos retângulos envolventes (R) das peças de cada um dos modelos permite ter uma ideia sobre o grau de irregularidade das peças envolvidas. Em ambos os modelos, a razão entre (A) e (R) varia entre 41% e 99%. Valores mais baixos indiciam uma maior irregularidade da peça em causa enquanto que valores mais elevados apontam para uma peça com contornos mais regulares. As tabelas 1 e 2 apresentam os dados para as peças dos modelos 1 e 2.

Na nossa implementação, considerámos um número limitado de rotações para cada peça. Para que uma peça rodada resulte numa peça significativamente diferente da figura original, foi aplicado um conjunto de regras de rotação para diferentes tipos de peças. Para as peças grandes e médias com várias concavidades, foram aplicadas 7 rotações múltiplas de 45°. Para as peças com poucas concavidades, foram aplicadas apenas rotações múltiplas de 90°. Para as peças com dois eixos de simetria (horizontal e vertical), foram considerados dois casos: se a razão entre a largura e o

comprimento do retângulo envolvente for muito diferente de 1, então são consideradas 3 rotações (45°, 90° e 135°). Caso contrário, é considerada apenas uma rotação de 45°. As peças de pequenas dimensões são rodadas apenas em 90°. No final, essas regras dão origem a 75 rotações para o modelo 1 e a 68 rotações para o modelo 2. Contando com a representação original (não rodada) das peças, os modelos 1 e 2 acabam por ter respectivamente 98 e 90 peças diferentes.

Tabela 1: Caracterização das peças do modelo 1

Peças	Zonas de qualidade	Área (cm ²) (A)	Perímetro (cm) (P)	Área do retângulo envolv. (cm ²) (R)	Razão A/R
1	B, D	337.21	85.81	342.19	0.99
2	B, D	598.55	116.66	869.19	0.69
3	A, D	858.80	165.21	1311.78	0.65
4	A, D	648.54	117.60	1038.15	0.62
5	B, D	983.56	137.66	1425.91	0.69
6	B, D	804.93	173.76	1356.72	0.59
7	B, D	598.79	114.68	914.06	0.66
8	A, D	$\frac{1607.4}{4}$	216.60	2138.60	0.75
9	A, D	92.55	39.68	119.52	0.77
10	A, D	70.06	37.36	110.92	0.63
11	A, D	70.06	37.36	110.92	0.63
12	A, D	92.55	39.68	119.52	0.77
13	A, D	58.15	38.44	115.51	0.50
14	A, D	58.15	38.44	115.51	0.50
15	A, D	778.61	118.91	1165.80	0.67
16	B, D	469.46	90.33	498.32	0.94
17	A, D	417.73	97.43	521.01	0.80
18	A, D	115.04	49.62	149.58	0.77
19	A, D	778.61	118.91	1165.80	0.67
20	A, D	417.73	97.43	519.26	0.80
21	A, D	$\frac{1607.4}{4}$	216.60	2135.64	0.75
22	B, D	835.62	169.54	1378.87	0.61
23	B, D	851.76	123.40	1242.99	0.69

Tabela 2: Caracterização das peças do modelo 2

Peças	Zonas de qualidade	Área (cm ²) (A)	Perímetro (cm) (P)	Área do retângulo envolv. (cm ²) (R)	Razão A/R
1	A, B, C	426.81	80.00	503.44	0.85
2	A, C	205.79	70.39	265.41	0.78
3	A, C	377.01	105.72	678.03	0.56
4	A, C	205.79	70.39	265.41	0.78
5	A, C	377.01	105.72	678.03	0.56
6	A, B, C	660.46	109.24	717.43	0.92
7	A, B, C	660.46	109.24	718.87	0.92
8	A, B, C	872.18	132.02	1082.06	0.81
9	A, C	362.35	83.23	504.35	0.72
10	A, C	418.77	125.73	1013.83	0.41
11	A, C	362.35	83.23	504.35	0.72
12	A, C	418.22	122.02	987.18	0.42
13	A, B, C	1878.19	206.64	2391.15	0.79
14	A, C	489.57	94.11	637.26	0.77
15	A, C	489.57	94.11	637.26	0.77
16	A, B, C	795.16	123.98	1082.79	0.73
17	A, B, C	766.67	122.66	1019.66	0.75
18	A, B, C	593.51	121.87	974.66	0.61
19	A, B, C	593.54	121.82	974.83	0.61
20	A, C	279.84	68.01	362.98	0.77
21	A, C	499.28	95.36	696.01	0.72
22	A, B, C	529.91	148.37	993.97	0.53

Para cada modelo, foram criadas várias instâncias a partir de diferentes combinações de peças que correspondem a ordens de fabrico. O número de peças usadas para cortar essas ordens é igual a 3. A Tabela 3 descreve os detalhes de cada uma dessas ordens de fabrico: o número de peças diferentes é indicado na coluna DPeças e o número total de peças na coluna NPeças.

O primeiro conjunto de experiências foi realizado com a ordem de fabrico 1 do modelo 1. Cada uma das classes de estratégias foi avaliada de forma individual. Uma vez que a combinação de todas as estratégias origina um número elevado de algoritmos diferentes, a avaliação de uma determinada

categoria de estratégias é efetuada restringindo o conjunto de estratégias pertencentes às outras categorias. As estratégias foram avaliadas com base no número de peças colocadas nas peles, na percentagem de aproveitamento de cada uma das peles, no tempo de computação necessário para gerar cada um dos planos de corte, e na área útil da última pele. A descrição exaustiva dos resultados obtidos pode ser consultada em (Alves *et al.*, 2012a).

Tabela 3: Instâncias (ordens de fabrico)

MODELO	ORDEM	DPeças	NPeças
1	1	23	60
	2	8	80
	3	23	80
	4	8	100
	5	23	100
	6	8	120
	7	23	120
2	1	22	60
	2	9	80
	3	22	80
	4	22	100

Essas experiências permitiram identificar as combinações de estratégias que dão origem aos algoritmos mais rápidos (Tabela 4). O critério usado nesse caso foi o tempo necessário para a construção do plano de corte referente à primeira pele. Além disso, foram identificadas também as combinações de estratégias que conduzem às melhores soluções em termos de aproveitamento obtido na primeira pele. Na Tabela 4, o índice de cada uma dessas combinações é indicado na coluna IND. As heurísticas mais rápidas correspondem às combinações 1 a 10. As heurísticas que geram as melhores soluções correspondem às combinações 11 a 20. Na Tabela 4, são indicados os casos em que se procuram pontos de posicionamento que maximizam ou minimizam as funções de avaliação (para os casos em que possam existir ambiguidades).

Tabela 4: Combinações de estratégias que resultam nas heurísticas mais rápidas (1-10) e nas heurísticas que geram os melhores planos de corte (11-20)

IND	Agrupamento	Próxima peça	Região de posicionamento admissível	Avaliação dos pontos de posicionamento
1	(G1)	(S2.D.1.a)	-	(E1)
2	(G5)	(S2.D.1.a)	-	(E1)
3	(G6)	(S2.D.1.a)	-	(E1)
4	(G1)	(S2.D.1.c)	-	(E1)
5	(G5)	(S2.D.1.c)	-	(E1)
6	(G5)	(S2.D.1.c)	-	(E12.MAX)
7	(G5)	(S2.D.1.c)	-	(E14.MIN)
8	(G5)	(S2.D.1.c)	-	(E14.MAX)
9	(G5)	(S2.D.1.c)	-	(E13.MIN)
10	(G5)	(S2.D.1.c)	-	(E13.MAX)
11	(G5)	(S1.D)	(P1)	(E1)
12	(G5)	(S1.D)	(P6)	(E1)
13	(G1)	(S1.D)	(P7)	(E1)
14	(G5)	(S1.D)	(P7)	(E1)
15	(G6)	(S1.D)	(P7)	(E1)
16	(G5)	(S1.D)	(P9)	(E1)
17	(G5)	(S1.D)	(P12)	(E1)
18	(G2)	(S2.D.2)	(P2)	(E1)
19	(G1)	(S2.D.2)	(P7)	(E1)
20	(G1)	(S2.D.2)	(P12)	(E1)

Na segunda série de experiências computacionais, foram usadas todas as ordens de fabrico descritas na Tabela 3. Uma descrição exhaustiva dos resultados obtidos pode ser consultada em (Alves *et al.*, 2012a). Todos os algoritmos mais rápidos (1 a 10) geraram planos de corte dentro do limite de 600 segundos. O tempo de computação mais elevado para a geração de um plano de corte na primeira pele foi de 148,36 segundos. O tempo de computação dos algoritmos depende do número de peças diferentes na ordem de fabrico, e do número de NFP e IFP que têm de ser calculados entre cada par de peças e entre as peças e a pele.

Para algumas ordens de fabrico do modelo 1 com 8 peças diferentes, esses algoritmos geraram planos de corte para a primeira pele em menos de 20 segundos. O algoritmo 7 é o mais rápido em 3 (de 11) ordens. Para as restantes instâncias, o algoritmo mais rápido foi o algoritmo 9. Os algoritmos

1 a 5 demoram geralmente mais tempo a completar as ordens do que os algoritmos 6 a 10 devido ao peso computacional envolvido no cálculo dos *offsets*.

Entre os algoritmos mais rápidos (1 a 10), o algoritmo 5 é o que gera os planos de corte com os melhores aproveitamentos na maior parte dos casos. Esse algoritmo combina as estratégias (G5) com (S2.D.1.c) e (E1). Para 7 ordens de fabrico, esse algoritmo conduz à melhor solução quer em termos de aproveitamento na primeira pele, quer em termos de percentagem de área útil que sobra na última pele. Para a ordem 2 do modelo 1, esse algoritmo necessita apenas de duas peles para completar a ordem, enquanto os restantes (do conjunto dos algoritmos mais rápidos) necessitam das três peles. O tempo máximo de computação requerido por este algoritmo para gerar um plano de corte para a primeira pele foi de 126.80 segundos (ordem 7 do modelo 1).

Os resultados obtidos para os algoritmos 11 a 20 mostram que a maioria destes algoritmos não é capaz de gerar planos de corte em menos de 600 segundos. A exceção é o algoritmo 12 que combina as estratégias (G5), (S1.D), (P6) e (E1), e que consegue gerar planos de corte dentro do limite de tempo definido para todas as ordens exceto uma. A utilização da estratégia (P6) para a seleção da região de posicionamento admissível permite restringir o conjunto de pontos de posicionamento que têm de ser inspecionados. O critério usado para a seleção dessa região baseia-se nas características dos IFP entre a peça e a pele. Quando combinada com a estratégia (S1.D) de seleção da próxima peça a colocar, essa estratégia permite obter soluções de boa qualidade sem comprometer o tempo de computação.

O algoritmo 12 conduz às melhores soluções em termos de aproveitamento da primeira pele em 7 ordens de fabrico. Para além disso, para 4 ordens de fabrico, esse algoritmo gera também as melhores soluções em termos de percentagem de área útil que sobra na última pele.

Para avaliar o desempenho do algoritmo de pesquisa por vizinhança variável, foram realizadas também duas séries de experiências computacionais: a primeira teve por objetivo a afinação dos parâmetros do

algoritmo e a comparação de estratégias alternativas, enquanto que o segundo conjunto de experiências consistiu na avaliação do desempenho das melhores abordagens identificadas na fase de afinação do algoritmo.

Os objetivos específicos da primeira série de experiências computacionais foram os seguintes:

- a) avaliar o impacto de cada vizinhança na qualidade dos planos de corte gerados;
- b) comparar diferentes estratégias para a definição dos intervalos de aproveitamento (u_{min}, u_{max}) com base nos quais as peças de P são selecionadas;
- c) avaliar valores alternativos para os parâmetros associados ao tamanho dos conjuntos P a partir dos quais são selecionadas as peças p que irão ser substituídas, permutadas ou eliminadas da sequência;
- d) avaliar valores alternativos para os parâmetros associados ao tamanho dos conjuntos P' a partir dos quais são selecionadas as peças a serem inseridas na sequência;
- e) comparar a qualidade dos planos de corte obtidos através da exploração de diferentes sequências de vizinhanças.

Foram usadas 24 instâncias definidas a partir dos modelos 1 e 2 descritos acima. Para o modelo 1, foram criadas 4 ordens de fabrico com diferentes combinações das 23 peças disponíveis. O número de peças diferentes em cada ordem foi de 8, 12, 15 e 23, respetivamente, enquanto que o número total de peças nessas ordens foi igual a 100. Para cada ordem, o algoritmo foi repetido com 3 peles diferentes. No total, são assim 12 instâncias que foram definidas a partir do modelo 1. Para o modelo 2, foram criadas 12 ordens distintas a partir de um conjunto de 22 peças disponíveis. O número de peças diferentes em cada ordem varia entre 5 e 22 (5,7,8,9,12,14,20,22). O número total de peças por ordem foi igual a 100. Foi usada a mesma pele para todas essas ordens de fabrico. A execução do algoritmo foi limitada a 600 segundos.

Os resultados obtidos nessa fase de afinação podem ser consultados em detalhe em (Alves *et al.*, 2012b). Essas experiências permitiram identificar as

estratégias promissoras para o algoritmo de pesquisa por vizinhança variável. Essas estratégias passam por usar as 4 estruturas (gerais) de vizinhança que é possível definir a partir dos movimentos descritos na Seção 7. Para cada uma dessas estruturas gerais, foram identificadas estruturas de vizinhança promissoras que são definidas a partir dos parâmetros seguintes: $q = 1, r = 3, (u_{min}, u_{max}) = (10\%, 50\%)$; $q = 3, r = 3, (u_{min}, u_{max}) = (10\%, 50\%)$.

Esses parâmetros e estratégias foram usados na segunda série de experiências computacionais num conjunto de 107 instâncias descritas na Tabela 5.

Tabela 5: Instâncias

MODELO	Ordem de fabrico	N.º de peças diferentes
1	1,..., 5	5
	6,...,10	8
	11,...,15	10
	16,...,18	12
	19,...,21	15
	22	23
2	1,...,10	5
	11,...,20	8
	21,...,30	10
	31,...,35	12
	36,...,40	15
	41	22

A lista exaustiva de resultados é descrita em (Alves *et al.*, 2012b). Lembramos que o plano de corte inicial é gerado usando a heurística construtiva apresentada neste artigo com as estratégias (G1), (S2.D.1.c), (P6) e (E11). O algoritmo de pesquisa por vizinhança variável conseguiu melhorar na grande maioria dos casos a solução gerada inicialmente por essa heurística construtiva. Após 600 segundos de computação, o algoritmo permitiu uma melhoria da solução inicial em cerca de 2.74% em termos de aproveitamentos. Em cerca de metade das instâncias, o algoritmo gerou um plano de corte de melhor qualidade do que o inicial nos primeiros 200 segundos de computação. Após 400 e 600 segundos, foram registradas melhorias

relativamente à solução inicial em cerca de 84% e 96% dos casos, respetivamente. A Figura 8 mostra a evolução do nível médio de aproveitamento ao longo da execução do algoritmo.

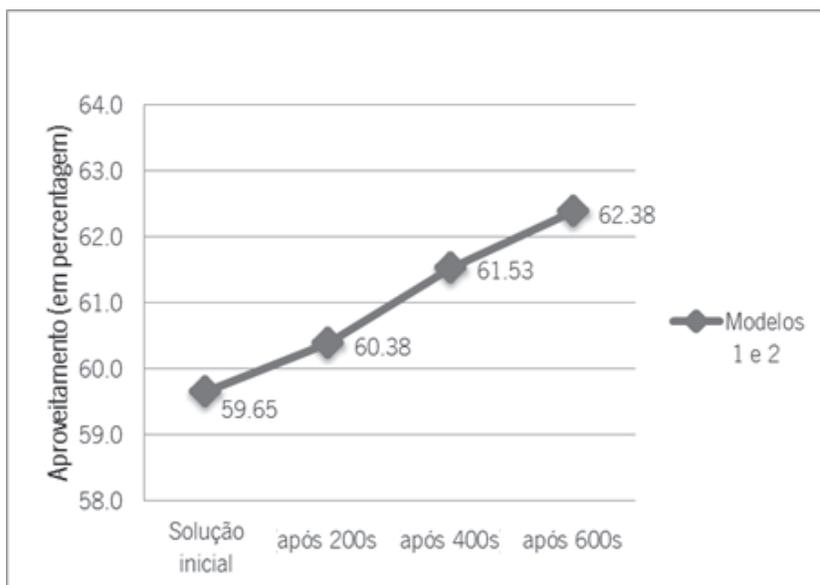


Figura 8: Evolução dos níveis médios de aproveitamento ao longo da execução do algoritmo de pesquisa por vizinhança variável

Os níveis de aproveitamento das peles assim como as melhorias que o algoritmo de pesquisa por vizinhança variável permite alcançar dependem do modelo de carro que é considerado. A Figura 9 apresenta a evolução das soluções em termos de percentagem de aproveitamento para os dois modelos. Os gráficos ilustram a qualidade das soluções iniciais e a capacidade do algoritmo em melhorar essas soluções.

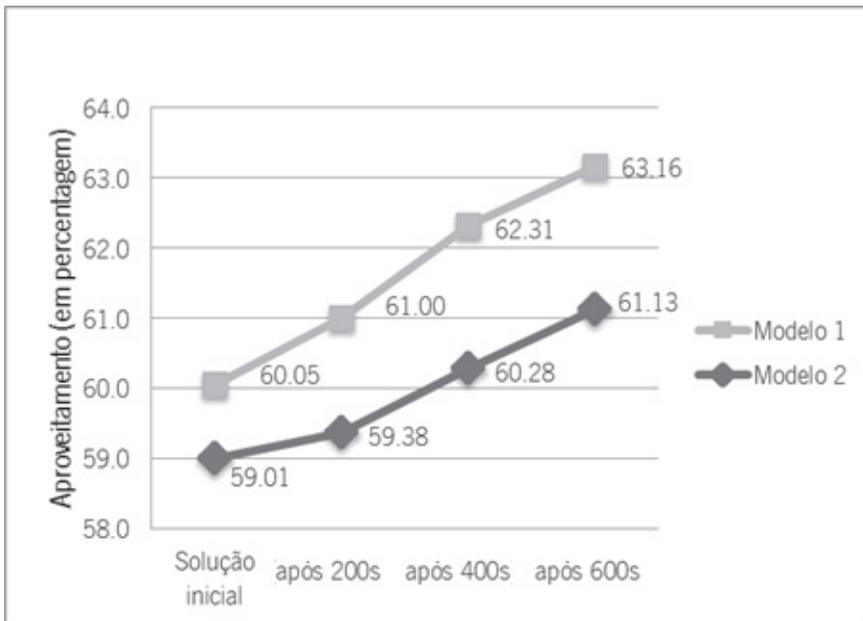


Figura 9: Evolução dos níveis de aproveitamento para os modelos 1 e 2

É importante salientar que os níveis de aproveitamento do couro que são alcançados por estes algoritmos são competitivos com os resultados obtidos pelos operadores manuais da Coindu. Além disso, os resultados são obtidos no mesmo intervalo de tempo que é necessário a dois operadores para montarem um plano de corte de forma manual. Na Coindu, o tempo médio de ciclo para o corte manual de couro é de aproximadamente 15 minutos por pele. Esse valor inclui a fase de revista da pele (5 minutos em média), e poderá variar consideravelmente, dependendo do tipo de couro, tamanho das peles, nível de qualidade e ordem de fabrico. O processo de colocação dos cortantes é executado, em média, nos 10 minutos restantes. Um único operador necessitaria do dobro do tempo (1200 segundos) para construir planos de corte com os mesmos níveis de eficiência.

9. Conclusões

O problema de corte em peles de couro é claramente um dos problemas de corte e empacotamento a duas dimensões mais difíceis que pode ser encontrado na indústria. Apesar da sua relevância prática, poucas abordagens foram propostas na literatura, e mesmo aquelas que tratam de problemas similares assentam na realidade em pressupostos e restrições que resultam numa simplificação do problema.

O valor financeiro desta matéria-prima e a dimensão económica do setor de mercado que a comercializa confere uma grande importância ao problema de corte em peles de couro. Para uma empresa como a Coindu cujo setor de corte de couro tem elevados níveis de produção, pequenos ganhos no aproveitamento desta matéria-prima convertem-se numa redução significativa dos custos totais. No ano de 2010, por exemplo, no conjunto das três unidades fabris, os níveis de produção da Coindu implicaram um consumo anual que superou o milhão de metros quadrados de couro, o que se traduziu num investimento de algumas dezenas de milhões de euros.

Neste trabalho, procurámos avaliar o potencial de duas famílias de algoritmos heurísticos para a elaboração automática de planos de corte. Analisámos um conjunto de heurísticas construtivas e um algoritmo de pesquisa por vizinhança variável. Para avaliar o desempenho desses algoritmos, foram conduzidos extensos conjuntos de experiências computacionais em instâncias reais da empresa.

Dois tipos de testes foram realizados para identificar e explorar as estratégias da heurística construtiva com melhor desempenho. Numa primeira fase, comparámos o desempenho das diferentes estratégias. A partir dos resultados destes testes, concluiu-se que a área e as zonas de qualidade das peças são fatores importantes para o agrupamento de peças. As estratégias de seleção da próxima peça a posicionar e da região de posicionamento admissível que geram os melhores resultados são aquelas que consideram os IFP das peças. Por fim, os testes computacionais revelam que as funções de avaliação dos pontos de posicionamento que envolvem o cálculo de *offsets* estão na origem de alguns dos melhores resultados obtidos.

A segunda fase de testes consistiu na aplicação das melhores e mais rápidas combinações de estratégias a todas as instâncias. Dos resultados obtidos, destacam-se os algoritmos 5 e 12 (Tabela 4). O algoritmo 5, que inclui algumas das estratégias mais rápidas, apresenta níveis competitivos de aproveitamento na generalidade dos testes realizados e envolve tempos de computação muito reduzidos. O algoritmo 12, apesar de requerer tempos de computação mais elevados, consegue alcançar níveis de aproveitamento que são na sua generalidade os mais elevados de todos os resultados apresentados. Importa também destacar o conjunto de algoritmos 6 a 10 pela sua rapidez na geração de planos de corte e pelos níveis de aproveitamento razoáveis. Este conjunto de algoritmos são uma boa alternativa para a inicialização de métodos de melhoria.

O algoritmo de pesquisa por vizinhança variável foi testado com base num conjunto significativo de instâncias com peças provenientes de modelos de carros que se encontram em produção na Coindu. As experiências realizadas mostram que o algoritmo foi capaz de melhorar a solução inicial em 96% das instâncias testadas. Em cerca de 48% dos casos, foi encontrada uma solução melhor do que a inicial nos primeiros 200 segundos de execução. Em todos os testes foi considerado um limite no tempo de computação de 600 segundos. Esse valor corresponde ao tempo médio necessário para uma equipa de dois operadores preencher uma pele de forma manual. Em termos globais, os testes confirmaram a qualidade dos planos de corte quando comparados com os resultados médios alcançados pelos operadores manuais da empresa.

BIBLIOGRAFIA

Alves, C., Brás, P., Valério de Carvalho, J., & Pinto, T. (2012). New constructive algorithms for leather nesting in the automotive industry. *Computers & Operations Research*, 39(7), 1487-1505.

Alves, C., Brás, P., Valério de Carvalho, J. M., & Pinto, T. (2012). A variable neighborhood search algorithm for the leather nesting problem. *Mathematical Problems in Engineering*, 2012, 1-28.

Babu, A., & Babu, N. (2001). A generic approach for nesting of 2-D parts in 2-D sheets using genetic and heuristic algorithms. *Computer-Aided Design*, 33(12), 879-891.

Bennell, J. A., & Dowsland, K. A. (1999). A tabu thresholding implementation for the irregular stock cutting problem. *International Journal of Production Research*, 37(18), 4259-4275.

Bennell, J. A., & Oliveira, J. F. (2008). The geometry of nesting problems: A tutorial. *European Journal of Operational Research*, 184(2), 397-415.

Crispin, A., Clay, P., Taylor, G., Bayes, T., & Reedman, D. (2005). Genetic algorithm coding methods for leather nesting. *Applied Intelligence*, 23(1), 9-20.

Hansen, P., Mladenović, N., & Pérez, J. A. M. (2010). Variable neighbourhood search: methods and applications. *Annals of Operations Research*, 175(1), 367-407.

Heckmann, R., & Lengauer, T. (1995). A simulated annealing approach to the nesting problem in the textile manufacturing industry. *Annals of Operations Research*, 57(1), 103-133.

Heistermann, J., & Lengauer, T. (1995). The nesting problem in the leather manufacturing industry. *Annals of Operations Research*, 57(1), 147-173.

Wäscher, G., Haußner, H., & Schumann, H. (2007). An improved typology of cutting and packing problems. *European Journal of Operational Research*, 183(3), 1109-1130.

Yuping, Z., Shouwei, J., & Chunli, Z. (2005). A very fast simulated re-annealing algorithm for the leather nesting problem. *The International Journal of Advanced Manufacturing Technology*, 25(11-12), 1113-1118.

Yuping, Z., & Caijun, Y. (2009). A generic approach for leather nesting. *Proceedings of the Fifth International Conference on Natural Computation*, 5, 303-307.

Parte

III

SETOR da

LOGÍSTICA E

TRANSPORTES

(Página deixada propositadamente em branco)

CASO 11

PROJETO DE UMA REDE LOGÍSTICA PARA A RECOLHA DE EQUIPAMENTOS ELÉTRICOS E ELETRÓNICOS

Maria Isabel Gomes

CMA – FCT - Universidade Nova de Lisboa
mirg@fct.unl.pt

Ana Paula Barbosa Póvoa

CEG – IST, Instituto Superior Técnico
Universidade de Lisboa
apovoa@tecnico.ulisboa.pt

RESUMO

A Amb3E tem por missão a definição e gestão de um sistema integrado de recolha, tratamento e valorização dos resíduos de equipamentos elétricos e eletrónicos. A rede de recolha para estes equipamentos encontra-se em funcionamento desde 2006 sem, no entanto, ter havido qualquer estudo científico relativamente à estrutura da rede e ao seu planeamento ao nível tático. Foi pois estudada não só a definição da rede de custo total mínimo, como também da avaliação do impacto ambiental da mesma e a determinação da estrutura ótima de acordo com este segundo objetivo. O estudo efetuado tem por base um modelo linear inteiro misto especialmente desenvolvido para o efeito. Este modelo é um modelo de localização de instalações, multi-período e multi-produto. Numa primeira fase foram definidas as redes ótimas segundo cada um dos objetivos: económico e ambiental. Numa segunda fase, efetuou-se uma análise multi-objetivo recorrendo ao método das restrições. Dadas as dimensões do problema, esta última análise foi efetuada apenas para a região de Lisboa e Vale do Tejo.

PALAVRAS-CHAVE

Logística inversa, Localização de instalações, Multi-objectivo, Equipamentos eléctricos e eletrónicos

1. Introdução

O rápido crescimento da produção de resíduos de equipamentos elétricos e eletrônicos (REEE) transformou a sua deposição num problema transnacional. A presença de elevadas quantidades de substâncias tóxicas nestes produtos tem causado, nos últimos anos, graves problemas ambientais. Sendo os objetivos da política ambiental da União Europeia a

“preservação, proteção e melhoria da qualidade do ambiente, a proteção da saúde das pessoas e a utilização prudente e racional dos recursos naturais”

foi publicada em 2003, a diretiva 2002/96/CE que impõe aos países membros a criação de estruturas que permitam uma “recolha separada” dos REEE e o seu posterior tratamento e valorização dentro de parâmetros estabelecidos a nível da União Europeia, de modo a reduzir a deposição destes produtos em aterros sanitários. Todos os países deveriam transpor esta diretiva para os respetivos quadros jurídicos até ao final do ano de 2004. Na transposição para a lei portuguesa, o dia 31 de Dezembro de 2006 foi estabelecido como prazo para que fosse atingido um conjunto de metas quantitativas, garantindo o cumprimento da diretiva europeia.

Neste sentido, um conjunto de produtores de equipamentos elétricos e eletrônicos (EEE) associou-se criando a Amb3E – Associação Portuguesa de Gestão de Resíduos. Esta associação tem por missão a definição e gestão de um sistema integrado de recolha, tratamento e valorização dos REEE. No processo de definição da rede de recolha destes equipamentos surgiram diversas questões das quais se destacam: Onde localizar os centros de receção? Quantos centros abrir em cada localização? Qual deve ser a capacidade de processamento de cada centro de modo a garantir o cumprimento das metas legais? Quais os tipos de transporte que devem ser subcontratados?

Aquando da realização deste trabalho, a rede de recolha da Amb3E já se encontrava em operação. No entanto, não tinha havido qualquer estudo sistematizado relativo à estrutura da rede e ao seu planeamento a nível tático. Era pois do interesse da associação a definição da rede de custo total mínimo considerando-se entre outros os custos de transporte, de armazenamento e de

processamento. Por outro lado, um dos pontos que deu origem à diretiva europeia prendia-se com questões ambientais. Assim, considerou-se uma avaliação dos impactos ambientais da rede onde foram integrados diversos impactos, entre os quais os causados pela abertura das instalações, pelo transporte e pelo processamento.

Para efetuar o estudo descrito desenvolveu-se um modelo linear inteiro misto (MILP) com as características de um modelo de localização de instalações, multi-período e multi-produto. Neste a rede de recolha modelada é formada por três níveis: as fontes geradoras de resíduos, os centros de recepção (CR) e as unidades de tratamento e valorização (UTV). Consideraram-se todos os 278 municípios de Portugal continental como potenciais localizações dos centros de recepção assim como fontes geradoras de resíduos. Dado que a associação pretendia estabelecer protocolos com as unidades de tratamento e valorização, foram consideradas neste estudo as oito unidades a operar em Portugal.

Numa primeira fase foi definida a rede de custo mínimo e comparada a estrutura de custos obtida com a realmente verificada pela Amb3E. Uma análise semelhante foi efetuada tendo por base a minimização de impactos ambientais. Com recurso ao método das restrições, efetuou-se também uma análise multi-objetivo considerando-se os objetivos económico e ambiental. Dadas as dimensões do problema, esta última análise foi efetuada apenas para a região de Lisboa e Vale do Tejo. Os resultados obtidos fornecem informações importantes ao decisor para a avaliação da rede atual quando comparada com a rede ótima quer em termos de custos, quer em termos de impactos ambientais, permitindo ainda analisar os níveis de recolha conseguidos.

2. Características da rede de recolha de REEE da Amb3E

A Amb3E é uma das entidades responsáveis pela gestão do sistema de recuperação de REEE em Portugal. Esta associação sem fins lucrativos tem por missão gerir eficientemente um sistema integrado de gestão de resíduos de equipamentos elétricos e eletrónicos. O problema em estudo surge na sequência do enorme crescimento da sua estrutura logística, nomeadamente,

do aumento do número de centros de recepção e de unidades de tratamento e valorização. Este crescimento foi motivado não só pelo aumento do consumo de EEE e, conseqüente geração de resíduos, mas também pela crescente consciencialização dos cidadãos para a problemática dos REEE. É assim necessário repensar a estrutura da rede estabelecida e planear adequadamente a rede de recolha dos REEE para a Amb3E considerando não só a questão económica mas também a questão do impacto ambiental deste sistema.

2.1 A situação atual

No final do ano de 2008, a Amb3E contava com 640 produtores aderentes (Figura 1) que colocaram no mercado, nesse mesmo ano, cerca de 30 000 unidades de EEE perfazendo um total de 116 000 t (Figura 2).

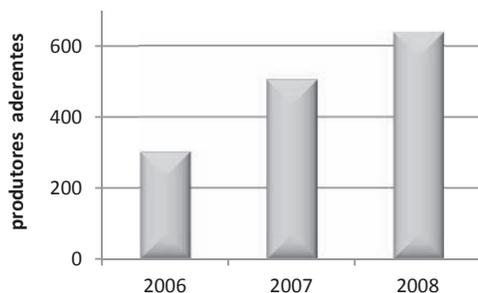


Figura 1: Número de produtores aderentes à Amb3E (Amb3E, 2009)

De acordo com o n° 10 do artigo 9° do Decreto-lei n° 230/2004, a meta legal para a recolha seletiva de REEE foi estabelecida em, pelo menos, 4 kg/(habitante*ano). Em 2008, a Amb3E assegurou a recolha de cerca de 32 990 t, correspondendo a apenas 3.11 kg/habitante. No entanto, apesar de à primeira vista não ter sido atingida a meta, a licença da Amb3E prevê um mecanismo de ajustamento da meta de recolha em função da quota de mercado real da associação, isto é, da quota de mercado detida pelos produtores associados. Considerando a meta e a população nacionais, o quantitativo nacional de recolha é cerca de 42 470 t de REEE provenientes de particulares. Sendo a quota de mercado dos produtores aderentes à Amb3E, em peso de EEE colocados no mercado, de cerca de 70%, a Amb3E tinha

como objetivo de recolha, para esse ano, aproximadamente 29 650 t, o que correspondia a 2.8 kg/habitante. Pelo que a meta foi plenamente atingida.

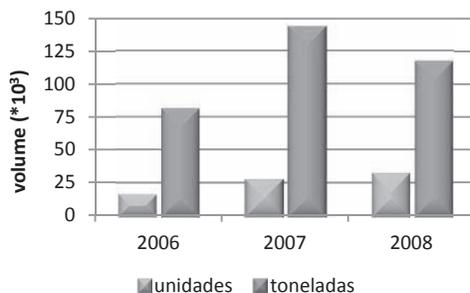


Figura 2: Volume de EEE colocados no mercado (Amb3E, 2009)

Em termos de quantidade recolhida podemos ver na Figura 3 que esta tem aumentado significativamente nos três primeiros anos de operação, tendo-se observado um crescimento de 61% de 2007 para 2008.

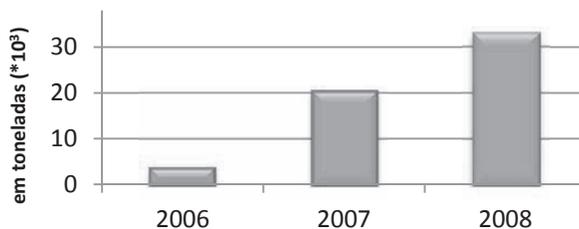


Figura 3: Quantidade recolhida pela Amb3E (Amb3E, 2009)

2.2 Categorização dos produtos

A diretiva europeia WEEE (2002/96/EC), bem como a legislação nacional, separam os equipamentos elétricos e eletrónicos em 10 categorias legais: (1) Grandes eletrodomésticos, (2) Pequenos eletrodomésticos, (3) Equipamentos informáticos e de telecomunicações, (4) Equipamentos de consumo, (5) Equipamentos de iluminação, (6) Ferramentas elétricas e eletrónicas, (7) Brinquedos e equipamentos de desporto e lazer, (8) Aparelhos médicos, (9) Instrumentos de monitorização e controlo e (10) Distribuidores automáticos. Contudo, por uma questão de operacionalidade e maior facilidade no controlo

dos resíduos tratados, a Amb3E reformulou as 10 categorias legais em cinco fluxos operacionais (de acordo com o apresentado na Tabela 1). Em termos de distribuição percentual, vemos que o fluxo A (grandes equipamentos) representa 35.52% do total de toneladas recolhidas, seguindo-se o fluxo B (equipamentos de frio). Com um volume de recolha menor estão os equipamentos de iluminação (fluxo E) que representam apenas 1.45% do peso total recolhido.

Tabela 1: Fluxos operacionais da Amb3E (Amb3E, 2009)

Fluxos operacionais	Categorias legais	Distribuição percentual
A – Grandes equipamentos	1	35.52
B – Equipamentos de refrigeração e congelação contendo CFC	1, 8, 10	25.90
C – Pequenos equipamentos	2, 3, 4, 5, 6, 7, 8, 9, 10	23.37
D – Aparelhos de televisão e monitores de tubos de raios catódicos (CRT)	3,4	13.76
E – Equipamentos de iluminação	5	1.45

2.3 A rede logística da recolha

A rede logística da Amb3E é formada por três elementos principais: os locais de recepção, os operadores logísticos e as unidades de tratamento e valorização. Estas entidades são responsáveis pelas três atividades principais da rede: recepção, recolha e triagem, e tratamento e valorização (Figura 4).

No final de 2008, a rede de locais de recepção cobria todo o território nacional contabilizando 315 locais, tendo-se registado um crescimento de 3.2 vezes relativamente ao ano anterior. Dependendo do papel que desempenham na rede, os locais de recepção podem ser: centros de recepção, locais de recolha e pontos eletrão.

Os centros de recepção são as entidades centrais desta rede logística. Apesar de serem externos à Amb3E contratam com esta um conjunto de

atividades por contrapartida financeira. Nomeadamente, é sua obrigação receber gratuitamente os REEE de proveniência particular e não-particular, separá-los nos cinco fluxos operacionais e armazená-los até que possam ser transportados para uma unidade de tratamento e valorização. É a atividade de triagem feita nestes centros que os torna fundamentais para a Amb3E e para o problema em estudo.

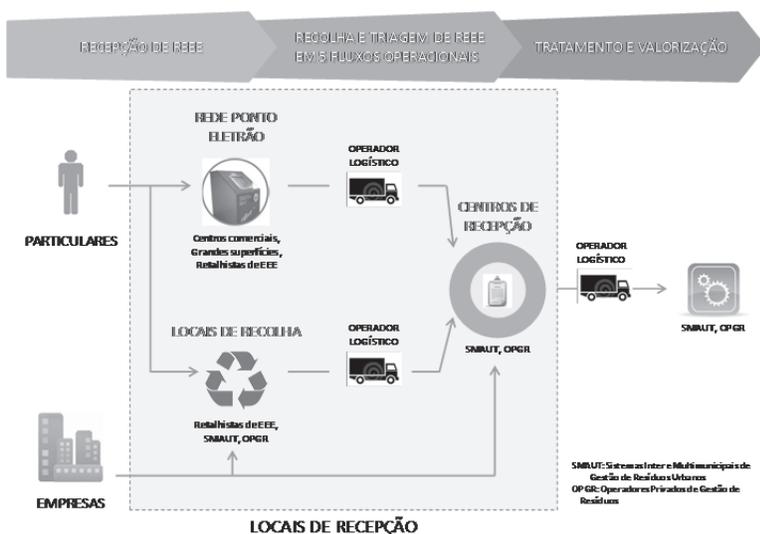


Figura 4: Rede de recolha da Amb3E (baseado em Amb3E, 2009)

Os locais de recolha são instalações onde se pode entregar gratuitamente REEE para que sejam armazenados e, posteriormente, expedidos. Estes locais funcionam de forma complementar aos centros de recepção, pelo que não efetuam a operação de triagem. Estas instalações são constituídas por serviços municipais e intermunicipais de gestão de resíduos urbanos, grandes produtores de REEE ou, mais recentemente, entidades mais próximas da população, como associações humanitárias de bombeiros voluntários.

Finalmente, os pontos eletrão são, muito provavelmente, a face mais visível do trabalho desenvolvido pela Amb3E. Estes são pequenos contentores que podem ser encontrados em centros comerciais, bombas de gasolina, ou outras grandes superfícies e que se destinam a receber REEE até média dimensão.

Disponíveis desde o final de 2007, tiveram como grande objetivo aproximar os consumidores da iniciativa, consciencializá-los para a problemática dos REEE e promover uma maior recolha deste tipo de resíduos.

Adicionalmente, a Amb3E disponibiliza um sistema de recolhas especiais aos seus associados, utentes e operadores com importância estratégica. Tal como o nome do serviço indica, a Amb3E recolhe os REEE junto dos seus produtores/detentores, assegurando o encaminhamento adequado para as unidades de tratamento e valorização. Além de ser um serviço de elevada mais-valia para os seus aderentes, estas recolhas especiais permitem à Amb3E aceder a quantidades adicionais de resíduos de forma integrada e otimizada com o restante processo.

Todos os REEE que chegam aos locais de recepção da Amb3E têm de ser transportados até às unidades de tratamento e valorização por operadores logísticos. A rede nacional de operadores logísticos efetua o transporte dos REEE para os destinos definidos pela Amb3E num regime de *outsourcing*, assegurando assim o eficiente e adequado tratamento dos resíduos e a sua valorização.

O tratamento dos REEE de todas as categorias legais e a totalidade das quantidades recebidas neste sistema integrado é feito através da rede de unidades de tratamento e valorização. Esta desenvolve práticas eficientes de tratamento dos REEE nos cinco fluxos operacionais definidos, o que permite recuperar o valor material e energético contido nos resíduos, assegurando a sua reciclagem e valorização. Os fluxos operacionais em que são processados os REEE englobam todos os equipamentos das 10 categorias legais anteriormente descritas.

Tendo iniciado as operações com duas UTVs, a Amb3E aumentou em 2008 a sua rede para 12 unidades de modo a assegurar o tratamento adequado de todos os REEE geridos no âmbito do sistema. Destas 12 unidades, apenas oito têm atividade em Portugal continental, pelo que serão estas as consideradas neste trabalho. Apesar do elevado número de lâmpadas recolhido, o seu peso diminuto torna este fluxo num fluxo de difícil tratamento logístico, tanto em questões de armazenamento como de tratamento. Devido a estas características, o tratamento deste resíduo é efectuado apenas por uma UTV

que se lhe dedica em exclusividade. As restantes sete unidades processam os outros quatro fluxos.

Neste contexto e dado que até à data de realização deste trabalho a Amb3E não tinha ainda efetuado nenhum estudo sistematizado sobre a estrutura da sua rede logística, definiram-se os seguintes objetivos para este trabalho: (1) determinar as localizações ótimas para os centros de recepção (2) propor um planeamento tático de apoio à gestão dos fluxos que integram a rede e (3) proceder à avaliação dos impactos ambientais da rede. Dada a existência de dois tipos de métricas para a definição da rede (económica e ambiental) considerou-se ainda uma abordagem multi-objetivo da solução “ótima” para este problema.

3. Contexto teórico e modelo adotado

De há cerca de 20 anos a esta data têm sido publicados alguns estudos relacionados com a recolha de equipamentos elétricos e eletrónicos. Anteriormente à diretiva europeia, diversos produtores de EEE tinham-se já debruçado sobre esta questão tendo desenvolvido sistemas de recolha para os seus produtos em fim-de-vida. Muitos destes trabalhos foram efetuados em parceria com a academia. Um bom exemplo desta parceria é o trabalho publicado por Flapper et al. (2005) onde são apresentados (numa perspetiva de gestão) diversos estudos de caso relacionados com EEE em fim-de-vida.

Nesta revisão da literatura focaremos a nossa análise em trabalhos que, de alguma forma, se relacionam com a otimização do projeto de redes de recolha de equipamentos elétricos e eletrónicos. Sugerimos a consulta de Rogers et al. (2012) e Sheriff et al. (2012) onde se apresentam revisões da literatura sobre as questões da modelação em logística inversa, área na qual se insere o projeto de redes de recolha.

Do conhecimento dos autores deste trabalho, um dos primeiros artigos publicados relacionado com o desenho de redes de equipamentos elétricos e eletrónicos foi o trabalho de autoria de Krikke et al. (1999). Neste artigo os autores otimizam a rede de refabrico de um tipo de máquina produzida por uma empresa de máquinas copiadoras. Os autores propõem um modelo MILP onde as decisões principais se colocam ao nível da localização das fábricas e

do fluxo de produtos. O modelo baseia-se num modelo de localização e afetação onde se consideram também decisões ao nível do processo de refabrico a instalar em cada fábrica. O problema foi resolvido com recurso ao *solver* LINDO.

Walther e Spengler (2005) desenvolveram um modelo linear que permite determinar as frações de REEE que devem ser desmontadas, enviadas para reciclagem ou para deposição final, com o objetivo de maximizar o lucro marginal anual. Relativamente ao método de resolução, os autores apenas referem que recorreram a um “algoritmo baseado em programação linear”. Sharma et al. (2007) recorrem a uma formulação MILP para determinar quais as empresas com quem efetuar contratos de *leasing* assim como quanto tempo estes contratos deverão ser mantidos. Este modelo foi aplicado a um conjunto diverso de equipamentos eletrónicos de uma empresa sediada na Alemanha, tendo sido resolvido recorrendo ao *solver* Xpress-MP. Lieckens e Vandaele (2007) propõem uma formulação não linear onde integram decisões relativas à localização das instalações com decisões de inventário. A não linearidade provém da modelação dos tempos de aprovisionamento efetuada por um modelo de filas de espera G/G/1. As soluções do problema são obtidas com recurso a um algoritmo de *differential evolution*. Grunow e Gobbi (2009) desenvolvem um modelo para a afetação dos centros de recolha às diversas empresas de recolha em operação na Dinamarca. Esta formulação é posteriormente estendida com base em modelos de cobertura de modo a minimizar o número de empresas de recolha a operar num determinado município. Como método de resolução os autores recorrem ao *solver* CPLEX.

Recentemente, Alumur et al. (2012) propuseram uma formulação MILP para o projeto de redes de recolha. O modelo contempla produtos e tipos de instalações diferentes, num contexto multi-período. A formulação baseia-se nos modelos de localização de instalações multi-período onde as capacidades das instalações são definidas por módulos. Isto é, aquando da abertura da instalação, uma determinada capacidade (de entre um número limitado de opções) tem que ser escolhida, podendo apenas ser expandida em unidades de tempo posteriores. Este modelo foi resolvido recorrendo ao *solver* CPLEX.

Relativamente ao estudo de sistemas nacionais de recolha de REEE considerando todas as categorias mencionadas na diretiva europeia, os trabalhos publicados são escassos. Podemos, no entanto, referir o trabalho de Chang et al. (2006) onde se desenvolve um modelo linear inteiro misto para o projeto de uma rede de recolha na China cuja resolução é efetuada recorrendo ao solver LINGO. Queiruga et al. (2008) estudam a localização de recicladores de EEE em Espanha. Devido principalmente a questões relacionadas com a independência administrativa das regiões espanholas, os autores não puderam otimizar o projeto da rede de recolha a nível nacional. Contudo, após a seleção de um conjunto de potenciais localizações para a instalação de recicladores, estas foram caracterizadas de acordo com diferentes critérios, tendo-se recorrido à metodologia PROMETHEE para selecionar as melhores alternativas.

O trabalho aqui descrito prossegue e complementa o trabalho anteriormente realizado pelos autores (Gomes et al., 2011) onde se abordou este tema pela primeira vez. O modelo adaptado (ver secção 3.2) é baseado no modelo publicado neste trabalho pelo que aqui será apresentado apenas de forma resumida. Para uma descrição mais pormenorizada do mesmo aconselhamos a leitura do trabalho mencionado.

3.1 Sustentabilidade

A bibliografia sobre a sustentabilidade da cadeia de abastecimento tem vindo a crescer nos últimos anos encontrando-se dispersa por diversas áreas. Turkay et al. (2004) exploram o problema de uma cadeia de abastecimento colaborativa entre várias empresas em que não se consideram apenas objetivos económicos, mas são também incorporados alguns aspetos ambientais. Matos e Hall (2007) analisam a integração dos conceitos de desenvolvimento sustentável na cadeia de abastecimento e propõem uma estrutura que deve guiar a utilização de métodos de análise do ciclo de vida na cadeia de abastecimento. Duque et al. (2009), baseando-se na metodologia Eco-Indicator 99, quantificam os impactos ambientais de uma cadeia de recuperação de resíduos de produtos industriais. Quariguasi Frota Neto et al. (2009) propõem uma heurística de duas fases para tratar problemas multi-

objetivo. No exemplo apresentado são minimizados objetivos de custos, procura acumulada de energia e geração de resíduos numa rede de recolha. Recentemente Kannan et al. (2012) desenvolvem um modelo de localização de instalações para o projeto de redes de recolha. Um dos aspetos mais interessantes deste trabalho prende-se com a quantificação da pegada ecológica no projeto de uma rede de recolha. Os autores incluem este termo na função objetivo (custo total da rede) como o custo associado aos créditos de emissão de CO₂. O modelo desenvolvido foi resolvido com recurso ao *solver* LINGO 8.

Apesar dos trabalhos aqui mencionados, a extensão das cadeias de abastecimento ao desenvolvimento sustentável é ainda um assunto pouco explorado (Nikolopoulou e Ierapetrinou, 2012). A construção de modelos em que se consideram tanto os custos como os impactos ambientais é ainda escassa.

Neste trabalho, a sustentabilidade da cadeia de abastecimento será modelada segundo a metodologia Eco-Indicator 99, proposta por PRÉ Consultants (2008). Este indicador de soma ponderada tem por objetivo refletir o impacto ambiental de produtos ou processos através da quantificação dos efeitos nocivos para o ambiente de todas as atividades que se relacionam com os produtos ou processos em análise. O método baseia-se na análise do ciclo de vida, sendo o valor final (o indicador) calculado em três etapas. A primeira prende-se com a inventariação das atividades envolvidas, reunindo-se informação relativa aos recursos usados, à utilização da superfície terrestre (*land-use*) e às emissões produzidas. Numa segunda fase reúne-se toda a informação anterior e calcula-se o “efeito” que o produto/processo tem nos recursos, *land-use* e no ecossistema. Por último, calcula-se o indicador final, determinando-se os pesos a usar na soma ponderada dos três termos anteriormente calculados.

3.2 O modelo matemático

O problema em estudo é um problema de localização de instalações num contexto multi-produto e multi-período. Com uma formulação linear inteira mista, este modelo contempla diversas questões diretamente relacionadas com

a gestão de um sistema logístico. Começaremos por definir as variáveis utilizadas assim como a forma como se modelou o tempo. Depois apresentaremos as duas funções objetivo e finalizaremos com uma descrição das restrições consideradas.

Como mencionado anteriormente, a rede da Amb3E é uma rede de três níveis onde se pretende avaliar a localização dos centros de receção que formam o segundo nível da rede (Figura 5). Entre cada nível, define-se um fluxo: F_1 modela o fluxo entre os locais de recolha e os centros de receção, enquanto F_2 modela os fluxos entre estes últimos e as unidades de tratamento e valorização. Como referido aquando da descrição da rede logística da Amb3E, F_1 corresponde ao fluxo de equipamentos indiferenciados, sendo esta diferenciação efetuada apenas nos centros de receção. Consequentemente, o fluxo F_2 representa o conjunto dos 5 fluxos distintos previamente caracterizados, cada um associado com cada uma das categorias definidas pela associação. A estes dois fluxos acresce ainda o fluxo que modela o tratamento/reciclagem dos equipamentos (F_p) e o fluxo dos equipamentos que, por diversas razões (contaminação, económica, entre outras), não podem ser reciclados sendo enviados para aterro (F_d). A modelação destes fluxos é feita com recurso a uma variável contínua X_{mijt} que representa a quantidade de produto m , enviado pela entidade i para a entidade j , no período (t, v) .

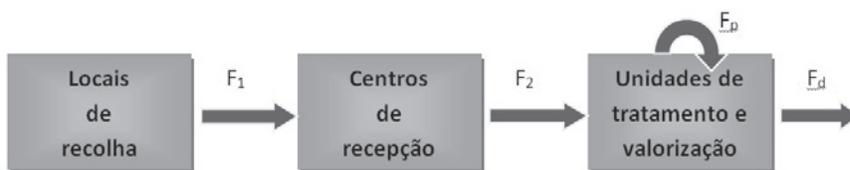


Figura 5: Representação esquemática da rede logística de recuperação de REEE

O tempo é modelado através da definição de uma escala dupla (Figura 6). Esta envolve uma unidade macro (t) onde a recolha dos equipamentos disponíveis nos locais de recolha deverá ser satisfeita e uma unidade micro (v) que permite um maior detalhe no planeamento desta recolha. De forma genérica, esta opção de modelação permite aproximar o modelo às diferentes

escalas temporais de tomada de decisão num contexto de gestão: decisões de natureza operacional (de curto prazo), tática (de médio prazo) e estratégica (de longo prazo). Sendo os modelos de localização tradicionalmente modelos de decisão estratégica, o recurso a uma dupla escala de tempo permite considerar também decisões de nível tático (ou mesmo operacional). Com este tipo de escalas é possível associar decisões anuais com mensais, ou mensais com diárias, ou qualquer outra combinação que melhor se ajuste ao problema em análise.

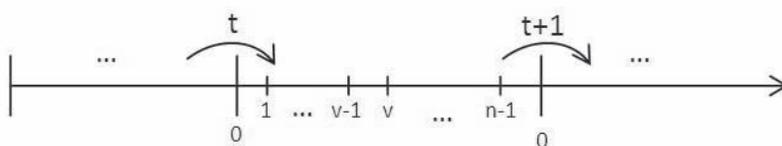


Figura 6: Dupla escala de tempo: relação entre a unidade macro (t) e a unidade micro (v)

A decisão relativa à localização dos centros de receção é feita recorrendo à variável binária Y_i que toma o valor 1 quando a localização i é escolhida para a abertura de um centro de receção. Estas decisões não dependem das unidades de tempo adotadas, pelo que uma determinada instalação permanecerá aberta (ou fechada) durante todo o horizonte temporal. O modelo contempla ainda decisões relativas às quantidades armazenadas nas entidades, modeladas pela variável S_{mitv} (quantidade do produto m , armazenado na entidade i , na unidade de tempo (t,v)).

Finalmente, uma outra característica que diferencia este modelo dos de localização tradicionais é o facto de se permitir que a quantidade disponível nos locais de recolha não seja totalmente transportada para os centros de receção. Pretende-se, desta forma, avaliar qual a quantidade de produto que é economicamente viável para ser recolhida. Para tal definiu-se a variável U_{mit} que representa a quantidade de produto m que não foi recolhida na entidade i na unidade de tempo t . Esta variável está definida apenas para a maior unidade de tempo. Note-se que apesar de estarmos a lidar com equipamentos, o que faria supor a modelação com variáveis inteiras, a Amb3E não considera

o número de equipamentos mas sim o seu peso, pelo que, excetuando a variável binária, nenhuma das outras é inteira.

A função objetivo económica é modelada pela equação [1]. O primeiro termo refere-se ao custo fixo de abertura dos centros de recepção, o segundo termo diz respeito aos custos diretamente relacionados com os fluxos anteriormente definidos, o que inclui os custos de reciclagem e de envio dos equipamentos não recicláveis para aterro. O terceiro termo é referente à taxa de compensação que é paga a cada centro de recepção por cada tonelada de REEE enviado para uma UTV. O quarto termo diz respeito aos custos de armazenamento que são contabilizados nos centros de recepção e nas unidades de tratamento e valorização. Finalmente, o último termo introduz na função objetivo uma penalização caso não seja recolhida a totalidade dos equipamentos disponíveis nos locais de recolha.

$$\begin{aligned} \text{Min } F_c = & \sum_i f_i Y_i + \sum_{mijtv} c_{mijtv} X_{mijtv} + \sum_{mijtv} c_i^c X_{mijtv} \\ & + \sum_{mitv} c_{mitv}^s S_{mitv} + \sum_{mit} c_{mit}^u U_{mit} \end{aligned} \quad [1]$$

A função objetivo que modela os impactos ambientais [2] é definida por quatro parcelas. O primeiro termo refere-se aos impactos relativos dos fluxos, ou seja, impactos do transporte entre as três entidades existentes, impactos dos processos de reciclagem e os impactos da eliminação quando os equipamentos não são recicláveis. O segundo termo diz respeito aos impactos causados pela abertura de uma nova instalação (*land-use*). O terceiro termo é referente aos impactos causados pelo armazenamento de produtos. Finalmente, o quarto termo penaliza a função objetivo pela não recolha de produtos que tenham sido colocados nos locais de recolha.

$$\text{Min } F_e = \sum_{mijtv} p_{ij} X_{mijtv} + \sum_i p_i Y_i + \sum_{mitv} p_i^s S_{mitv} + \sum_{mit} p_i^u U_{mit} \quad [2]$$

O modelo contempla as seguintes restrições: (1) garante o balanço de material em cada centro de recepção e nas UTVs; os centros de recepção têm duas atividades principais: separar os produtos recolhidos nos cinco fluxos

operacionais definidos e armazená-los para ganhar economias de escala; esta restrição relaciona os produtos que chegam dos locais de recolha e os que são enviados para as UTVs com os equipamentos armazenados; (2) o total dos equipamentos existentes em cada local de recolha ou é enviado para os centros de recepção ou é contabilizado como não recolhido e consequentemente penalizado na função objetivo; (3) garante o cumprimento da meta legal; (4) capacidades máximas e mínimas de processamento e armazenamento nos centros de recepção e nas UTVs; (5) capacidades máximas e mínimas no fluxo entre os centros de recepção e as UTVs. Para um maior detalhe sobre as opções de modelação sugere-se a consulta do artigo Gomes et al. (2011).

3.3 Metodologia multi-objetivo

Num problema de programação linear inteira mista com apenas um objetivo, que não seja impossível ou ilimitado, garante-se a existência de solução ótima, mesmo que computacionalmente possa ser difícil determiná-la ou provar que se trata da solução ótima. Quando existe mais do que um objetivo, a questão coloca-se logo no que se entende por “ótimo” dado que os objetivos podem ser conflituosos entre si. Isto é, quando uma solução é a melhor, para um dos objetivos, pode ser “muito pouco boa” para o outro (no caso de termos dois objetivos). Neste contexto, em vez do conceito de “solução ótima” recorre-se o conceito de “solução eficiente”. Informalmente, diz-se que uma solução é eficiente (ou não dominada) quando não é possível melhorar o seu valor num dos objetivos sem o piorar em pelo menos um dos outros. À projeção das soluções eficientes no espaço dos objetivos dá-se o nome de frente de Pareto.

De acordo com Clímaco et al. (2003) existem diversas técnicas para determinar o conjunto de soluções eficientes. Destaquemos três: (i) otimização de uma das funções objetivo transformando as outras em restrições (também conhecido como método das restrições ou ε -*constraint*), (ii) otimização da soma ponderada das funções objetivo e (iii) minimização da distância L_∞ a um ponto de referência. Todos estes métodos apresentam vantagens e inconvenientes. A título de exemplo refira-se que, no método dos pesos, a

escala de valores em que variam as funções objetivo pode ter uma influência determinante nas soluções obtidas, sendo por isso necessário reescalar as funções. Por outro lado, diversos conjuntos de pesos podem conduzir à mesma solução ótima, o que pode tornar pouco eficiente a geração da frente de Pareto. Estas desvantagens são ultrapassadas pelo método das restrições. No entanto, este método pode determinar soluções não eficientes quando alguma das funções objetivo tem soluções ótimas alternativas (Mavrotas, 2009). Para mais detalhes sobre os métodos de otimização multi-objetivo para problemas combinatórios sugere-se a leitura de Ehrgott e Gandibleux (2003).

Neste trabalho, e devido à sua facilidade de implementação, optou-se pelo método das restrições em detrimento de outros. A implementação foi efetuada em duas etapas. Primeiramente, os pontos extremos da frente de Pareto foram obtidos considerando a minimização independente de cada uma das funções objetivo. De seguida, otimizou-se uma das funções, tornando a outra numa restrição de menor ou igual, com o termo independente fixado num valor adequado (ϵ). Este valor ϵ foi sendo alterado de modo a permitir calcular diferentes pontos da curva entre os dois pontos extremos inicialmente obtidos. Alerta-se, no entanto, para o facto de não se ter investigado a existência de soluções ótimas alternativas, pelo que não podemos garantir que as soluções obtidas sejam, efetivamente, eficientes. Nesse sentido, sempre que nos referirmos à frente de Pareto, referiremos que se trata de uma aproximação.

4. O caso de estudo

O modelo anteriormente apresentado baseia-se num modelo de localização discreto pelo que se assumiram, como possíveis localizações para os centros de receção, as 278 sedes dos concelhos de Portugal continental. Dada a impossibilidade de considerar no modelo as localizações reais dos locais de recolha, estes foram agrupados por concelho pelo que totalizam também 278 localizações. Foram consideradas as oito UTVs existentes associando-se a sua localização aos concelhos onde estão sediadas, nomeadamente, Tondela, Setúbal, Loures, Nelas, Seixal, Alenquer, Avis e Vila Nova de Gaia.

Relativamente à quantidade a recolher pela Amb3E esta foi estimada considerando-se o volume anual a recolher de 4 kg/habitante (nº 10, artigo 9º

do Decreto-Lei nº 230/2004) tomando por base o número de habitantes de cada concelho. Este facto permitiu modelar de forma mais realista as necessidades de recolha das grandes áreas urbanas em relação às zonas rurais, por exemplo. Considera-se ainda um crescimento de 3% na quantidade a recolher ao longo do tempo, acompanhando a tendência de geração crescente de REEE que se verifica um pouco por todo o mundo (Gomes et al., 2011).

Para o cumprimento da legislação vigente, pelo menos 80% dos produtos recolhidos terão de ser reciclados ou valorizados. Os restantes 20% podem ser reciclados/valorizados ou enviados para eliminação, o que neste caso pode significar o envio para aterro sanitário ou para incineração.

Em termos de planeamento, o horizonte temporal de cinco anos foi dividido em cinco períodos de um ano (unidade macro) e estes em semestres (unidades micro).

Para modelar os produtos que fluem na rede assumiu-se que os equipamentos recolhidos pelos centros de recepção são distintos dos equipamentos enviados por estes para tratamento e valorização. Assim os primeiros serão denominados por *produtos sem triagem* sendo os segundos denominados de acordo com a classificação atribuída pela Amb3E (fluxos A a E, conforme indicado na Tabela 1).

Em termos da capacidade da rede foram definidos alguns limites: sobre os fluxos, o processamento e o inventário. Não existem limites nos fluxos entre locais de recolha e centros de recepção, já que estes estão fora do controlo da Amb3E, mas existem todavia entre centros de recepção e unidades de tratamento e valorização. Estas últimas entidades têm também limites de processamento e de inventário, em toneladas (Tabela 2).

Tabela 2: Limites de capacidade da rede de recolha

		Mínimo	Máximo
Fluxo entre CR e UTV		0	30 000
UTV	Processamento	0	-
	Inventário	0	30 000
CR	Processamento e Inventário	70	30 000

No que se refere aos custos, estes foram divididos em 6 tipos distintos, tendo alguns dos valores sido fornecidos pela Amb3E e outros estimados de acordo com indicações recebidas da associação. Para garantir a confidencialidade da informação, os valores que seguidamente se apresentam foram alterados, mantendo-se no entanto a coerência com o problema real.

- Custos de transporte ($c_{mijtv}: (m, i, j) \in F_1 \cup F_2$): transporte dos equipamentos em fim de vida entre as diferentes entidades, isto é, entre os locais de recolha e os centros de receção, e destes para unidades de tratamento e valorização. A Amb3E apenas comporta os custos de transporte para as UTVs (1 €/km) não sendo o transporte para os centros de receção da sua responsabilidade. No entanto, a associação pretende ter uma noção do impacto que estes têm na rede pelo que foi assumido um valor idêntico ao anterior, 1 €/km.
- Custos de armazenamento (c_{mitv}^s): este custo modela a existência de equipamentos numa instalação entre duas unidades de tempo micro, ou seja, equipamentos que permaneçam na instalação por um período superior a seis meses. O custo unitário foi estimado em 200 €/t para os centros de receção e em 500 €/t para as UTVs.
- Custos de processamento ($c_{mijtv}: (m, i, j) \in F_p$): montante pago às unidades de tratamento e valorização pelo processamento de uma tonelada de material. Estes custos são dependentes não só da quantidade, mas também da categoria a que pertence o material a tratar conforme Tabela 3.

Tabela 3: Custos de processamento por categoria

	A	B	C	D	E
Custo (€/t)	218	396	272	510	900

- Taxa de compensação (c_i^c): montante pago a cada centro de receção por tonelada de equipamentos em fim de vida processada e enviada para as UTVs. Este valor recompensa os

centros de recepção pelo trabalho de triagem e armazenamento que realizam e tem o valor unitário de 26.5 €/t.

- Custo de abertura de centro de recepção (f_i): custo fixo que ocorre aquando da abertura de um centro cujo valor médio é de 15 000 €.
- Custo de penalização (c_{mit}^u): ocorre se e quando uma tonelada de REEE não é recolhida nos locais de recolha. A Amb3E não incorre realmente neste custo, mas este reflete não só o custo social de deixar por recolher equipamentos depositados nos locais de recolha, mas também a taxa que tem de ser paga à EU se as metas legais não forem cumpridas. O seu valor foi estimado em 1 000 €/t.
- Custo de eliminação (c_{mitv} : $(m, i, j) \in F_d$): custo de enviar uma tonelada de REEE para aterro ou incineração. Este custo aplica-se quando os produtos são enviados para eliminação pelas UTVs e reflete o custo ambiental de recorrer à eliminação dos produtos, em vez de recuperar o material que neles esteja contido. Este valor foi estimado em 50 €/t.

Relativamente aos diversos impactos ambientais envolvidos e modelados, os valores apresentados foram obtidos através do *software* SimaPro 7.1 (PRé Consultants, 2008). Escolheu-se este *software* pois faz uma ponderação dos diferentes impactos segundo um sistema de pontos, o que permite a sua análise e comparação sistemática e consistente. Como principais tipos de impactos ambientais na rede de recolha da Amb3E têm-se:

- Impacto da abertura de instalações (p_i): este valor reflete o impacto resultante da abertura de instalações em termos da utilização de espaço, edifícios e instalações (máquinas, cabos, etc.), bem como as respetivas atividades de fim de vida inerentes às instalações. Este impacto foi estimado tanto para os centros de recepção como para as UTVs, considerando que cada instalação comporta um impacto ambiental ajustado ao horizonte temporal de 5 anos assumido neste estudo. Os valores estimados são 8 880 pontos (pts) para os centros de recepção e 72 600 pts para as UTVs.

- Impacto do transporte ($p_{ij}: (i, j) \in F_1 \cup F_2$): valor associado aos impactos gerados pelas atividades de transporte deste sistema, que é função dos quilómetros percorridos e da quantidade a transportar (em toneladas). Este parâmetro integra o consumo de combustível, a emissão de gases para a atmosfera e a emissão de metais pesados para o solo e para as águas, tendo sido estimado em 0.0464 pts/(t*km).
- Impacto do processamento ($p_{ij}: (i, j) \in F_p$): impacto imputado às UTVs por cada tonelada de material que estas processem, refletindo o impacto causado pela operação das UTVs em termos de consumo energético e de emissões atmosféricas, estimando-se o seu valor de 3.31 pts/t.
- Impacto da manutenção de inventário (p_i^s): valor do impacto ambiental correspondente à utilização do espaço e às emissões atmosféricas geradas pelos produtos mantidos em inventário. Para a manutenção de inventário em centros de recepção ou UTV correspondem os valores de 1.8 pts/t e 2 pts/t, respetivamente.
- Impacto da eliminação ($p_{ij}: (i, j) \in F_d$): impacto associado à eliminação dos produtos que não são recuperáveis. Esta eliminação pode ser feita de várias formas, por exemplo, através do envio dos produtos para aterro ou para incineração. Esta operação nem sempre é nefasta para o ambiente, já que a incineração de produtos gera energia. Considerou-se assim que 60% dos equipamentos a eliminar seriam enviados para aterro sanitário, enquanto a eliminação dos restantes 40% seria feita por incineração. Os impactos inerentes a cada um destes processos foram ajustados ao horizonte temporal considerado e à capacidade dos aterros e dos incineradores escolhidos da base de dados do *software* SimaPro. Estes foram combinados – na razão 60/40 descrita acima – de modo a definir um único impacto de eliminação, dado por tonelada de produto eliminado. O valor final estimado foi de 7.44 pts/t.

- Impacto de penalização pela não recolha dos equipamentos (p_i^u): para o cálculo deste valor considerou-se novamente que 60% dos produtos não recolhidos seriam enviados para aterro sanitário juntamente com os resíduos sólidos urbanos e os restantes 40% permaneceriam a céu aberto. O *software* SimaPro não dispõe de informação que permita aferir o valor relativo à permanência a céu aberto, pelo que se considerou que este deveria ser notavelmente superior ao do envio para aterro, tendo-se estimado o valor final para este impacto em 1 607.38 pts/t.

5. Análise de resultados

Esta análise de resultados será efetuada em três partes. A primeira analisará a rede de custo mínimo, seguindo-se a análise da rede de impacto mínimo, com a apresentação de uma pequena comparação entre as duas. Finalmente apresentar-se-á a análise multi-objetivo que, devido a dificuldades computacionais, incidirá apenas sobre os concelhos de Lisboa e Vale do Tejo.

5.1 Rede de custo mínimo

A rede de custo mínimo (Figura 7) é formada por 20 centros de receção que recebem a totalidade dos equipamentos disponíveis nos locais de recolha. Como seria expectável, a abertura de um centro de receção ocorre em cada um dos sete concelhos onde se localizam as unidades de tratamento e valorização que maior volume processam. Os restantes 13 centros ficam dispersos de forma mais ou menos uniforme pelo país. Na Figura 7 apresentam-se as ligações entre os centros de receção e as unidades que processam os quatro fluxos (A, B, C, e D). O fluxo E não está representado na figura dado ser totalmente enviado para a unidade que lhe está dedicada (em Setúbal). Assim, havendo apenas uma unidade de tratamento e valorização para onde são enviados os equipamentos do fluxo E, todos os centros de receção são afetos a esta unidade.

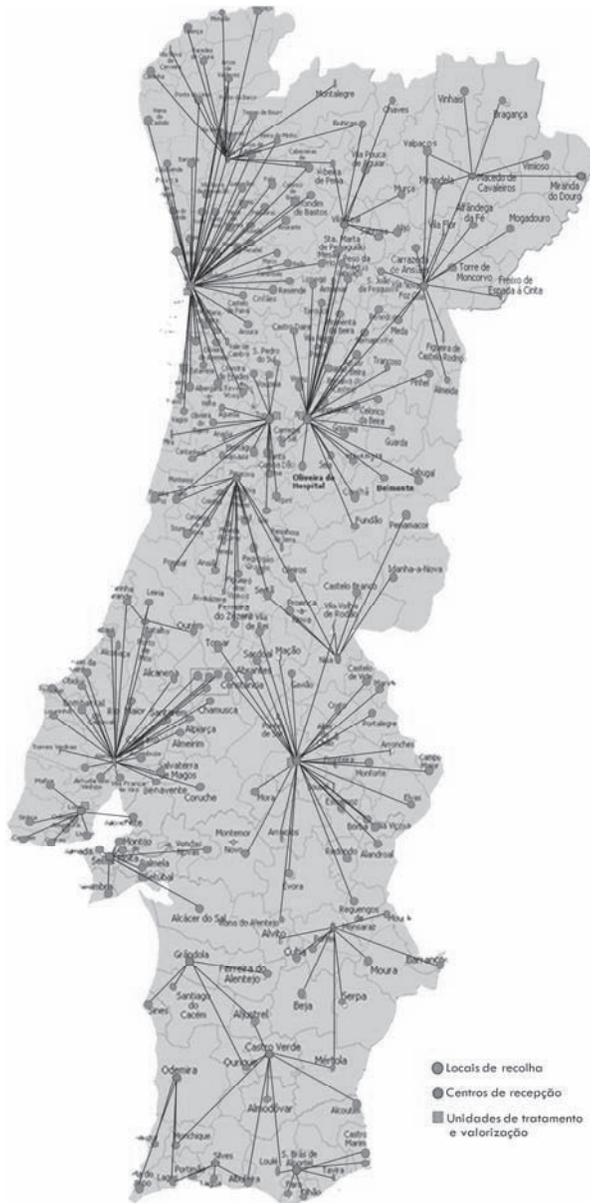


Figura 7: Localização dos centros de receção e afetação dos locais de recolha – minimização de custos

Os volumes reciclados pelas unidades de tratamento e valorização são apresentados na Figura 8. Podemos observar que a UTV com maior volume reciclado é a que opera em Vila Nova de Gaia, seguindo-se a localizada em

Loures. Esta situação é explicada pela proximidade dos grandes centros habitacionais do Porto e de Lisboa, respetivamente.

O reduzido peso das lâmpadas (fluxo E) encontra-se bem evidenciado na comparação entre os volumes processados pela unidade de Setúbal e pelas restantes unidades. É novamente o peso que torna o fluxo A (grandes equipamentos) o mais representativo na figura em análise. Relativamente ao volume que estas unidades enviam para eliminação (não mostrado na figura) este representa 20% do volume total recebido nas mesmas. Ou seja, apenas é reciclada a percentagem mínima exigida por lei.

Esta situação realça o facto deste sistema ser imposto por uma medida legal e não por uma oportunidade de negócio percebida pelos produtores de equipamentos elétricos e eletrónicos.

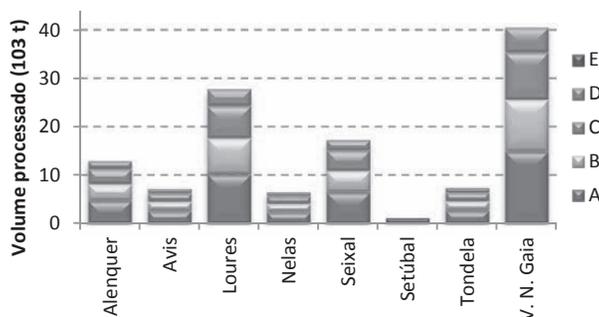


Figura 8: Volumes de equipamentos reciclados nas UTVs – minimização de custos

No que se refere à política de armazenamento podemos observar na Figura 9 que existe um elevado volume de produtos armazenados nos centros de recepção no fim do horizonte temporal, não se tendo registado qualquer armazenamento até ao ano 4. O armazenamento de lâmpadas (fluxo E) inicia-se no primeiro semestre do penúltimo ano do horizonte temporal. Esta diferença de comportamento é justificada pelo baixo peso destes equipamentos.

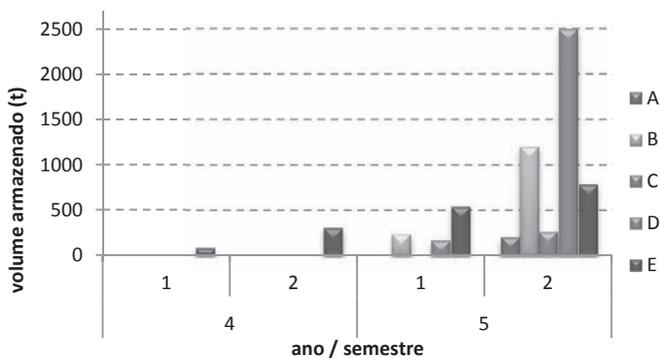


Figura 9: Volume total de equipamentos armazenados nos CRs – minimização de custos

O custo total desta rede é cerca de 55 410 mil euros. Os custos de processamento representam aproximadamente 70% do custo total, seguindo-se os custos de transporte que totalizam 20%. O custo de abertura dos centros representa apenas 1% do custo total (Figura 10).

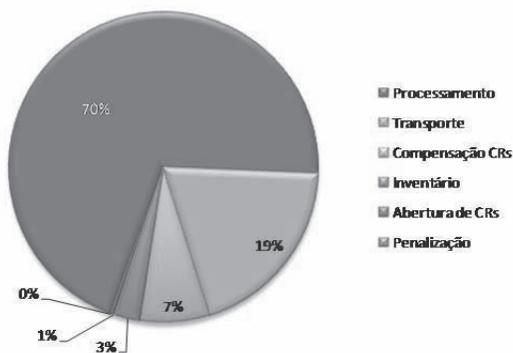


Figura 10: Estrutura de custos da rede de custo mínimo

5.2 Rede de impacto mínimo

A rede de impacto mínimo é formada por 15 centros de recepção, menos 5 que no cenário de custos (Figura 11).



Figura 11: Localização dos centros de receção e afetação dos locais de recolha – minimização de impactos

Realce-se que em duas das localizações onde estão instaladas unidades de tratamento e valorização não foram abertos centros de recepção. Este facto pode ser justificado pela proximidade geográfica das duas instalações e pela redução em 25% do número de centros de recepção relativamente ao cenário anterior.

Quando comparada com a rede de custo mínimo observa-se que, na de impacto mínimo, a afetação dos locais de recolha aos centros de recepção provoca um efeito mais estrelar. Isto é, os centros de recepção são abertos em concelhos cuja posição geográfica se aproxima do centro geográfico da área formada pelos concelhos que lhe estão afetos.

Em termos de volumes processados, estes são superiores aos verificados no cenário anterior (143 922 t vs. 120 019 t). Este aumento de processamento é justificado pela diminuição das quantidades enviadas para eliminação, que neste segundo cenário não ultrapassaram os 5%. Este aspeto mostra as vantagens ambientais da recuperação dos equipamentos por oposição à deposição em aterro.

Na Figura 12 mostram-se os volumes totais processados por cada uma das UTV, por fluxo de equipamentos. Repare-se que a unidade instalada em Vila Nova de Gaia mantém-se como sendo a instalação que processa mais equipamentos (cerca de 35% do volume total) seguindo-se a unidade instalada em Loures (cerca de 23%).

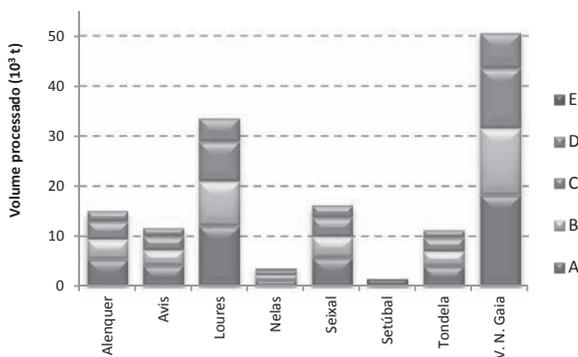


Figura 12: Volume total processado nas UTVs – minimização de impactos

No entanto, comparando os volumes processados em cada uma das unidades nos dois cenários verifica-se que as unidades de Avis e de Tondela têm um aumento de processamento acima dos 50%. Por contraponto, a unidade localizada em Nelas vê o volume de equipamentos que recebe reduzido em cerca de 50%. A unidade instalada no Seixal também vê o seu volume de processamento decrescer mas apenas em 6%.

O armazenamento dos produtos nos centros de recepção (Figura 13) segue um padrão idêntico ao do cenário anterior. Apesar de originado pela limitação necessária no horizonte temporal, este processo de armazenamento tem início logo no segundo semestre do primeiro ano (note-se que, dado o valor armazenado no dois primeiros anos ser inferior a 100 t, a correspondente coluna na Figura 13 é de muito reduzida dimensão não se conseguindo visualizar adequadamente). Novamente são as lâmpadas (fluxo E) que mais cedo começam a ser armazenadas. Outro aspeto distinto é o facto do principal fluxo armazenado ser o dos grandes equipamentos (fluxo A) e não o de aparelhos de televisão e monitores observado no cenário anterior (fluxo D).

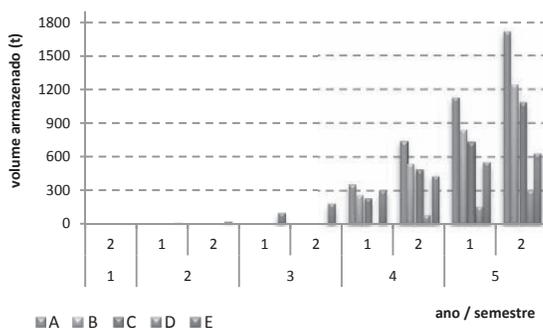


Figura 13: Volume total de equipamentos armazenados nos CRs – minimização de impactos

A estrutura de impactos apresentada na Figura 14 mostra que é o impacto provocado pela abertura das UTVs que tem maior contribuição para o impacto total do sistema (35%). Seguem-se o processamento e o transporte com 28% e 25%, respetivamente.

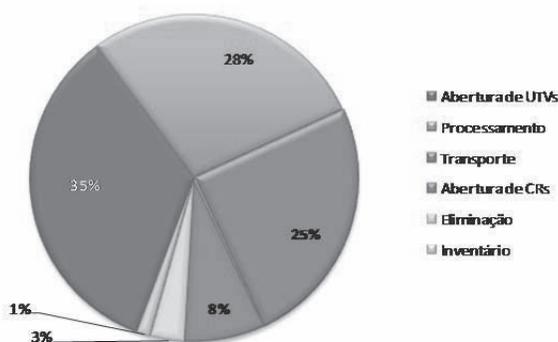


Figura 14: Estrutura de impactos da rede de impacto mínimo

Um dos aspetos que seria muito importante analisar relaciona-se com o fluxo de saída das unidades de tratamento e valorização. Contudo, estas unidades não são da responsabilidade da Amb3E mas sim dos seus parceiros, pelo que a associação não dispõe de informação relativa a este fluxo.

Na Tabela 4 mostram-se os custos e os impactos ambientais obtidos nos dois cenários. De realçar o facto de que uma redução de 7% no impacto ambiental da rede corresponde um aumento de 14% no custo total. Este aumento é causado principalmente pelo aumento do custo dos transportes, consequência da redução do número de centros de recepção.

Tabela 4: Custos e impactos totais da rede de recolha da Amb3E

	Custo (€)	Impacto ambiental (pts)	Nº de CR
Min. de custo	55 410 628	1 828 030	20
Min. do imp. ambiental	63 569 548	1 688 834	15

Para obtenção dos resultados apresentados recorreu-se ao *software* GAMS conjuntamente com o *solver* CPLEX num computador Pentium 1.86 GHz. As estatísticas computacionais dos dois cenários anteriormente analisados encontram-se expressas na Tabela 5. O modelo de impacto ambiental é de bastante mais difícil resolução que o de custo, não se tendo conseguido provar, ao fim de 11 horas de computação, que o ótimo havia sido atingido. Este facto é todavia difícil de explicar mas tudo leva a crer que este aumento

de dificuldade se prende com a alteração do valor dos parâmetros envolvidos na função objetivo deste último problema. Este aspeto carece de uma maior investigação, a qual sai fora do âmbito deste trabalho.

Tabela 5: Estatísticas computacionais

	Variáveis		Total de restrições	Iterações	CPU (s)	Gap (%)
	total	binárias				
Min. custo	821 164	278	789 979	297 473	3 992	0.04
Min. impacto				822 880	39 149	2.6

5.3 Comparação com a situação atual

À data da realização deste estudo, a rede contava com centros de receção nos 278 concelhos de Portugal continental, o que representava um custo total de aproximadamente 58 940 mil euros e um impacto ambiental de 3 972 225 pts.

A solução apresentada no cenário de custo mínimo permite uma redução de cerca de 6% no custo total da rede. Todavia é em termos do impacto ambiental que os resultados se mostram mais interessantes. As reduções obtidas no impacto ambiental da rede são de 54% quando comparada com o valor obtido com a solução de menor custo e de 57% quando comparada com a solução de menor impacto. Esta redução é conseguida principalmente devido à redução do número de centros de receção. Realce-se, no entanto, que a solução de menor impacto tem um custo total superior em cerca de 8% quando comparado com o custo total da rede atual.

5.4 Análise multi-objetivo dos concelhos de Lisboa e Vale do Tejo

A fim de analisar conjuntamente os dois objetivos considerados, abordou-se o problema utilizando uma metodologia multi-objetivo. Os valores obtidos nas análises anteriores, expressos na Tabela 4, foram tomados como uma aproximação aos dois pontos extremos da frente de Pareto. Para determinar algumas aproximações aos pontos intermédios recorreu-se ao método das restrições. Tomou-se como objetivo a minimizar o económico e limitaram-se

os impactos ambientais com recurso a uma restrição. Os pontos pretendidos foram obtidos fazendo variar o limite imposto de acordo com a restrição [3],

$$F_e \leq \varepsilon * I_{max} \quad [3]$$

sendo $\varepsilon \in [0,1]$ e I_{max} o valor do impacto da rede de menor custo.

O objetivo inicial era a aplicação da metodologia multi-objetivo a Portugal continental. Contudo, dadas as dimensões do modelo, foi computacionalmente impossível fazer este estudo. Assim, optou-se por reduzir a área geográfica, tendo-se analisado os concelhos de Lisboa e Vale do Tejo (LVT) por duas razões principais. A primeira prende-se com o seu elevado número de concelhos (53 dos 278 concelhos de Portugal continental). A segunda por se tratar da área geográfica onde estão localizadas quatro das oito unidades de tratamento e valorização (nomeadamente, nos concelhos de Alenquer, Loures, Seixal e Setúbal).

Na Tabela 6 apresentam-se os valores relativos ao custo e ao impacto das aproximações aos pontos extremos da frente de Pareto relativos a LVT. Note-se que a minimização da função ambiental conduziu a um valor que é apenas 10% inferior ao impacto ambiental da rede de custo mínimo. A esta redução está associado um aumento de 13% no custo total da rede logística.

Tabela 6: Custos e impactos da aproximação dos pontos extremos da frente de Pareto para Lisboa e Vale do Tejo

	Custo (€)	Impacto ambiental (pts)	Nº de CR
Opt. custo	19 533 728	669 042	6
Opt. impacto	22 190 000	608 362	4

A Figura 15 mostra uma aproximação à frente de Pareto. Para a obtenção destes pontos, fez-se variar o valor do parâmetro ε entre 0.9 e 1, com incrementos de 0.02. Note-se que a sua forma de exponencial negativa sugere que os benefícios em termos de impactos ambientais geram um incremento menor nos custos totais perto da solução de menor custo. À medida que a solução se aproxima do impacto ambiental de menor valor, melhorias quase

marginais ao nível dos impactos exigem investimentos financeiros consideráveis.

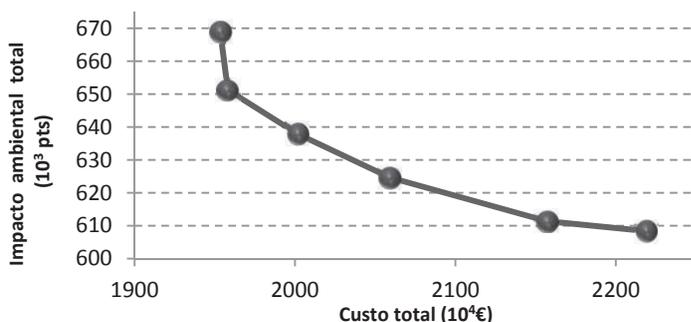


Figura 15: Aproximação da frente de Pareto para a região de Lisboa e Vale do Tejo

Analisando os impactos ambientais nas três vertentes que compõem o *Eco-Indicator 99* podemos observar que é no consumo dos recursos naturais que esta rede tem um maior impacto (Figura 16). As atividades da rede que mais influenciam o ambiente são a abertura das instalações, as atividades de processamento e as de transporte, todas elas, atividades “consumidoras” de recursos naturais. Segue-se o impacto na saúde humana e, por fim, a qualidade do ecossistema.

Foi utilizado um modelo linear inteiro misto, tendo todas as instâncias corridas para determinar esta aproximação da frente de Pareto mais de 100 mil variáveis, das quais 53 são binárias, e um total de 98 500 restrições. Todas correram em menos de uma hora com *gaps* entre 0.01% e 1.37%, tendo-se obtido um *gap* médio de 0.53%.

6. Conclusões

Neste trabalho apresentou-se o caso de estudo que envolve o estabelecimento da rede logística de recolha de equipamentos elétricos e eletrónicos da Amb3E. O objetivo foi o re-projecto da rede tendo em atenção métricas económicas e ambientais. Um modelo linear inteiro misto foi utilizado tendo revelado que uma redução de custos ou de impactos ambientais passará pela redução no número de centros de receção em operação. A análise económica da rede propõe que se recicle apenas o mínimo permitido por lei,

sendo os restantes equipamentos enviados para aterro. Já o objetivo ambiental mostrou que os equipamentos deverão ser reciclados uma vez que produzem menos impacto que quando enviados para aterro.

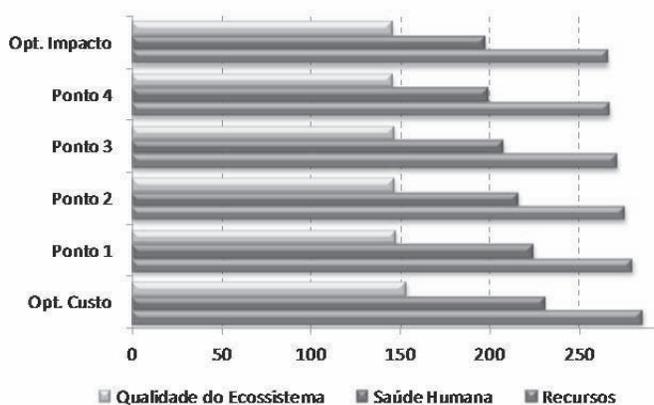


Figura 16: Valores por categoria de impacto segundo o indicador *Eco-Indicator 99* (em 10^3 pts)

Um estudo multi-objetivo efetuado para os municípios de Lisboa e Vale do Tejo mostrou que se pode melhorar a qualidade ambiental da rede desta região com um pequeno acréscimo no custo total da mesma. Contudo, para se atingirem os níveis ambientais ótimos, existe um aumento significativo nos custos totais.

Como trabalho futuro, seria muito interessante não só contemplar o fluxo de saída das unidades de tratamento e valorização, como também estender este estudo considerando todo o ciclo de vida dos equipamentos, desde a produção à sua valorização em fim-de-vida. Relativamente às análises apresentadas, pretende-se aprofundar o estudo multi-objetivo considerando todo o país ou, pelo menos, a maior extensão possível. No que se refere ao método usado para a determinação da frente de Pareto, estamos a implementar melhorias na qualidade das soluções apresentadas.

Finalmente importa salientar a excelente colaboração que existiu entre a equipa de investigação e a Amb3E no desenvolvimento do trabalho descrito. Os resultados obtidos permitiram analisar aspetos não ainda considerados

previamente pela empresa, possibilitando à mesma obter informações de como pode a rede responder às necessidades e como pode ainda equilibrar objetivos económicos e ambientais. Este trabalho necessita todavia de mais investimento no desenvolvimento de uma ferramenta que possa ser utilizada pela empresa. Este investimento deve abarcar os aspetos acima referidos e o estudo de métodos de resolução do modelo bem como o desenvolvimento de uma interface informática.

Agradecimentos

Este trabalho foi parcialmente suportado pela Fundação para a Ciência e Tecnologia no âmbito do projeto PTDC/SEN-ENR/102869/2008 e pelo CMA/FCT/UNL, no âmbito do projeto PEstOE/MAT/UI0297/2011.

Agradece-se ainda a colaboração do Pedro Furtado que no âmbito da sua tese de mestrado em Engenharia e Gestão Industrial no IST contribuiu para algum do trabalho reportado neste estudo.

REFERÊNCIAS

AMB3E (2009). Relatório de Contas 2008, Amb3E – Associação Portuguesa de Gestão de Resíduos de Equipamentos Eléctricos e Electrónicos.

ALUMUR, S. A., NICKEL, S., SALDANHA-DA-GAMA, F., and VERTER, V. (2012). Multi-period reverse logistics network design. *European Journal of Operational Research*, 220(1), 67–78.

CHANG, X., HUO, J., and CHEN, S. (2006). Study on integrated logistics network model and network design for waste electrical and electronic equipment. *SOLI'06. IEEE International Conference on Service Operations and Logistics, and Informatics*, 654–658.

CLÍMACO, J. N., ANTUNES, C. H., and ALVES, M. J. G. (2003). Programação Linear Multiobjectivo, Imprensa da Universidade de Coimbra.

DUQUE, J., BARBOSA-PÓVOA A. P. F. D., and NOVAIS, A. Q. (2009). Effect of life cycle impact assessment on the design and scheduling of a recovery network for industrial polluted waste. *Computer Aided Chemical Engineering*, 26, 1117-1182.

EHRGOTT, M., and GANDIBLEUX, X. (2003). Multiobjective Combinatorial Optimization - theory, methodology, and applications. In *Multiple Criteria Optimization: State of the Art Annotated Bibliographic Survey*, M. Ehrgott and X. Gandibleux (Eds.) Kluwer's International Series in Operations Research and Management Science, 52, 369-444.

FLAPPER, S. D. P., VAN NUNEN J., and VAN WASSENHOVE, L. N. (2005). *Managing Closed-Loop Supply Chains*. Springer-Verlag.

GOMES, M. I., BARBOSA-POVOA, A. P., and NOVAIS, A. Q. (2011). Modelling a recovery network for WEEE: a case study in Portugal. *Waste Management (New York, N.Y.)*, 31(7), 1645–1660.

- GRUNOW, M., and GOBBI, C. (2009). Designing the reverse network for WEEE in Denmark. *CIRP Annals-Manufacturing Technology*, 58(1), 391–394.
- KANNAN, D., DIABAT, A., ALREFAEI, M., GOVINDAN, K., and YONG, G. (2012). A carbon footprint based reverse logistics network design model. *Resources, Conservation and Recycling*, 67, 75–79.
- KRIKKE, H. R., VAN HARTEN, A., and SCHUUR, P. C. (1999). Business case Océ: reverse logistic network re-design for copiers. *OR Spektrum*, 21, 381–409.
- LIECKENS, K., and VANDAELE, N. (2007). Reverse logistics network design with stochastic lead times. *Computers and Operations Research*, 34(2), 395–416.
- MATOS, S., and HALL, J. (2007). Integrating sustainable development in the supply chain: The case of life cycle assessment in oil and gas and agricultural biotechnology. *Journal of Operations Management*, 25(6), 1083–1102.
- MAVROTAS, G. (2009). Effective implementation of the ϵ -constraint method in Multi-Objective Mathematical Programming problems. *Applied Mathematics and Computation*, 213(2), 455–465.
- NIKOLOPOULOU, A., and IERAPETRITOU, M. G. (2012). Optimal design of sustainable chemical processes and supply chains: A review. *Computers and Chemical Engineering*, 44, 94–103.
- PRÉ CONSULTANTS (2008), “Introduction to LCA with SimaPro 7”. <http://www.pre-sustainability.com/download/manuals/SimaPro7IntroductionToLCA.pdf> (accedido em novembro de 2012)
- QUARIGUASI FROTA NETO, J., WALTHER, G., BLOEMHOF, J., VAN NUNEN, J. A. E. E., and SPENGLER, T. (2009). A methodology for assessing eco-efficiency in logistics networks. *European Journal of Operational Research*, 193(3), 670–682.
- QUEIRUGA, D., WALTHER, G., GONZÁLEZ-BENITO, J., and SPENGLER, T. (2008). Evaluation of sites for the location of WEEE recycling plants in Spain. *Waste Management (New York, N.Y.)*, 28(1), 181–190.
- ROGERS, D. S., MELAMED, B., and LEMBKE, R. S. (2012). Modeling and analysis of reverse Logistics. *Journal of Business Logistics*, 33(2), 107–117.
- SHARMA, M., AMMONS, J. C., and HARTMAN, J. C. (2007). Asset management with reverse product flows and environmental considerations. *Computers and Operations Research*, 34(2), 464–486.
- SHERIFF, K. M. M., GUNASEKARAN, A., and NACHIAPPAN, S. (2012). Reverse logistics network design: a review on strategic perspective. *International Journal of Logistics Systems and Management*, 12(2), 171.
- TURKAY, M., ORUC, C., FUJITA, K., and ASAKURA, T. (2004). Multi-company collaborative supply chain management with economical and environmental considerations. *Computers and Chemical Engineering*, 28, 985-992.
- WALTHER, G., and SPENGLER, T. (2005). Impact of WEEE-directive on reverse logistics in Germany. *International Journal of Physical Distribution and Logistics Management*, 35(5), 337–361.

CASO 12

REESTRUTURAÇÃO DE ÁREAS DE INFLUÊNCIA E DE ROTAS DE VEÍCULOS NUM SISTEMA DE RECOLHA DE RESÍDUOS RECICLÁVEIS

Tânia Rodrigues Pereira Ramos

BRU, Instituto Universitário de Lisboa (ISCTE-IUL)
tania.ramos@iscte.pt

Maria Isabel Gomes

CMA - FCT, Universidade Nova de Lisboa
mirg@fct.unl.pt

Ana Paula Barbosa Póvoa

CEG - IST, Instituto Superior Técnico, Universidade de Lisboa
apovoa@tecnico.ulisboa.pt

RESUMO

Este trabalho versa sobre o planeamento de sistemas logísticos associados à recolha de resíduos recicláveis. Pretende-se definir as áreas de influência de depósitos e as rotas de recolha para três tipos de resíduos de embalagens – vidro, papel/cartão e plástico/metálico. A empresa em estudo, a Valnor, define as áreas de recolha (ou de influência) dos depósitos com base nos limites geográficos dos municípios, sendo todos os materiais recicláveis em cada local de recolha recolhidos a partir do mesmo depósito. A empresa deseja agora reestruturar as rotas efetuadas pelos veículos com o objetivo de reduzir a distância total percorrida e, conseqüentemente, diminuir os seus custos operacionais. De modo a melhor apoiar esta tomada de decisão da empresa, estudaram-se três cenários relativamente à configuração das áreas de recolha dos depósitos: (i) manter as áreas de recolha atuais e otimizar as rotas de recolha; (ii) reestruturar as áreas de recolha tornando-as distintas para cada material reciclável (cada depósito poderá ter três áreas de recolha diferentes); (iii) reestruturar as áreas de recolha por depósito (cada depósito terá apenas uma área de recolha comum aos três materiais). Com vista à resolução do problema, foi desenvolvido um método híbrido que combina procedimentos exatos com procedimentos heurísticos. A aplicação de técnicas de investigação operacional para a resolução deste caso permitiu encontrar soluções mais vantajosas face à situação atual implementada na empresa.

PALAVRAS-CHAVE

Recolha de resíduos recicláveis, Áreas de influência, Roteamento, Planeamento.

1. Introdução

A produção de resíduos tem aumentado significativamente nos últimos anos. Segundo dados do Instituto Nacional de Estatística (2010), a produção de resíduos em Portugal aumentou cerca de 20% entre 2004 e 2009, tendo atingido o valor de 5.185 mil toneladas em 2009 (aproximadamente 511 kg por habitante). Com este crescimento torna-se evidente a necessidade de gerir os resíduos de uma forma eficaz, e de se tomarem medidas no sentido de incentivar a sua redução e a sua reutilização e reciclagem por fileiras. A reciclagem possibilita a valorização de matérias-primas que, de outra forma, teriam um ciclo de vida mais reduzido e, conseqüentemente, a preservação dos recursos naturais, economizando-se energia e reduzindo-se a quantidade de resíduos que são depositados em aterros sanitários.

Relativamente aos resíduos de embalagens – vidro, papel/cartão e plástico/metálico - a União Europeia impôs metas de reciclagem aos estados-membros, que os obrigou ao desenvolvimento de novos sistemas de recolha adequados às especificidades deste tipo de resíduos – o sistema de recolha seletiva. Em Portugal, o sistema integrado de gestão de resíduos de embalagens (SIGRE) é gerido pela Sociedade Ponto Verde, que conta com 32 sistemas aderentes para a recolha e triagem deste tipo de resíduos produzidos no território nacional.

Neste contexto, surge o caso em estudo da empresa responsável pela recolha e triagem dos resíduos produzidos em 19 municípios do norte Alentejo e Ribatejo, servindo cerca de 180 mil habitantes – a Valnor¹. A sua área de intervenção não se esgota na recolha seletiva e triagem de resíduos de embalagens. Diversos tipos de resíduos são alvo de recolha, triagem, valorização e tratamento por parte da Valnor: resíduos sólidos urbanos indiferenciados, óleos alimentares usados, pneus usados, equipamentos elétricos e eletrónicos, veículos em fim de vida, resíduos de construção e demolição, resíduos volumosos, entre outros. O sistema de recolha seletiva de resíduos de embalagens é composto por 1275 vidrões, 1018 papelões e 1017 embalões, dispersos por 167 localidades. A empresa detém uma frota de

¹ Este trabalho foi realizado no início do ano 2011, com dados relativos ao ano 2010, quando a área de intervenção da Valnor incluía 19 municípios.

veículos que se encontra dividida entre dois depósitos, sendo um deles a estação de triagem. A existência de mais do que um depósito implica a definição da área de influência dos depósitos, delimitando geograficamente a responsabilidade pela recolha dos contentores. Cada depósito é responsável por definir as rotas de recolha por material, uma vez que estes têm de ser recolhidos separadamente. Neste processo são considerados dois tipos de fluxos: o transporte primário dos resíduos recicláveis dos contentores de recolha para os depósitos, e o transporte secundário entre os depósitos e a estação de triagem.

Atualmente, a empresa em estudo define as áreas de influência dos depósitos com base nos limites geográficos dos municípios. Consequentemente, todos os materiais recicláveis em cada local são recolhidos a partir do mesmo depósito. A empresa deseja reestruturar estas áreas de influência com o objetivo de reduzir a distância total percorrida, diminuindo assim os seus custos operacionais. Complementarmente, a empresa pretende aferir o impacto de ter áreas de influência por material reciclável (onde a recolha dos diferentes materiais pode ser efetuada a partir de depósitos diferentes) em vez das atuais áreas definidas por depósito. O redimensionamento da frota de veículos e a determinação do número de motoristas estão também em análise.

Tendo em conta que se pretende definir as áreas de influência dos depósitos e as rotas de recolha para três tipos de produtos de modo a minimizar a distância total percorrida (incluindo o transporte primário e secundário), o problema apresentado pela empresa é modelado como um *Multi-Product, Multi-Depot Vehicle Routing Problem (MP-MDVRP)*. Neste caso, a definição das rotas é considerado um problema de natureza tática, uma vez que as decisões tomadas serão mantidas durante um período de médio prazo dada a estabilidade existente nas taxas de enchimento dos contentores. Assim, pretende-se definir rotas estáticas (e não dinâmicas) não havendo, por isso, fortes condicionantes relativamente ao tempo computacional, uma vez que a solução encontrada será repetida ao longo do tempo e apenas revista por razões de sazonalidade e/ou aumento/diminuição significativo das quantidades depositadas nos contentores. Por este motivo, a abordagem seguida para a resolução do problema passou pela combinação de procedimentos exatos

(mais morosos) com procedimentos heurísticos (mais ágeis), de forma a encontrar uma solução com boa qualidade num tempo computacional razoável para um nível de decisão tático.

Este capítulo está estruturado da seguinte forma. Na secção 2 será realizada uma breve revisão da literatura sobre problemas de rotas de veículos com múltiplos depósitos. Na secção 3 será descrito o problema *MP-MDVRP* e detalhado o método híbrido desenvolvido para a resolução do mesmo. A aplicação do método ao caso real será apresentada na secção 4, onde se pormenorizará todo o processo de recolha dos dados necessários à aplicação em estudo. Finalmente, a discussão e considerações finais serão apresentadas na secção 5.

2. Problemas de rotas de veículos com múltiplos depósitos

O problema de rotas de veículos (*VRP – Vehicle Routing Problem*, na literatura inglesa) pode ser descrito como a definição de rotas ótimas de entrega ou de recolha a partir de um depósito, para um conjunto de cidades ou clientes dispersos geograficamente, tendo em conta um conjunto de restrições (Laporte, 1992). O problema de rotas de veículos com múltiplos depósitos (*MDVRP – Multi-Depot Vehicle Routing Problem*, na literatura inglesa) aparece como uma generalização do *VRP* onde, para além da sequência de visita de cada veículo, é também necessário definir qual o depósito que serve cada ponto de entrega ou de recolha. Assim, o *MDVRP* estabelece simultaneamente as áreas de influência de cada depósito e as rotas de veículos associadas. Estas rotas são definidas de forma a: (1) cada rota começar e terminar no mesmo depósito; (2) cada cliente ser visitado apenas uma vez por um veículo; (3) a procura total de cada rota não exceder a capacidade máxima do veículo; (4) a duração total de cada rota, incluindo os tempos de deslocação e os tempos de serviço, não exceder o limite pré-definido. A solução ótima será aquela que minimiza o custo total, podendo este custo ser traduzido em unidades monetárias, tempo ou distância.

O *MDVRP* é geralmente definido num grafo $G = (V, A)$ constituído pelo conjunto de vértices $V = \{1, \dots, n + w\}$ e pelo conjunto de arcos $A = \{(i, j) : i, j \in V, i \neq j\}$. O conjunto de vértices V está particionado em

dois subconjuntos $V_c = \{1, \dots, n\}$ e $V_d = \{n+1, \dots, n+w\}$ representando, respectivamente, o conjunto de clientes e o conjunto de depósitos. O conjunto de rotas de veículos K está particionado entre os vários depósitos, $K = K_1 \cup \dots \cup K_i$, onde K_i representa o conjunto de rotas de veículos afetos ao depósito $i \in V_d$. Cada rota de veículo k tem uma capacidade Q . Cada vértice i tem uma procura p_i e um tempo de serviço t_i . Ao conjunto A está associada uma matriz distância $D = (d_{ij})$.

Para a resolução do *MDVRP* foram desenvolvidos vários modelos, explorando abordagens exatas e aproximadas. Tendo em conta a natureza combinatoria deste tipo de problema, os modelos propostos na literatura são predominantemente heurísticos. Existem, no entanto, alguns trabalhos que propõem abordagens exatas. Laporte *et al.* (1984), assim como Laporte *et al.* (1988) desenvolveram algoritmos exatos de *branch and bound* que só são aplicáveis a instâncias de pequena dimensão. Recentemente, Baldacci e Mingozzi (2009) desenvolveram um método exato para a resolução do problema de rotas com heterogeneidade nas capacidades dos veículos, que também é capaz de resolver, entre outros, o *MDVRP*. Relativamente aos métodos aproximados, várias abordagens têm sido desenvolvidas: (i) abordar o problema em duas fases, em que na primeira os clientes são afetos ao depósito de acordo com um determinado critério e na segunda fase, para cada depósito, é resolvido um *VRP* (Golden *et al.* (1977), Lim e Wang (2005)); (ii) abordar o problema de forma integrada, em que a afetação aos depósitos e a construção das rotas é realizada em simultâneo (Tillman e Cain (1972), Salhi e Sari (1997)); (iii) construção de uma solução inicial e aplicação de meta-heurísticas (Renaud *et al.* (1996), Pisinger e Ropke (2007), Ho *et al.* (2008)).

Existem algumas variantes do *MDVRP* estudadas na literatura. Crevier *et al.* (2007) estudaram o *MDVRP* com rotas entre depósitos (*Multi-Depot Vehicle Routing Problem with Inter-Depot Routes*), Hadjiconstantinou e Baldacci (1998) e Vidal *et al.* (2012) estudaram o *MDVRP* periódico (*Multi-Depot Periodic Vehicle Routing Problem*), Bettinelli *et al.* (2011) investigaram o *MDVRP* com janelas temporais de entrega e frota de veículos heterogénea (*Multi-Depot Heterogeneous Vehicle Routing Problem with Time Windows*). Parthanadee e Logendran (2006) consideram um problema de distribuição com múltiplos

depósitos e produtos num contexto periódico (*Multi-Product, Multi-Depot Periodic Distribution Problem*). Neste problema, as encomendas colocadas pelos clientes contemplam diversos produtos, sendo necessário definir as rotas de distribuição ótimas para servir esses clientes a partir de vários depósitos. Os autores consideram o cenário do cliente ser abastecido por depósitos diferentes, se existir rotura de *stock* de algum produto no depósito que habitualmente o abastece. Portanto, no problema estudado por estes autores, os clientes podem ser visitados por diferentes depósitos para abastecimento dos múltiplos produtos encomendados. Assim, as áreas de influência dos depósitos podem ser definidas considerando cada produto individualmente, ou seja, resolvendo de forma independente um *MDVRP* para cada produto. No *MP-MDVRP* considerado neste trabalho, a definição das áreas de influência dos depósitos tem em conta os múltiplos produtos em simultâneo e obriga a que, em cada cliente, todos os produtos tenham de ser abastecidos (ou recolhidos) a partir do mesmo depósito.

3. O *Multi-Product, Multi-Depot Vehicle Routing Problem*

3.1 Descrição do Problema

O *MDVRP* considera que a procura do cliente i (p_i) é de apenas um tipo de produto. O *MP-MDVRP* considera que a procura do cliente i pode ser constituída por um conjunto de produtos (conjunto M), ou seja, p_{im} representa a procura do cliente i para o produto m . Combinando a multiplicidade de produtos com a multiplicidade de depósitos, duas situações podem ocorrer relativamente à definição das áreas de influência dos depósitos: *a)* todos os produtos em cada cliente são abastecidos (ou recolhidos) a partir do mesmo depósito – pelo que cada depósito tem uma área de influência comum a todos os produtos (áreas de influência por depósito); *b)* os produtos em cada cliente podem ser abastecidos (ou recolhidos) a partir de depósitos diferentes – pelo que cada depósito tem um número de áreas de influência correspondente ao número de produtos existentes (áreas de influência por produto). No primeiro caso estamos perante um *MP-MDVRP* que consiste na definição das rotas ótimas de distribuição (ou recolha) a partir de múltiplos depósitos, impondo-se que cada cliente tenha todos os produtos abastecidos

(ou recolhidos) pelo mesmo depósito. No segundo caso, estamos perante um *MDVRP* dado que o problema pode ser resolvido de forma independente para cada produto.

A formulação matemática para o *MP-MDVRP* considera as seguintes variáveis de decisão:

- X_{ijmk} , variável binária que representa a sequência de visita:
= 1, se o nó j é visitado imediatamente após o nó i , para recolher o material m , pela rota k ; 0, caso contrário;
- Y_{ijmk} , variável contínua que representa a carga de material m , na rota k , quando o arco (i, j) é atravessado;
- Z_{imk} , variável binária que afeta os locais de recolha às rotas:
=1, se o local de recolha i , com o material m , é visitado pela rota k ; 0, caso contrário.

O *MP-MDVRP* consiste na definição da sequência de visita que minimiza a distância total a percorrer garantindo: (i) que cada local de recolha i com o material m é visitado uma só vez, (ii) que cada rota k começa e termina no seu depósito de origem, (iii) que a capacidade do veículo para transportar o material m não é ultrapassada, (iv) que a duração da rota respeita o tempo disponível de um dia de trabalho, (v) a continuidade da rota, ou seja, quando a rota k visita um local de recolha, tem de sair do mesmo, (vi) que a rota k não pode recolher mais do que um tipo de material, (vii) que todos os materiais em cada local i são recolhidos a partir do mesmo depósito.

A formulação desenvolvida considera $2 \times |V| \times |V| \times |M| \times |K| + |V| \times |M| \times |K|$ variáveis de decisão, conseguindo-se apenas resolver problemas de pequena dimensão. Para problemas de média e grande dimensão foi desenvolvido um método de resolução híbrido, que será detalhado na secção seguinte.

3.2 Método de resolução híbrido

Para a resolução do problema em causa, optou-se por uma abordagem híbrida, ou seja, uma abordagem que combina formulações matemáticas com um procedimento heurístico de decomposição. Neste caso, o problema original (*MP-MDVRP*) é decomposto em quatro sub-problemas, sendo estes modelados com formulações lineares inteiras mistas (ver Figura 1).

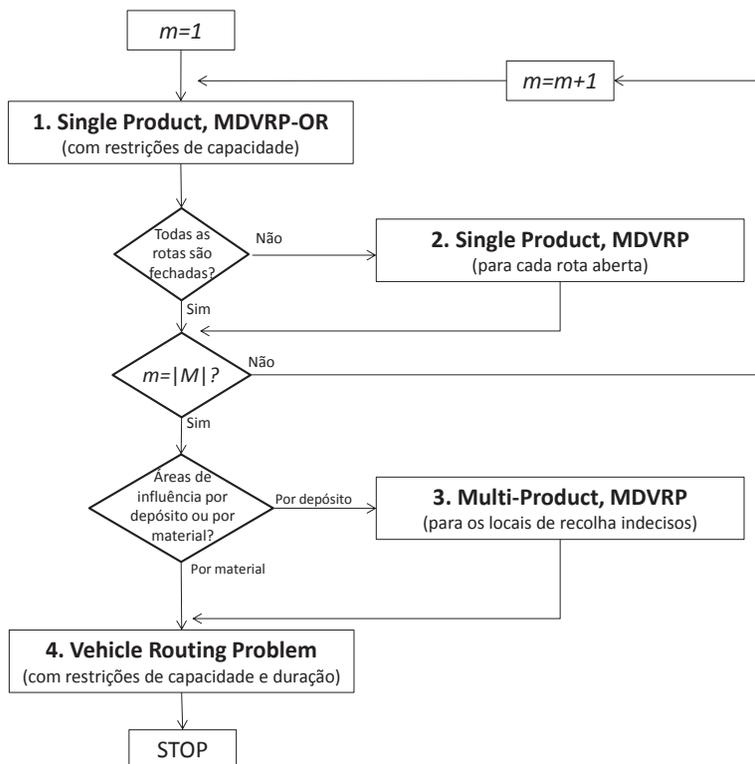


Figura 1: Fluxograma do método híbrido

Como referido anteriormente, pretende-se avaliar duas situações relativas à definição das áreas de influência: *a)* áreas de influência por depósito e *b)* áreas de influência por material. O método desenvolvido está preparado para lidar com ambas as situações. Se o objetivo for definir áreas de influência por depósito, os quatro módulos são executados; se o objetivo for definir áreas de influência por material, então o módulo 3 não será executado (ver Figura 1).

No método de decomposição representado na Figura 1, o primeiro módulo contempla uma simplificação do problema original (*MP-MDVRP*) para o problema *Single-Product, Multi-Depot Vehicle Routing Problem with Open Routes (Single-Product, MDVRP-OR)*. No problema original existem múltiplos produtos e todas as rotas têm de começar e terminar no mesmo depósito (rotas fechadas); no problema do módulo 1 considera-se apenas um produto de cada

vez e são permitidas rotas abertas entre depósitos (rotas que não começam e terminam no mesmo depósito). Este módulo é executado tantas vezes quanto o número de produtos a recolher ($|M|$). Como solução obtêm-se rotas que começam e terminam no mesmo depósito (rotas fechadas) e rotas que não começam e terminam no mesmo depósito (rotas abertas). Dado que no problema original apenas são admitidas rotas fechadas, se existirem rotas abertas na solução final, seguir-se-á a execução do segundo módulo.

Os locais de recolha que pertencem às rotas abertas serão o *input* deste segundo módulo, onde o problema *MDVRP*, formulado em programação linear inteira mista (PLIM), é resolvido. No *MDVRP* são definidas apenas rotas fechadas, pelo que os locais de recolha que pertencem a rotas abertas serão naturalmente afetos a um único depósito.

Se o módulo 1 não gerar rotas abertas ou terminada a execução do módulo 2, coloca-se a questão sobre o critério para a definição das áreas de influência. Se o objetivo for a definição de áreas de influência por material, o quarto módulo é executado onde, para cada depósito e para cada material, é resolvida uma formulação em PLIM para o *VRP*. Se o objetivo for a definição de áreas de influência por depósito, é necessário analisar os resultados obtidos nos módulos anteriores (módulos 1 e 2) considerando todos os materiais em simultâneo e verificar se, para cada local de recolha, todos os materiais são recolhidos a partir do mesmo depósito. Quando existem divergências na afetação entre os materiais, esses locais são designados de “indecisos” e o módulo 3 é executado antes do módulo 4 de forma a decidir a afetação final desses locais, considerando todos os materiais em simultâneo.

Os quatro módulos envolvem o desenvolvimento de formulações matemáticas que tiveram em conta as características do problema em estudo. O detalhe de cada módulo é descrito de seguida.

Módulo 1

A formulação proposta para o *Single-Product, MDVRP-OR* é baseada na *two-commodity flow formulation* proposta por Baldacci *et al.* (2004) para o *Capacitated Vehicle Routing Problem (CVRP)*. Todas as rotas começam num depósito real e terminam num depósito réplica.

Cada rota é definida por dois fluxos: um fluxo com origem no depósito real e destino no depósito réplica, definido pela variável y_{ij} (que representa a carga no veículo e que irá aumentando ao longo da rota); o outro fluxo, representando o caminho inverso, começa no depósito réplica e termina no depósito real; este é definido pela variável y_{ji} (que representa o espaço vazio no veículo e que irá diminuindo ao longo da rota). A Figura 2 ilustra os dois caminhos que formam uma rota.

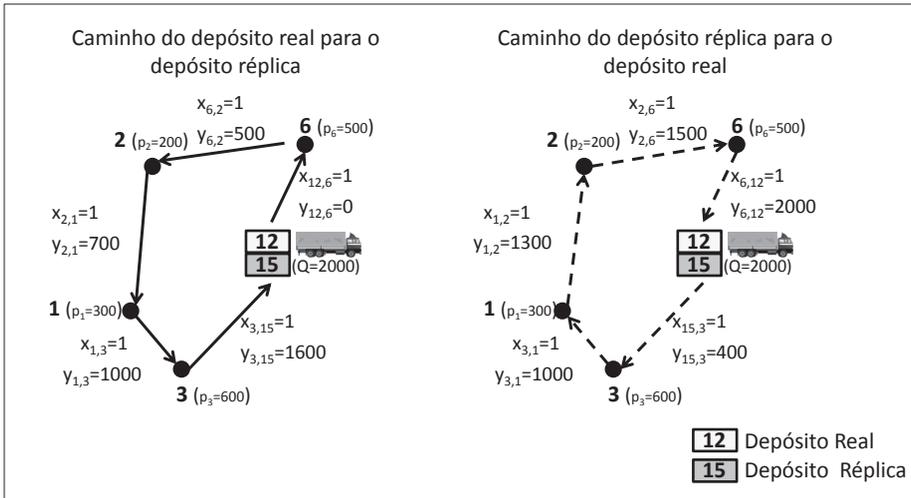


Figura 2: Ilustração dos dois caminhos que formam uma rota

No *Single-Product, MDVRP-OR* as rotas têm de começar e terminar num depósito mas não necessariamente no mesmo, como acontece no *MDVRP*. Os índices, conjuntos, parâmetros e variáveis de decisão da formulação proposta para o *Single-Product, MDVRP-OR* são apresentados de seguida:

Índices

i, j Índice dos nós

Conjuntos

\bar{V} Conjunto dos nós $\bar{V} = \{1, \dots, n + 2w + s\}$; $\bar{V} = V_c \cup V_d \cup V_f \cup V_s$

V_c Conjunto dos locais de recolha $V_c = \{1, \dots, n\}$

V_d Conjunto dos depósitos reais $V_d = \{n + 1, \dots, n + w\}$

V_f Conjunto dos depósitos réplica $V_f = \{n + w + 1, \dots, n + 2w\}$

V_s Conjunto das estações de triagem $V_s = \{n + 2w + 1, \dots, n + 2w + s\}$

Parâmetros

d_{ij}	Distância entre o nó i e o nó j (em metros)
p_i	Quantidade a recolher no local de recolha i (em kg)
c_i	Número de contentores no local de recolha i
fr	Frequência de recolha no horizonte temporal
Q	Capacidade do veículo para o transporte primário (em kg)
QT	Capacidade do veículo para o transporte secundário (em kg)
E	Distância média entre dois contentores dentro de um local de recolha (em metros)
R	Número máximo de rotas

Variáveis de Decisão

x_{ij}	Variável binária que representa a sequência de visita ($x_{ij} = 1$ se o nó j é visitado após o nó i ; $x_{ij} = 0$ caso contrário)
y_{ij}	Variável contínua que representa a carga no veículo quando atravessa o arco (i,j) . A variável y_{ji} representa o espaço vazio no veículo quando o arco (i,j) é atravessado.

Modelo Matemático

$$\text{Min } H = \frac{1}{2} \sum_{i \in \bar{V}} \sum_{\substack{j \in \bar{V} \\ j \neq i}} x_{ij} d_{ij} fr + \sum_{i \in \bar{V}_c} c_i E fr + 2 \sum_{i \in \bar{V}_f} \sum_{j \in \bar{V}_c} \sum_{i \in \bar{V}_s} \frac{y_{ji} fr}{QT} d_{hi} \quad (1)$$

$$\sum_{\substack{j \in \bar{V}_s \\ j \neq i}} (y_{ij} - y_{ji}) = 2p_i, \quad \forall i \in \bar{V}_c \quad (2)$$

$$\sum_{i \in \bar{V}_c} \sum_{j \in \bar{V}_f} y_{ij} = \sum_{i \in \bar{V}_c} p_i \quad (3)$$

$$\sum_{j \in \bar{V}_f} \sum_{i \in \bar{V}_c} y_{ji} \leq RQ - \sum_{i \in \bar{V}_c} p_i \quad (4)$$

$$\sum_{i \in \bar{V}_c} \sum_{j \in \bar{V}_d} y_{ij} \leq RQ \quad (5)$$

$$\sum_{\substack{i \in \bar{V}_s \\ i \neq j}} x_{ij} = 2, \quad \forall j \in \bar{V}_c : p_j \neq 0 \quad (6)$$

$$y_{ij} + y_{ji} = Qx_{ij} \quad \forall i, j \in \bar{V} \setminus \bar{V}_s, i \neq j \quad (7)$$

$$x_{ij} \in \{0,1\} \quad \forall i, j \in \bar{V} \setminus V_s \quad (8)$$

$$y_{ij} \geq 0, y_{ji} \geq 0 \quad \forall i, j \in \bar{V} \setminus V_s \quad (9)$$

A função objetivo, equação (1), modela a minimização da distância total percorrida, que inclui a distância percorrida no transporte primário (primeira e segunda parcelas) e a distância percorrida no transporte secundário (terceira parcela).

O transporte primário considera a distância entre os locais de recolha e os depósitos sendo, como referido anteriormente, as rotas são definidas por dois caminhos. Isto significa que cada arco é contado duas vezes na solução, duplicando a distância real percorrida. Assim, esta parcela tem de ser dividida por dois para eliminar o valor do segundo caminho. A segunda parcela do transporte primário está relacionada com a distância a percorrer dentro de cada local para recolher todos os contentores existentes. Este valor não depende de nenhuma variável de decisão pelo que é constante, mas foi incluído na função objetivo para contabilizar a distância total percorrida. De referir que esta segunda parcela contempla o número total de contentores existentes em cada local de recolha (c_i) e não ($c_i - 1$), de modo a considerar a distância a percorrer desde a entrada na localidade (local de recolha) até ao primeiro contentor ou desde o último contentor até à saída da localidade.

O transporte secundário consiste em deslocações de ida e volta entre os depósitos e as estações de triagem, razão pela qual o terceiro termo encontra-se multiplicado por dois.

A equação (2) garante que, para cada local de recolha, a diferença entre o fluxo de saída e o fluxo de entrada é igual a duas vezes a quantidade a ser recolhida nesse local (devido aos dois caminhos que definem as rotas).

As restrições (3), (4) e (5) modelam os fluxos de entrada e saída dos depósitos reais e dos depósitos réplica.

A equação (6) garante que, em qualquer solução admissível, cada local de recolha tem dois arcos incidentes.

A equação (7) garante que a capacidade de cada rota não é ultrapassada.

As restrições (8) e (9) definem os domínios das variáveis.

Módulo 2

O módulo 2 é executado apenas se o módulo 1 produzir uma solução que contenha rotas abertas. O objetivo deste módulo é afetar a um único depósito os locais de recolha que pertencem a este tipo de rotas, utilizando-se, para tal, uma formulação em PLIM para o *MDVRP*. Dado que o *MDVRP* apenas admite soluções com rotas fechadas, os nós são naturalmente afetados a um único depósito. Neste módulo apenas se consideram os locais de recolha que pertencem a cada rota aberta e o módulo é executado considerando cada rota separadamente. Como resultado, estes locais de recolha podem ser todos afetados a um só depósito ou divididos pelos vários depósitos existentes.

A formulação em PLIM para o *MDVRP* considera um conjunto de rotas de veículos K , a que corresponde o índice k , e a variável binária z_{ik} que afeta a rota k ao local de recolha i . As variáveis x_{ij} e y_{ij} da formulação do módulo 1 são redefinidas de modo a contemplarem a rota a que estão afetadas - x_{ijk} e y_{ijk} . A variável binária x_{ijk} toma o valor 1 se o nó j é visitado imediatamente após o nó i , pela rota k e a variável contínua y_{ijk} representa a carga transportada na rota k quando o arco (i, j) é atravessado. Para que exista sequência de visita entre o nó i e j ($x_{ijk} = 1$) e o respetivo fluxo de carga (y_{ijk}), os nós i e j têm de estar afetados à mesma rota k ($z_{ik} = 1 \wedge z_{jk} = 1$). As restrições (1) a (9) são reescritas de forma a contemplarem estas alterações e são adicionadas restrições que garantem que as rotas começam no depósito real e terminam no depósito réplica correspondente.

Módulo 3

O módulo 3 é executado apenas se se pretender definir áreas de influência por depósito. Neste caso, em cada local de recolha, todos os materiais recicláveis têm de ser recolhidos pelo mesmo depósito. Assim, os resultados obtidos nos módulos anteriores para os diferentes materiais recicláveis são agora analisados em simultâneo, podendo existir situações em que o mesmo local é visitado por veículos provenientes de depósitos diferentes, consoante o material a recolher. Considerando, por exemplo, três materiais recicláveis (m_1 , m_2 , m_3), o local de recolha i pode ter o material m_1 recolhido pelo depósito d_1 , mas ser o depósito d_2 a recolher o material m_2 e o depósito d_3 a recolher

o material m_3 . Deste modo, três situações podem acontecer: (i) não existir acordo entre os materiais; (ii) existir acordo entre dois materiais; (iii) existir acordo entre todos os materiais. Os locais de recolha que se encontram nos casos (i) e (ii) são designados de “indecisos”. Neste módulo 3 é resolvida a formulação em PLIM para o *MP-MDVRP* descrita na secção 3.1 apenas considerando os locais de recolha “indecisos”. Esta redução da dimensão do problema (tomando apenas o conjunto dos locais de recolha indecisos, em vez do conjunto original – levando a redução da cardinalidade dos conjuntos V e K) permite a utilização da formulação original. Como mencionado, esta formulação considera um conjunto de materiais recicláveis M , a que corresponde o índice m , pelo que as variáveis de decisão passam a ser x_{ijmk} , y_{ijmk} e z_{imk} . Consequentemente, as restrições são reescritas de modo a contemplarem o conjunto M , sendo adicionada uma restrição que garante que, em cada local de recolha, todos os materiais são recolhidos pelo mesmo depósito.

Módulo 4

Uma vez que as áreas de influência já estão definidas, o módulo 4 irá definir as rotas finais para cada depósito e para cada material. Neste caso serão consideradas restrições de duração das rotas, o que até agora não foi tido em conta. Esta última formulação baseia-se na formulação apresentada para o módulo 1, embora exista apenas um depósito real e, consequentemente, um depósito réplica (em vez de múltiplos). Para modelar a duração das rotas foram definidas duas novas variáveis de decisão: a_{ij} que representa o tempo de chegada ao nó j com origem no nó i e e_{ij} que representa o tempo de saída do nó i com destino ao nó j . Foram também considerados sete novos parâmetros:

t_i *Tempo necessário para recolher o material no local de recolha i (em minutos)* $t_i = c_i \left(U + \frac{E}{vw} \right)$

b_{ij} *Tempo de deslocação entre o nó i e o nó j (em minutos)* $b_{ij} = \frac{d_{ij}}{vb}$

vb *Velocidade entre locais de recolha (em metros por minuto)*

v_w	<i>Velocidade dentro dos locais de recolha (em metros por minuto)</i>
U	<i>Duração da recolha de um contentor (em minutos)</i>
L	<i>Tempo de descarga do veículo no fim de uma rota (em minutos)</i>
T	<i>Duração máxima de um dia de trabalho por motorista (em minutos)</i>

Para além das situações anteriormente referidas, esta formulação modela também as seguintes situações: (i) a duração de uma rota não pode ultrapassar $T - L$ minutos; (ii) o tempo de chegada ao nó j corresponde ao tempo de saída do nó i mais o tempo de deslocação entre i e j ; (iii) o tempo de saída do nó i corresponde ao tempo de chegada ao nó i mais o tempo necessário para proceder à recolha neste nó.

As formulações matemáticas desenvolvidas para cada um dos quatro módulos foram escritas em Gams 23.7 e resolvidas de forma independente através do método de *branch-and-cut* disponibilizado no ILOG CPLEX 12.3 (ILOG-CPLEX, 2011) em versão paralela, num Two Intel Xeon X5680 @ 3.33GHz, com 24 GB de RAM, com 6 núcleos em cada processador.

4 . Caso de estudo - Valnor

A Valnor é responsável pela recolha seletiva de três tipos de resíduos de embalagens (vidro, papel/cartão, plástico/metal) em 19 municípios do norte Alentejo e Ribatejo. Os 3.310 contentores (1.275 vidrões, 1.018 papelões e 1.017 embalões) que compõem o seu sistema de deposição são recolhidos a partir de dois depósitos situados nos municípios de Avis e Abrantes. Toda a gestão das operações da Valnor baseia-se numa lógica municipal, ou seja, existe um circuito por concelho, exceto nos concelhos mais populosos (Ponte de Sôr, Portalegre e Elvas) onde existem dois circuitos, e no concelho de Abrantes onde existem 3 circuitos, totalizando 24 circuitos. Por outro lado, também as áreas de intervenção de cada depósito são definidas com base nos limites dos municípios (ver Figura 3).

Para a recolha de vidro, os 24 circuitos foram repetidos anualmente num total de 373 circuitos por ano (considerando o período de Novembro 2009 a Outubro 2010). Em média, cada circuito foi percorrido 15,54 vezes, dando origem a um intervalo médio entre recolhas de 3,35 semanas (ver Tabela 1). Para a recolha de papel, a repetição dos 24 circuitos totalizou 954 circuitos

anuais. Cada circuito foi executado, em média, 39,75 vezes no período em análise, dando origem a um intervalo médio entre recolhas de 1,31 semanas. Finalmente, para o plástico/metal, em média, cada um dos 24 circuitos foi repetido 40,7 vezes num ano (num total de 977 circuitos), o que perfaz um intervalo médio entre recolhas de 1,28 semanas.



Figura 3: Áreas de influência atuais

Tabela 1: Intervalo entre recolhas por tipo de material, no período de Novembro 2009 a Outubro de 2010 (em semanas)

	<i>Vidro</i>	<i>Papel</i>	<i>Plástico/Metal</i>
Máximo	5,84	2,9	3,06
Média	3,35	1,31	1,28
Moda	4	1,6	1,7
Mínimo	1,18	0,5	2,04

Verificou-se também que a taxa média de enchimento dos contentores no momento da recolha foi de 21% para o vidro e de 23% para o papel e plástico/metal (ver Tabela 2). Apenas no papel se verificou uma taxa máxima de enchimento de 100%.

Tabela 2: Peso recolhido por contentor e respetiva taxa de enchimento

	<i>Vidro</i>		<i>Papel</i>		<i>Plástico/Metal</i>	
	<i>Peso (kg)</i>	<i>Tx.Enchim.</i>	<i>Peso (kg)</i>	<i>Tx.Enchim.</i>	<i>Peso (kg)</i>	<i>Tx.Enchim.</i>
Máximo	464	77%	200	100%	68	57%
Média	127,96	21%	46,71	23%	28	23%
Mínimo	37,77	6%	8,75	4%	1	1%

A Valnor pretende agora reestruturar as suas áreas de influência dos depósitos e redefinir os circuitos de recolha de modo a diminuir a distância total percorrida, reduzindo assim os seus custos operacionais. Para cumprir este objetivo, o método apresentado na secção 3 foi aplicado ao caso Valnor. Não existindo nenhuma imposição legal ou contratual relativa à periodicidade dos circuitos, e tendo em conta a reduzida taxa média de enchimento atual, a Valnor pretende também redefinir o intervalo entre recolhas, garantindo que a referida taxa de enchimento não ultrapasse 80%. Sendo a periodicidade das rotas um parâmetro do modelo, esta questão será detalhada na secção seguinte, assim como todo o processo de recolha de dados necessários ao modelo.

4.1 Recolha e tratamento de dados

Os dados necessários foram obtidos através de entrevistas ao gestor de operações e pela análise da base de dados dos circuitos realizados. A Valnor facultou a informação relativa aos circuitos efetuados durante Novembro de 2009 a Outubro de 2010. Para cada circuito existe informação sobre a data em que o mesmo ocorreu, qual o veículo utilizado, o tipo de resíduo que foi recolhido, o número de contentores recolhidos, o número de quilómetros percorridos, a duração do circuito e o peso recolhido. Neste período foram percorridos cerca de 277.300² km, congregando tanto o transporte primário como o secundário.

Os valores e a forma como se estimaram os vários parâmetros do modelo são apresentados de seguida.

² Por questões de confidencialidade, as distâncias apresentadas ao longo do capítulo não são as reais (foi aplicado um coeficiente a todas as distâncias).

4.1.1 Locais de recolha

Os 3.310 contentores estão localizados em 167 localidades diferentes. Assim, um local de recolha agrega um ou mais contentores, de um ou mais materiais recicláveis. Por exemplo, a localidade “Vila de Rei” é um local de recolha com 12 contentores para a deposição de papel, 12 contentores para o plástico/metal e 16 contentores para o vidro. Assim, para determinar a distância percorrida e o tempo despendido dentro de cada localidade/local de recolha considerou-se a distância média entre os contentores, a velocidade média dentro das localidades e o tempo médio de recolha por contentor. Realce-se também o facto de nem todos os locais de recolha terem os três materiais para serem recolhidos. A título de exemplo, o local de recolha “Monte do Duque”, no concelho de Nisa, tem um contentor para o papel e um contentor para o vidro, não existindo contentores para a recolha do plástico/metal.

Comparando a quantidade a recolher em cada local e a capacidade máxima dos veículos para o transporte primário, observaram-se algumas situações em que a quantidade a recolher excede a capacidade do veículo. Nestes casos, os locais de recolha foram replicados o número de vezes necessário para respeitar a capacidade máxima dos veículos. A título de exemplo, o local de recolha “Ponte de Sôr” tem 49 contentores para papel e um peso total a recolher de 3.284 kg em cada visita. Sendo a capacidade do veículo para a recolha do papel de 2.000 kg, este local de recolha foi replicado, repartindo-se o número de contentores e peso a recolher por dois locais: “Ponte de Sôr-1” com 29 contentores (correspondendo um peso de 1.944 kg) e “Ponte de Sôr-2” com os restantes 20 contentores (com um peso de 1.340 kg). Considerando a recolha do plástico/metal, “Ponte de Sôr” necessita também de 2 réplicas. Para além de “Ponte de Sôr”, existem ainda outras 5 localidades onde foi necessário proceder de forma análoga: Abrantes (5 réplicas), Campo Maior (2 réplicas), Elvas (3 réplicas), Nisa (2 réplicas) e Portalegre (8 réplicas). Assim, em vez de 167 localidades consideraram-se 189 locais de recolha, o que corresponde a 178 locais com o material papel, 183 com o material plástico/metal e 174 com o material vidro.

4.1.2 Quantidade a recolher e periodicidade da recolha

Para determinar a quantidade a recolher em cada local e a periodicidade da recolha calculou-se previamente a quantidade média diária depositada em cada contentor. Este cálculo baseou-se no intervalo de tempo entre duas realizações consecutivas do mesmo circuito, assim como na quantidade média depositada em cada contentor (ver Figura 4).

Data	Circuito	Material	Nº Contentores	Peso (kg)	Peso/Contentor (kg)	Intervalo Tempo (dias)	Peso/Contentor/Dia (kg)
27/10/2010	2	Vidro	58	4760	82,07	32	2,56
25/09/2010	2	Vidro	59	5740	97,29	28	3,47
28/08/2010	2	Vidro	58	6380	110	29	3,79
30/07/2010	2	Vidro	57	5860	102,81	-	...
...

Média= 3,27 kg

Figura 4: Exemplo do cálculo do peso médio por contentor por dia

De referir que para o cálculo da quantidade média diária depositada em cada contentor foram retirados alguns *outliers* detetados. A título exemplificativo apresenta-se na Figura 5 o caso do circuito do concelho do “Crato” para recolha de vidro. O *outlier* identificado corresponde a um aumento da deposição do material vidro, consequência da realização do Festival do Crato entre os dias 25 a 28 de Agosto de 2010. Assim, a quantidade recolhida antes do início do festival encontra-se num nível normal (dia 24 de Agosto de 2010), tendo a quantidade recolhida após o festival (dia 30 de Agosto de 2010) apresentado um valor extraordinário justificado pela realização do mesmo. Neste sentido, para o cálculo da quantidade média depositada diariamente por contentor retirou-se o valor do *outlier* identificado. Note-se que o valor médio com o *outlier* é de 5,18 kg/contentor/dia, sendo o valor médio sem *outlier* de 2,57 kg/contentor/dia.

Este procedimento para o cálculo do peso por contentor por dia assume que todos os contentores visitados em cada circuito apresentam a mesma deposição média diária. A quantidade a recolher em cada local de recolha/localidade irá depender do número de contentores existentes que, por sua vez, dependerá da população existente em cada localidade. Este cálculo da

quantidade média depositada por dia em cada contentor é um passo intermédio para o cálculo da periodicidade da recolha.

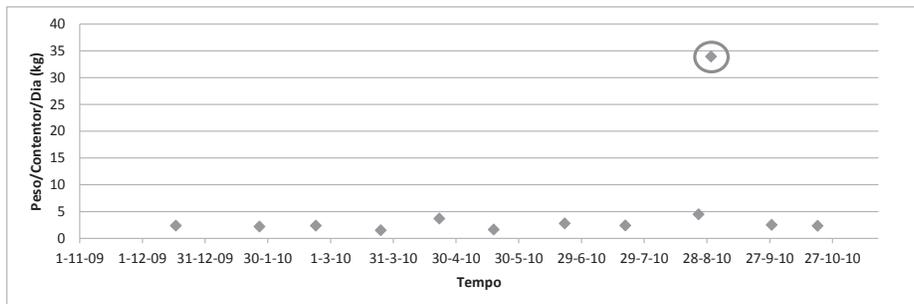


Figura 5: Peso por contentor por dia no circuito Crato para a recolha de vidro durante um ano

O intervalo de tempo entre recolhas para cada local depende da taxa de enchimento dos contentores que, por sua vez, depende da quantidade média depositada por dia, calculada anteriormente. O maior intervalo de tempo ditará o período de planeamento a considerar. Foi, no entanto, imposto um máximo de 6 semanas (1 semana = 7 dias) o que corresponde ao intervalo máximo entre recolhas praticado atualmente pela Valnor. Desta forma, e assumindo um intervalo mínimo entre recolhas (t) de 1 semana, aplicou-se o procedimento apresentado na Figura 6 de modo a determinar qual o padrão de recolha a efetuar para que nenhum contentor ultrapasse uma taxa de enchimento de 80%.

O modelo desenvolvido exige que o intervalo de tempo entre recolhas por tipo de resíduo seja igual em todos os locais de recolha. Pela aplicação da metodologia apresentada obtiveram-se intervalos entre recolhas diferentes em vários locais de recolha com o mesmo tipo de resíduo. Para cumprir este requisito, a frequência a implementar para cada tipo de material corresponderá à frequência que permite que a taxa de enchimento de todos os contentores não ultrapasse os 80%. Assim, para o material papel o intervalo entre recolhas determinado é de $t = 2$ (14 dias), para o plástico/metál é de $t = 3$ (21 dias) e para o vidro de $t = 6$ (42 dias). Isto significa que no horizonte temporal de 6 semanas (intervalo máximo encontrado), a frequência de recolha de vidro é de 1 vez, a frequência de recolha do papel é de 3 vezes e, para o plástico/metál, esta frequência é de 2 vezes.

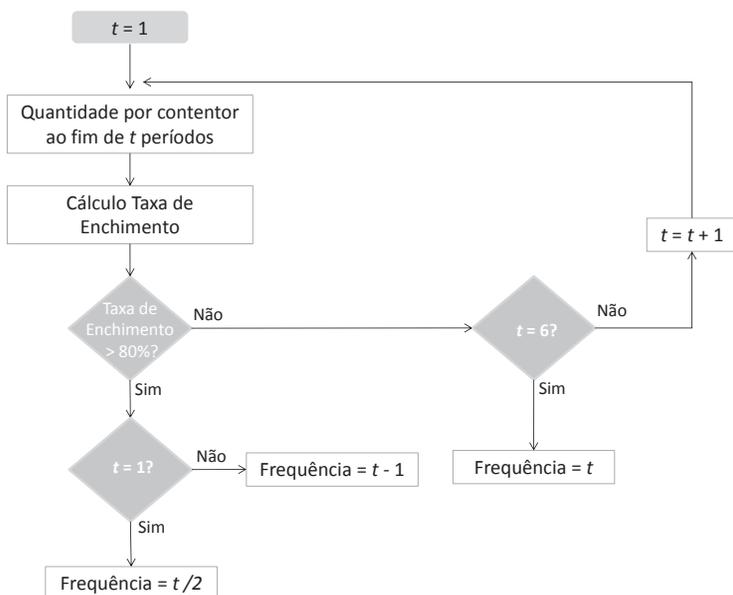


Figura 6: Procedimento para cálculo do intervalo de tempo entre recolhas

A quantidade diária por contentor multiplicada pelo intervalo entre recolhas de cada tipo de material indica o valor a recolher em cada contentor em cada visita. O peso a recolher em cada local resulta da agregação de todos os contentores que compõem esse local de recolha.

4.1.3 Capacidade de recolha

Devido à diferente densidade dos materiais, a capacidade dos veículos difere consoante o material a recolher. A capacidade máxima dos veículos (Q) é de 2.000 kg para a recolha do papel, para o plástico/metálico é de 1.000 kg e para o vidro é de 8.500 kg. Para o transporte secundário são utilizados veículos de maior dimensão e, conseqüentemente, a capacidade de transporte é mais elevada (QT). Assim, para o transporte secundário de papel a capacidade máxima a considerar é de 5.000 kg, para o plástico/metálico é de 3.000 kg e para o vidro é de 12.000 kg. A frota de veículos, quer para o transporte primário, quer para o secundário, é homogénea.

4.1.4 Tempo

A duração máxima de uma rota é de 7 horas ($T = 420$ minutos), onde se inclui o tempo de descarga do veículo quando este regressa ao depósito ($L = 10$ minutos). Esta duração máxima corresponde a um dia de trabalho para os motoristas. No entanto, o tempo disponível para a realização dos circuitos de recolha é de 63 horas por semana dado que em 4 dos 5 dias da semana realizam-se dois turnos (14h x 4 dias + 7h x 1 dia). Isto significa que cada veículo pode operar um total de 378 horas no horizonte temporal de 6 semanas.

Para determinar o tempo de recolha dos contentores e a velocidade dos veículos dentro das localidades foi efetuado um modelo de regressão linear múltipla, em que a duração do percurso é explicada através do número de contentores recolhidos e da distância percorrida. No entanto, o coeficiente de correlação obtido não revelou um resultado satisfatório ($R^2 = 0,34$), pelo que os parâmetros estimados pelo modelo de regressão não foram utilizados. Com base na experiência do gestor de operações da Valnor foram definidos os seguintes valores médios:

- Duração da recolha de um contentor (U): 5 minutos,
- Velocidade dentro das localidades (v_w): 10 km/hora.

Para a velocidade entre localidades (v_b) foi considerado um valor de 60 km/hora.

4.1.5 Distâncias

Existem dois tipos de distâncias a considerar. A distância a percorrer dentro de cada local de recolha e a distância a percorrer entre locais de recolha. A primeira corresponde ao número de contentores existentes em cada local multiplicado pela distância média entre dois contentores (E). O valor apurado para E foi de 500 metros, correspondendo à média de todas as distâncias entre contentores nas 167 localidades. A distância a percorrer entre locais de recolha e depósitos (d_{ij}) foi obtida através de uma ferramenta de georreferenciação com base na rede viária disponível no Sistema de Informação Geográfica "ArcGis Online" (ArcGis, 2011).

4.2 Resultados obtidos

De modo a apoiar a decisão por parte da empresa Valnor foram estudados três cenários relativamente à configuração das áreas de influência dos depósitos: (1) manter as áreas de influência atuais e otimizar as rotas de recolha; (2) reestruturar as áreas de influência por material reciclável (cada depósito poderá ter três áreas de influência distintas); (3) reestruturar as áreas de influência por depósito (cada depósito terá apenas uma área de influência comum aos três materiais). De referir que, como se pretendeu também determinar qual o número mínimo de veículos necessários, o valor do parâmetro R , nos módulos 1 e 4, e a cardinalidade do conjunto K , nos módulos 2 e 3, foram suficientemente elevados para não limitarem a solução.

4.2.1 Cenário 1: manter as áreas de influência atuais

As áreas de influência definidas pela Valnor respeitam os limites geográficos dos municípios. Assim, o depósito de Abrantes é responsável pelas localidades situadas nos concelhos de Abrantes, Sardoal, Vila de Rei, Mação e Gavião, totalizando 57 locais de recolha. O depósito e estação de triagem de Avis é responsável pelos restantes 14 municípios, totalizando 132 locais de recolha. Neste cenário, como são assumidas as áreas de influência da Valnor, apenas foi necessário executar o módulo 4 de modo a otimizar as rotas de recolha de cada depósito.

Os resultados obtidos indicam que a distância total a percorrer no horizonte temporal de 6 semanas é de 29.399 km, dos quais 25.751 km (cerca de 88%) correspondem ao transporte primário e 3.648 km (cerca de 12%) ao secundário. O depósito de Abrantes é responsável por 36% do total da distância percorrida, que inclui a totalidade do transporte secundário (ver Figura 7).

A título exemplificativo apresenta-se na Figura 8 a solução obtida para a recolha do material papel no depósito de Abrantes. Pode-se observar que foram definidas oito rotas, três das quais (#2, #7 e #8) correspondem a rotas diretas à localidade de Abrantes dado o volume de material a recolher. Ao contrário das rotas atuais que se cingem aos limites dos concelhos, na solução proposta para este cenário vários concelhos são integrados numa mesma rota (por exemplo, a rota #1 passa por 5 municípios diferentes). Como o intervalo

entre recolhas para o papel é de 2 semanas, as rotas definidas serão repetidas três vezes durante o horizonte temporal.

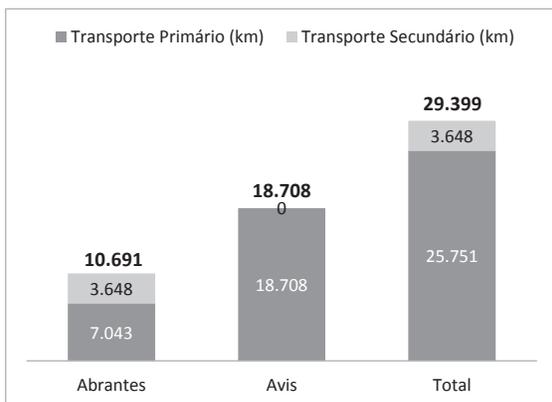
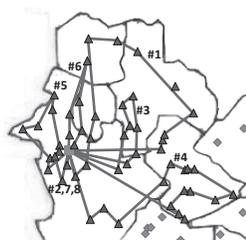


Figura 7: Distância total a percorrer por depósito no cenário 1



Rota	Sequência de Visita	Carga (kg)	NºCont.	Duração* (m)
#1	191 - 123 - 126 - 127 - 138 - 136 - 141 - 140 - 144 - 143 - 153 - 154 - 191	1.918	34	393
#2	191 - 146 - 191	1.975	23	235
#3	191 - 149 - 147 - 133 - 134 - 135 - 131 - 132 - 148 - 191	1.945	36	362
#4	191 - 150 - 152 - 151 - 124 - 116 - 121 - 120 - 118 - 125 - 119 - 128 - 129 - 130 - 117 - 191	1.734	38	398
#5	191 - 156 - 167 - 166 - 165 - 159 - 160 - 161 - 162 - 157 - 155 - 191	1.934	36	364
#6	191 - 163 - 142 - 145 - 139 - 158 - 164 - 187 - 191	1.993	38	384
#7	191 - 185 - 191	1.975	23	235
#8	191 - 186 - 191	1.975	23	235
Total		15.450	251	2.606

* Não está incluído o tempo de descarga do veículo no fim de cada rota

Figura 8: Rotas de recolha de papel no depósito de Abrantes

A solução obtida para este cenário implica 307 horas de trabalho para o depósito de Abrantes e 893 horas para o depósito de Avis, considerando apenas o número de horas necessárias às rotas de recolha e o respetivo tempo de descarga (excluiu-se o tempo relativo ao transporte secundário). Atendendo a que um motorista labora 210 horas no horizonte temporal considerado (35 horas/semana x 6 semanas) são necessários 2 motoristas para o depósito de Abrantes e 5 motoristas para o depósito de Avis. Quanto ao número de veículos e assumindo que cada veículo pode operar 378 horas em 6 semanas, é necessário pelo menos 1 veículo para o depósito de Abrantes e 3 veículos para o depósito de Avis.

Os resultados computacionais do módulo 4 para este cenário são apresentados na Tabela 3. O limite máximo para o tempo de computação varia consoante a dificuldade em encontrar uma boa solução (com um *gap* reduzido ou idealmente de 0%). Assim, no módulo 4, o tempo de computação é inicialmente limitado a 1 hora (3.600 s). Se o *gap* obtido for inferior ou igual a 5% não é necessário aumentar o tempo de computação. Se o *gap* for superior a 5% o tempo de computação é alterado para 4 horas (14.400 s) e a instância é resolvida novamente. Se a solução obtida ao fim de 4 horas apresentar um *gap* superior a 10% então o limite máximo de computação é novamente aumentado, desta vez para 8 horas. As soluções encontradas apresentam um desvio entre a solução inteira e a solução linear relaxada (*gap*) de 1,4% a 8% tendo sido necessário entre 3.600 e 18.000 s para o seu processamento (os 18.000s incluem o tempo de execução de 3.600 s, em que se obteve um *gap* superior a 5% e os seguintes 14.400 s em que se encontrou um *gap* inferior a 10%).

Tabela 3: Resultados computacionais do módulo 4 no cenário 1

Material Reciclável	Depósito Avis			Depósito Abrantes		
	<i>FO (km)</i>	<i>Tempo (s)</i>	<i>Gap</i>	<i>FO (km)</i>	<i>Tempo (s)</i>	<i>Gap</i>
Papel	8.393	18.000	4,3%	4.328	18.000	3,6%
Plástico/Metal	7.672	18.000	3,7%	4.421	3.600	1,4%
Vidro	2.643	18.000	8,0%	1.942	18.000	6,1%

4.2.2 Cenário 2: reestruturar as áreas de influência por material reciclável

Neste cenário, os módulos 1, 2 e 4 do método híbrido foram executados de modo a construir as áreas de influência por material e as rotas de recolha associadas. Este cenário permite que, em cada local, cada material seja recolhido por depósitos diferentes. Assim, e dado existirem três materiais, cada depósito poderá ter três áreas de influência distintas.

O módulo 1 foi executado para cada material tendo-se obtido os resultados apresentados na Tabela 4. Para cada material identifica-se o número total de rotas abertas, o número de locais de recolha que pertencem às rotas abertas, o número de locais de recolha que foram afetos aos dois depósitos através das rotas fechadas e a distância total percorrida. De referir que o módulo 1 gerou

rotas abertas para os materiais papel e vidro, não tendo sido gerada nenhuma para o material plástico/metalo.

Tabela 4: Caracterização da solução obtida para o módulo 1 para cada material

<i>Material</i>	<i>Total Rotas</i>	<i>Rotas Abertas</i>	<i>Locais afetos às Rotas Abertas</i>	<i>Locais afetos ao Dep. Avis</i>	<i>Locais afetos ao Dep. Abrantes</i>	<i>Distância Total (km)</i>
Papel	33	2	23	134	21	10.817
Plástico/Metal	56	0	0	143	40	10.140
Vidro	28	4	28	139	7	3.600

A título demonstrativo apresenta-se na Figura 9 a solução do módulo 1 para os materiais vidro e papel com a identificação das rotas abertas.

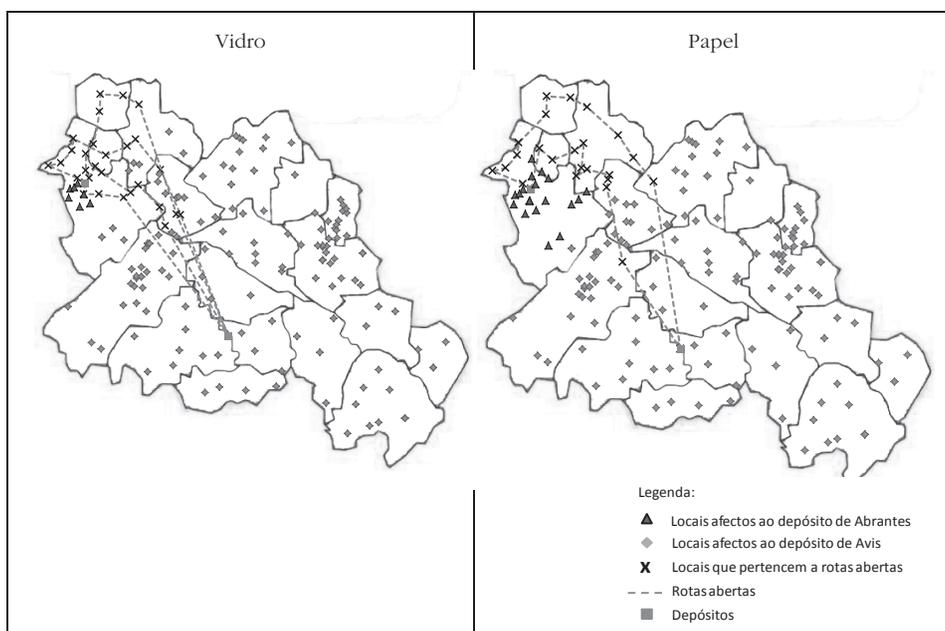


Figura 9: Solução do módulo 1 para os materiais vidro e papel

Dada a existência de rotas abertas para os materiais vidro e papel, o módulo 2 foi executado para estes materiais para cada rota aberta gerada (num total de 6). Dos resultados obtidos conclui-se que os locais de recolha pertencentes às quatro rotas abertas do material vidro foram todos afetos ao depósito de Avis, o que significa que este depósito fica responsável por 167 locais de recolha

para o material vidro, restando apenas 7 para o depósito de Abrantes. No caso do papel, 12 locais de recolha foram afetos ao depósito de Abrantes e 11 ao depósito de Avis, totalizando os depósitos 33 e 145 locais, respetivamente. As áreas de influência finais por material reciclável estão representadas na Figura 10. Note-se que para o plástico/metalo, as mesmas ficaram definidas logo no módulo 1.

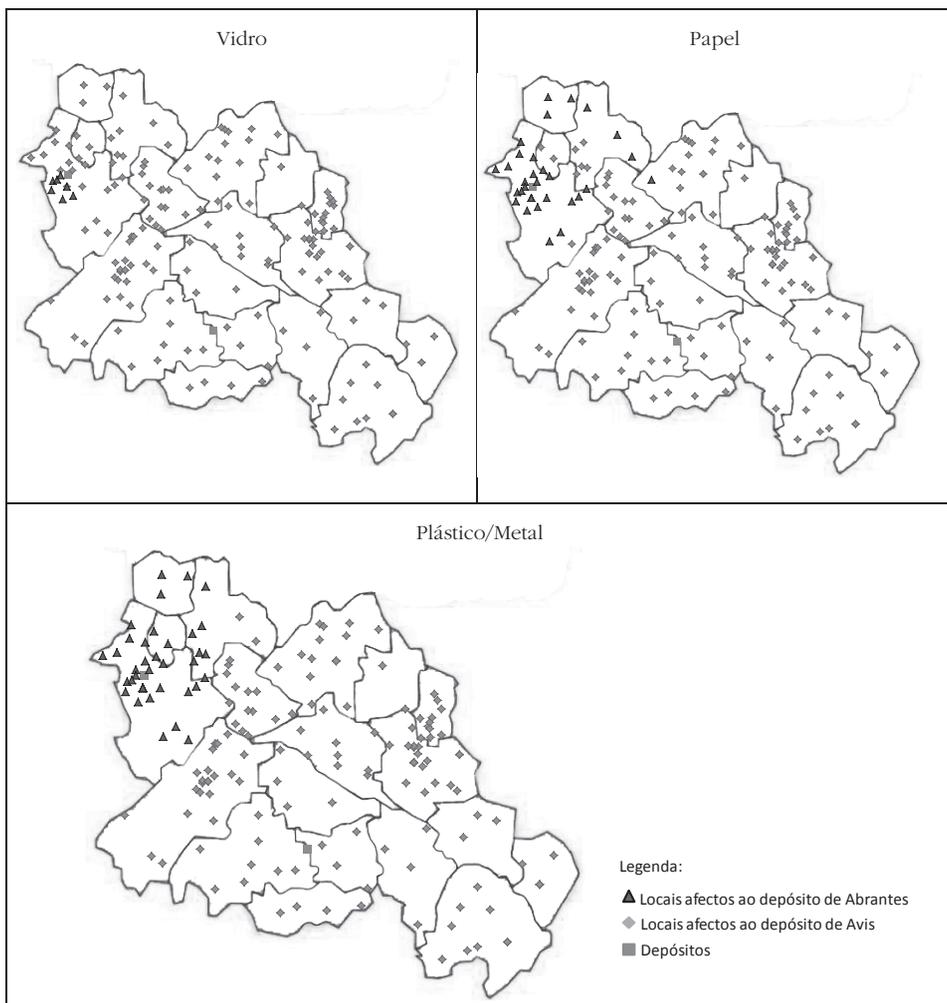


Figura 10: Áreas de influência por material no cenário 2

Pela observação da Figura 10 constata-se que as áreas de influência são significativamente diferentes entre si. O depósito de Abrantes é apenas responsável por 7 locais de recolha para o vidro, sendo responsável por 33 locais para o papel e 40 para a recolha do plástico/metall.

Após a execução do módulo 4 obteve-se uma distância total percorrida de 24.802 km no horizonte temporal considerado (ver Figura 11). Comparando com o cenário 1, o depósito de Avis tem de percorrer mais 1.603 km (o que corresponde a um aumento de 9%). Consequentemente, o depósito de Abrantes diminui a distância a percorrer em 58%. A distância total percorrida diminui 16%, tendo o transporte primário sido reduzido em 14% e o transporte secundário em 27%.

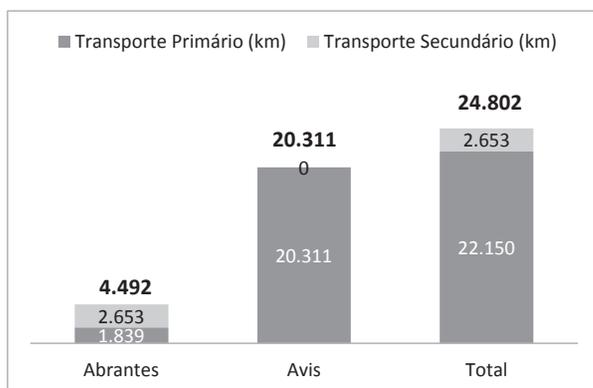


Figura 11: Distância total a percorrer por depósito no cenário 2

Este cenário envolve 1.009 horas de trabalho para o depósito de Avis e 167 horas para o depósito de Abrantes, o que corresponde a uma redução de 2% no total de horas necessárias face ao cenário 1. Assim, são necessários pelo menos 6 motoristas e 4 veículos para operacionalizar esta solução.

Os resultados computacionais dos módulos 1 e 4 estão apresentados nas Tabelas 5 e 6, respetivamente. Como o módulo 1 lida com a instância de maior dimensão (uma vez que considera os 189 locais de recolha em simultâneo) e vai condicionar os resultados dos módulos seguintes, optou-se por alargar o tempo de computação a 8 horas (28.800 s) de forma a encontrar soluções mais próximas do ótimo. Assim, obtiveram-se soluções que apresentam um *gap*

inferior a 5% ao fim de 8 horas de processamento. Relativamente ao módulo 2, as 6 corridas realizadas (correspondente às 6 rotas abertas geradas pelo módulo 1) foram resolvidas até à otimalidade em menos de 1 minuto. O módulo 4 conseguiu provar a otimalidade em 3 instâncias e nas restantes 3 encontrou soluções com *gaps* entre 3,7% e 9,1%. Este último valor corresponde à instância de maior dimensão, com 167 locais de recolha.

Tabela 5: Resultados computacionais para o módulo 1 no cenário 2

Material Reciclável	FO (km)	Tempo (s)	Gap
Papel	10.817	28.800	4,9%
Plástico/Metal	10.140	28.800	3,3%
Vidro	3.600	28.800	4,5%

Tabela 6: Resultados computacionais para o módulo 4 no cenário 2

Material Reciclável	Depósito Avis			Depósito Abrantes		
	FO (km)	Tempo (s)	Gap	FO (km)	Tempo (s)	Gap
Papel	8.968	18.000	8,3%	2.070	22	0,0%
Plástico/Metal	7.977	3.600	3,7%	2.163	43	0,0%
Vidro	3.366	18.000	9,1%	259	1	0,0%

4.2.3 Cenário 3: reestruturar as áreas de influência por depósito

Neste cenário, dado que se pretende definir áreas de influência por depósito, todos os módulos do método híbrido proposto foram executados. A solução final contempla apenas uma área de influência para cada depósito, comum a todos os materiais.

As soluções obtidas para os módulos 1 e 2 são comuns às do cenário anterior, pelo que para este cenário apenas foi necessário executar o módulo 3 antes do módulo 4. Sobrepondo as três áreas de influência obtidas no cenário 2 (Figura 12), observa-se que 32 locais de recolha estão na situação designada de “indecisos”, ou seja, pertencem a depósitos diferentes consoante o material a recolher.

Estes 32 locais de recolha “indecisos” com os três materiais são o *input* do módulo 3, onde uma formulação em PLIM para o *MP-MDVRP* é resolvida,

impondo-se que todos os materiais de um mesmo local pertençam ao mesmo depósito. Como resultado, os 32 locais de recolha foram afetos ao depósito de Abrantes. Assim, este depósito fica responsável por 43 locais de recolha e o depósito de Avis por 146. As áreas de influência finais para o cenário 3 são apresentadas na Figura 13.

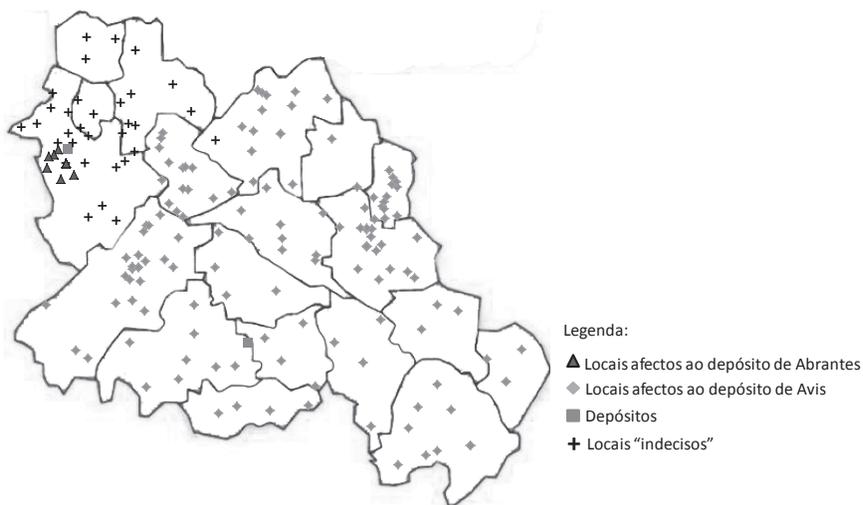


Figura 12: Resultado da sobreposição das três áreas de influência do cenário 2

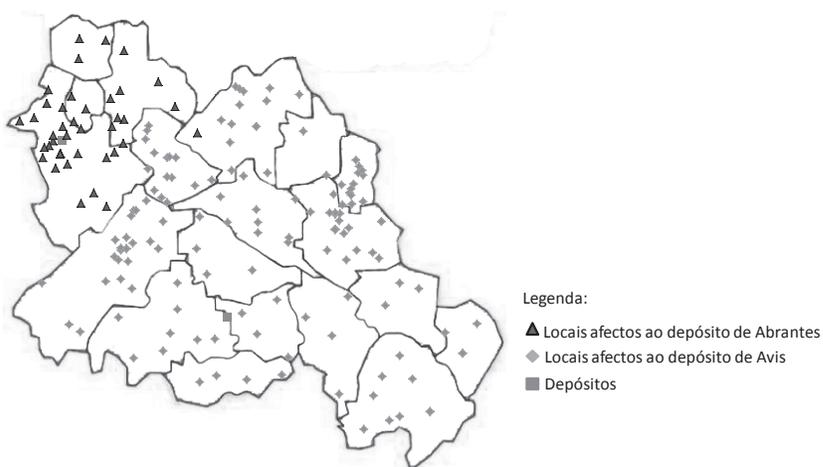


Figura 13: Áreas de influência por depósito no cenário 3

A solução proposta para o cenário 3 implica uma distância total percorrida de 25.290 km (ver Figura 14). Isto significa que uma reestruturação das áreas de influência por depósito permite uma redução de 14% no total da distância a percorrer face às atuais áreas de influência (cenário 1).

Relativamente às horas de trabalho, este cenário envolve um total de 1.165 horas, repartidas por 213 horas para o depósito de Abrantes e 952 horas para o depósito de Avis. Tal como no cenário anterior, são necessários pelo menos 6 motoristas e 4 veículos.

A Tabela 7 apresenta os resultados computacionais para o módulo 4. Novamente é no depósito de Avis e para os materiais papel e vidro que se verificam os maiores *gaps*, mesmo com um tempo computacional de 8 horas (os tempos de computação apresentados na Tabela 7 consideram os tempos que não permitiram obter um *gap* satisfatório). Estas são, como já foi referido, as instâncias de maior dimensão. Relativamente ao módulo 3, o problema com os 32 locais de recolha foi repartido em cinco instâncias, todas elas resolvidas até à otimalidade em menos de 1 minuto.

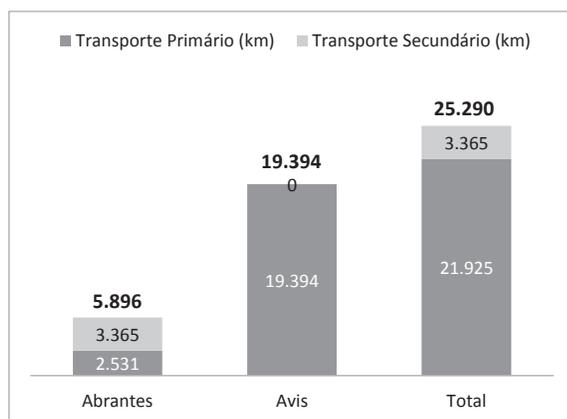


Figura 14: Distância total a percorrer por depósito no cenário 3

Tabela 7: Resultados computacionais para o módulo 4 no cenário 3

Material Reciclável	Depósito Avis			Depósito Abrantes		
	FO (km)	Tempo (s)	Gap	FO (km)	Tempo (s)	Gap
Papel	8.693	46.800	8,4%	2.422	565	0,0%
Plástico/Metal	7.923	3.600	3,7%	2.275	53	0,0%
Vidro	2.779	46.800	9,7%	1.200	3.600	1,8%

5. Discussão Final

O objetivo deste trabalho prende-se com a utilização de técnicas de Investigação Operacional de forma a encontrar melhores soluções para o sistema de recolha de resíduos de embalagens gerido pela Valnor. Entre as várias abordagens possíveis para a resolução deste caso (abordagens exatas, heurísticas ou híbridas) optou-se por uma abordagem híbrida, em que o problema original modelado como um *Multi-Product, Multi-Depot Vehicle Routing Problem* é decomposto em quatro sub-problemas, sendo cada um modelado através de uma formulação PLIM.

A aplicação do método proposto exigiu uma extensa análise e tratamento dos dados de forma a extrair os necessários para alimentar o modelo. Por outro lado, em aplicações reais de problemas de definição de rotas de veículos é impreterível fornecer ao modelo as distâncias reais, o que exige a utilização de um sistema de informação geográfica capaz de construir uma matriz distância de dimensão elevada.

A aplicação do método ao caso Valnor permitiu encontrar soluções mais vantajosas face à situação atualmente implementada, onde são percorridos cerca de 31.000 km no horizonte temporal de 6 semanas (ver Figura 15).

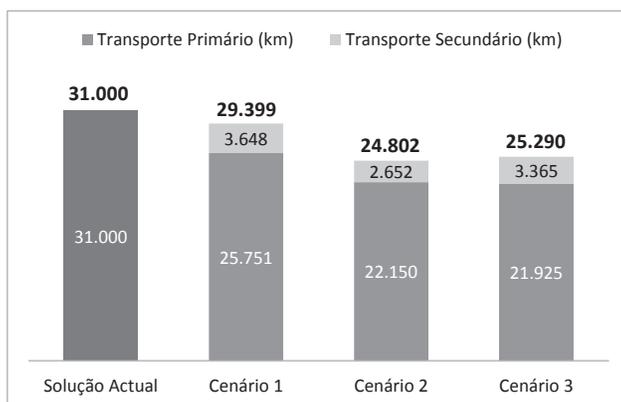


Figura 15: Comparação da distância percorrida nos cenários estudados

Optando por um cenário mais conservador (cenário 1) em que se mantêm as atuais áreas de influência dos dois depósitos, a Valnor conseguirá uma redução de cerca de 5% no total da distância percorrida, se proceder apenas ao

reprojeto e otimização das rotas de recolha. Esta solução representa uma alteração nos circuitos atualmente em vigor e na periodicidade dos mesmos. Será, portanto, a alteração onde deverão existir menos barreiras à sua implementação dado que necessitará apenas de formar os motoristas para novos circuitos de recolha. A taxa de utilização agregada para o número mínimo de veículos necessários para esta solução é de 81% para Abrantes e 79% para Avis (considerando o número de horas de trabalho *versus* número de horas disponíveis).

Optando por um cenário intermédio, em que as áreas de influência são reestruturadas, mantendo-se a recolha de todos os materiais de um mesmo local da responsabilidade de um único depósito (cenário 3), é possível obter, face à solução atual, uma redução de 18% na distância total a percorrer, o que se refletirá diretamente numa redução dos custos operacionais. Neste cenário, um maior número de locais de recolha são afetos ao depósito de Avis (146 contra 132 na situação atual) que opera também como estação de triagem. Assim, neste cenário existe uma redução tanto no transporte primário como no transporte secundário. Face ao cenário 1, a reestruturação das áreas de influência por depósito (cenário 3) gera uma poupança total de 14%. A solução do cenário 3 implica uma alteração das áreas de influência atuais, transferindo-se locais de recolha que atualmente são da responsabilidade de Abrantes para Avis. Esta transferência implica um aumento da carga de trabalho de Avis, resultando num aumento da taxa de utilização dos veículos para 89% e numa diminuição da taxa de utilização de Abrantes para 44%. Ao contrário do cenário 1, o cenário 3 apresenta taxas de utilização dos veículos mais desequilibradas entre depósitos. No entanto, os ganhos relativos à redução da distância total a percorrer são significativos.

O cenário mais afastado da situação atual é aquele em que a lógica da definição das áreas de influência é mais alterada, levando a que cada depósito tenha de gerir três áreas diferentes, uma por cada material (cenário 2). Este é o cenário com a menor distância a percorrer, reduzindo em 20% o valor atualmente verificado. Comparando os cenários 2 e 3, o primeiro implica uma distância menor em apenas 2% face ao último, mas envolve uma mudança mais complexa na gestão das operações no sistema em causa. Por outro lado, esta

solução implica que o sistema de informação de recolha de dados esteja integrado entre depósitos. Note-se que atualmente quando um contentor de um material é recolhido, é registada informação sobre a taxa de enchimento dos restantes contentores que compõe um ecoponto (um ecoponto é constituído por três contentores, um para cada material). Como neste cenário o depósito que visita um tipo de material pode não ser o mesmo que visita os restantes materiais, se o sistema de informação de dados não estiver integrado entre depósitos, perder-se-á informação sobre as taxas de enchimento reais. Esta informação é essencial para se poder replanear as rotas em caso de observação de situações atípicas. No cenário 2, as taxas de utilização para o número mínimo de veículos é de 56% para Abrantes e 84% para Avis, existindo um desequilíbrio menor entre depósitos do que o observado no cenário 3.

Como trabalho futuro pretende-se adicionar um módulo de calendarização das rotas ao método proposto. Uma vez que as rotas estão definidas, para operacionalizar a solução encontrada é necessário definir o dia em que cada uma irá ocorrer, garantindo a frequência de recolha definida e minimizando o número de veículos e motoristas. Por outro lado, dever-se-á estudar a robustez das áreas de influência propostas face ao aumento das quantidades a recolher, quer devido à sazonalidade, quer ao aumento da consciencialização das populações para a reciclagem, o que gerará maiores quantidades depositadas nos contentores.

Por último, a Valnor encontra-se neste momento num processo de alargamento da sua área de abrangência, tendo integrado recentemente seis novos concelhos: Castelo Branco, Idanha-a-Nova, Oleiros, Sertã, Proença-a-Nova e Vila Velha de Rodão. Neste sentido, será necessário estudar esta nova realidade que passará por, numa primeira fase, definir as rotas para estes novos concelhos a partir do novo depósito de Castelo Branco e, numa segunda fase, estudar uma solução integrada para os 25 concelhos e 3 depósitos.

Como comentário final é importante notar como a aplicação de metodologias de Investigação Operacional contribuiu para a resolução de um conjunto de problemas que se colocam à Valnor, permitindo encontrar e comparar soluções alternativas à situação atual que melhoram a sua performance operacional. Por outro lado, as ferramentas desenvolvidas são

suficientemente genéricas, o que permite a sua aplicação a outros casos com características semelhantes, contribuindo assim para melhorar a eficiência em outros sistemas de recolha de resíduos de embalagens.

BIBIOGRAFIA

- ARCGIS, 2011. <http://www.arcgis.com>, consultada em Fevereiro 2011.
- BALDACCI, R., HADJICONSTANTINO, E. e MINGOZZI, A. 2004. *An exact algorithm for the capacitated vehicle routing problem based on a two-commodity network flow formulation*. Operations Research, 52, 723-738.
- BALDACCI, R. e MINGOZZI, A. 2009. *A unified exact method for solving different classes of vehicle routing problems*. Mathematical Programming, 120, 347-380.
- BETTINELLI, A., CESELLI, A. e RIGHINI, G. 2011. *A branch-and-cut-and-price algorithm for the multi-depot heterogeneous vehicle routing problem with time windows*. Transportation Research Part C-Emerging Technologies, 19, 723-740.
- CREVIER, B., CORDEAU, J. F. e LAPORTE, G. 2007. *The multi-depot vehicle routing problem with inter-depot routes*. European Journal of Operational Research, 176, 756-773.
- GOLDEN, B. L., MAGNANTI, T. L. e NGUYEN, H. Q. 1977. *Implementing Vehicle Routing Algorithms*. Networks, 7, 113-148.
- HADJICONSTANTINO, E. & BALDACCI, R. 1998. *A multi-depot period vehicle routing problem arising in the utilities sector*. Journal of the Operational Research Society, 49, 1239-1248.
- HO, W., HO, G. T. S., JI, P. & LAU, H. C. W. 2008. *A hybrid genetic algorithm for the multi-depot vehicle routing problem*. Engineering Applications of Artificial Intelligence, 21, 548-557.
- ILOG-CPLEX, 2011. <http://www-01.ibm.com/software/integration/optimization/cplex-optimizer/>, consultada em Fevereiro 2011.
- INSTITUTO NACIONAL DE ESTATÍSTICA, 2010. *Gestão dos Resíduos em Portugal 2004-2009*. <http://www.ine.pt>, consultada em Agosto 2010.
- LAPORTE, G. 1992. *The Vehicle-Routing Problem - an Overview of Exact and Approximate Algorithms*. European Journal of Operational Research, 59, 345-358.
- LAPORTE, G., NOBERT, Y. e ARPIN, D. 1984. *Optimal solutions to capacitated multi-depot vehicle routing problems*. Congressus Numerantium, 44, 283-292.
- LAPORTE, G., NOBERT, Y. e TAILLEFER, S. 1988. *Solving a Family of Multi-Depot Vehicle-Routing and Location-Routing Problems*. Transportation Science, 22, 161-172.
- LIM, A. & WANG, F. 2005. *Multi-depot vehicle routing problem: A one-stage approach*. Ieee Transactions on Automation Science and Engineering, 2, 397-402.
- PARTHANADEE, P. e LOGENDRAN, R. 2006. *Periodic product distribution from multi-depots under limited supplies*. IIE Transactions, 38, 1009-1026.
- PISINGER, D. e ROPKE, S. 2007. *A general heuristic for vehicle routing problems*. Computers & Operations Research, 34, 2403-2435.
- RENAUD, J., LAPORTE, G. & BOCTOR, F. F. 1996. *A tabu search heuristic for the multi-depot vehicle routing problem*. Computers & Operations Research, 23, 229-235.
- SALHI, S. e SARI, M. 1997. *A multi-level composite heuristic for the multi-depot vehicle fleet mix problem*. European Journal of Operational Research, 103, 95-112.

TILLMAN, F. A. e CAIN, T. M. 1972. *Upperbound Algorithm for Single and Multiple Terminal Delivery Problem*. Management Science Series a-Theory, 18, 664-682.

VIDAL, T., CRAINIC, T. G., GENDREAU, M., LAHRICHI, N. e REI, W. 2012. *A Hybrid Genetic Algorithm for Multidepot and Periodic Vehicle Routing Problems*. Operations Research, 60, 611-624.

(Página deixada propositadamente em branco)

CASO 13

TRANSPORTE MULTIPRODUTO POR OLEODUTO E GESTÃO DE PARQUE DE TANQUES: O CASO DA COMPANHIA LOGÍSTICA DE COMBUSTÍVEIS

Susana Relvas

CEG-IST - Centro de Estudos de Gestão do IST, Instituto Superior Técnico
Universidade de Lisboa
susana.relvas@tecnico.ulisboa.pt

Ana Paula F.D. Barbosa Póvoa

CEG-IST -- Centro de Estudos de Gestão do IST, Instituto Superior Técnico
Universidade de Lisboa
apovoa@tecnico.ulisboa.pt

Henrique A. Matos

CERENA-Centro de Recursos Naturais e Ambiente, Instituto Superior Técnico
Universidade de Lisboa
henrimatos@tecnico.ulisboa.pt

RESUMO

A cadeia de abastecimento do petróleo (CAP) engloba várias entidades e funções, desde a exploração e transformação de crude, até à distribuição final. A competitividade no setor e os custos crescentes têm levado os seus decisores a procurar novas estratégias e melhorias nos processos a vários níveis, nomeadamente na vertente logística desta cadeia.

É exemplo de atividade logística na CAP a Companhia Logística de Combustíveis (CLC), que é responsável pela operação do oleoduto multiproduto entre a Refinaria de Sines e o Parque de Tanques em Aveiras de Cima bem como pela gestão de inventário dos derivados de petróleo que chegam ao seu centro de distribuição, em Portugal. O planeamento do transporte de produtos pelo oleoduto requer a integração da competição entre produtos no escalonamento do transporte, com o fornecimento do mercado e a produção de derivados a montante.

Neste sentido, surgiu em 2005 a colaboração científico-tecnológica entre a CLC e o Instituto Superior Técnico (IST), com o objetivo de sistematizar e otimizar o planeamento de operações a cargo desta empresa. Este objetivo passa pelo estudo do escalonamento do transporte de produtos através de oleodutos multiproduto com integração da gestão de inventário no Parque de Tanques. A metodologia adotada consiste no desenvolvimento de modelos matemáticos de programação linear inteira mista (MILP – Mixed Integer Linear Programming) suportados por heurísticas desenhadas para reduzir a complexidade do espaço de soluções e obter soluções em períodos de tempo útil. O objetivo final desta investigação é o desenvolvimento de um Sistema de Apoio à Decisão para o planeamento das operações neste sistema logístico, cujo processo de validação, desenvolvimento e atualização decorre atualmente junto da CLC.

PALAVRAS-CHAVE

Oleoduto multiproduto, Gestão de inventário, Escalonamento, Reescalonamento, Programação matemática, Tempo e volume contínuos

1. Introdução

A Cadeia de Abastecimento do Petróleo (CAP) apresenta um impacto elevado na economia mundial. Desde os primeiros anos em que o petróleo começou a ser explorado, que tem sido usado como combustível fóssil e como matéria-prima para a produção de produtos de interesse nas diferentes fileiras industriais de forma a satisfazer as mais diversas necessidades. Inicialmente a investigação centrou-se essencialmente no desenvolvimento de novos processos industriais, seguindo-se a automação e controlo de processos e, mais recentemente, o desenvolvimento de ferramentas de apoio à decisão na procura de uma gestão otimizada dos processos produtivos nos diferentes níveis da cadeia com o objetivo de melhorar as margens de lucro globais (Shah, Li e Ierapetrinou, 2011).

Nesta última vertente, um dos níveis críticos da cadeia é o transporte e distribuição de produtos petrolíferos. Uma das necessidades identificadas como importantes prende-se com o desenvolvimento de metodologias de otimização que permitam a definição de estratégias sustentadas de planeamento do transporte e distribuição. Nesta área, os oleodutos aparecem como um equipamento essencial no transporte de volumes elevados, com custos relativamente baixos, ao longo de variadas distâncias e configurações em rede, garantido uma operação segura e fiável (Trench, 2001).

É neste contexto que se enquadra a história e operação da Companhia Logística de Combustíveis (CLC), que opera em Portugal o oleoduto multiproduto que interliga a Refinaria de Sines ao Parque de Tanques da CLC situado em Aveiras de Cima (Figura 1) e que abastece a zona centro do território continental do país.



Figura 1: Oleoduto da CLC

Com o aumento de complexidade da operação e dinamismo do mercado, a CLC assumiu a necessidade de sistematizar o planeamento da sua operação. Surge então a oportunidade de colaboração com o Instituto Superior Técnico (IST), por intermédio de uma equipa de investigadores do Grupo de Operações e Logística do Centro de Estudos de Gestão (CEG-IST) e do Grupo de Integração e Otimização de Processos do Centro de Processos Químicos (CPQ).

O objetivo delineado para esta colaboração é o desenvolvimento de um sistema de apoio à decisão que permita desenvolver o planeamento mensal das operações de transporte de produtos por oleoduto e a associada gestão de inventário no parque de tanques, ambos geridos pela CLC.

Este artigo irá detalhar um estudo de caso que explora o escalonamento da operação do oleoduto operado pela CLC. O capítulo 2 é reservado à definição do problema em estudo, onde será feito um enquadramento real (secção 2.1) e científico (secção 2.2) do problema. A secção 2.3 consiste numa definição do problema em estudo e, posteriormente, a secção 2.4 apresenta os principais pressupostos e considerações que levam à construção dos modelos de otimização. O método de solução é então explorado no capítulo 3, começando a secção 3.1 por apresentar o enquadramento geral e cronológico do processo de desenvolvimento do método proposto. A secção 3.2 apresenta os pressupostos base à construção dos modelos MILP. As secções 3.3 e 3.4 resumem várias estratégias adotadas na representação do problema e redução do espaço de solução do mesmo. O capítulo 4 é inteiramente dedicado à apresentação sucinta da formulação do modelo base do método de solução (modelo I). O capítulo 5 apresenta vários cenários de validação e resultados obtidos pela implementação do método desenvolvido. O capítulo 6 está reservado para a etapa de transferência de tecnologia, onde foi desenvolvida uma interface com o utilizador (secção 6.1) que foi caracterizada e avaliada (secções 6.2 e 6.3). Por fim, serão apontadas as principais conclusões e direções para trabalho futuro no capítulo 7.

2. Definição do Caso de Estudo

Nesta secção, com base no enquadramento real do problema e na literatura da área, será definido o problema a estudar. Esta definição passará por uma primeira fase em que a definição é genérica e numa segunda fase onde serão feitas algumas considerações que permitem moldar o problema real ao problema definido.

2.1. Enquadramento do Problema Real

A CLC é a principal empresa Portuguesa de distribuição logística a operar em Portugal. Garante a operação do oleoduto multiproduto, que interliga a refinaria localizada em Sines (sul de Portugal) com o seu parque de tanques localizado em Aveiras de Cima. Este oleoduto tem um total de 147 km e possui 18.000 m³ de capacidade.

A origem da CLC remonta à década de 90 do século XX, onde por motivos de realização da EXPO 98 na zona oriental de Lisboa, se decidiu dismantlar a Refinaria de Cabo Ruivo. Por um lado, a produção aí realizada foi deslocalizada para as restantes duas Refinarias da Petrogal (hoje Galp). No que respeita à componente logística de distribuição de produtos foi tomada a decisão sobre uma alternativa que mantivesse o abastecimento da zona centro de Portugal. Foi então decidido constituir a CLC, localizada em Aveiras de Cima, onde um Parque de Tanques seria abastecido por oleoduto a partir da Refinaria de Sines. Além de abastecer a mesma zona que a anterior alternativa, permitiu retirar o tráfego diário de cerca de 500 camiões cisterna com combustíveis a circular dentro da cidade de Lisboa.

O oleoduto foi construído entre 1996 e 1997 e a operação da CLC teve início em Abril desse ano (CLC, 2011).

As instalações da CLC cobrem uma área de cerca de 60 ha, incluindo, entre outros, tanques e esferas de armazenagem, num total atual de 14 esferas e 22 tanques, e a estação de receção do oleoduto.

O oleoduto transporta atualmente sete produtos (P1-P7), sendo cinco deles produtos brancos e os restantes dois gases liquefeitos. Dado que o transporte é feito sem elementos separadores entre batches (um batch ou lote consiste no volume de um único produto bombeado e delimitado por dois batches de

outros produtos um a montante e outro a jusante) de sucessivos de produtos, é conhecida a matriz de sequências possíveis entre pares de produtos que garantem a manutenção da qualidade dos mesmos.

Os produtos são fornecidos exclusivamente pela Galp, enquanto a CLC possui vários clientes. Sendo que o consumo da CLC representa cerca de 40% da expedição da refinaria de Sines, o planeamento de operações é feito numa base mensal, que pode ser integrada no plano de produção da Refinaria da Galp em Sines. Para este efeito, a CLC recebe a meio de cada mês uma previsão mensal (com quantidades numa base diária) dos consumos previstos pelos seus clientes. Com base nesta previsão, os chefes de turno da sala de controlo elaboram um plano de transporte e gestão de tanques para o mês de calendário seguinte, que validam junto da Galp.

O plano de transporte deve garantir as sequências possíveis de produtos dentro do oleoduto e garantir ainda que são respeitados os limites mínimos e máximos de caudal de bombagem e de capacidade de armazenagem no parque de tanques. No que respeita ao parque de tanques, há que garantir que em todo o momento existe produto para satisfazer os clientes. Este exercício implica que a gestão de inventário no Parque de Tanques da CLC seja coordenada com o plano de transporte de produtos pelo oleoduto. Assim, em cada mês, os chefes de turno elaboram, com base na sua experiência, dois documentos baseados em folha de cálculo e que denominam por Ordem de Bombagem (plano de transporte de produtos através do oleoduto) e Movimentação de Tanques (gestão de inventário no parque de tanques).

A primeira versão de cada um destes documentos é feita usualmente num turno da noite, onde existe menor atividade no parque, com base nas restrições físicas do sistema e também em algumas restrições operacionais. Por exemplo, caso tenham de ocorrer paragens, tenta-se que estas ocorram com o mínimo de produtos dentro do oleoduto. Por fim, tenta-se que um batch de um determinado produto encha por completo um ou mais tanques, a fim de manter a rastreabilidade dos produtos e também minimizar o efeito de alguma fração de interface que seja diluída nesse batch.

A CLC identificou que este método não é sistemático e é unicamente baseado na experiência dos seus chefes de turno. Não são exploradas sinergias

no planejamento que um método de otimização pode permitir. Além disso, com a evolução da sua operação e do próprio mercado, que atualmente é muito dinâmico e competitivo, mais restrições e direções operacionais surgem. Assim, o método original de planejamento da CLC torna-se pouco flexível e apenas permite enumerar uma solução possível.

2.2. Enquadramento do Problema em Estudo

A literatura existente na área da gestão da cadeia de abastecimento do petróleo foca-se, desde os primeiros anos, nas operações de refinação, uma vez que são a atividade principal desta cadeia. No entanto, nos últimos anos têm surgido trabalhos focados quer na cadeia como um todo (Más e Pinto, 2004), permitindo a integração de operações, quer na distribuição (aqui desde a distribuição rodoviária até ao uso de oleodutos), onde várias vantagens competitivas podem ser atingidas. Na área de transporte e distribuição, os oleodutos aparecem como um equipamento essencial para o transporte de volumes elevados, com custos relativamente baixos, ao longo de distâncias variadas e configurações em rede, garantido uma operação segura e fiável. Na Europa a 27, a densidade de oleodutos era, em 2008, de 8 km/1000 km², correspondendo a mais de 30000 km de tubagem, representando um transporte de 135637 milhões de km de crude e seus derivados (Eurostat Pocket Book, 2008; Noreland, 2008). Apesar de o planejamento de operações neste tipo de equipamentos ser complexo, a falta de soluções comerciais leva a que a maioria das empresas se mantenha a utilizar ferramentas simples baseadas em folha de cálculo, experiência dos programadores e iterações sucessivas. Neste sentido, existe uma enorme oportunidade nesta área para a comunidade científica.

As metodologias de planejamento para a operação de oleodutos têm na programação matemática o método com maiores progressos (de entre um conjunto abrangendo também a simulação e heurísticas – Sasikumar *et al.* 1997 - de índole mais avançado), sendo o trabalho pioneiro de Shah datado de 1996. Os oleodutos multiproduto têm motivado duas linhas de modelos de representação baseados em programação matemática, em que o tempo é

modelado de forma discreta (Magatão *et al.*, 2004, Rejowski e Pinto, 2004) ou contínua (Cafaro e Cerdá, 2004, 2008 e Rejowski e Pinto, 2008).

Mais recentemente os métodos de solução expandiram-se para vários tipos de topologias de oleodutos, desde o oleoduto simples (uma entrada e saída, Relvas *et al.*, 2006), até às redes de oleodutos (Boschetto *et al.*, 2010). Outros trabalhos têm integrado a operação do oleoduto com operações a montante ou jusante (Relvas *et al.*, 2006 e 2007; Relvas, 2008; Relvas, Barbosa-Póvoa e Matos, 2009).

A revisão bibliográfica na área permite concluir que na maioria dos casos têm sido usados modelos MILP para resolução destes problemas, recorrendo a algoritmos comerciais com elevado potencial de resolução na procura de soluções. Os modelos MILP permitem combinar a programação linear contínua com programação inteira, onde diversos tipos de decisões podem ser modelados. Em alguns casos, nomeadamente em aplicações reais, recorre-se adicionalmente e de forma integrada a heurísticas que permitam reduzir o espaço de soluções e permitam encontrar soluções em tempo real, como no caso do trabalho de Boschetto *et al.* (2010).

2.3. Definição do Problema em Estudo

O problema em estudo considera o transporte de produtos derivados do petróleo, desde uma refinaria até um parque de tanques, através de um oleoduto multiproduto (Figura 2). No oleoduto os produtos são bombeados sem que exista uma separação física entre produtos, limitando o mapa de sequências possíveis de modo a garantir a qualidade do produto final. O centro de distribuição é responsável por satisfazer as necessidades do mercado local, que se traduz por uma carteira específica de clientes com procura conhecida e definida sobre uma escala temporal diária. Cada produto, quando alcança o final do oleoduto, é armazenado em um ou mais tanques os quais armazenam sempre esse mesmo produto. Cada tanque que recebe produto deve repousar durante um determinado período de tempo, após o qual pode fornecer os clientes até ao estado vazio. O ciclo repete-se ao longo do tempo.

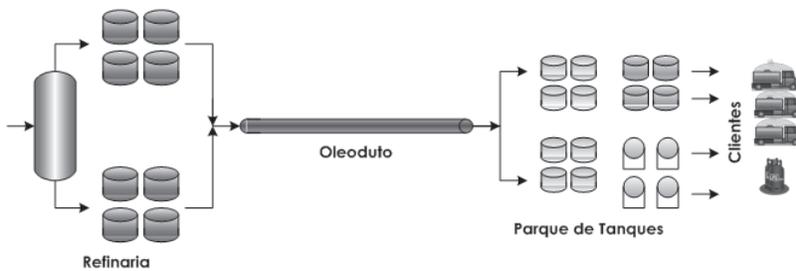


Figura 2 – Esquematização do Sistema em Estudo

O problema pode então ser definido na seguinte estrutura:

Dado: i) o número de produtos a bombear; ii) a matriz de sequências possíveis entre pares de produtos consecutivos; iii) o volume do oleoduto; iv) os caudais mínimo e máximo de bombagem; v) o número máximo de batches a transportar; vi) a extensão do horizonte temporal; vii) o serviço de tanque (alocação de tanques a produtos); viii) os inventários iniciais de cada produto em cada tanque; ix) a capacidade máxima de armazenagem por tanque; x) os produtos e quantidades que se encontram inicialmente no oleoduto; xi) as previsões de saídas de produtos para clientes; e xii) as condições temporais para a localização de possíveis paragens pré-estabelecidas do oleoduto;

Determinar: 1) o escalonamento da bombagem de produtos no oleoduto, nomeadamente a sequência de produtos e o número de batches a bombear, volumes e caudais de bombagem por batch, início e final da bombagem de cada batch para o oleoduto, a localização e duração das paragens pré-estabelecidas do oleoduto; e 2) a gestão de inventário no parque de tanques, incluindo o perfil de inventário por produto (e por tanque) e a monitorização das chegadas de produtos, aprovação e repouso de batches e expedição de produtos para clientes;

A fim de otimizar um objetivo representado por uma função de indicadores de operação.

2.4. Considerações sobre o Problema em Estudo

De modo a propor um método de resolução do problema descrito que se adequa não só ao problema de investigação em estudo, tal como definido na

secção 2.3, mas também ao problema real, devem ser analisadas algumas questões operacionais que suportem a metodologia a desenvolver.

No caso real da CLC, o seu parque de tanques é abastecido exclusivamente a partir da refinaria de Sines, pelo que é relevante o volume requerido pela CLC para o planeamento da produção. O planeamento da produção é feito numa base mensal, o que implica que o planeamento da operação de transporte pelo oleoduto seja feito também na mesma base. Face à literatura da área, esta questão aparece como um desafio, pois usualmente os horizontes típicos das soluções propostas até à data de início da colaboração (2005) nunca ultrapassaram 1 semana. Além deste facto, existe uma condicionante do próprio problema genérico que sugere a utilização de horizontes temporais mais extensos, apesar do planeamento ser ao nível operacional. Neste caso, o transporte sequencial de produtos no oleoduto gera um diferencial de tempo considerável desde que o produto começa a ser bombeado até que está disponível para os clientes finais. Como tal, para que as necessidades do parque de tanques se façam sentir no planeamento do transporte de produtos, o horizonte de planeamento deverá ser, no mínimo, o tempo requerido para um batch de um produto percorrer todo este processo. No caso da CLC, em média este tempo ronda as 75 h, com casos em que as 100 h poderão ser atingidas. Considerando que são bombeados 7 produtos na operação da CLC e que o processo é sequencial (existe sobreposição dos períodos requeridos por produto), para repor pelo menos 1 batch de cada produto é requerida em média uma semana de operação. Assim sendo, um horizonte temporal de um mês permite bombear vários batches do mesmo produto e repor o consumo efetuado.

Por outro lado, considerando que i) a refinaria produz e armazena quantidades superiores ao requerido pela CLC, ii) o oleoduto bombeia apenas um produto de cada vez e iii) existe continuidade da produção ao longo do período de planeamento, então pode assumir-se que não existe limite de quantidade de produto para bombear à entrada do oleoduto.

Outra dificuldade no problema a abordar é a dimensão do parque de tanques da CLC. À data de início da colaboração, existiam 32 tanques (passando a 36 no ano de 2007), que se encontram divididos entre os vários

produtos. Numa abordagem inicial considera-se que cada produto é armazenado num tanque único, agregando a capacidade real de vários tanques num só tanque. Com esta abordagem pretende-se reduzir a complexidade da representação e resolução do problema, estabelecendo um compromisso com a decisão de adotar um horizonte temporal de 1 mês. Esta simplificação implica que no volume de produto determinado pelo modelo haja produto disponível para fornecer clientes e produto que ainda está em fase de repouso, o que irá requerer o controlo de dois níveis diferentes de inventário.

No que respeita à representação temporal do problema, existem duas perspetivas. A perspetiva do cliente, que abastece produtos numa base diária e que disponibiliza com antecedência a sua previsão de consumo diária. A outra perspetiva é a do transporte de produtos no oleoduto, que se caracteriza por ser uma operação contínua. Assim sendo, será necessário lidar com esta dualidade temporal do problema.

Considerando estas questões, o primeiro foco do desenvolvimento da metodologia centra-se na operação do oleoduto, deixando para uma segunda fase a operação detalhada de gestão de inventário.

3. Proposta de Método de Solução

Com base na revisão da literatura e considerando a data do início da colaboração, a larga maioria das abordagens propostas ao problema em estudo consideram a utilização de modelos de otimização usando modelos do tipo MILP e usando para resolução algoritmos exatos comerciais validados pela comunidade científica, como o caso do CPLEX (comercializado à data pela ILOG, atualmente pela IBM). Como tal, optou-se por adotar a mesma abordagem, de modo a tirar partido do conhecimento existente. No que respeita à linguagem de modelação do problema, optou-se por adotar a linguagem de alto nível usada no *software* GAMS (*General Algebraic Modeling System*), uma vez que esta era já conhecida e dominada pela equipa do IST.

Todas estas opções foram sempre validadas junto da CLC, tendo sido discutidas as suas vantagens e desvantagens a nível operacional e a nível financeiro.

3.1. Enquadramento Geral e Cronológico do Método de Solução

O método de solução proposto iniciou-se com o desenvolvimento de um conjunto de modelos matemáticos MILP que representassem o problema em estudo. A estrutura do sistema desenvolvido apresenta-se na Figura 3. Primeiramente, foi desenvolvido o Modelo I (Relvas et al. 2006, Relvas 2008), tal como denominado na Figura 3. Este modelo representa o sistema em estudo (oleoduto multiproduto e parque de tanques), onde a armazenagem é tratada de forma agregada. Os resultados da implementação deste modelo usando dados reais da CLC (Inputs) para horizontes temporais de 1 mês não permitiram obter nenhuma solução após mais de 24 h de computação. Estes resultados iniciais permitiram identificar dificuldades e concluir pela necessidade de melhorar a formulação e simplificar o processo de pesquisa de soluções. É neste sentido que surge a heurística de inicialização, desenvolvida com base no problema em estudo e validada com o caso real (Relvas, Barbosa-Póvoa e Matos 2009). Os resultados obtidos da implementação da heurística permitiram obter soluções com gaps relativos inferiores a 2% em tempos de computação entre 2 e 48 min, num conjunto de testes que incluíram 6 meses distintos. Por gap relativo entenda-se o rácio (em valores absolutos) entre a diferença entre a solução atual e o melhor limite inferior (no caso de minimização) e a solução atual.

O escalonamento reativo surgiu em seguida de modo a poder responder a situações não previstas aquando do momento do planeamento inicial, permitindo assim que sejam acomodados novos dados e situações que alterem o modelo inicial.

Por fim, desenvolveu-se o Modelo II, cuja diferença para o Modelo I reside apenas na forma de representação do inventário no parque de tanques.

Nas secções seguintes serão descritas com mais detalhe as várias componentes do método proposto.

Por questões de limitação da extensão deste artigo, apenas será apresentada a formulação do Modelo I no capítulo 4, sendo que o Modelo II, a Heurística de Inicialização e o Escalonamento Reativo podem ser consultados em publicações dos autores (Relvas et al. 2006, 2007, Relvas 2008, Relvas et al. 2009) e serão brevemente discutidos nas secções deste capítulo.

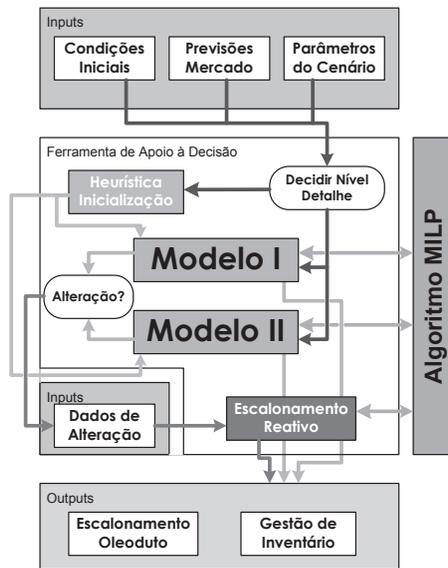


Figura 3 – Arquitectura da Metodologia desenvolvida

3.2 Pressupostos dos Modelos de Programação Linear Inteira Mista

Os trabalhos anteriormente publicados por Relvas et al. (2006) e Relvas (2008) descrevem o modelo MILP de representação agregada de tanques (Modelo I), tendo sido aplicado a problemas de curto e médio prazo. O modelo de representação desagregada (Modelo II) foi construído como uma alternativa em que a diferença reside na representação dos tanques. Os principais pressupostos a ter em conta na conceção dos modelos focam-se nas seguintes questões essenciais:

- (1) Representação temporal e volumétrica;
- (2) Representação da procura;
- (3) Representação de tanques;
- (4) Tempo de repouso mínimo.

- Representação temporal e volumétrica

A escala de tempo utilizada considera uma estratégia do tipo contínuo, utilizando instantes temporais nos quais se finaliza a bombagem de cada batch para o oleoduto. A formulação matemática recorre a variáveis sequenciais para

definir a escala contínua de tempo, baseada nestes eventos. Analogamente, sabendo o volume de cada batch, através de um balanço volumétrico é calculado o volume descarregado pelo oleoduto e a evolução dos produtos ainda no interior do mesmo. Todas as restantes ocorrências no sistema (nomeadamente descarga de produtos ou atualização de perfis de inventário) são atualizadas nesta escala temporal.

- Representação da procura

A procura é fornecida numa base mensal ou semanal e será apresentada em intervalos discretos numa base diária. No modelo, esta informação discreta é traduzida em informação contínua por intermédio de um operador temporal que faz a correspondência entre as escalas temporais.

- Representação de tanques

A gestão de inventário é o fator diferenciador entre o Modelo I e o Modelo II (ver Figura 3). No caso do Modelo I, cada produto é armazenado num tanque virtual que possuiu uma capacidade igual à capacidade agregada dos tanques reais. No entanto esta agregação requer a gestão do nível de inventário global, que tem de respeitar a capacidade disponível, e o nível de inventário disponível para clientes. Este último é constituído por todo o volume de um produto armazenado no parque de tanques que já cumpriu o tempo de repouso mínimo e que deve ser o suficiente para satisfazer a procura atual daquele produto. Assim, cada batch recebido pelo oleoduto é atualizado nos dois perfis de inventário em momentos temporais diferentes e que diferem, pelo menos, no tempo mínimo de repouso (Figura 4, onde a cinzento-escuro se representa um novo batch que permanece em repouso e não disponível para clientes até cumprir o repouso mas que deve ser contabilizado para aferir se há capacidade livre ou não no tanque agregado). Note-se que o tempo de repouso depende do produto e dado a escala temporal ser contínua, o repouso de um batch pode decorrer de um a vários períodos de tempo.

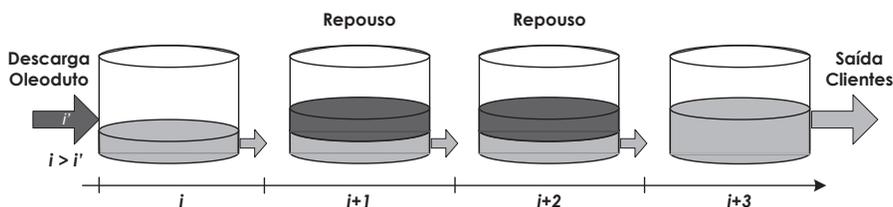


Figura 4 – Ciclo operacional de um tanque agregado (Modelo I)

O Modelo II, por sua vez, adota uma representação desagregada de tanques. Para esta formulação são definidos quatro estados de operação por tanque: E1) cheio e em repouso, E2) a disponibilizar produto para clientes, E3) vazio e E4) a receber produto do oleoduto, tal como descrito na Figura 5. Este ciclo é repetido continuamente quantas vezes necessárias no horizonte temporal. De notar que nos estados dinâmicos (E2 e E4) duas situações podem acontecer num dado intervalo de tempo: i) o tanque fornece/recebe produto e altera a sua ocupação de capacidade ou ii) o tanque está num estado em espera ou estático, mas sem ter atingido o estado cheio ou vazio. Os estados estáticos (E1 e E3) são atingidos quando o nível de inventário ultrapassa determinados níveis pré-definidos da capacidade.



Figura 5 – Ciclo operacional de um tanque de repouso (Modelo II)

O desenvolvimento dos dois modelos reside na sua aplicabilidade. O Modelo I é aconselhável para horizontes temporais de curto a médio prazo (uma semana a um mês) e permite obter resultados que alimentem um plano de produção ao nível da refinaria. Por outro lado, o Modelo II tem caráter mais operacional, utilizando horizontes temporais curtos (semana) e fornecendo esquemas de rotação de tanques nas atividades de receção e expedição de produtos.

- Tempo de repouso mínimo

O tempo mínimo de repouso considera duas situações que devem ser verificadas aquando da descarga de um novo batch para o parque de tanques: tarefas de aprovação do lote e repouso do lote, por questões standard de qualidade. Estes tempos usualmente variam com o tipo de produto. A representação do tempo de repouso é essencial num modelo matemático deste tipo, pois esta tarefa limita consideravelmente a disponibilidade de produtos no parque de tanques.

A representação do tempo de repouso varia entre os dois modelos. No caso de os tanques serem representados de forma agregada, existe a necessidade de monitorizar dois níveis de inventário e a sua viabilidade: o inventário total e o inventário disponível para clientes. Conceptualmente o inventário total permite garantir que um novo batch pode ser armazenado no parque de tanques e o inventário disponível para clientes apenas considera o volume do produto que já efetuou repouso e que está apto a ser comercializado. No que respeita a uma representação desagregada de tanques, o tempo de repouso é representado de uma forma semelhante à realidade: o tanque recetor de um novo batch entra em período de repouso até estar completo o tempo mínimo

3.3 Método de Simplificação da Solução – Heurística de Inicialização

Os primeiros testes ao Modelo I permitiram verificar que a definição da sequência de produtos a bombear tem elevado impacto na performance computacional. Tal facto deve-se à utilização de variáveis sequenciais que são determinadas com base na definição da sequência de produtos, que terão impacto na escala contínua de tempo (usando a duração da bombagem de cada batch em sequência) e na escala volumétrica (através do volume de cada batch bombeado). Adicionalmente, estas questões entram em conflito com aplicações de carácter real, em que é requerida uma boa solução num curto período de tempo (máximo 2 horas).

Esta situação motivou o desenvolvimento de uma heurística de sequenciação a ser utilizada antes da utilização do modelo matemático. Esta heurística analisa os dados operacionais do cenário, nomeadamente inventários iniciais e procura de produtos, bem como as restrições operacionais do período

temporal de modo a estabelecer direções de operação. A análise destes dados permite identificar prioridades na bombagem e limites ao número de batches a bombear. As prioridades mais elevadas de produtos a bombear são determinadas considerando que o oleoduto permanece parado (ausência de novos batches) e verificando qual o produto que entra primeiro em rutura de stock. No que respeita ao número de batches a bombear, o critério consiste em considerar que se pretende repor todo o volume que é consumido por clientes durante o horizonte temporal. Considera-se ainda como dados o tempo previamente conhecido como disponível para operar e os limites de caudal de bombagem. A heurística pode ser utilizada com dois fins: geração de várias sequências fixas a utilizar no modelo para posterior comparação de resultados ou apenas inicialização da sequência e fornecimento de dados relativos aos números mínimos e máximos de batches a bombear.

A aplicação desta heurística mostrou-se essencial para diminuir a complexidade do método de resolução do problema em estudo, melhorando significativamente os tempos de computação. Verificou-se ainda que a hipótese mais viável é a geração de um conjunto de sequências fixas, direcionadas para o cenário em questão. Mais informação sobre a heurística de inicialização pode ser consultada em Relvas (2008), Relvas et al. (2009).

3.4 Escalonamento Reativo

Num ambiente industrial real, sujeito aos efeitos dinâmicos da cadeia de abastecimento, de mercado e operacionais, ocorrem frequentemente situações imprevistas face ao planeado. A CLC não é exceção, até porque para um período de planeamento com duração de um mês, existe uma elevada probabilidade de ocorrência de situações imprevistas ou de instabilidade dos dados usados no planeamento. Assim sendo, um dos requisitos necessários para solucionar o problema real foi o desenvolvimento de uma metodologia que permitisse refazer o planeamento com base em situações não previstas inicialmente. Para tal usou-se o escalonamento reativo, ou seja, face a uma alteração, o plano atual seria alterado para um novo plano. De modo a tornar este método viável no caso real da CLC e do contexto onde se insere, é necessário que a alteração a fazer ao planeamento seja a mínima possível,

minimizando o impacto no sistema e na fronteira que interage com o mesmo. Para o sistema em estudo foram analisadas as situações imprevistas que ocorrem com maior frequência. A Tabela 1 apresenta a classificação adotada e a descrição como as suas causas mais comuns.

De modo a utilizar os modelos MILP desenvolvidos para o escalonamento reativo, modelaram-se matematicamente as várias potenciais ocorrências imprevistas correspondentes às situações descritas na Tabela 1.

Em seguida, foi estruturado um algoritmo de reescalonamento em que os dados operacionais relativos à situação imprevista são introduzidos, bem como os dados temporais relativos à mesma, sendo ainda fornecido o planeamento em vigor (isto porque, em situações reais, o planeamento anterior poderá já conter alterações efetuadas pelos operadores). É necessário definir o instante em que a situação é identificada e o instante em que a situação terá efeito no planeamento. O instante em que a situação é identificada será tomado pelo modelo como o instante presente. A informação proveniente do escalonamento anterior associada a instantes temporais anteriores ao instante atual é tomada como fixa (já passada). O instante ao qual a situação tem efeito será o instante a partir do qual os novos dados terão efeito no planeamento futuro.

A função objetivo além de ser função de um conjunto de indicadores de operação, terá termos de penalização cuja função é garantir desvios mínimos nas variáveis de decisão selecionadas. Deste modo garante-se que o novo planeamento considera as situações imprevistas, alterando o mínimo possível o planeamento em vigor, de modo a garantir que quer na origem quer no destino as condições de operacionalidade são mantidas.

O escalonamento reativo foi aplicado com sucesso a vários exemplos. No que respeita à aceitação por parte da CLC, esta foi extremamente positiva pois a revisão de um planeamento tem sido, segundo o seu método manual, a tarefa mais complicada de resolver. Para mais informação sobre o escalonamento reativo devem ser consultadas as publicações dos autores (Relvas et al. 2007, Relvas 2008).

Tabela 1 - Situações imprevistas no planeamento da CLC

	Situação	Descrição
S1	Variação da previsão da procura	Revisões periódicas ao longo do horizonte temporal
S2	Alteração à sequência de produtos	Essencialmente devido a alterações operatórias não planeadas ou indicações provenientes de outros níveis organizacionais
S3	Paragem do oleoduto imprevista	Usualmente devido a rutura de <i>stock</i> na origem ou razões operacionais (avaria nas bombas, entre outras)
S4	Alteração a volumes de <i>batches</i>	Usualmente devido a rutura de <i>stock</i> na origem
S5	Ajustes de caudal	Usualmente devido a questões operacionais relativas à bombagem de produtos ou para satisfazer pedidos imprevistos de clientes
S6	Variação na capacidade de armazenagem	Poderá ocorrer em situações de manutenção do tanque

4. Formulação Matemática do Modelo I

Neste capítulo apresenta-se sucintamente a formulação matemática do Modelo I, que representa a operação do oleoduto e a gestão de inventário de forma agregada.

4.1. Nomenclatura

- Índices/Conjuntos

$i \in I$	=	<i>Batches</i>
$p \in P$	=	Produtos
$lt \in LT$	=	Volumes discretos de <i>batches</i>
$k \in K$	=	Dias
$I^{new} \subset I$	=	<i>Batches</i> cuja bombagem ocorre durante o horizonte temporal
$T^{stop} \in T$	=	Instante de paragem do oleoduto (h)
$T^{start} \in T$	=	Instante de arranque do oleoduto (h)

- Parâmetros

h_{max}	=	Horizonte Temporal (h)
vb_{min}	=	Caudal Mínimo (uv/h) ¹
vb_{max}	=	Caudal máximo (uv/h)

¹ uv – unidades volumétricas

D_{\max}	=	Limite superior de volume a descarregar do oleoduto num período de tempo (uv)
D_{\min}	=	Limite inferior de volume a descarregar do oleoduto num período de tempo (uv)
σ	=	Coordenada volumétrica do ponto final do oleoduto (uv)
M_{vol}	=	<i>big-M</i> , relacionado com o volume do oleoduto
$T_{\text{rep},p}$	=	Tempo de repouso mínimo do produto p (h)
$ID_{\max,p}$	=	Capacidade de armazenagem do produto p (uv)
Wo_i	=	Volume dos <i>batches</i> que estão inicialmente no oleoduto (uv)
$TDem_k$	=	Escala temporal diária (k) para atualização da procura de clientes (h)
$lots_{p,lt}$	=	Matriz de volumes lt de <i>batches</i> possíveis para cada produto p (uv)
$Dem_{p,k}$	=	Procura do produto p no dia k (uv)
$sequence_{p,p'}$	=	Matriz de sequências possíveis entre pares de produtos p e p'
ID_p^0	=	Inventário inicial do produto p no parque de tanques (uv)
w_1, w_2, w_3	=	Pesos dos termos da função objetivo

- Variáveis Não Negativas

C_i	=	Tempo ao qual o <i>batch</i> i termina de ser injectado no oleoduto (h)
L_i	=	Duração da injeção do <i>batch</i> i (h)
Q_i	=	Volume do <i>batch</i> i (uv)
$F_i^{i'}$	=	Coordenada volumétrica superior do <i>batch</i> i enquanto é injectado o <i>batch</i> i' (uv)
$W_i^{i'}$	=	Fração volumétrica do <i>batch</i> i ainda no interior do oleoduto ao injectar o <i>batch</i> i' (uv)
$D_i^{i'}$	=	Fração volumétrica do <i>batch</i> i descarregada do oleoduto ao injectar o <i>batch</i> i' (uv)
$D_{\text{clients},i}^{i'}$	=	Volume do <i>batch</i> i disponibilizado para clientes após repouso aquando da injeção do <i>batch</i> i' (uv)
qd_p^i	=	Procura do produto p enquanto é injectado o <i>batch</i> i (uv)
$DV_{i,p}^{i'}$	=	Fração de volume do <i>batch</i> i de produto p descarregada do oleoduto aquando da injeção do <i>batch</i> i' (uv)
$DV_{\text{clients},i,p}^{i'}$	=	Volume do <i>batch</i> i de produto p disponibilizada para clientes aquando da injeção do <i>batch</i> i' (uv)
$ID_{\text{total},p}^i$	=	Inventário total do produto p disponível no parque de tanques aquando da injeção do <i>batch</i> i (uv)
ID_p^i	=	Inventário de produto p disponível para clientes aquando da injeção do <i>batch</i> i (uv)
$T_{\text{set},i}^{i'}$	=	Tempo cumulativo de repouso do <i>batch</i> i aquando da injeção do <i>batch</i> i' (h)

$T_{setp,i,p}^{i'}$ = Tempo cumulativo de repouso do *batch* i de produto p aquando da injeção do *batch* i' (h)
minid = Variável maior que qualquer dos inventários finais de todos os produtos

- Variáveis Binárias

$y_{i,p}$ = O *batch* i contém o produto p caso $y_{i,p} = 1$
 $x_i^{i'}$ = Uma fração do *batch* i pode ser descarregada aquando da injeção do *batch* i' caso $x_i^{i'} = 1$
 $ls_{i,p,lt}$ = O *batch* i tem o volume lt do produto p caso $ls_{i,p,lt} = 1$
 $dme_{i,k}$ = O dia k (0 horas) está no intervalo temporal $[0, C_i]$ caso $dme_{i,k} = 1$.
 $dm_{i,k}^{Final}$ = As 0 horas do dia k estão localizadas no intervalo temporal $[C_{i-1}, C_i]$ caso $dm_{i,k}^{Final} = 1$
 $x_{set,i}^{i'}$ = O *batch* i já efetuou tempo de repouso aquando da injeção do *batch* i' caso $x_{set,i}^{i'} = 1$
 $x_{available,i}^{i'}$ = O *batch* i já pode ser disponibilizado para clientes aquando da injeção do *batch* i' caso $x_{available,i}^{i'} = 1$
 $x_{clients,i}^{i'}$ = O *batch* i está disponível para clientes aquando da injeção do *batch* i' caso $x_{clients,i}^{i'} = 1$
 $pmatch_i$ = O tempo de paragem do oleoduto (T^{stop}) é menor que C_i caso $pmatch_i = 1$
 $pstop_i$ = A paragem do oleoduto ocorre no intervalo temporal i caso $pstop_i = 1$

4.2. Formulação Matemática do Modelo I

Por restrições de dimensão do artigo, apenas a Função Objetivo é descrita com detalhe. As restrições do modelo apenas são apresentadas e agrupadas por função. Para maior detalhe consultar Relvas et al. (2006) e Relvas (2008).

- Função Objetivo

$$\min \left(w_1 \times \frac{\sum_{i \in I} \sum_{i' \in I^{new}} D_i^{i'} - \sum_{p \in P} \sum_{k \in K} Dem_{p,k}}{h_{\max} \times vb_{\max}} - w_2 \times \frac{\sum_{i \in I^{new}} L_i}{h_{\max}} - w_3 \times \text{minid} \right) \quad (1)$$

$$\sum_{i \in I} \sum_{i' \in I^{new}} D_i^{i'} - \sum_{p \in P} \sum_{k \in K} Dem_{p,k} \geq 0 \quad (1a)$$

$$\text{minid} \geq ID_{total,p}^i \Big|_{i=|I|} \quad (1b)$$

A função objetivo minimiza um conjunto de indicadores de operação que por relacionarem grandezas diferentes se encontram adimensionalizados e com valores restringidos ao intervalo [0,1]. De modo a poder dar-se diferentes pesos a cada indicador, cada termo pode ser multiplicado por um valor (w_1 , w_2 e w_3), associados a diferentes modos operacionais. O primeiro termo da função objetivo pretende minimizar a diferença entre o volume total bombeado e o volume total consumido, garantindo com a restrição (1a) que esta diferença é sempre positiva ou nula. O segundo termo maximiza o tempo total de bombagem de produtos. Estes dois termos em conjunto permitem minimizar o caudal de bombagem mas garantindo a reposição de inventário. O caudal é minimizado indiretamente pois por questões de linearidade não é uma variável do modelo. Por fim o último termo maximiza o menor dos inventários finais relativos por produto, por uma estratégia *minmax* usando a restrição adicional (1b). Assim, existe alguma garantia de flexibilidade deste inventário para inicializar o horizonte temporal seguinte sem que ocorram ruturas de *stock*.

- Sequenciação de *Batches*

$$C_i - L_i \geq C_{i-1} \quad \forall i \in I^{new} \quad (2)$$

$$L_i \leq C_i \quad \forall i \in I^{new} \quad (3)$$

$$C_i \leq h_{\max} \quad \forall i \in I^{new} \quad (4)$$

- Relação entre volume do *batch* e duração da sua bombagem para o oleoduto

$$vb_{\min} \times L_i \leq Q_i \leq vb_{\max} \times L_i \quad \forall i \in I^{new} \quad (5)$$

$$l_{\min} = \frac{\min(lots_{p,lt})}{vb_{\max}} \quad (6)$$

$$l_{\max} = \frac{\max(lots_{p,lt})}{vb_{\min}} \quad (8)$$

$$\sum_p y_{i,p} \times l_{\min} \leq L_i \leq \sum_p y_{i,p} \times l_{\max} \quad (9)$$

- Sequências proibidas

$$y_{i-1,p} + y_{i,p'} \leq 1 + \text{sequence}_{p,p'} \quad \forall i \in I^{new}, \forall p, p' \in P \quad (10)$$

- Limites superior e inferior do *batch* i no interior do oleoduto

$$F_{i+1}^{i'} + W_i^{i'} = F_i^{i'} \quad \forall i \in I, \forall i' \in I^{new}, i' > i \quad (11)$$

- Gestão das operações que ocorrem no final do oleoduto

$$Q_i = W_i^i + D_i^i \quad \forall i \in I^{new} \quad (12)$$

$$F_i^i = W_i^i \quad \forall i \in I^{new} \quad (13)$$

$$W_i^{i'} = W_i^{i'-1} - D_i^{i'} \quad \forall i \in I, \forall i' \in I^{new}, i' > i \quad (14)$$

$$D_i^{i'} \leq D_{\max} \times x_i^{i'} \quad \forall i \in I, \forall i' \in I^{new}, i' \geq i \quad (15)$$

$$F_i^{i'} \geq \sigma \times x_i^{i'} \quad \forall i \in I, \forall i' \in I^{new}, i' \geq i \quad (16)$$

$$F_{i+1}^{i'-1} \leq \sigma - D_i^{i'} + (1 - x_i^{i'}) \times M_{vol} \quad \forall i \in I, \forall i' \in I^{new}, i' > i \quad (17)$$

$$W_i^{i'-1} \geq D_{\min} \times x_i^{i'} \quad \forall i \in I, \forall i' \in I^{new}, i' > i \quad (18)$$

$$D_i^{i'} \leq W_i^{i'-1} \quad \forall i \in I, \forall i' \in I^{new}, i' > i \quad (19)$$

- Restrições de alocação de produtos a *batches*

$$\sum_{p \in P} y_{i,p} \leq 1 \quad \forall i \in I^{new} \quad (20)$$

$$\sum_{p \in P} y_{i,p} \leq \sum_{p \in P} y_{i-1,p} \quad \forall i \in I^{new} \quad (21)$$

$$\sum_p y_{i,p} = 1 \quad \forall i \in I^{new} \quad (22)$$

- Seleção do volume do *batch*

$$\sum_{lt \in LT} ls_{i,p,lt} = y_{i,p} \quad \forall i \in I^{new}, p \in P \quad (23)$$

$$Q_i = \sum_{p \in P} \sum_{lt \in LT} (ls_{i,p,lt} \times lots_{p,lt}) \quad \forall i \in I^{new} \quad (24)$$

- Balanço volumétrico global ao oleoduto ao injetar o *batch* i'

$$Q_{i'} = \sum_{i \in I, i \leq i'} D_i^{i'} \quad \forall i' \in I^{new} \quad (25)$$

- Condições iniciais do oleoduto

$$W_i^{i'-1} = W_{O_i} \quad \forall i \notin I^{new}, i' = \text{first}(I^{new}) \quad (26)$$

- Gestão de inventário no parque de tanques

$$DV_{i,p}^{i'} \leq D_{\max} \times y_{i,p} \quad \forall i' \in I^{new}, p \in P, i' \geq i \quad (27)$$

$$\sum_p DV_{i,p}^{i'} = D_i^{i'} \quad \forall i' \in I^{new}, i' \geq i \quad (28)$$

$$ID_{total,p}^{i'} = ID_{total,p}^{i'-1} + \sum_{i \in I, i \leq i'} DV_{i,p}^{i'} - qd_p^{i'} \quad \forall p \in P, i' \in I^{new} \quad (29)$$

$$ID_{total,p}^i = ID_p^0 \quad i = 1, \forall p \in P \quad (30)$$

$$ID_{total,p}^{i'} \leq ID_{max,p} \quad \forall p \in P, i' \in I^{new} \quad (31)$$

- Gestão da procura de produtos

$$C_i \geq TDem_k \times dme_{i,k} \quad \forall i \in I^{new}, k \in K \quad (32)$$

$$C_i \leq TDem_k + h_{max} \times dme_{i,k} \quad \forall i \in I^{new}, k \in K \quad (33)$$

$$dm_{i,k}^{Final} = dme_{i,k} - dme_{i-1,k} \quad \forall i \in I^{new}, k \in K \quad (34)$$

$$qd_{p,i} = \sum_{k \in K} Dem_{p,k} \times dm_{i,k}^{Final} \quad \forall i \in I^{new}, p \in P \quad (35)$$

- Tempo de repouso

$$W_i^{i'-1} \geq (1 - x_{set,i}^{i'}) \times D_{min} \quad \forall i \in I, i' \in I^{new}, i' > i \quad (36)$$

$$x_{set,i}^{i'} \leq 1 - x_i^{i'} \quad \forall i \in I, i' \in I^{new}, i' > i \quad (37)$$

$$T_{set,i}^{i'} \leq x_{set,i}^{i'} \times h_{max} \quad \forall i \in I, i' \in I^{new}, i' > i \quad (38)$$

$$T_{set,i}^{i'} \leq T_{set,i}^{i'-1} + C_i^{i'} - C_{i-1}^{i'-1} \quad \forall i \in I, i' \in I^{new}, i' > i \quad (39)$$

$$T_{set,i}^{i'} \geq T_{set,i}^{i'-1} + C_i^{i'} - C_{i-1}^{i'-1} - (1 - x_{set,i}^{i'}) \times \quad \forall i \in I, i' \in I^{new}, i' > i \quad (40)$$

$$\sum_{p \in P} T_{setp,i,p}^{i'} = T_{set,i}^{i'} \quad \forall i \in I, i' \in I^{new}, i' > i \quad (41)$$

$$T_{setp,i,p}^{i'} \leq y_{i,p} \times h_{max} \quad \forall i \in I, i' \in I^{new}, i' > i \quad (42)$$

$$T_{setp,i,p}^{i'-1} \geq T_{rep,p} \times (y_{i,p} + x_{available,i}^{i'} - 1) \quad \forall i \in I, i' \in I^{new}, i' > i \quad (43)$$

$$T_{setp,i,p}^{i'-1} \leq T_{rep,p} + x_{available,i}^{i'} \times h_{max} \quad \forall i \in I, i' \in I^{new}, i' > i \quad (44)$$

$$x_{clients,i}^{i'} = x_{available,i}^{i'} - x_{available,i}^{i'-1} \quad \forall i \in I, i' \in I^{new}, i' > i \quad (45)$$

$$D_{clients,i}^{i'} \leq Q_i \quad \forall i \in I, i' \in I^{new}, i' > i \quad (46)$$

$$D_{clients,i}^{i'} \geq Q_i - (1 - x_{clients,i}^{i'}) \times D_{max} \quad \forall i \in I, i' \in I^{new}, i' > i \quad (47)$$

$$D_{clients,i}^{i'} \leq x_{clients,i}^{i'} \times D_{max} \quad \forall i \in I, i' \in I^{new}, i' > i \quad (48)$$

$$ID_p^{i'} = ID_p^{i'-1} + \sum_{i \in I, i \leq i'} DV_{clients,i,p}^{i'} - qd_p^{i'} \quad \forall p \in P, i' \in I^{new} \quad (49)$$

$$ID_p^i = ID_p^0 \quad i = 1, \forall p \in P \quad (50)$$

$$ID_p^{i'} \leq ID_{max,p} \quad \forall p \in P, i' \in I^{new} \quad (51)$$

$$\sum_p DV_{clients,i,p}^{i'} = D_{clients,i}^{i'} \quad \forall i \in I, i' \in I^{new}, i' > i \quad (52)$$

$$DV_{clients,i,p}^{i'} \leq D_{max} \times y_{i,p} \quad \forall i \in I, i' \in I^{new}, p \in P, i' > i \quad (53)$$

- Parâmetros do Oleoduto

$$C_{i-1} \geq T^{stop} \times pmatch_{i-1} \quad \forall i \in I^{new} \quad (54)$$

$$C_{i-1} + L_i \leq T^{stop} + pmatch_i \times h_{max} \quad \forall i \in I^{new} \quad (55)$$

$$pstop_i = pmatch_i - pmatch_{i-1} \quad \forall i \in I^{new} \quad (56)$$

$$C_i \geq C_{i-1} + L_i + (T^{start} - T^{stop}) \times pstop_i \quad \forall i \in I^{new}, i' \in I^{new}, i > i' \quad (57)$$

- Restrições Auxiliares

$$F_i^{i'} \geq F_i^{i'-1} \quad \forall i \in I, \forall i' \in I^{new}, i' > i \quad (58)$$

$$F_i^{i'} - W_i^{i'} \geq F_i^{i'-1} - W_i^{i'-1} \quad \forall i \in I, \forall i' \in I^{new}, i' > i \quad (59)$$

$$F_i^{i'} \geq W_i^{i'} \quad \forall i \in I, \forall i' \in I^{new}, i' \geq i \quad (60)$$

$$\sum_{i \in I, i \leq i'} D_i^{i'} = \sum_p \sum_{l \in L} (ls_{i',p,l} \times lots_{p,l}) \quad \forall i' \in I^{new} \quad (61)$$

$$\sum_{i \in I^{new}} dm_{i,k}^{Final} = 1 \quad k \in K \quad (62)$$

$$\sum_{i' \in I^{new}, i'' \geq i} x_i^{i''} \geq x_{set,i}^{i'} \quad k \in K \quad (63)$$

5. Validação do Método Desenvolvido

Nesta secção serão brevemente apresentados resultados que ilustram a utilização do sistema desenvolvido. Para tal, foi adotado um exemplo real baseado na operação da CLC e com horizonte temporal de um mês. De notar que para o caso de estudo aqui apresentado apenas são considerados 6 produtos, sendo este o lote de produtos em vigor à data da recolha de dados.

Todos os modelos desenvolvidos foram implementados em GAMS 22.4 (Brooke et al., 2006) e solucionados usando o CPLEX 10.0. instalado num Pentium D820 com 2 GB RAM. Foi assumido como critério de paragem uma das seguintes opções: duas horas de computação ou a obtenção de uma solução com gap relativo inferior a 5%.

Os resultados serão apresentados sequencialmente, primeiro na definição e aplicação de sequências de bombagem no Modelo I, desenvolvimento do plano de gestão de inventário com o modelo II e escalonamento reativo.

5.1. Definição de Sequências de Bombagem e Escalonamento com o Modelo I

O primeiro passo para usar a metodologia passa por definir sequências de produtos a serem bombeadas no oleoduto. Foram propostas cinco sequências usando a heurística de inicialização, com números de batches a bombear que variam entre 33 e 41. Estas sequências foram construídas com o pressuposto que os batches dos diferentes produtos bombeados deverão repor o volume previsto a ser consumido no mês em estudo. Após obter as sequências, estas foram testadas no Modelo I, tendo-se obtido os resultados apresentados na Tabela 2. Verifica-se que a sequência menor torna a solução inviável (poderá não permitir manter o inventário nos limites admissíveis apesar de prever a reposição do volume consumido), enquanto as restantes têm solução. Apenas a solução para 35 batches requereu um tempo de computação inferior a duas horas para cumprir o critério de paragem imposto. Nas restantes, o gap relativo é inferior a 5% pois é adicionada a possibilidade de o CPLEX aplicar a opção de polishing à solução encontrada durante alguns segundos.

Tabela 2 – Comportamento do Modelo I para o cenário em estudo

Sequência	33	35	37	39	41
Variáveis contínuas	19088	21228	23480	25844	28320
Variáveis binárias	4685	5111	5553	6011	6485
Restrições	34101	38030	42171	46524	51089
Nós explorados	-	1082	16349	9807	73745
Iterações	-	797306	10139853	9211173	6600480
CPU (s)	33.578	521.640	7230.562	7230.781	7231.531
Função Objectivo	-	-2.229111	-2.231544	-2.231544	-2.234025
Gap Relativo (%)	-	0.11	0.52	0.52	0.41

Na Tabela 3 incluem-se os resultados operacionais que cada uma das seqüências gerou ao ser implementada no Modelo I. Incluiu-se também na última coluna da tabela os valores calculados para os mesmos indicadores tendo como base a solução proposta pelos chefes de turno da CLC.

Tabela 3 – Resultados Operacionais do Modelo I para o cenário em estudo

Seqüência	33	35	37	39	41	CLC (38)
Caudal Médio (uv/h)	-	511	511	511	510	558
Δ Inventário (uv)	-	+380	+680	+680	+280	+19580
Utilização do Oleoduto (%)	-	99	99	99	99	95
Nível de Inventário Final (%)	-	55	56	56	55	65
Inventário Mínimo (%)	-	16.4 (P3)	18.4 (P4)	18.8 (P3)	21.2 (P3)	5.3 (P3)
Inventário Mínimo Final (%)	-	51.7 (P1)	52.5 (P1)	52.5 (P1)	52.5 (P1)	20.5 (P3)

Observando a Tabela 3 verifica-se que todas as soluções viáveis apresentam um caudal médio inferior ao proposto pela solução real. Tal deve-se ao facto de o modelo minimizar o caudal de bombagem mas com a ressalva que deve ser repostos o volume consumido, sendo que as soluções propostas apenas transportam algumas uv's a mais face ao consumido no parque de tanques da CLC. A utilização do oleoduto e o nível de inventário final global também se mantiveram semelhantes entre as soluções propostas. As duas últimas linhas incluem outros indicadores de inventário que são referentes a produtos e não ao inventário global. O inventário mínimo contabiliza o nível relativo de inventário que qualquer um dos produtos verificou ao longo do horizonte temporal, sendo o produto P3 aquele que mais frequentemente apresenta o menor valor. Além disso, o valor mais baixo verificou-se na solução real. Estes valores são essenciais para verificar a flexibilidade da solução, pois caso ocorram situações imprevistas, um nível de inventário mais baixo tem menor flexibilidade para acomodar alterações. Já no caso dos inventários finais, também a solução real é menos flexível. Neste caso trata-se de flexibilizar o início do próximo período de operação.

A sequência selecionada para este mês é a sequência com 41 *batches*, uma vez que reúne o maior conjunto de indicadores operacionais com melhores valores. Nesta sequência, a título exemplificativo, são bombeados 11 *batches* do produto P1, 5 *batches* de P2, 1 *batch* de P3, 9 *batches* de P4, 10 *batches* de P5 e 5 *batches* de P6.

5.2. Gestão de Inventário - Modelo II

Com a sequência selecionada usou-se o Modelo II para obter um escalonamento mais detalhado da 1ª semana de operação, cujos resultados se encontram nas tabelas 4 e 5. Para esta solução considerou-se que os 41 *batches* a bombear representam em média 1.32 *batches* por dia. Em 7 dias de operação seriam bombeados 9.26 *batches*. Assim, assumiu-se um valor máximo de 15 *batches* e, caso cada *batch* fosse usado na sequência, seguiria a sequência pré-determinada com a heurística de inicialização. Os resultados foram obtidos em menos de 4 minutos de computação, com um valor de função objetivo penalizado face aos valores obtidos pelo Modelo I. De notar que a função objetivo (1) usou pesos para w_1 de 3, para w_2 de 2 e para w_3 de 0,5 para os diferentes termos (nesta ordem). Os resultados computacionais encontram-se na Tabela 4.

Tabela 4 – Comportamento do Modelo II para a 1ª semana do cenário em estudo (Sequência com 41 *batches*)

	Modelo II
Variáveis contínuas	5311
Variáveis binárias	2199
Restrições	11474
Nós explorados	7429
Iterações	462125
CPU (s)	191.953
Função Objectivo	-1.969278
Gap Relativo (%)	0.04

No caso dos resultados operacionais apresentados na Tabela 5, verifica-se um ligeiro agravamento dos indicadores analisados quando se passa do Modelo I para o Modelo II (por exemplo, aumento do caudal médio),

implicando algum ajustamento nos resultados uma vez que se tem o sistema mais detalhado. De qualquer forma, todos os indicadores são ainda melhores que os indicadores obtidos na solução real. De notar que aqui apenas se mostram os resultados operacionais para o Modelo I referentes à 1ª semana, referentes aos resultados obtidos da otimização mensal.

Tabela 5 – Resultados Operacionais do Modelo II para a primeira semana do cenário em estudo

Solução	Modelo II	Modelo I	CLC (38)
Caudal Médio (UV/h)	537.43	527.84	550.00
Δ Inventário (UV)	+30	-171	-23063
Utilização do Oleoduto (%)	92.59	92.59	79.76
Nível de Inventário Final (%)	55.68	55.05	42.88
Inventário Mínimo (%)	23.63 (P6)	23.58 (P3)	18.19 (P6)
Inventário Mínimo Final (%)	23.65 (P3)	23.58 (P3)	23.89 (P3)

5.3. Escalonamento Reativo

Além dos resultados iniciais, ou seja, planeamento inicial, também se acompanhou o desenrolar deste mês em detalhe a fim de registar e documentar todas as alterações passíveis de reajustar o planeamento em vigor. Ao todo houve necessidade de usar o método de planeamento reativo em 5 ocasiões distintas. De notar que o método desenvolvido permite que mais do que uma das situações seja incorporada no mesmo reescalonamento. Por exemplo, a última revisão do planeamento incluiu novas previsões de clientes referentes à última semana (mais detalhadas que as originais), aumento de capacidade de um dos produtos devido à entrada em funcionamento de um novo tanque e inclusão de uma paragem obrigatória no oleoduto para ligar o novo tanque à linha de abastecimento do oleoduto.

Neste caso, a função objetivo original foi alterada, incluindo agora termos de penalização. Estas penalizações permitem que sejam minimizadas alterações ao planeamento original. Consideraram-se alterações de sequência e alterações de volume de batches. Ambas as ocorrências são controladas por variáveis binárias e, como tal, incluíram-se na função objetivo dois termos que penalizam alterações ao valor inicial de cada uma destas variáveis. Os pesos

destes termos são unitários. No entanto, cada alteração conta com o valor de 1, o que poderá prejudicar bastante o valor mínimo da função objetivo. Em termos de computação, os critérios de paragem usados foram os mesmos que na aplicação dos Modelos I e II. Os resultados computacionais dos vários reescalamentos encontram-se na Tabela 6.

Verifica-se que conforme se avança no horizonte temporal, o facto de se ter um menor período para refazer o planeamento, contribui largamente para que o tempo de computação diminua drasticamente. Em termos de alterações, a primeira revisão do planeamento (R1) foi aquela que teve mais alterações. Neste caso, a situação que ocorreu implicava, ela própria, alteração da sequência de produtos por falta de inventário na refinaria. Em termos de dimensão do modelo, esta permanece relativamente estável porque, apesar de apenas se reajustar uma parte do planeamento, todo o mês é contabilizado. A parte inicial do planeamento é fixada. A solução ótima é obtida para as três últimas revisões, enquanto na primeira não se atingiu o gap relativo desejado ao final de duas horas de computação.

Tabela 6 – Resultados computacionais para o reescalamento

	Plano Inicial	R1	R2	R3	R4	R5
Variáveis contínuas	28320	28320	28320	28320	28316	28356
Variáveis binárias	6485	5625	6358	6230	6230	6014
Restrições	51089	51089	51089	51090	51076	51937
Nós explorados	73745	438395	51880	558	2182	1
Iterações	6600480	9537433	1407067	6043	31084	69
CPU (s)	7231.5	7230.6	1294.468	15.4	29.1	1.8
Função Objectivo	-2.2340	5.9160	-1.0992	-2.0992	-0.0992	-2.0240
Gap Relativo (%)	0.41	8.07	4.89	0.00	0.00	0.00
Número de Penalizações	-	8	1	0	2	0

Por fim, na Figura 6 mostra-se a comparação entre os perfis de inventário após a última revisão feita a partir da solução proposta neste caso de estudo pelo Modelo I e a solução adotada na CLC. Apesar de os perfis não serem semelhantes, os padrões seguidos e as ordens de grandeza são comparáveis.

Os desfasamentos nas receções de produto (declives positivos) são devidos ao produto ser recebido em pontos temporais diferentes, devido a se terem seqüências e volumes bombeados diferentes.

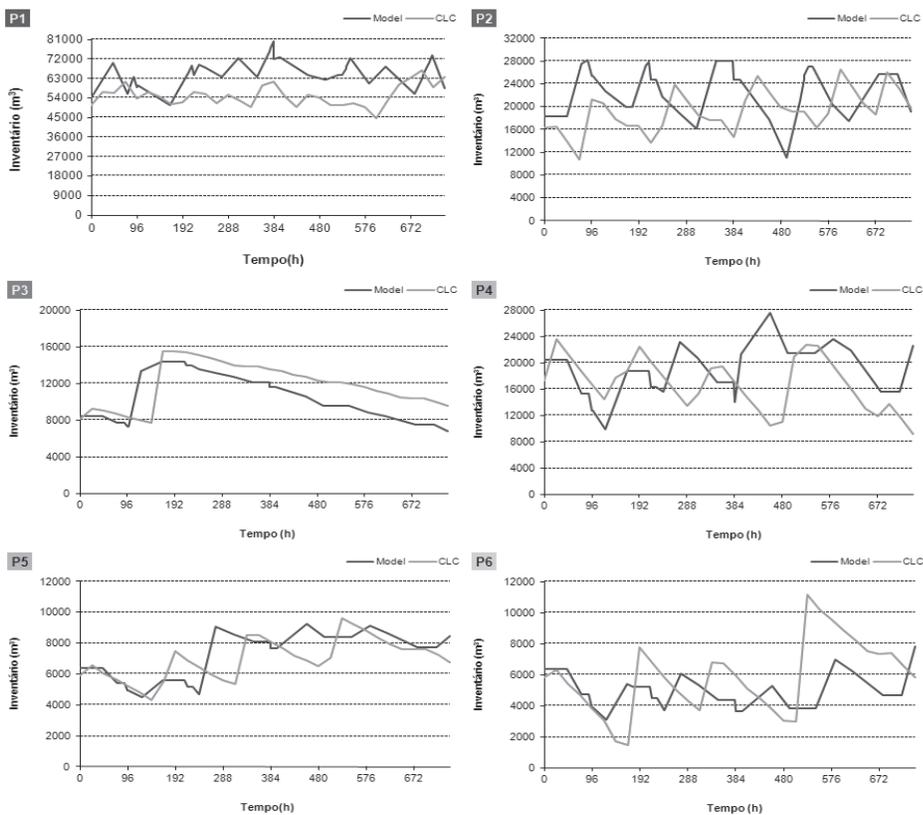


Figura 6 – Perfis de Inventário por produto após revisões: Solução Proposta vs. Real

6. Transferência de Tecnologia

Com a finalidade de permitir que um utilizador com conhecimento sobre o sistema mas sem conhecimento sobre a metodologia de solução desenvolvida possa usufruir de uma solução otimizada em detrimento de uma solução baseada em conhecimento pessoal foi desenvolvida uma interface para o sistema descrito. Esta interface e sistema de apoio à decisão associado permitiu não só consolidar a sistematização da solução desenvolvida, através da ligação a bases de dados com a informação necessária para correr a solução, mas também sistematizar o armazenamento de resultados obtidos.

Esta interface foi desenvolvida por um aluno de mestrado do IST de Engenharia Informática, sob orientação da restante equipa do IST. O software ao ser desenvolvido foi sendo apresentado na CLC para customização conjunta da ferramenta.

6.1 Desenvolvimento do Software de Interface

O ponto inicial para a fase de desenvolvimento do software passou pela compreensão do tipo de interface mais adequada para a futura aplicação. Neste caso, identificou-se que a interface requerida é direcionada para o utilizador. Neste ponto, foi necessário confrontar o software onde foram desenvolvidos os modelos base – GAMS – e o formato da interface. Selecionou-se um formato de aplicação baseada em browser (aplicação Web). Esta permite acessibilidade a qualquer utilizador, bastando para tal que a aplicação esteja sediada unicamente num servidor acedido remotamente. Além disso, este tipo de aplicação elimina potenciais incompatibilidades de sistema operativo e os privilégios a atribuir a cada utilizador podem ser definidos na aplicação. Todos os dados são assim instalados apenas do lado do servidor, o que simplifica também a fácil manutenção e atualizações.

Como o GAMS é uma aplicação multiplataforma, a única restrição para o programador é o sistema operativo do servidor. Esta restrição é essencial no processo de escolha da linguagem de programação e no sucesso global da aplicação. Outra restrição é o próprio programador, que deve conhecer profundamente a linguagem a adotar. Por fim, existe o critério financeiro, que neste caso direcionou a escolha para uma linguagem de livre acesso e gratuita.

Outras ferramentas essenciais são as bibliotecas desenvolvidas por terceiros. Existe uma larga variedade de bibliotecas e muitas delas também de acesso livre. Assim, a qualidade do projeto pode ser assegurada usando este tipo de bibliotecas. Um bom exemplo é a necessidade de gerar páginas Web compatíveis com todos os browsers.

Com esta base, a escolha da linguagem de programação e bibliotecas a utilizar é essencial para assegurar a qualidade e sucesso da aplicação a desenvolver. Para a interface desenvolvida foram escolhidas várias aplicações

de acesso livre, tais como JBoss RichFaces, JBoss Ajax4jsf, JBoss Seam 2.1.1, JBoss Tools 3.0.0.CR1, EJB3, JFreeCharts and MySql 5.1.

6.2 Instalação do Software no Cliente

O desenvolvimento da interface para um método de escalonamento com utilização num ambiente empresarial real requer a análise de várias questões que vão desde a sua arquitetura conceptual até à implementação real ou integração de sistemas e validação. O principal objetivo para o processo de desenvolvimento da interface e transferência de tecnologia foi garantir a sua fiabilidade e a facilidade de utilização, sem esquecer que a navegação e os resultados apresentados pela interface tenham significado para os utilizadores. O objetivo da ferramenta no seu global sempre foi a obtenção de boas soluções para o problema real em estudo e em tempo de computação reduzido.

A implementação em ambiente real de uma ferramenta deste tipo requer um elevado nível de interação com o cliente e uma clara definição temporal dos recursos necessários em cada fase da passagem de tecnologia. O nível inicial requer o levantamento dos sistemas de informação presentes no cliente, nomeadamente ao nível da recolha de dados a usar pelo software em desenvolvimento, formato de relatórios de resultados, manutenção de base de dados, formatação, segurança, requisitos em termos de recursos em sistemas de informação, compatibilidade ou ainda procedimentos do cliente a ter em conta. Também se deve pensar na manutenção futura, que pode incluir não só a atualização da interface mas também das metodologias de resolução adotadas (modelos ou heurísticas). Finalmente, a aplicabilidade da ferramenta desenvolvida deve ser testada no cliente, usando os seus dados e limites temporais reais, para comparação dos resultados com aqueles obtidos pelo método de planeamento até então em vigor.

6.3 Avaliação da Utilização da Solução Real

A interface foi instalada na CLC durante o ano de 2010, tendo passado pelas várias fases descritas na secção anterior. O processo de instalação permitiu atualizar e corrigir questões pontuais de utilização e modificar algum aspeto criticado e discutido junto do cliente. O processo geral baseou-se sempre numa

proposta por parte da equipa da CLC que era apresentada e instalada posteriormente na CLC e, após algum tempo de teste e validação, era atualizada ou alterada para um formato acordado entre as partes.

O processo de validação da ferramenta desenvolvida permitiu não só o melhoramento da interface mas também dos próprios modelos. Algumas das soluções obtidas nos testes de validação mostraram que para operacionalizar os resultados obtidos seria necessário melhorar o modelo base, nomeadamente ao nível da representação do sistema, hipóteses base de construção, simplificações e estratégias de modelação. Um exemplo é a escala de tempo adotada. Esta escala depende do final da bombagem de cada batch, que tem, conforme o produto envolvido, diferentes durações de bombagem. Uma vez que qualquer outra operação no sistema só pode ser atualizada ao final de cada um destes intervalos, no caso dos intervalos maiores existe pouca flexibilidade operacional para as restantes operações.

O Modelo II ainda não foi incorporado na ferramenta atualmente implementada no cliente. Em alternativa, foi construído um algoritmo que distribui os produtos bombeados pelos tanques disponíveis, permitindo analisar dessa forma os recursos a gerir na CLC ao longo do horizonte temporal. Este algoritmo segue as mesmas regras usadas pelos chefes de turno da CLC na gestão de inventário.

Os dados do cenário em estudo a introduzir pelo utilizador incluem o horizonte temporal para o escalonamento, o inventário inicial disponível por produto e as previsões de procura diária por produto. Este último dado foi integrado com bases de dados já disponíveis na empresa.

Em seguida a heurística de sequenciação é aplicada ao cenário em estudo disponibilizando ao utilizador várias sequências possíveis de produtos a bombear. O utilizador pode selecionar quais das sequências vai testar e definir um tempo máximo de computação para obter uma solução do Modelo I com cada sequência. Quando toda a execução termina, os resultados são visionados a partir das páginas de resultados específicos (Figuras 7 e 8) podendo-se aceder ainda a uma página de resumo dos resultados (Figura 9), que contém uma breve descrição de cada solução obtida por sequência, incluindo o nome da sequência, o estado final do modelo, o tempo de computação e data de

computação. Além dos resultados computacionais, esta página também mostra resultados operacionais. Estes podem apoiar a decisão sobre qual a sequência que deve ser adotada. Estes resultados são ilustrados por indicadores que incluem o caudal médio de bombagem (que se pretende baixo, pois indiretamente minimiza o consumo energético), duração de paragens, caso existam (as paragens, a menos que obrigatórias, são indesejáveis, pois permitem que a interface entre os produtos se alargue), balanço entre consumo no parque e volume reabastecido pelo oleoduto e, por fim, o nível de inventário global final.

Os resultados de cada sequência podem ser acedidos pela página resumo e incluem os perfis de inventário por produto (Figura 7), o perfil de caudal de bombagem e o escalonamento da bombagem (Figura 8). Este último tem informação sobre a sequência de produtos, os volumes de cada batch, os instantes temporais de início e fim de cada bombagem de um batch, os instantes iniciais e finais de cada descarga de um batch, e os tempos em trânsito de cada batch.

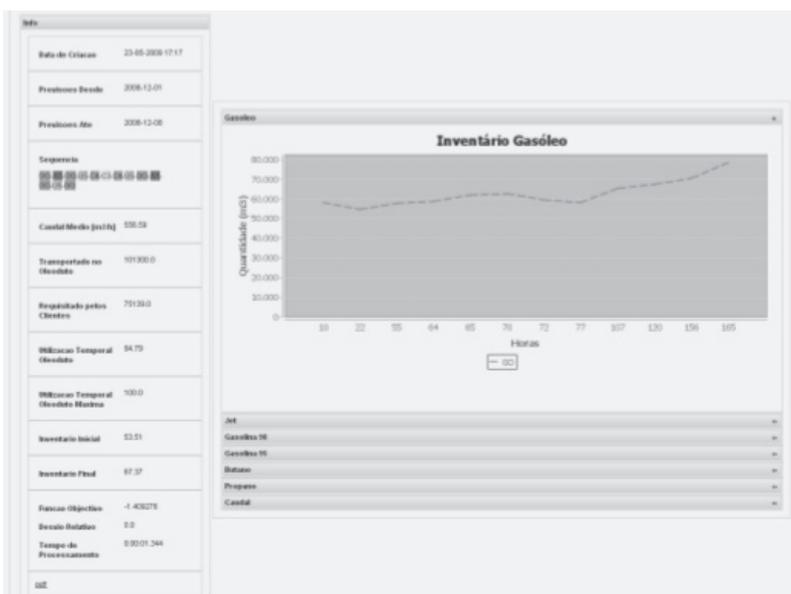


Figura 7 – Página de resultados 1: Perfis de Inventário.

Ordens de Bombagem												
Numero	Partidos	Quantidade [m3]	Capacidade [m3/h]	Tempo de Bombagem	Desde	Até	Tempo de Trânsito	Relatório Bom.	Desde	Até	Tempo de Resposta	Utilização PE
1	██████	18000.0	0.0	18	01-12-2008 00:00	01-12-2008 10:00	111	111	111	111	111	Op.
2	██████	8000.0	680.0	12.31	01-12-2008 10:00	01-12-2008 22:18	111	111	111	111	111	Op.
3	██████	17000.0	513.68	33.88	01-12-2008 22:18	03-12-2008 07:59	111	111	111	111	111	Op.
4	██████	3000.0	450.0	0.44	03-12-2008 07:59	03-12-2008 16:26	111	111	111	111	111	Op.
5	██████	880.0	571.43	1.4	03-12-2008 16:26	03-12-2008 17:50	111	111	111	111	111	Op.
6	██████	3000.0	680.0	4.37	03-12-2008 17:50	03-12-2008 23:36	111	111	111	111	111	Op.
7	██████	880.0	571.43	1.4	03-12-2008 23:36	04-12-2008 00:00	111	111	111	111	111	Op.
8	██████	3000.0	680.0	5.85	04-12-2008 00:00	04-12-2008 05:51	111	111	111	111	111	Op.
9	██████	17000.0	579.04	29.95	04-12-2008 05:51	05-12-2008 11:42	111	111	111	111	111	Op.
10	██████	8000.0	680.0	12.31	05-12-2008 11:42	05-12-2008 00:00	111	111	111	111	111	Op.
11	██████	17000.0	485.32	36.34	05-12-2008 00:00	07-12-2008 12:56	111	111	111	111	111	Op.
12	██████	3000.0	450.0	0.44	07-12-2008 12:56	07-12-2008 21:23	111	111	111	111	111	Op.
13	██████	17000.0	680.0	26.62	07-12-2008 21:23	09-12-2008 00:00	111	111	111	111	111	Op.

Figura 8 – Página de resultados 2: Sequência de Bombagem

Lista de Ordens de Bombagem									
Data de Coluna	Partidas	Sequencia	Capacidade Média [m3/h]	Tempo Respost. [m3]	Utilização Temporal Oleoduto Máx.	Inv. Médio Inv. Final	Pico do Objetivo	Desvio Relativo	Tempo de Proc.
23-05-2009 17:27	2009-12-01	██████	450.0	75400.0	85.43%	53.51%	-1.802075	0.0	0.00:00.281
	2009-12-08	██████		79199.0	93.62%	94.19%			
23-05-2009 17:17	2009-12-01	██████	558.59	101308.8	94.79%	53.51%	-1.409276	0.0	0.00:01.344
	2009-12-08	██████		79199.0	108.8%	87.37%			

Figura 9 – Página de resumo de resultados

7. Conclusões e Direções Futuras

O presente estudo de caso pretende ilustrar a aplicação de técnicas de investigação operacional na resolução de um caso real de elevada complexidade - transporte multiproduto por oleoduto e gestão de inventário num parque de tanques. Importa referir que para a resolução deste caso a existência de uma forte colaboração entre a empresa, fornecedora do caso, e a comunidade académica que desenvolveu as metodologias de modelação e resolução foi essencial. Como resultado ambas as equipas envolvidas e as instituições beneficiaram positivamente em termos de resultados. Conclui-se que é necessário dedicar tempo e os recursos humanos necessários e interdisciplinares para que uma parceria deste tipo alcance o sucesso desejado. O desenvolvimento do método base para o sistema de apoio à decisão requereu conhecimentos de Investigação Operacional e *Process Systems Engineering*, enquanto o desenvolvimento da interface com o utilizador requer conhecimentos de Engenharia Informática.

No que respeita à contribuição científica, os modelos MILP desenvolvidos têm uma componente inovadora no que respeita à integração do transporte de produtos por oleoduto multiproduto com a gestão de inventário destes mesmos produtos. Tal abordagem permite contribuir cientificamente não só para a área das cadeias de abastecimento de petróleo, nomeadamente escalonamento de oleodutos, mas também para a área de *Enterprise-Wide Optimization* (Grossmann, 2005), onde o foco está em integrar diferentes níveis de decisão, processos/operações e metodologias de solução. No caso da modelação da operação do oleoduto o trabalho individualiza-se por se focar em soluções para prazos temporais intermédios (um mês). No caso específico da gestão de inventário, o trabalho desenvolvido é dos primeiros a considerar a operação de um parque de tanques, onde cada tanque segue um ciclo de operações, sendo cada estado e a transição entre estados modelada.

No ponto atual em que se encontra a colaboração já foi transferido para a empresa o *software* desenvolvido. No entanto, verifica-se que o ciclo de colaboração tem potencial para continuar. As principais forças motrizes para que a colaboração continue são não só a validação e atualização do *software*, mas também a incorporação de novas ideias ou métodos com melhores resultados para o problema em questão. Tal justifica-se por dois motivos: a dinâmica do próprio sistema em estudo, que é altamente dependente de várias externalidades, e a evolução científica na área em geral.

Neste ponto, é importante referir algumas direções futuras. A equipa do IST desenvolveu recentemente um modelo para a operação do oleoduto que permite obter soluções para o mesmo horizonte temporal mas que não recorre a soluções híbridas com heurísticas (Relvas *et al.*, 2012). Além de ser um modelo holístico no que respeita à obtenção de solução, permite obter soluções em tempo comparável à solução anterior. Este modelo carece ainda de mais validação para aferir a sua adequação ao problema real.

Outro ponto com potencial para melhoria é a representação da gestão de inventário no parque de tanques, de modo a que o modelo integrado de oleoduto e parque de tanques possa ser otimizador também para períodos de um mês (Relvas, 2008).

Por fim, existe ainda potencial para se melhorar a modelação integrada a fim de permitir uma utilização flexível e adaptada à realidade atual, uma vez que neste momento se verifica ainda que existe pouca flexibilidade no modelo. Tal deve-se ao facto de se ter adotado uma única escala temporal sob a qual todos os eventos ocorrem no sistema. Dado que se estão a considerar operações diferentes, a especificidade das várias ocorrências também é diferente. Tal implica que haja mais pontos temporais de avaliação do sistema, ou seja, ou uma escala mais detalhada ou um outro tipo de escala temporal, mas que não adicione complexidade ao método de solução.

Agradecimentos

Os autores agradecem todo o apoio financeiro e operacional disponibilizado pela CLC, através dos Projetos de Cooperação Científico Tecnológica, celebrados entre a CLC e o IST, e a Bolsa de Doutoramento em Empresas SFRH/BDE/15523/2004, em parceria com a FCT.

BIBIOGRAFIA

- BOSCHETTO, S. N.; Magatão, L.; Brondani, W. M.; Neves, F., Jr.; Arruda, L. V. R.; Barbosa-Póvoa, A. P. F. D.; Relvas, S. (2010). An Operational Scheduling Model to Product Distribution through a Pipeline Network. *Industrial & Engineering Chemistry Research*, 49, 5661–5682;
- BROOKE, A.; Kendrick, D.; Meeraus, A.; Raman, R. (2006). *GAMS — A User's Guide*; GAMS Development: Washington, DC;
- CAFARO, D. C., Cerdá, J., (2004), Optimal Scheduling of Multiproduct Pipeline Systems using a Non-Discrete MILP Formulation, *Comput. Chem. Eng.*, 28 2053-2068;
- CAFARO, D. C., Cerdá, (2008), Dynamic scheduling of multiproduct pipelines with multiple delivery due dates, *Comput. Chem. Eng.*, 32, 728–753;
- CLC (2011), Site Oficial da CLC, Consultado em Janeiro de 2012 em www.clc.pt;
- EUROSTAT POCKET BOOK (2008), Energy, transport and environment indicators, 78-79;
- GROSSMANN, I. (2005), Enterprise-wide Optimization: A New Frontier in Process Systems Engineering. *AIChE Journal*, 51, 1846-1857;
- MAGATÃO, L., Arruda, L. V. R., Neves, Jr F., (2004), A Mixed Integer Programming for Scheduling Commodities in a Pipeline, *Comput. Chem. Eng.*, 28 171-185;
- MÁS, R., Pinto, J.M. (2004), A Mixed-Integer Optimization Strategy for Oil Supply in Distribution Complexes, *Opt. Eng.*, 4, 23–64;
- NORELAND, J. (2008) Eurostat Statistics in Focus: Transport, 35;
- REJOWSKI, Jr. R., Pinto, J. M., (2004), Efficient MILP Formulations and Valid Cuts for Multiproduct Pipeline Scheduling, *Comput. Chem. Eng.*, 28, 1511-1528;

REJOWSKI, Jr. R., Pinto, J. M., (2008), A novel continuous time representation for the scheduling of pipeline systems with pumping yield rate constraints, *Comput. Chem. Eng.*, 32, 1042–1066;

RELVAS, S., Matos, H. A., Barbosa-Póvoa, A. P. F. D., Fialho, J., Pinheiro, A. S. (2006), Pipeline Scheduling and Inventory Management of a Multiproduct Distribution Oil System, *Ind. Eng. Chem. Res.*, 45, 7841-7855;

RELVAS, S., Matos, H. A., Barbosa-Póvoa, A. P. F. D., Fialho, J. (2007), Reactive Scheduling Framework for a Multiproduct Pipeline with Inventory Management, *Ind. Eng. Chem. Res.*, 46, 5659-5676;

RELVAS, S. (2008), Optimal Pipeline Scheduling and Inventory Management of a Multiproduct Oil Distribution Centre, PhD Thesis, Instituto Superior Técnico, Technical University of Lisbon;

RELVAS, S., Barbosa-Póvoa, A. P. F. D., Matos, H. A. (2009), Heuristic batch sequencing on a multiproduct oil distribution system, *Comput. Chem. Eng.*, 33, 712–730;

RELVAS, S., Boschetto Magatão, S.N., Barbosa-Póvoa, A.P.F.D., Neves Jr., F. (2013), Integrated Distribution Scheduling and Inventory Management of an Oil Products Distribution System, *Omega*, 41 (6), 955–968;

SASIKUMAR, M., Prakash, P.R., Patil, S.M., Ramani, S. (1007) PIPES: A heuristic search model for pipeline schedule generation, *Knowledge-Based Systems*, 10, 169-175;

SHAH, N. (1996) Mathematical programming techniques for crude oil scheduling, *Comput. & Chem. Eng.*, 20, S1227-S1232;

SHAH, N.K; Li, Z.; Ierapetritou, M.G. (2011), Petroleum Refining Operations: Key Issues, Advances, and Opportunities, *Ind. Eng. Chem. Res.*, 50 (3), 1161–1170;

TRENCH, Ch. J. (2001), How pipelines make the oil market work their networks, operation, and regulation; Allegro Energy Group, Association of Oil Pipe Lines: New York, 1-20.

(Página deixada propositadamente em branco)

CASO 14

AFETAÇÃO MULTICRITÉRIO DE RECURSOS A PORTEFÓLIOS DE ATIVIDADES DA REFER

Ricardo Mateus

Centro de Estudos de Gestão do IST (CEG-IST), Instituto Superior Técnico
Universidade de Lisboa,
ricardo.mateus@sapo.pt

José A. Antunes Ferreira

CESUR-IST, Instituto Superior Técnico
Universidade de Lisboa,
antunes.ferreira@tecnico.ulisboa.pt

João Lourenço

Centro de Estudos de Gestão do IST (CEG-IST), Instituto Superior Técnico
Universidade de Lisboa,
joao.lourenco@tecnico.ulisboa.pt

RESUMO

O Plano de Atividades de uma organização é o instrumento de gestão que define, para um determinado período de tempo, os objetivos a atingir, as atividades a desenvolver e os recursos a afetar a essas atividades. Os recursos disponíveis para a implementação do Plano tipicamente não permitem realizar todas as atividades de interesse para a organização, o que obriga à existência de um processo de seleção de atividades. Neste contexto de restrição de recursos, a elaboração de um Plano de Atividades consiste num processo de seleção do portefólio de atividades que mais benefícios pode gerar para a organização face aos recursos que consome. Neste artigo apresenta-se a metodologia multicritério de apoio à decisão de afetação de recursos preconizada para a elaboração do Plano de Atividades da Direção Geral de Engenharia da REFER. A metodologia apresentada integra técnicas e ferramentas de índole técnica (tais como, os mapas cognitivos, a análise multicritério, o método MACBETH e a afetação de recursos com o Equity) e social (como é o caso das conferências de decisão). Neste enquadramento, o papel do apoio à decisão é o de clarificar, modelar e integrar as diferentes perspetivas existentes no processo de seleção do portefólio de atividades, através do desenvolvimento de uma linguagem de comunicação comum e de uma estrutura formal de interação entre os diversos atores da REFER (decisores e técnicos especialistas), seguindo uma abordagem construtiva e de aprendizagem, conducente à recomendação das atividades a incluir no Plano.

PALAVRAS-CHAVE

Análise de decisão multicritério; Afetação multicritério de recursos; Análise de portefólios de atividades.

1. Introdução

A REFER é a empresa pública responsável pela administração da rede ferroviária portuguesa, que em 2005 se estendia ao longo de 3.613 km, dos quais 2.839 km com tráfego ferroviário. A REFER desenvolve a sua atividade em duas áreas principais: a gestão da infraestrutura, que inclui a sua conservação, manutenção e exploração; e a construção, instalação e renovação da infraestrutura.

Como entidade pública, a REFER está obrigada à elaboração de planos de atividade anuais e plurianuais, levando em consideração os objetivos estratégicos estabelecidos pela tutela para o setor em que opera. Um Plano de Atividades estabelece os objetivos a alcançar, as atividades em curso e planeadas e os recursos necessários para as desenvolver.

No final de outubro de 2005, tomou posse um novo Conselho de Administração, que teve de definir e aprovar num curto espaço de tempo o Plano de Atividades da REFER para o ano seguinte. Para o efeito, o Conselho solicitou à Direção Geral de Engenharia (DGEN), responsável pela área de construção, instalação e renovação da infraestrutura (Figura 1; REFER, 2005a), a listagem das atividades em curso e planeadas para 2005, juntamente com uma breve caracterização das mesmas. O novo Conselho transmitiu a necessidade de definir objetivos, custos e indicadores para a seleção e priorização das atividades a incluir no respetivo Plano de Atividades (Monteiro, 2006). De acordo com as diretrizes da tutela, o Conselho transmitiu também como orientação prioritária a dinamização de uma política de contenção de custos (REFER, 2005b).

De acordo com as normas internas e procedimentos vigentes (REFER, 2004), o Plano de Atividades da DGEN (daqui em diante designado por Plano) deveria ser preparado tendo por base as orientações estratégicas emanadas pelo Conselho. A partir das orientações e informações recebidas, o Diretor Geral da DGEN deveria identificar as prioridades de investimento em conjunto com as diversas Direções da DGEN e em estreita colaboração com os restantes órgãos da empresa (Figura 1; REFER, 2004). Após a sua elaboração, o Plano deveria ser submetido para aprovação ao Conselho de Administração.

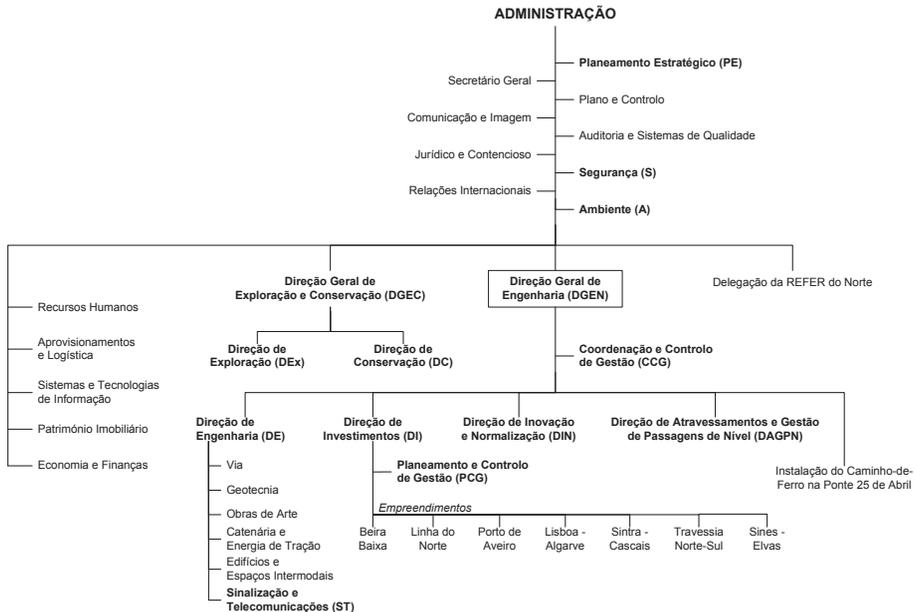


Figura 1: Estrutura organizacional da REFER

O procedimento para a elaboração de cada Plano anual consistia genericamente em adicionar novas atividades, propostas pelos diversos órgãos da empresa e previamente aprovadas a nível superior, ao conjunto de atividades em curso ou planeadas em anos anteriores. Embora o procedimento adotado pela REFER pareça intuitivamente correto, de facto padece de um conjunto de problemas identificados na literatura da análise de decisão relativa à afetação multicritério de recursos, tais como:

- Não coloca em causa atividades selecionadas anteriormente, mantendo atividades com benefícios marginais apenas porque no passado já se investiu nelas, ao invés de considerar em cada avaliação apenas os custos e benefícios que as atividades podem gerar no futuro (Hammond et al. (2006) designam esta prática como a armadilha dos custos afundados);
- Não quantifica de forma clara os benefícios das diversas atividades de forma a criar as condições para uma melhor tomada de decisão;

- Não fundamenta, perante terceiros (externos ou internos), os benefícios que as diversas atividades podem gerar o que vai contra a transparência e responsabilização necessárias em decisões que afetam múltiplos atores;
- Considera apenas os benefícios que as atividades podem gerar, sem atender aos custos dessas atividades, promovendo um conjunto de atividades que não garante a melhor afetação dos recursos;
- Fomenta a criação de «silos» onde cada Direção propõe o maior número de atividades na esperança de conseguir a aprovação de uma parte proporcional. Por outro lado, a melhor afetação de recursos disponíveis na ótica de cada uma das Direções não implica que essa afetação seja a melhor na perspetiva global da DGEN (Phillips e Bana e Costa, 2007).

Estamos assim perante um problema típico de afetação multicritério de recursos, cuja resolução deve levar em consideração os benefícios das atividades (avaliados de acordo com os objetivos estratégicos da REFER), os *trade-offs* entre os objetivos, os recursos consumidos por essas atividades e o limite dos recursos disponibilizados à REFER pelo Estado para o efeito.

Face a esta realidade, o Diretor Geral da DGEN aproveitou a necessidade de elaborar o Plano de Atividades de 2006 como uma oportunidade para melhorar qualitativamente o processo de elaboração dos Planos de Atividades. Para o efeito, convidou os autores deste artigo para desenvolver e aplicar uma metodologia de apoio à decisão adequada à resolução deste problema. A metodologia preconizada baseia-se numa abordagem sociotécnica que inclui aspetos de natureza:

- técnica – sustentada num modelo de afetação de recursos, priorização e orçamentação (Phillips e Bana e Costa, 2007) que utiliza de forma integrada diversos métodos da análise de decisão, tais como a teoria de valor multicritério (Belton e Stewart, 2002), a estruturação de problemas centrada nos valores (Keeney, 1992) e a modelação do risco e incerteza (Clemen, 1996);
- e social – sustentada num processo construtivo de conferências de decisão (Phillips, 2007) com os atores-chave designados pela REFER,

em que os autores agem exclusivamente como facilitadores, intervindo apenas no contexto e não nos conteúdos, modelando e integrando dados factuais e juízos de valor expressos pelos atores-chave.

A escolha deste tipo de abordagem de apoio à decisão fundamenta-se nas seguintes convicções (Bana e Costa, 1993b; Phillips e Bana e Costa, 2007):

- a) uma convicção de base é que a estruturação e modelação do problema centrada nos valores e preferências dos decisores, considerando o risco e a incerteza, potenciam a tomada de decisões de qualidade;
- b) a outra convicção fundamental reside no processo de apoio à decisão, que visa a construção de uma linguagem de comunicação comum e de uma estrutura formal de avaliação construtiva e de aprendizagem entre os diversos atores deste problema de decisão.

Como resultado da aplicação desta metodologia obteve-se um modelo multicritério de afetação de recursos que serviu para a DGEN utilizar como ferramenta de apoio à seleção do portefólio de atividades a incluir no seu Plano de Atividades.

A metodologia preconizada fundamentou-se em três fases fundamentais (Bana e Costa, 1992b e 1993a), decompostas nas seguintes etapas (ver Figura 2) que são desenvolvidas nas secções indicadas entre parênteses:

- *Fase de Estruturação* do problema de uma forma interativa e partilhada entre os diversos intervenientes no processo de decisão:
 - o Contextualização do problema (Secção 2);
 - o Identificação, estruturação e caracterização das atividades (Secção 3);
 - o Estimação dos custos de investimento das atividades (Secção 4);
 - o Identificação e estruturação dos objetivos estratégicos (Secção 5);
 - o Definição de descritores de impactos (Secção 6);
 - o Estimação dos impactos das atividades (Secção 7);

- *Fase de Avaliação* seguindo uma abordagem transparente, construtiva e de aprendizagem, sem a pretensão de assumir um posicionamento otimizador e normativo:
 - Avaliação dos benefícios parciais das atividades (Secção 8);
 - Ponderação dos objetivos (Secção 9);
 - Avaliação dos benefícios globais das atividades (Secção 10);
- *Fase de Recomendações* relativamente à seleção do portefólio de atividades a incluir no Plano:
 - Seleção de um portefólio de atividades eficiente que não ultrapasse a capacidade orçamental disponível para 2006 (Secção 11).

Sublinhe-se porém que em qualquer momento do processo de apoio à decisão é possível voltar a uma etapa anterior, para incluir elementos adicionais de informação ou rever juízos de valor.

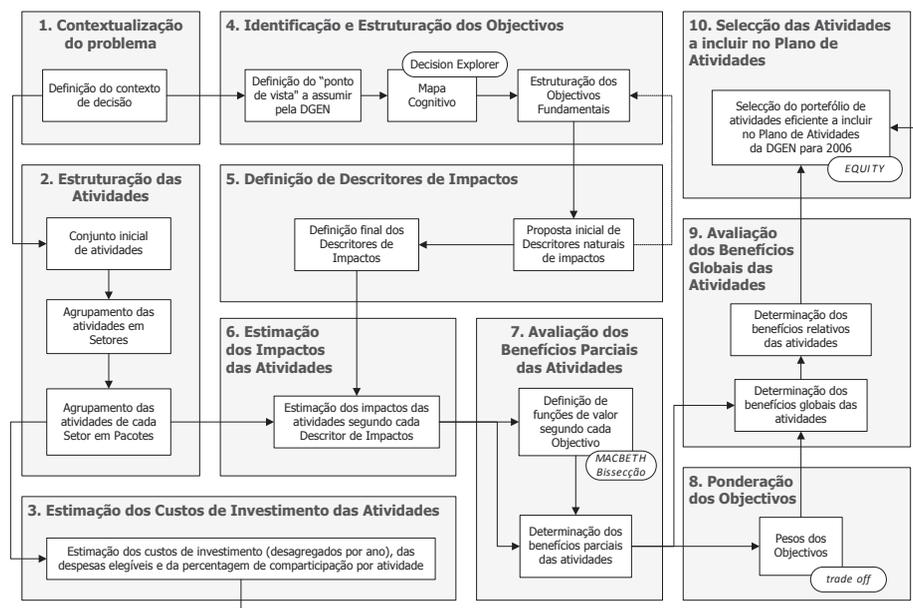


Figura 2: Síntese da metodologia

2. Contextualização do problema

A fase de estruturação iniciou-se com a caracterização do contexto de decisão. Os principais atores envolvidos no processo de elaboração do Plano incluíam o Conselho de Administração, o Diretor Geral da DGEN (DG), os principais Diretores da REFER e vários técnicos da empresa com conhecimento tácito relativamente ao negócio ou com acesso a informação essencial para o processo de avaliação. Dois técnicos qualificados do PCG foram designados para apoiar a tempo inteiro os facilitadores deste processo sob a supervisão do DG. De entre as várias condicionantes (temporais, financeiras, administrativas, tecnológicas e legais) existentes, a conclusão do Plano no prazo de um mês foi desde logo identificada como um dos fatores críticos de sucesso deste processo de apoio à decisão.

O Plano de Atividades de 2005 constituiu a principal base de trabalho para a elaboração do Plano de Atividades de 2006. A estrutura desse Plano espelhava de certa forma a estrutura organizacional da empresa e o grau de responsabilidade de cada órgão relativamente à proposição e execução das diversas atividades da empresa. Cada órgão era responsável por um conjunto de programas. Cada programa consistia num conjunto de atividades agregadas em função da sua natureza. No total, o Plano continha assim 34 programas de atividades da responsabilidade de seis Direções: 9 da DI; 4 da DAGPN; 10 da ST; 5 da DE; 3 da DIN; 2 da DGEN; e 1 da responsabilidade conjunta da DI e DE. Os Diretores destas seis Direções, sob a liderança do DG, representam assim os Decisores neste processo de decisão (doravante identificados simplesmente por Decisores).

Após a apreensão do contexto de decisão estabeleceu-se o processo social que devia ser seguido para alcançar os resultados pretendidos. A conceção deste processo é essencial em contextos organizacionais como o da REFER, caracterizado pela existência de múltiplos intervenientes que podem influenciar as decisões (Administração, tutela, Diretores, técnicos, operadores ferroviários, sociedade em geral, etc.) com agendas, valores, preferências e poderes bastante diversos e por vezes conflituantes. Observe-se que, por um lado, numa estrutura organizacional como a da REFER, um único decisor não tem a capacidade de conhecer a multiplicidade de efeitos e consequências que cada

atividade pode gerar, uma vez que a informação e conhecimento necessários para o efeito se encontram disseminados pelas diversas estruturas organizacionais, respetivos técnicos, e sistemas de informação da empresa. Por outro lado, os decisores competem entre si pelos mesmos recursos, que são limitados. Para que os decisores estejam alinhados e comprometidos com a implementação das atividades pelas quais são responsáveis é essencial que lhes seja demonstrada a racionalidade das decisões tomadas, de uma forma transparente e rastreável (suportada em evidências documentais) e desejavelmente através de um processo participativo e colaborativo de decisão em que a sua voz seja ouvida e levada em consideração.

3. Identificação, estruturação e caracterização das atividades

Considerou-se vantajoso agregar as atividades em setores. Um setor corresponde a um órgão (por exemplo, uma Direção) ou uma área estratégica (por exemplo, uma Linha) que compete por recursos e que funciona de forma independente em relação aos outros setores, ou seja, a proposição e execução das atividades de um determinado setor não dependem das iniciativas tomadas nos restantes setores. Por essa razão, as atividades foram agrupadas em 15 setores distintos da seguinte forma (a Direção responsável por cada setor é indicada entre parênteses):

- Setor S1 – Intervenção na Linha do Norte (DI);
- Setor S2 – Intervenção na Beira Baixa (DI);
- Setor S3 – Ligação ao Algarve (DI);
- Setor S4 – Melhoria das condições de exploração ferroviária (DI);
- Setor S5 – Missão para intervenções artísticas nas estações e interfaces (DGEN);
- Setor S6 – Consolidação das acessibilidades urbanas na margem Sul (DI);
- Setor S7 – Intervenção na Linha de Cintura e Linha de Sintra (DI);
- Setor S8 – Intervenção na Linha de Cascais (DI);
- Setor S9 – Ligação ferroviária Porto de Sines - Espanha (DI);
- Setor S10 – Acessibilidades Portuárias e Industriais (DI);

- Setor S11 – Atividades da Direção de Atravessamentos e Gestão de PN's (DAGPN);
- Setor S12 – Atividades da Direção de Sinalização e Telecomunicações (ST);
- Setor S13 – Atividades da Direção de Engenharia (DE);
- Setor S14 – Atividades da Direção de Inovação e Normalização (DIN);
- Setor S15 – Atividades da Direção Geral de Engenharia (DGEN).

As atividades só têm sentido para uma organização como um meio para alcançar os seus objetivos. No entanto, atividades como a elaboração de um estudo ou do projeto de execução de uma linha ferroviária apenas geram impactos após a execução da respetiva obra. Por conseguinte, foi necessário adicionar atividades que incorporassem no modelo os impactos que só ocorrem após a efetiva implementação da respetiva obra. Por exemplo, quando existia apenas a atividade relacionada com a fase de estudos de uma linha ferroviária, foram adicionadas duas novas atividades, uma relativa ao projeto de execução e outra à execução da obra propriamente dita. Note-se que, no exemplo apresentado, todas as atividades consomem recursos, i.e. têm custos de investimento específicos, mas apenas a última atividade (execução da obra) gera impactos positivos (benefícios).

Desde o início, algumas Direções manifestaram a sua preocupação relativamente à possibilidade de o modelo de apoio à decisão vir a recomendar uma solução que passasse, por um lado, pela seleção de algumas atividades que contrariassem ordens e prioridades de nível superior existentes ao nível de cada setor e, por outro lado, pela não realização de quaisquer atividades (e inclusive o seu cancelamento) em alguns dos setores.

A primeira preocupação foi resolvida considerando como regra, por setor, a que o modelo deveria obedecer, que as atividades seriam agrupadas em pacotes de atividades, de acordo com a classificação que se apresenta na Tabela 1, pela respetiva Direção, e que só seriam selecionadas atividades de um pacote de nível superior após a seleção de todas as atividades incluídas em pacote(s) de nível(is) inferior(es). Note-se porém que isto não significa que se

tenha de se selecionar todas as atividades incluídas num determinado pacote em simultâneo. Cada atividade será selecionada individualmente com base numa avaliação do seu próprio mérito.

A segunda preocupação foi atendida impondo como regra adicional que as atividades incluídas no Pacote P1, de cada setor, nunca seriam colocadas em causa, devido aos custos colaterais que uma decisão nesse sentido teria para a REFER (por exemplo, o pagamento de indemnizações por quebra de contrato ou a quebra da reputação da REFER como entidade de bem e credível na qual o mercado pode confiar). Face à imposição desta regra, deixou de ser relevante a determinação dos benefícios destas atividades, no entanto os respetivos custos a incorrer foram considerados pois tinham impacto no orçamento disponível.

Tabela 1: Pacotes e atividades

Pacote	Descrição do pacote	N.º atividades
P1	Atividades em curso e/ou indispensáveis	101
P2	Atividades a realizar por indicação do Estado	10
P3	Atividades importantes ao nível da segurança	18
P4	Atividades com compromissos externos já assumidos	21
P5	Atividades importantes com prioridade de primeira ordem	29
P6	Atividades importantes com prioridade de segunda ordem	9
P7	Atividades importantes com prioridade de terceira ordem	11
Total		199

Uma vantagem essencial deste tipo de estruturação em pacotes (Bana e Costa et al., 2002) é a de que permite modelar prioridades resultantes quer de preferências dos Decisores em relação à realização de algumas atividades em detrimento de outras quer da existência de vários tipos de restrições temporais entre atividades, em especial os quatro tipos de relações de sequencialidade entre pares de atividades (Fim-Princípio, Princípio-Princípio, Fim-Fim e Fim-Princípio). Por exemplo, no setor S1, existe uma relação de sequencialidade do tipo Fim-Princípio entre a realização das atividades de projeto (P01-0-180) e de execução da obra (P01-0-181), ambas da estação do Entroncamento, pelo que a primeira atividade (projeto de execução) foi classificada no pacote P6 e a

segunda atividade (execução da obra) foi classificada no pacote P7, ou seja, a segunda atividade só poderia iniciar-se após a conclusão da primeira. Outro exemplo, ainda relativo ao setor 1, consiste numa relação de sequencialidade do tipo Fim-Fim existente entre a atividade de implementação da sinalização, telecomunicações e edifício técnico de Santa Cita e a atividade de implementação do convél, ambas do ramal de Tomar. Neste caso, ambas as atividades foram agrupadas num mesmo pacote para que estas pudessem terminar no mesmo período temporal.

Adicionalmente, as atividades foram caracterizadas pelas Direções a um nível de pormenor que incluía:

- A sua descrição inequívoca em termos do objeto e âmbito que compreendiam (O quê?);
- A delimitação territorial abrangida pela intervenção (Onde?);
- Uma estimativa plausível dos custos totais de implementação (desagregados em função das suas fontes de financiamento, nomeadamente: recursos públicos centrais, comunitários ou recursos próprios) e de exploração e manutenção (Quanto?);
- Uma estimativa plausível da sua duração, incluindo relações de sequencialidade (Quando?);
- A definição explícita do órgão responsável pela sua implementação (Quem?).

4. Estimação dos custos de investimento das atividades

A decisão de realização de qualquer atividade implica um custo de investimento que deve deseavelmente ser conhecido *a priori*.

Para esse efeito, foram estimados os custos totais de investimento de todas as atividades do Plano, desagregados temporalmente (por ano) e em função da sua fonte de financiamento (comunitária ou nacional), tal como se apresenta, a título de exemplo, na Tabela 2, para algumas atividades do Plano.

O custo total de uma atividade corresponde ao investimento total associado à realização dessa atividade. No entanto, *a priori* só é possível obter uma estimativa desse custo total. Neste processo admitiram-se dois tipos de estimativas para esse custo: pontual (o custo total esperado); ou através de um

intervalo que compreendesse todos os custos plausíveis (sendo neste caso posteriormente utilizado o ponto médio do intervalo como estimativa do custo total). Esta última possibilidade revelou-se taticamente bastante útil, nomeadamente para desbloquear situações em que não foi possível obter de uma ou várias fontes de informação um custo total esperado.

Tabela 1: Custos totais de investimento por atividade (em milhares de euros)

Código da Atividade	Pacote	Custos a incorrer				Custo total	Despesas elegíveis	% de compart.
		2006	2007	2008	anos seg. indeterm.			
P01-0-010	P1	205				205		
P01-0-015	P6		4.041	5.600	5.200	14.841		
P01-0-020	P4	500	500			1.000		
P01-0-025	P1	641				641		
P01-0-030	P1	374				374		
P01-0-035	P1	700				700		
P01-0-036	P3	709	3.180			3.889		
P01-0-040	P1	37.038	1.500	400		108.000	99.900	80%

Como já foi referido na Secção 1, não é relevante levar em consideração, neste processo de decisão, os custos afundados, pois já foram incorridos e são irre recuperáveis se a empresa interromper a atividade, mas somente os custos a incorrer no futuro. Os custos a incorrer correspondem à parcela do custo total de investimento a realizar no futuro. Naturalmente, a restante parcela corresponde ao investimento já realizado, que será necessariamente positivo para aquelas atividades que já se encontram em curso (pacote P1) e que para as quais já houve lugar ao pagamento de despesas (ver na Tabela 2, por exemplo, a atividade com o código P01-0-040). Os custos a incorrer com cada atividade foram também desagregados temporalmente por ano (2006, 2007, 2008 e anos seguintes) com base na calendarização prevista até à sua completa implementação.

Uma vez que o financiamento de algumas das atividades poderia vir a incluir uma comparticipação comunitária, analisou-se também a hipótese de considerar apenas como custo de investimento dessas atividades os custos a incorrer (caso esse cofinanciamento efetivamente ocorresse). Os custos a

incorrer (com financiamento comunitário) foram calculados de acordo com a seguinte expressão:

$$\text{Custos a incorrer (com financiamento)} = \text{custos a incorrer} \times \left(\frac{\text{custo total} - \text{despesas elegíveis} \times \text{perc. de participação}}{\text{custo total}} \right)$$

A Tabela 3 apresenta os custos totais de investimento a incorrer (em euros) com a realização do total das atividades inscritas no Plano de acordo com os pacotes (apresentados na Tabela 1) e os setores (apresentados no início da Secção 3) a que se encontram afetos.

Tabela 3: Custos totais de investimento a incorrer por setor e pacote (em milhares de euros)

Setor	P1	P2	P3	P4	P5	P6	P7	Total
S1	218.212		30.904	18.000	5.900	84.841	276.040	633.897
S2	5.936	118.248		46.542		14.095		184.821
S3	14.727	143.100	4.742	46.617	555	81		209.821
S4	25.108	5.058			29.738			59.903
S5	13			1.114	126	40		1.292
S6	11.686			8.375			16.433	36.493
S7	7.563		1.500	3.562	51.480	30.500	63.455	158.060
S8	13.341				139.914			153.255
S9	20.308	133.894					594.248	748.449
S10		68.504		14.833				83.338
S11			76.438					76.438
S12	6.808				3.250			10.058
S13	1.060		0,5		1.284			2.345
S14	2.048		488		1.260			3.796
S15	325				95			420
Total	327.133	468.804	114.072	139.043	233.602	129.557	950.175	2.362.385

É possível destacar o seguinte em relação a estes montantes: o custo a incorrer com a realização de todas as atividades inscritas no Plano soma cerca de 2,3 mil milhões de euros (embora esse montante baixe para 1,7 mil milhões de euros com o eventual cofinanciamento comunitário); apesar de cerca de metade das atividades ter sido classificada no pacote P1 relativo às atividades em curso e/ou indispensáveis (ver Tabela 1), este pacote corresponde apenas a

cerca de 13% do custo total; os custos das atividades inseridas no pacote P7 relativo às atividades importantes com prioridade de terceira ordem representam cerca de 40% do custo total; os setores S1 (Intervenção na Linha do Norte) e S9 (Ligação Porto de Sines – Espanha) representam por si só cerca de 60% do total dos custos de investimento a incorrer. Estes valores são bem demonstrativos do tipo de atividades em análise, com longos prazos até à sua implementação (nomeadamente a construção) e prazos ainda mais longos de exploração, o que reforçam a necessidade de se planear cuidadosamente a sua realização numa perspetiva de médio/longo prazo através de planos plurianuais.

5. Identificação e estruturação dos objetivos estratégicos

Esta etapa metodológica foi iniciada pelo DG, logo após a identificação dos atores-chave deste processo de decisão, convocando-os para uma conferência de decisão. A conferência de decisão iniciou-se com uma descrição das principais linhas de orientação do trabalho a realizar, das condicionantes existentes e da metodologia de apoio à decisão, quer do ponto de vista técnico como social, que iria apoiar esse processo com vista à produção dos resultados desejados.

Em contextos organizacionais com múltiplos atores que têm diferentes interesses, valores, preferências e agendas em jogo é natural que existam diversas perspetivas individuais acerca do problema em mãos, que deverão estar representadas no processo de análise de decisão (Russo e Schoemaker, 1989; Keeney, 2007) e que combinadas permitem obter uma perspetiva de grupo (Phillips, 1990). Por essa razão, os Decisores foram questionados relativamente ao objetivo global que a DGEN pretendia atingir com a realização das atividades propostas no Plano. A resposta consubstanciou-se através da definição explícita do objetivo global «investir eficientemente nos próximos anos na REFER, considerando o orçamento disponível da DGEN para 2006». A definição, de uma forma clara e partilhada, de um único objetivo global e perspetiva de avaliação, permitiu alinhar os diversos Decisores em direção a um objetivo comum durante todo o processo subsequente de elaboração do Plano.

As atividades propostas no Plano visam naturalmente satisfazer diversos objetivos fundamentais da REFER. A sua identificação e estruturação deve especificar o objetivo global anteriormente estabelecido, com base numa abordagem centrada nos valores (Keeney, 1992). Recorrendo à construção de mapas cognitivos (Eden, 1988; Eden e Simpson, 1989), começou-se por indagar os Decisores acerca dos aspetos e preocupações que consideravam mais relevantes relativamente a este contexto de decisão. Para o efeito, através de um procedimento dinâmico de perguntas-respostas, os autores foram colocando questões aos Decisores do seguinte tipo: O que consideram importante alcançar/atingir a médio e a longo prazo? Por que é que a realização de determinada atividade é importante/relevante? Como é que isso pode ser alcançado?. As respostas foram sendo introduzidas, em tempo real, no *software Decision Explorer* (Banxia, 2002), e projetadas num ecrã, resultando desse procedimento um mapa cognitivo (Figura 3) com conceitos interligados entre si por relações causa-efeito. Analisando o mapa cognitivo da Figura 3 podem encontrar-se várias cadeias de conceitos. Por exemplo, efetuar a «30 Supervisão técnica de infraestruturas» permite uma maior «14 Rapidez de intervenção», que possibilita aumentar a «15 Disponibilidade» da infraestruturas, que contribui para garantir a «16 Continuidade do serviço», que por sua vez é relevante para a satisfação do «6 Operador / Cliente final». Desta cadeia resultaram o objetivo fundamental «2.3. Garantir a continuidade do serviço» e a área de preocupação «2. Satisfação do operador/cliente final» referidos na estrutura arborescente de objetivos que se apresenta mais adiante.

Pela sua relevância destacam-se dois tipos de conceitos:

- Objetivos-fins: que resultam de respostas à pergunta «Por que é que esse aspeto é importante?»;
- Objetivos-meios: que resultam de respostas à pergunta «Como é que esse aspeto pode ser alcançado?».

Os objetivos-fins representam os objetivos fundamentais a atingir pela REFER (por exemplo, «3 Diminuir tempo de trajeto») e os objetivos-meios (como por exemplo, melhorar as «24 Telecomunicações») representam objetivos intermédios, meios ou atividades específicas necessários para atingir os objetivos-fins.

Em simultâneo, de uma forma progressiva e resultante da interação entre os diversos Decisores, foi-se construindo e projetando num outro ecrã uma estrutura arborescente dos objetivos da REFER que, após algumas revisões, estabilizou na seguinte forma:

1. Fatores intrínsecos à REFER
 - 1.1. Reduzir os custos de manutenção
 - 1.2. Reduzir os custos de operação
 - 1.3. Melhorar a eficiência da gestão
 - 1.4. Evitar acidentes
 - 1.4.1. Evitar acidentes pessoais
 - 1.4.2. Evitar acidentes materiais
 - 1.5. Garantir a segurança de pessoas e bens
2. Satisfação do operador/cliente final
 - 2.1. Aumentar a utilização da infraestrutura
 - 2.2. Reduzir o tempo de trajeto
 - 2.3. Garantir a continuidade do serviço
 - 2.4. Aumentar a atratividade do sistema
 - 2.4.1. Passageiros
 - 2.4.2. Mercadorias
3. Cumprimento das obrigações legais (nacionais e comunitárias)
 - 3.1. Acessibilidade
 - 3.2. Ruído
 - 3.3. Intervenções artísticas
 - 3.4. Passagens de nível

Os objetivos fundamentais estão representados no nível mais baixo da estrutura arborescente de objetivos. Os objetivos de nível superior representam áreas de preocupação que agrupam conjuntos de objetivos fundamentais. Os impactos e pontuações das atividades são determinados ao nível dos objetivos fundamentais (doravante designados simplesmente por objetivos).

Esta etapa de construção de uma forma partilhada dos objetivos estratégicos foi essencial, pois permitiu consolidar uma estrutura comum de avaliação e fundamentação dos benefícios das diversas atividades propostas, com base

na qual se estabeleceu uma linguagem formal de comunicação entre os vários atores do processo.

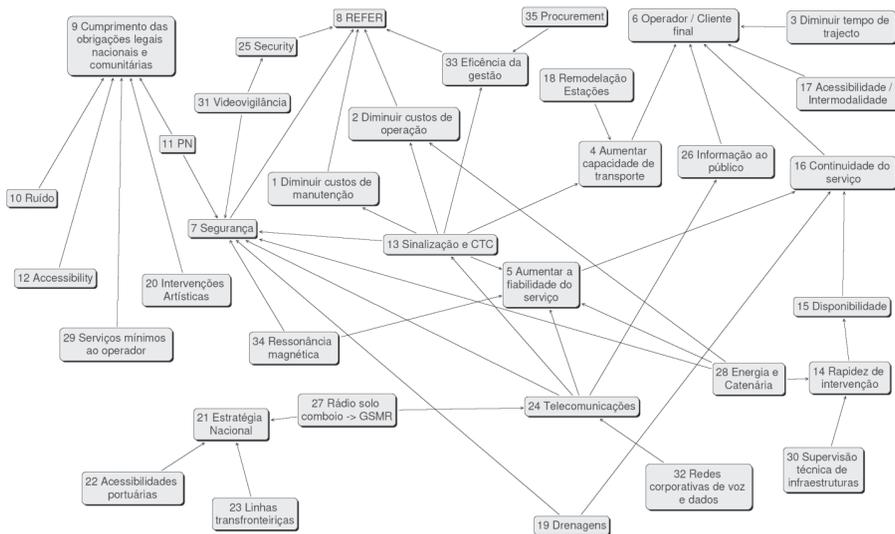


Figura 3: Mapa cognitivo

Existem vários modelos de agregação das preferências expressas em múltiplos objetivos. Neste caso, optou-se por utilizar o modelo aditivo simples de agregação com base na teoria de valor multicritério, onde a pontuação global de cada atividade é determinada pela soma dos produtos das pontuações parciais dessa atividade, segundo os diversos objetivos, pelos respetivos pesos, cuja expressão se mostra de seguida:

$$B(a) = \sum_{o=1}^n k_o \times b_o(i_o(a)), \text{ com } \sum_{o=1}^n k_o = 1 \text{ e } 0 < k_o < 1 \quad (1)$$

em que,

- $B(a)$ corresponde à pontuação global da atividade a , que representa o valor relativo do seu benefício global, tendo em consideração todos os objetivos fundamentais;
- n corresponde ao número de objetivos fundamentais;

- $i_o(a)$ corresponde ao impacto da atividade a segundo o objetivo o , obtido de acordo com um descritor de impactos $i_o(\cdot)$;
- $b_o(i_o(a))$ corresponde à pontuação parcial do impacto da atividade a , segundo o objetivo o , que representa o valor do seu benefício parcial segundo esse objetivo;
- k_o corresponde ao peso do objetivo o .

Identificar objetivos e organizá-los numa estrutura arborescente não garante porém a coerência intrínseca do modelo de avaliação aditivo definido em (1). Para o efeito, é necessário garantir o cumprimento de algumas propriedades fundamentais e procurar desejavelmente o cumprimento de outras (Dogson et al., 2000; Keeney, 1992 e 2007). Por exemplo, tal como referido anteriormente, é essencial identificar os objetivos-fins de forma a não introduzir no modelo redundâncias que iriam sobrevalorizar impactos que só deveriam ser considerados uma única vez. É também essencial garantir que é possível avaliar cada atividade segundo cada um dos objetivos fundamentais sem levar em consideração o seu desempenho nos restantes (propriedade que é conhecida como «independência preferencial» (Dyer e Sarin, 1979)). Outras propriedades importantes incluem a verificação da exaustividade, concisão (parcimónia), consensualidade, operacionalidade e inteligibilidade (não-ambiguidade). As três primeiras podem ser alcançadas através da sua estruturação com base em abordagens partilhadas e construtivas tais como as realizadas neste caso através dos mapas cognitivos. Já as duas últimas podem ser sobretudo garantidas através da definição de adequados descritores de impactos, tal como se descreve na próxima secção.

Outra vantagem adicional deste tipo de estrutura advém da possibilidade de os diversos órgãos da DGEN poderem identificar novas oportunidades de decisão com base num questionamento sobre quais as melhores formas de melhorar os objetivos fundamentais identificados. Desta forma é possível identificar novas atividades e justificar desde logo os seus eventuais benefícios. Se uma nova atividade não contribui para atingir ou melhorar pelo menos um objetivo, é porque ou não têm razão de ser ou então é necessário rever a estrutura de objetivos fundamentais anteriormente definida.

6. Definição de descritores de impactos

A determinação dos benefícios gerados pela realização das atividades exige que seja definida uma forma (quantitativa ou qualitativa) de medir ou estimar os seus impactos segundo cada um dos objetivos fundamentais identificados na fase anterior. Um descritor de impactos é um conjunto de níveis de impacto concretos ordenado por ordem de preferência segundo um determinado objetivo fundamental. O significado de um objetivo fundamental é descrito de uma forma clara e objetiva pelo seu descritor de impactos. Cada objetivo fundamental é operacionalizado por um e um só descritor de impactos. Um descritor de impactos serve assim para medir a contribuição de cada atividade para a consecução de um determinado objetivo fundamental.

As atividades propostas no Plano vão agir sobre uma realidade pré-existente. Há vantagens significativas em medir, estimar ou avaliar os impactos destas atividades em relação a uma situação base ou de referência (*baseline*) (Clemen e Smith, 2009), quer ela corresponda à situação atual (*status quo*) (Bana e Costa et al., 2014), à situação de manutenção da tendência atual (*business as usual*) (Neves et al., 2012) ou outra. São duas as vantagens inerentes à utilização dessa linha de referência: a primeira é a de permitir comparar o mérito de cada atividade em relação à situação de referência sem ter de determinar os seus impactos, possibilitando assim uma avaliação relativa das atividades em relação a essa situação de referência; a segunda é a de reduzir o esforço inerente à determinação dos impactos das atividades, uma vez que é normalmente mais fácil medir apenas o impacto (acrescido) da realização de uma determinada atividade (por exemplo, +1.000 passageiros/ano em relação ao *status quo*) do que medir o desempenho que resulta no sistema em análise após a realização dessa atividade (por exemplo, 5,1 milhões de passageiros/ano, resultantes dos 5 milhões de passageiros/ano do *status quo* mais os 1.000 passageiros/ano da atividade).

A definição dos descritores de impactos foi desenvolvida em colaboração com os técnicos da REFER e alguns dos Decisores da DGEN, sob a supervisão do DG. Existiu sempre a preocupação de identificar descritores naturais. Quando não foi possível usar um descritor deste tipo, por exemplo, devido a dificuldades em obter a informação necessária para quantificar os impactos de

todas as atividades, definiu-se em alternativa um descritor indireto quantitativo. Só quando não foi possível definir um descritor quantitativo (direto ou indireto), é que se optou por definir um descritor qualitativo construído.

A Tabela 4 apresenta uma descrição abreviada dos descritores de impactos definidos para os objetivos fundamentais identificados na fase anterior.

A responsabilidade pela determinação dos impactos das atividades, de acordo com o descritor definido para cada objetivo, foi distribuída ao nível dos diversos órgãos e técnicos especialistas da REFER em função do seu grau de atribuições, competências e responsabilidades dentro da estrutura organizacional da empresa (Tabela 4). Note-se que a informação necessária para a determinação dos impactos segundo a maioria dos objetivos existia em Direções fora da esfera de influência da DGEN, tais como a DEx, a DC, a S, o A e o PE (ver Figura 1). A responsabilidade pela determinação dos respetivos impactos foi assim atribuída a esses órgãos (este grupo juntamente com os Decisores serão doravante designados simplesmente por Representantes da REFER).

Esta fase de estruturação dos descritores seguiu igualmente um processo iterativo e recursivo de aprendizagem, de que resultaram algumas alterações, inclusive em relação à estrutura inicial de objetivos estratégicos fundamentais.

7. Estimação dos impactos das atividades

Uma importante convicção metodológica desta metodologia de apoio à decisão reside na «interrelação e inseparabilidade entre os elementos objetivos e subjetivos do contexto de decisão» (Bana e Costa, 1993b). A estimativa dos impactos é o resultado de um sistema de relações entre elementos de natureza objetiva e elementos de natureza subjetiva. Estes dois elementos são indissociáveis e nenhum deles deve ser negligenciado. A procura da objetividade deve ser uma preocupação presente, mas convém não esquecer que a subjetividade está sempre omnipresente nas decisões (é fundamental não esquecer que a tomada de decisão é antes do mais uma atividade humana) e, portanto, é essencial incorporar todo o conhecimento, não apenas os elementos de natureza objetiva mas também, de uma forma estruturada e

Tabela 4: Objetivos fundamentais e descritores de impacto

Objetivos	Descritor de Impactos
<i>1</i>	<i>Fatores intrínsecos à REFER</i>
1.1	Reduzir os custos de manutenção Variação dos custos de manutenção em relação ao <i>status quo</i> [€/ano] ⁽ⁿ⁾ (DC+DAGPN)
1.2	Reduzir os custos de operação Variação dos custos de operação em relação ao <i>status quo</i> [€/ano] ⁽ⁿ⁾ (DEX+DAGPN)
1.3	Melhorar a eficiência da gestão Conjunto ordenado de níveis qualitativos descrevendo impactos plausíveis distintos em termos do objetivo ^(c) (DG)
1.4.1	Evitar acidentes pessoais Variação do n.º de acidentes pessoais em relação ao <i>status quo</i> [n.º de mortos equivalentes ¹ /ano] ⁽ⁿ⁾ (S+DAGPN)
1.4.2	Evitar acidentes materiais Variação do n.º de acidentes sobre materiais sob a responsabilidade da REFER em relação ao <i>status quo</i> [n.º de acidentes materiais/ano] ^(p) 2 (DC+DAGPN)
1.5	Garantir a segurança de pessoas e bens Variação do n.º de atos de vandalismo sobre pessoas ³ ou bens ⁴ sob a responsabilidade da REFER em relação ao <i>status quo</i> [%/ano] ^(p) 5 (S)
<i>2</i>	<i>Satisfação do operador ferroviário/cliente final</i>
2.1	Aumentar a utilização da infraestrutura Aumento da capacidade de transporte em novas linhas ou linhas congestionadas em relação ao <i>status quo</i> [comboios.km/ano] ^(p) 6 (DEX)
2.2	Reduzir o tempo de trajeto Variação do total das poupanças de tempo de trajeto ⁷ em relação ao <i>status quo</i> [comboios.minutos/ano] ⁽ⁿ⁾ (DEX)
2.3	Garantir a continuidade do serviço Variação do total dos atrasos ⁸ por causas imputáveis à REFER em relação ao <i>status quo</i> [comboios.minutos/ano] ⁽ⁿ⁾ (DEX+DC)
2.4.1	Aumentar a atratividade do sistema (passageiros) Variação da utilização do sistema ferroviário por passageiros em relação ao <i>status quo</i> [passageiros/ano] ⁽ⁿ⁾ (PE)
2.4.2	Aumentar a atratividade do sistema (mercadorias) Variação da utilização do sistema ferroviário por mercadorias em relação ao <i>status quo</i> [toneladas/ano] ⁽ⁿ⁾ (PE)
<i>3</i>	<i>Cumprimento das obrigações legais (nacionais e comunitárias)</i>
3.1	Acessibilidade Variação do n.º de estações com acessibilidade a deficientes em relação ao <i>status quo</i> [n.º de estações] ⁽ⁿ⁾ (DI)
3.2	Ruído Variação do n.º de pessoas beneficiadas por medidas de proteção de ruído nos grandes eixos ferroviários em relação ao <i>status quo</i> [n.º de pessoas] ^(p) (A)
3.3	Intervenções Artísticas Variação do n.º de estações em incumprimento da legislação referente a intervenções artísticas em relação ao <i>status quo</i> [n.º de estações] ⁽ⁿ⁾ (DG)
3.4	Passagens de Nível Variação do n.º de passagens de nível em incumprimento da respetiva legislação em relação ao <i>status quo</i> [n.º de passagens de nível] ⁽ⁿ⁾ (DAGPN)

Legenda: ⁽ⁿ⁾descritor natural; ^(p)descritor indireto; ^(c)descritor qualitativo construído

¹ Mortos equivalentes = mortos + feridos graves / 10 + feridos ligeiros / 200

² *Descritor natural quantitativo*: variação dos prejuízos (em €/ano) causados por acidentes materiais, da responsabilidade da REFER (sobre a infraestrutura, material circulante, etc.), em relação ao *status quo*

³ Exemplos: assaltos e atos de violência física

⁴ Exemplos: furtos ou destruição de equipamento ferroviário, danos na infraestrutura, etc.

⁵ *Descritores naturais quantitativos*: variação do n.º de atos de vandalismo sobre as pessoas (em n.º de atos de vandalismo/ano), nas estações sob a responsabilidade da REFER, em relação ao *status quo*; variação dos prejuízos (em €/ano) causados por atos de vandalismo sobre bens, da responsabilidade da REFER, em relação ao *status quo*

⁶ *Descritor natural quantitativo*: variação da utilização da infraestrutura (em comboios.km/ano) em relação ao *status quo*

⁷ Nos trajetos onde não estão previstos investimentos em linhas de Alta Velocidade

⁸ Diretos, indiretos e considerados (em que o atraso máximo equivalente a uma supressão de comboio é igual a 1 hora)

rigorosa, todos os elementos subjetivos (experiência, conhecimentos, convicções, juízos de valor) de natureza qualitativa na posse dos técnicos da REFER.

Considerando que a estimativa da ocorrência de qualquer acontecimento futuro é por natureza incerta, a estimativa dos impactos das atividades não pode deixar de possuir uma componente de risco e incerteza. Esta componente foi incorporada da seguinte forma:

- por um lado, os impactos das atividades foram estimados através de um único valor (melhor estimativa do impacto) e/ou através de um intervalo plausível (sendo neste caso, à semelhança do que foi feito para os custos, posteriormente utilizado o ponto médio desse intervalo como estimativa desse impacto);
- por outro lado, sempre que o valor do impacto de uma atividade compreendia uma série de estimativas de impacto juntamente com as suas probabilidades de ocorrência, o valor esperado do impacto foi estimado através da soma ponderada desses fatores (Clemen, 1996).

A título de exemplo, refira-se como os impactos das atividades relativas à supressão de passagens de nível no objetivo «1.4.2 Evitar acidentes materiais» foram determinados: analisou-se o histórico dos últimos cinco anos do número de acidentes materiais, da responsabilidade da REFER, que ocorreram nas respetivas passagens de nível; ao limite inferior plausível foi atribuído o menor número de acidentes materiais anual, ao limite superior plausível correspondeu o maior número de acidentes materiais anual e o valor esperado desse impacto foi determinado com base na média dos acidentes materiais desses últimos cinco anos (ou seja, considerando que a probabilidade de ocorrência de cada acidente material anual se mantém igual ao longo dos anos). O resultado final desta etapa metodológica foi consolidado numa matriz com os impactos de cada atividade (em linha) segundo cada um dos objetivos (em coluna), tal como se ilustra na Tabela 5 para algumas atividades representativas do Plano.

Cada atividade pode ter impactos em vários objetivos, mas nem todos esses impactos são obrigatoriamente positivos. Por exemplo, algumas atividades (P01-0-190 e P03-3-005) podem ter impactos positivos em objetivos tais como

«1.4.1 Evitar acidentes pessoais» ou «2.3 Garantir a continuidade do serviço», que consequentemente impliquem impactos negativos em relação a outros objetivos como é o caso do objetivo «1.1 Reduzir os custos de manutenção». Repare-se também que não foram estimados quaisquer impactos para a atividade P01-0-080, relativa à elaboração do projeto de execução da Estação do Entroncamento, uma vez que esses impactos só se materializarão após a realização da respetiva Estação (atividade P01-0-081).

Tabela 5: Impactos estimados das atividades

Código	1.1	1.2	1.3	1.4.1	1.4.2	1.5	2.1	2.2	2.3	2.4.1	2.4.2	3.1	3.2	3.3	3.4
P01-0-015	-0,5				-1,8										
P01-0-036				-0,17	-1										
P01-0-075	-0,6	-0,1		-0,01	-2,4				-47	80	180	2	0,1		
P01-0-080															
P01-0-081	-0,5	-0,9		-0,06	-15,2				-7				1		
P01-0-190	0,1	-1,4		-2,76	-18,8				-58	551	180	14	4,1		
P02-0-080	-0,2	-0,1		-0,4	-1,8			68	-0,7	8	68	5			
P03-3-005	1,4				-0,20			7	-0,3	100					
P05-0-040															-1
P08-0-025				-0,69	-3,2		482	16		205		1	2,3		
P13-0-010	-0,2	-0,2		-6	-16,2										-80
P13-21-010							-25			175					
P16-34-020															

Unidades: 1.1 e 1.2 em milhões euros/ano; 2.1 em milhares de comboios.km/ano; 2.2 e 2.3 em milhares de comboios.minutos/ano; 2.4.1 em milhares de passageiros/ano; 2.4.2 em milhares de toneladas/ano; 3.2 em milhares de pessoas

Os impactos das atividades foram validados numa reunião conjunta com os Representantes da REFER. Uma técnica que auxiliou bastante este trabalho foi a de ordenar previamente todas as atividades por ordem decrescente dos seus impactos segundo cada um dos objetivos, tal como se apresenta, na Tabela 6, para o objetivo «2.3 Garantir a continuidade do serviço».

Os impactos das atividades segundo o objetivo «1.3 Melhorar a eficiência da gestão» foram determinados de forma distinta. Numa primeira fase, analisou-se o impacto previsto de cada atividade no objetivo em causa. Posteriormente, os

impactos foram classificados segundo um conjunto de níveis distintos ordenados por ordem decrescente de preferência (ver Tabela 7).

Tabela 6: Impactos por ordem decrescente segundo o objetivo 2.3

Código	Atividade	Δatrasos
P01-0-140	Modernização Alfarelos/Pampilhosa	-82.631
P01-0-190	Modernização Ovar/Gaia	-57.936
P01-0-075	Modernização Mato Miranda/Entroncamento	-46.541
P01-0-065	Modernização Vale de Santarém/Mato Miranda - Variante de Santarém	-29.619
P04-0-045	Troço Alfarelos - Louriçal - Figueira da Foz	-24.607
P02-0-020	Subst. Armamento de via, Piquetagem e Remodelação de 5 estações e Apeadeiros	-10.680
P01-0-081	Estação do Entroncamento	-7.306
P02-0-055	C. Civil, Via e Catenária, RCT+TP, Sinalização, Telecomunicações, Telecomando Catenária, RSC, Convel, SST Fatela	-6.384
P09-0-020	Modernização da Linha de Cascais	-3.052
P02-0-080	C. Civil, Via e Catenária, RCT+TP, Sinalização, Telecomunicações, Telecomando Catenária, RSC, Convel	-655
P03-3-005	Variante de Alcácer	-280

comboios.min/ano

A realização em conjunto de algumas atividades pode gerar impactos superiores (ou inferiores) à soma dos respetivos impactos individuais, pelo que é fundamental identificar e modelar as sinergias (ou antagonismos) existentes entre atividades. Por exemplo, durante a determinação dos impactos de algumas atividades, segundo o objetivo «2.1 Aumentar a utilização da infraestrutura», verificou-se que a ordem com que cada atividade era realizada tinha como efeito impactos diferentes.

Os impactos destas atividades foram assim ajustados com base na sequencialidade da sua realização. A Tabela 8 exemplifica um desses casos. A atividade P08-0-025 quando realizada em conjunto com a atividade P08-0-022 tinha um impacto estimado de 429.687 passageiros/ano. Como a segunda atividade está incluída no pacote P6 realizar-se-á sempre depois da primeira (que está incluída no pacote P5). Consequentemente, o impacto da segunda atividade foi ajustado de forma a incluir apenas o impacto acrescentado por esta atividade. Um procedimento semelhante foi realizado para a atividade P07-2-010.

Tabela 7: Níveis de impactos segundo o objetivo 1.3

Código	Atividade	Impacto	Nível
P16-34-010	Implementação de um sistema de Gestão de Qualidade	Diminuição dos custos da não-qualidade (tempos, custos, performance, riscos)	N1
P14-28-004	Implementação de um Sistema de Qualificação de Empreiteiros	Redução dos prazos de concursos e consultas para empreitadas, garantia acrescida de boa execução e avaliação contínua de desempenho	N2
P14-28-005	Implementação de um Sistema de Qualificação de Prestadores de Serviço de Fiscalização	Redução dos prazos de concursos e consultas para fiscalizações, garantia acrescida de boa execução e avaliação contínua de desempenho	N2
P16-34-020	Integração do Planeamento da DGEN em SAP/R3	Melhor gestão dos recursos financeiros	N2
P14-28-006	Assessoria para melhoria do desempenho recorrendo ao conceito RCM 2	Interligação atividades entre Engenharia e Conservação/Manutenção	N3
P14-30-006	Gestão documental - Indexação e digitalização de Desenhos Técnicos da DGEN	Facilidade no acesso à informação técnica da infraestrutura ferroviária	N3
P14-28-001	Novos equipamentos / Sistemas (GRP 3000, GPS e de <i>software</i> para projeto)	Melhorar o rigor dos projetos	N4
P14-28-003	<i>Software</i> do âmbito ferroviário (projeto de terraplenagem e drenagem)	Melhorar o rigor dos projetos	N4
P14-28-002	Projeto piloto de RFID (identificação por rádio frequência) nos AMVs	Melhorar a gestão de peças	N4
P14-30-007	Gestão documental - Custódia de documentação e tratamento documental arquivístico	Transferir para terceiros a gestão e custódia de documentação de reduzida consulta	N5
P13-23-010	Rede de Voz e Dados	Ligações de dados (aplicações informáticas de gestão / SAP, email, intranet) e ligações de voz (rede telefónica automática)	N6

Tabela 8: Impactos das atividades segundo o objetivo 2.1

Código	Atividade	Pacote	Passageiros/ano
P07-2-010	Quadruplicação Linha Cintura entre Chelas-B. Prata (estações de passageiros)	P7	22.116.650
P08-0-022	Remodelação estações passageiros Tercena Barcarena, Quadruplicação	P6	225.156
P08-0-025	Remodelação da Estação Cacém; Quadruplicação	P5	204.531

8. Avaliação dos benefícios parciais das atividades

Os benefícios parciais de cada atividade, segundo cada objetivo, foram avaliados com base em funções de valor. Uma função de valor transforma cada nível de impacto plausível de um descritor num único valor numérico que representa o valor (ou atratividade) desse impacto segundo o objetivo em questão. As funções de valor permitem transformar impactos estimados em escalas de medição muito diferentes (por exemplo, €/ano, passageiros/ano, toneladas/ano, comboios.minuto/ano) em escalas cardinais (numéricas) que representam esses impactos em unidades de valor (pontuações) nos objetivos em questão. Existem vários métodos que permitem construir funções de valor (Goodwin e Wright, 2004; von Winterfeldt e Edwards, 1986; Bana e Costa e Vansnick, 1994). Qualquer que seja o método escolhido, um decisor deve começar por estabelecer *a priori*, segundo cada descritor, as pontuações associadas a dois níveis de impacto de referência distintos, atribuindo ao impacto mais atrativo uma pontuação superior à do impacto menos atrativo. Caso os níveis de impacto escolhidos possuam um valor intrínseco (correspondendo, por exemplo, a um nível considerado «bom» – i.e. muito positivo – ou «neutro» – i.e. nem positivo nem negativo), os impactos poderão ser avaliados de forma absoluta, permitindo tirar conclusões sobre se uma atividade é boa, muito boa, ou má (Bana e Costa, 1992a; Bana e Costa et al., 2000). Caso esses níveis de impacto de referência sejam de valor relativo, então apenas se pode concluir que uma determinada atividade é melhor, pior ou indiferente a outra, mas nada se pode concluir relativamente ao seu mérito absoluto. No caso aqui apresentado, os impactos de referência adotados em cada descritor foram o «melhor» impacto e o «*status quo*», respetivamente com 100 e 0 unidades de valor.

Para os objetivos com descritores quantitativos os Decisores optaram por funções de valor lineares. O pressuposto da linearidade dessas funções foi validado, para cada objetivo, com base em questões colocadas diretamente aos atores-chave da REFER relativamente às diferenças de atratividade (preferência) entre incrementos de igual dimensão na unidade do descritor (Goodwin e Wright, 2004). Por exemplo, relativamente ao objetivo «2.4.2 Aumentar a atratividade do sistema (mercadorias)», os Decisores consideraram que o valor

do incremento de 0 toneladas/ano para 1 milhão de toneladas/ano era equivalente ao valor do incremento de 1 milhão de toneladas/ano para 2 milhões de toneladas/ano (ver Figura 4).

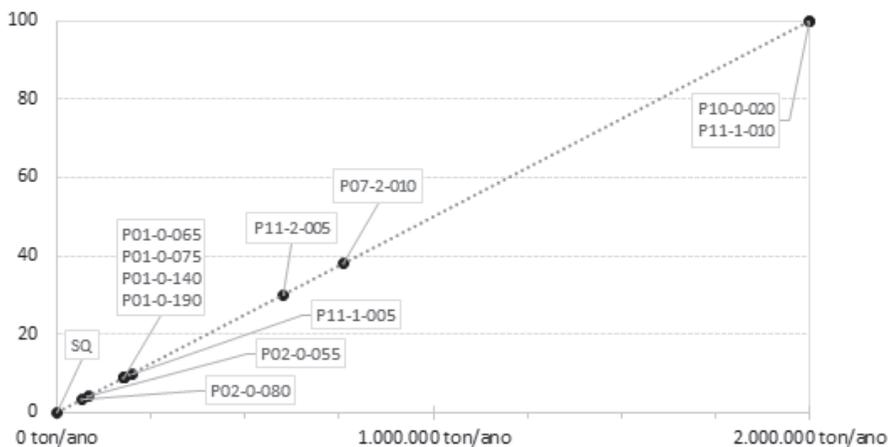


Figura 4: Função de valor linear do objetivo 2.4.2

A função de valor referente ao objetivo «1.3 Melhorar a eficiência da gestão» foi construída com base no método MACBETH (Bana e Costa et al., 2012) e respetivo sistema de apoio à decisão M-MACBETH (Bana e Costa et al., 2005). As vantagens desta solução em relação a outras incluem: a possibilidade de construir escalas de valor, segundo um determinado objetivo, com base apenas em juízos qualitativos; a deteção automática de inconsistências entre as respostas dos decisores; e a sólida base científica em que se baseia, nomeadamente a definição de valor sobre uma escala de intervalos. Para além disso, os autores têm utilizado com sucesso o método em diversos contextos de decisão (Oliveira e Lourenço, 2002; Mateus et al., 2008; Bana e Costa et al., 2008).

Este método começa por questionar as preferências de um decisor em relação às diferenças de atratividade entre os vários níveis de impacto de um descritor, utilizando para o efeito uma escala semântica com sete categorias

distintas de juízos de valor: nula; muito fraca; fraca; moderada; forte; muito forte; e extrema.

No caso presente, coube ao DG, como responsável pela determinação dos impactos bem como da realização das atividades, segundo este objetivo, a formulação do juízo qualitativo que no seu entender melhor representava a diferença de atratividade entre cada par de níveis de impacto do descritor associado ao objetivo 1.3 (ver Tabela 9). Esses juízos de valor são apresentados na matriz da Figura 5. Com base nesses juízos e na restrição de que ao melhor nível de impacto (N1) devem corresponder 100 unidades de valor e 0 unidades de valor ao *status quo*, o método MACBETH propõe uma escala de valor, atribuindo pontuações aos restantes níveis de impacto (ver coluna «Escala atual» da matriz apresentada na Figura 5).

Tabela 9: Níveis de impacto e pontuações parciais das atividades segundo o objetivo 1.3

Código	Atividade	Impacto	Pontuação
P16-34-010	Implementação de um sistema de gestão de qualidade	N1	100
P14-28-004	Implementação de um sistema de qualificação de empreiteiros	N2	87
P14-28-005	Implementação de um sistema de qualificação de empresas de fiscalização	N2	87
P16-34-020	Integração do planeamento da DGEN em SAP/R3	N2	87
P14-28-006	Assessoria para melhoria do desempenho recorrendo ao conceito RCM 2	N3	74
P14-30-006	Gestão documental: Indexação e digitalização de desenhos técnicos da DGEN	N3	74
P14-28-001	Novos equipamentos/sistemas (GRP 3000, GPS e de <i>software</i> para projeto).	N4	61
P14-28-003	<i>Software</i> do âmbito ferroviário (projeto de terraplenagem e drenagem).	N4	61
P14-28-002	Projeto piloto de RFID (identificação por rádio frequência) nos AMVs	N4	61
P14-30-007	Gestão documental: Custódia de documentação e tratamento arquivístico	N5	43
P13-23-010	Rede de voz e dados	N6	26
SQ	<i>Status quo</i>	SQ	0

Aumentar a Eficiência da Gestão								X	
	N1	N2	N3	N4	N5	N6	SQ	Escala actual	
N1	nula	fraca	moderada	forte	forte	mt. forte	extrema	100	extrema
N2		nula	fraca	moderada	forte	mt. forte	mt. forte	87	mt. forte
N3			nula	fraca	moderada	forte	mt. forte	74	moderada
N4				nula	moderada	moderada	mt. forte	61	fraca
N5					nula	moderada	forte	43	mt. fraca
N6						nula	moderada	26	nula
SQ							nula	0	

Julgamentos consistentes

Figura 5: Matriz MACBETH e pontuações parciais segundo o objetivo 1.3

O resultado final desta etapa metodológica foi consolidado numa matriz com as pontuações parciais de cada atividade (em linha) segundo cada um dos objetivos (coluna), tal como se ilustra na Tabela 10 para algumas atividades do Plano. Estas pontuações resultam da aplicação das funções de valor definidas para cada um dos objetivos à matriz dos impactos das atividades desenvolvida e apresentada como resultado da etapa anterior (ver Tabela 5).

Tabela 10: Pontuações parciais das atividades segundo os objetivos fundamentais

Código	Objetivos fundamentais														
	1.1	1.2	1.3	1.4.1	1.4.2	1.5	2.1	2.2	2.3	2.4.1	2.4.2	3.1	3.2	3.3	3.4
P01-0-015	33,3				2,6										
P01-0-036				2,8	1,4										
P01-0-075	39,5	5,6		0,1	3,4				56,3	0,4	9,0	14,3	1,3		
P01-0-080															
P01-0-081	34,0	60,6		1,1	21,7				8,8			7,1			
P01-0-190	-6,9	90,0		46,0	26,9				70,1	2,5	9,0	100	55,4		
P02-0-080	13,0	3,7		6,7	2,6			100	0,8	0,1	3,4	35,7			
P03-3-005	-96,3				0,3			10,7	0,3	0,5					
P05-0-040															50
P08-0-025				11,4	4,6		100	23,7		0,9		7,1	31,0		
P13-0-010	11,6	10,7		100	23,1										100
P13-21-010							100			0,8					
P16-34-020															87

9. Ponderação dos objetivos

A avaliação dos benefícios parciais das atividades permite retirar conclusões acerca da atratividade das atividades segundo cada objetivo, mas não permite avaliar a atratividade global dessas atividades.

Observe-se que no modelo de agregação aditivo definido em (1), os pesos dos objetivos não são mais que coeficientes de harmonização das diversas escalas de pontuação parcial dos objetivos, só tendo significado substantivo no contexto particular do modelo de avaliação em questão, isto é, do conjunto de objetivos, funções de valor e descritores nele considerados. Caso algum destes elementos seja alterado os pesos também se alteram. O que se mantém constante são as taxas de substituição (*trade-offs*) das pontuações parciais (unidades de valor) entre objetivos definidas pelo decisor. Os pesos operacionalizam assim uma noção de compensação entre objetivos, ou seja, é aceite pelo decisor que a perda de valor num determinado objetivo pode ser compensada pelo ganho de valor noutra objetivo, e vice-versa. A inobservância destes pressupostos em procedimentos de ponderação é considerado por Keeney (1992, pp. 147-148) como «o erro crítico mais comum», o que resulta em pesos completamente arbitrários, que não refletem as preferências dos decisores. De facto, os pesos dos objetivos são frequentemente mal interpretados como sendo tradutores de uma noção de importância intrínseca de valor dos objetivos, o que é incorreto no enquadramento do modelo aditivo de agregação. É exatamente para evitar este tipo de equívocos que os procedimentos de ponderação teoricamente corretos pressupõem que sejam desenvolvidos com referência aos intervalos de variação dos impactos, segundo cada um dos objetivos, e às preferências dos decisores em relação a esses impactos. Uma técnica eficaz para o efeito consiste em utilizar as respostas dos decisores a questões que requerem da parte destes a comparação de atividades fictícias, mas plausíveis, construídas com base na combinação de dois níveis de referência, como sejam o melhor e o pior níveis de impacto, segundo cada um dos objetivos. De entre os procedimentos de ponderação teoricamente corretos existentes, destacam-se os métodos *trade-off procedure* (Keeney e Raiffa, 1976), *swing weighting* (apresentado em Edwards (1977) e corrigido em Edwards e

Barron (1994)) e MACBETH (Bana e Costa e Vansnick (1999) e Bana e Costa et al. (2012)).

Tendo em consideração as características específicas dos Decisores, de formação de base com uma sólida componente quantitativa, os facilitadores escolheram o método *trade-off procedure* porque consideraram que era o que lhes iria permitir apreender mais facilmente a noção de compensação e o significado substantivo dos pesos dos objetivos (Mateus et al., 2010). Os pesos foram assim calculados através da metodologia que se descreve genericamente nos seguintes passos:

1. Em primeiro lugar, foram criadas 16 alternativas fictícias ao, tantas quanto o número de objetivos fundamentais existentes mais uma, tal que, uma delas mantivesse o *status quo* em todos os objetivos (ou seja, que tivesse o pior impacto em todos os objetivos) e cada uma das restantes tivesse o melhor impacto segundo um dos objetivos *o* e o pior impacto (equivalente ao *status quo*) segundo os restantes objetivos;
2. De seguida, os Decisores ordenaram essas alternativas fictícias por ordem decrescente de atratividade global. Este procedimento foi conduzido de uma forma interativa, através de questões do tipo «qual das alternativas fictícias é preferível?» ou «se apenas fosse possível melhorar um dos objetivos, do *status quo* para o melhor impacto, qual era o objetivo onde preferia que isso acontecesse?» até se ordenarem todas as alternativas fictícias;
3. Posteriormente, foi pedido aos Decisores que selecionassem uma das alternativas fictícias para servir de alternativa padrão de comparação com as restantes. É aconselhável que esta alternativa seja aquela que foi considerada preferida como resultado do procedimento anterior, de forma a que a atratividade do melhor impacto da alternativa padrão seja sempre preferível ao da alternativa com o qual vai ser comparado, e que tenha um descritor quantitativo contínuo para se poder determinar níveis intermédios entre o melhor impacto e o *status quo* (o que não seria prático, ou mesmo inexequível, se o objetivo escolhido tivesse um descritor

qualitativo discreto). Neste caso, os Decisores optaram por escolher o segundo objetivo preferido («1.1 Reduzir os custos de manutenção»), uma vez que consideraram ser mais fácil comparar impactos utilizando como escala de referência a variação dos custos de manutenção em euros/ano;

4. Os Decisores foram então interpelados acerca de «quanto é que a REFER estaria disposta a despende a mais em custos de manutenção (€/ano), para passar do *status quo* para o melhor impacto segundo cada objetivo?». Isto corresponde, para cada alternativa fictícia em avaliação, a variar o impacto segundo o objetivo 1.1 da alternativa padrão, até os Decisores considerarem as duas alternativas equivalentes em termos de preferência. Por exemplo, para o objetivo «1.2 Reduzir os custos de operação», os Decisores indicaram que a variação de 1,5 milhões de euros/ano em custos de manutenção era equivalente a uma variação de 1,3 milhões de euros/ano em custos de operação (ou seja, consideraram que é preferível reduzir 1 €/ano em custos de manutenção do que 1€/ano em custos de operação). Já para o objetivo «2.4.2 Aumentar a atratividade do sistema (mercadorias)» os Decisores consideraram que um aumento 2 milhões de toneladas/ano em mercadorias era equivalente a uma variação de 1,75 milhões de euros/ano em custos de manutenção (ver Tabela 11). Uma forma de auxiliar os Decisores nesta tarefa consistiu em determinar primeiro os *trade-offs* em euros/ano (de custos de manutenção) para aqueles objetivos que os Decisores consideravam mais fáceis, e para os restantes ir sucessivamente estabelecendo limites inferiores e superiores com base na ordenação relativa em termos de preferência anteriormente construída;
5. Desta forma, foi possível construir 14 relações de equivalência, tantas quanto o número total de objetivos fundamentais n menos um. Cada uma destas relações de equivalência pode ser formalmente enunciada através de uma equação linear do seguinte tipo:

$$\left(\frac{100 \times \text{tradeoff}_o}{1.500.000}\right) \times k_{1.1} = 100 \times k_o \quad \forall o \neq 1.1$$

em que, o índice o corresponde ao objetivo fundamental o ($o \neq 1.1$) de acordo com a Tabela 11; tradeoff_o corresponde a quanto é que a REFER estaria disposta a despende a mais em custos de manutenção (€/ano), para passar do *status quo* para o melhor impacto segundo o objetivo fundamental o ; e $k_{1.1}$ e k_o representam os pesos, respetivamente, dos objetivos fundamentais 1.1 e o ;

6. Finalmente, se se adicionar, ao conjunto de equações lineares definido anteriormente, uma restrição linear que impõe que a soma dos pesos k_o dos n objetivos fundamentais totalize 1, será então possível formar um sistema de equações lineares com um número de equações igual ao número de incógnitas, que são os pesos dos objetivos. A resolução deste sistema de equações permite assim calcular um conjunto de pesos consistente com os *tradeoffs* entre os objetivos estabelecidos pelos Decisores.

A Tabela 11 apresenta, para cada um dos objetivos fundamentais, os seguintes elementos:

- o *status quo* (SQ) e o melhor impacto segundo o objetivo em questão;
- as unidades de impacto segundo cada objetivo;
- a ordenação em termos de preferência de cada objetivo;
- quanto é que a REFER estaria disposta a despende a mais em custos de manutenção (€/ano), para passar do *status quo* para o melhor impacto segundo cada um dos objetivos;
- os pesos dos objetivos, resultantes da aplicação do método de ponderação descrito.

10. Avaliação dos benefícios globais das atividades

Após a determinação dos pesos dos objetivos foi possível agregar os benefícios parciais de cada atividade numa pontuação de benefícios globais, aplicando o modelo aditivo apresentado em (1). As pontuações globais das atividades em avaliação (incluídas nos pacotes P2 a P7 de todos os setores)

foram posteriormente analisadas por setor e depois confrontadas entre os vários setores. Destas análises resultaram algumas discordâncias, por exemplo, em relação à ordenação relativa das atividades por setor ou às pontuações globais das atividades de setores distintos.

Tabela 11: *Tradeoffs* e pesos dos objetivos fundamentais

Objetivos	SQ	Melhor	Unidades	Ord.	Tradeoff	Pesos
1.1 Reduzir os custos de manutenção	0	-1.500.855	€/ano	1.º	1.500	13,2%
1.2 Reduzir os custos de operação	0	-1.500.000	€/ano	4.º	1.300	11,5%
1.3 Melhorar a eficiência da gestão	SQ	N1		12.º	200	1,8%
1.4.1 Evitar acidentes pessoais	0	-6	mortos equiv./ano	6.º	850	7,5%
1.4.2 Evitar acidentes materiais	0	-70	acidentes/ano	4.º	1.300	11,5%
1.5 Garantir a segurança de pessoas e bens	0	-25%	atos vandalismo/ano	7.º	750	6,6%
2.1 Aumentar a utilização da infraestrutura	0	482.328	comboios.km/ano	8.º	500	4,4%
2.2 Reduzir o tempo de trajeto	0	68.112	comboios.minuto/ano	11.º	250	2,2%
2.3 Garantir a continuidade do serviço	0	-82.631	comboios.minuto/ano	8.º	500	4,4%
2.4.1 Aumentar a atratividade (passageiros)	0	22.116.650	passageiros/ano	3.º	1.300	11,5%
2.4.2 Aumentar a atratividade (mercadorias)	0	2.000.000	toneladas/ano	2.º	1.750	15,4%
3.1 Acessibilidade	0	14	estações	12.º	250	2,2%
3.2 Ruído	0	7.341	peessoas	12.º	350	3,1%
3.3 Intervenções artísticas	0	-2	estações	15.º	50	0,4%
3.4 Passagens de nível	0	-80	passagens de nível	10.º	500	4,4%

Tradeoffs em milhares de euros/ano

Uma das vantagens da construção de um modelo deste tipo é a de que permite, por um lado, justificar de forma transparente as razões que levaram à atribuição de cada pontuação e, por outro lado, resolver algumas das discordâncias que possam existir no decurso do processo de avaliação, nomeadamente através da revisão de alguns dos parâmetros (por exemplo, pesos, funções de valor ou impactos) que estiveram na base da avaliação efetuada. Uma convicção base da metodologia é a de que o modelo deve representar os dados objetivos do problema e as preferências dos Decisores. Nos casos em que os resultados do modelo não coincidiam com as

preferências holísticas iniciais dos Decisores em relação ao mérito das atividades, o modelo foi revisitado de forma a verificar se o mesmo estava a representar adequadamente as suas preferências, nomeadamente em termos de objetivos e juízos de valor, até considerarem válidas as recomendações sugeridas pelo modelo. Este processo culminou numa melhor compreensão do problema de decisão e das razões que levaram à pontuação global das atividades tal como sugerida pelo modelo.

As pontuações globais das atividades assim obtidas foram depois normalizadas para que o conjunto de todas as atividades em avaliação correspondesse a um benefício global relativo de 100%. Desta forma, foi possível determinar o benefício global relativo de cada atividade em relação ao *status quo* (medido em percentagem do total de benefícios) dividindo a pontuação global dessa atividade pela soma das pontuações globais de todas as atividades. A Figura 6 sintetiza, para cada setor, o somatório dos benefícios globais relativos das atividades incluídas em cada pacote (uma célula a cinzento indica que não foram classificadas quaisquer atividades no respetivo pacote). Assim, por exemplo, para o setor S1, as atividades classificadas no pacote P3 representam 1,9% do total dos benefícios, as atividades classificadas no pacote P3 (que inclui também as atividades do pacote P2) representam 4,5% do total dos benefícios. O conjunto das atividades deste setor S1 representam 31,6% do total dos benefícios.

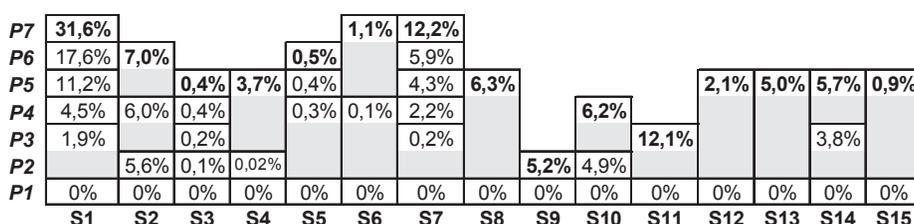


Figura 6: Benefícios globais relativos das atividades

11. Seleção das atividades a incluir no plano de atividades

Os recursos são escassos e limitados em qualquer organização. Em especial, a aplicação dos recursos públicos está sujeita a um escrutínio cada vez maior da sociedade, por via das instituições ou da opinião pública e publicada. A

afetação de recursos deve ser portanto baseada em critérios transparentes de avaliação da eficácia, eficiência e economicidade das atividades que é suposto sustentarem e fundamentada num processo participativo que garanta a qualidade das decisões tomadas e a sua fundamentação perante os restantes atores, internos ou externos à organização.

O método proposto para a determinação das atividades a incluir no Plano consistiu na seleção de um portefólio (subconjunto) de atividades eficiente que não excedesse o orçamento disponível da DGEN para 2006.

Os benefícios globais relativos e os custos a incorrer com cada atividade foram determinados nas etapas metodológicas anteriores. No âmbito de cada setor, a priorização das atividades foi realizada, a um primeiro nível, considerando os pacotes em que foram incluídas, sendo que, nesse âmbito, a realização de qualquer atividade incluída num determinado pacote pressupõe que todas as atividades incluídas em pacotes de nível inferior também sejam realizadas.

Neste contexto, cada estratégia de operacionalização do Plano de Atividades consiste numa qualquer combinação possível de atividades que respeite a regra de priorização anteriormente estabelecida. O número de diferentes estratégias que é possível assim construir compreende desde a estratégia de realizar apenas as atividades incluídas no pacote P1 de cada setor até à estratégia de realizar todas as atividades propostas pelos diversos órgãos da DGEN. O número total de combinações diferentes de atividades que é possível assim gerar é enorme (mais de $1,637 \times 10^{19}$ estratégias distintas), o que torna inexequível qualquer tentativa de comparar diretamente todas as estratégias entre si.

Existem várias abordagens para resolver este problema. Uma delas baseia-se no princípio da priorização das atividades, com base no seu rácio benefício/custo (Bana e Costa et al., 2006; Phillips e Bana e Costa, 2007): primeiro, as atividades são listadas por ordem decrescente do seu rácio benefício/custo; posteriormente, estas atividades são selecionadas pela ordem listada até que não seja possível selecionar mais nenhuma atividade sem que se ultrapasse o limite de recursos disponível. O portefólio de atividades assim construído é constituído pelas atividades com maiores rácios benefício/custo, e

é um portfólio eficiente, pois não existe outro portfólio de atividades que dê maior benefício global sem custar mais, ou que custe menos e dê o mesmo benefício global. Foi desta forma que se selecionou o portfólio de atividades para o Plano considerando o orçamento disponível pela DGEN para 2006.

Dada a sua simplicidade, este método pode ser facilmente implementado em Excel. Contudo, existe *software* específico no mercado (Lourenço et al., 2008) que, para além de apresentar variações em relação à forma como o portfólio eficiente de atividades é selecionado e quanto ao tipo de restrições e interações que é possível estabelecer, permite visualizar e interagir, de uma forma amigável, com os decisores, face a alterações em tempo real de qualquer um dos parâmetros do modelo de afetação de recursos (Goodwin e Wright, 2004). Por esta última razão, optou-se por selecionar o *software* Equity (Catalyze, 2008). A Figura 7 apresenta um dos *outputs* desta aplicação.

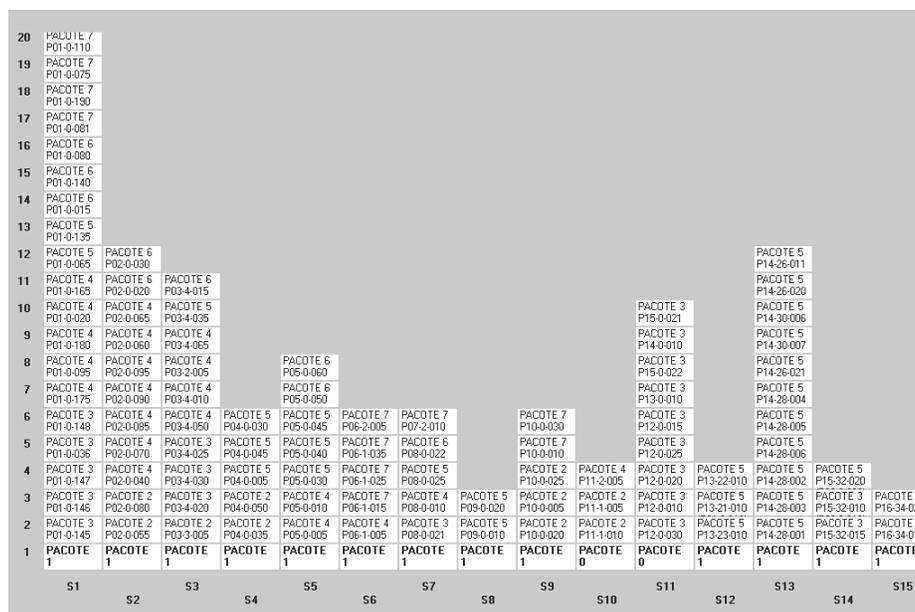


Figura 7: Ordenação das atividades por pacote e, dentro destes, pelo seu rácio benefício/custo

Na Figura 7, os setores são apresentados no eixo horizontal e as atividades, incluídas em cada um dos setores, são apresentadas no eixo vertical. Cada uma das atividades é representada por uma caixa retangular onde se encontram

descritos a sua designação e o respetivo pacote. A única exceção a esta regra diz respeito às atividades do pacote P1 que, para cada setor, foram agrupadas numa única caixa retangular onde constava simplesmente a designação do respetivo pacote, uma vez que a sua seleção não estava em avaliação. As restantes atividades, incluídas nos pacotes de nível superior, foram previamente priorizadas: primeiro, de acordo com o pacote em que tinham sido classificadas, sendo que as atividades inseridas em pacotes de nível inferior tinham prioridade sobre outras que tivessem sido inseridas em pacotes de nível superior; de seguida, para as atividades que tinham sido classificadas num mesmo pacote, de acordo com o seu rácio benefício/custo, sendo que as atividades com um maior rácio tinham prioridade sobre outras com um rácio inferior. Assim, por exemplo, as atividades do setor S1 foram assim priorizadas: em primeiro lugar, uma vez que nenhuma das atividades tinha sido classificada no pacote P2, as atividades do pacote P3 foram ordenadas em função do seu rácio benefício/custo. A atividade P01-0-145 tinha o maior rácio benefício/custo, pelo que deverá ser a primeira atividade a ser realizada segundo este setor. A atividade P01-0-146 apresentava o segundo melhor rácio, pelo que ficou ordenada em segundo lugar. Da mesma forma, as restantes três atividades deste pacote foram priorizadas por ordem decrescente do seu rácio de acordo a sequência P01-0-147, P01-0-036 e P01-0-148. De seguida, foram as atividades do pacote P4 a ser priorizadas, em função do seu rácio benefício/custo, resultando na ordenação P01-0-175, P01-0-095, P01-0-180, P01-0-020 e P01-0-165. Finalmente, as atividades dos pacotes P5, P6 e P7 foram priorizadas, sucessivamente, de acordo com a mesma regra.

Após a estruturação do problema desta forma no *software* Equity, esta aplicação permite representar graficamente os portefólios de atividades eficientes (ver Figura 8). A área a sombreado representa o «envelope» que define os limites convexos em termos dos benefícios relativos e custos totais a incorrer em 2006 do universo de portefólios distintos que é possível criar face aos dados deste problema. A curva que estabelece o limite superior deste envelope é definida pelo conjunto dos pontos que representam os portefólios eficientes que é possível gerar utilizando a abordagem da priorização anteriormente descrita. De acordo com esta abordagem, o portefólio eficiente a

selecionar para o Plano de Atividades é o que corresponde ao ponto que se situa imediatamente à esquerda da linha vertical a tracejado, que define o limite orçamental para 2006.

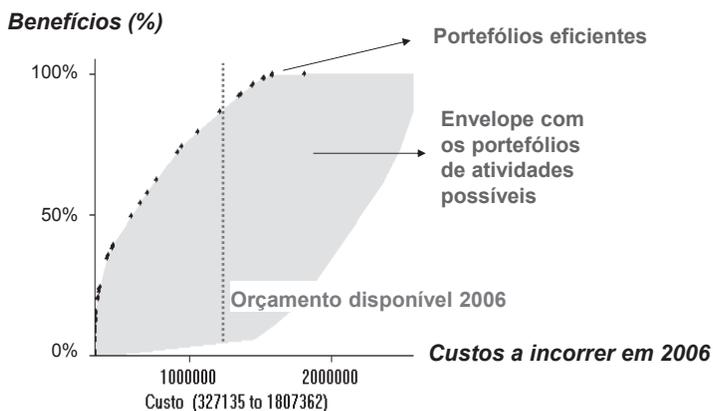


Figura 8: Representação gráfica dos portfólios eficientes no *software* Equity

A constituição dos portfólios eficientes encontrados pela priorização face a sucessivos aumentos dos limites orçamentais disponíveis pode ser observada numa janela do *software* Equity (Figura 9). O portfólio eficiente representado na Figura 9 é constituído pelas atividades do setor S1 até à atividade P01-0-065, pelas atividades do pacote P1 dos setores S2 e S3, pelas atividades do setor S4 até à atividade P04-0-005, por todas as atividades do setor S5, pelas atividades do pacote P1 do setor S6, pelas atividades do setor S7 até à atividade P08-0-010, pelas atividades do pacote P1 dos setores S8, S9 e S10, pelas atividades do setor S11 até à atividade P12-0-025, pelas atividades do setor S12 até à atividade P13-21-010 e por todas as atividades dos setores S13, S14 e S15.

12. Notas finais

O modelo desenvolvido é uma ferramenta que visa apoiar os decisores da REFER no processo de seleção do portfólio de atividades a incluir no Plano, tendo em consideração os recursos disponíveis na organização para esse efeito.

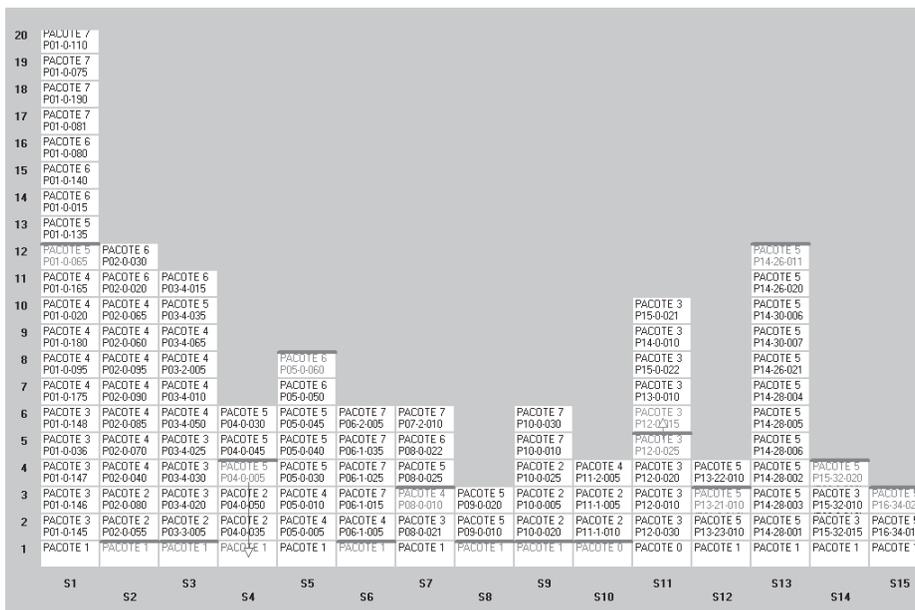


Figura 9: Portefólio de atividades eficiente

A aceitação do modelo e dos respetivos resultados por parte dos participantes é uma das virtudes da abordagem sociotécnica aplicada. Nas palavras de um Diretor da REFER «os resultados são atingidos por pessoas. Por conseguinte, se o grau de convergência entre os objetivos das pessoas e os objetivos da organização for elevado, o Plano de Atividades tem maiores hipóteses de se tornar uma realidade» (Diogo, 2007, p. 4). Neste âmbito, refira-se ainda, a título ilustrativo, uma das reações manifestadas por um dos Decisores relativamente ao processo de apoio à decisão: «Pela primeira vez, em muitos anos, vi numa mesma sala todos os Diretores da REFER, a interagir e discutir sobre os problemas comuns da organização». De facto, o envolvimento dos atores-chave da DGEN desde a definição dos objetivos até à análise e seleção das atividades a integrar no Plano permitiu que os participantes desenvolvessem um sentido de pertença do modelo desenvolvido (Phillips e Bana e Costa, 2007); na realidade, o modelo foi elaborado pelos atores-chave para eles próprios utilizarem como instrumento de afectação de recursos e, posteriormente, na sua componente de avaliação multicritério, como

ferramenta de monitorização. No entanto, o desenvolvimento de um processo deste tipo, envolvendo diversos participantes com múltiplas perspetivas sobre o problema da REFER, requer uma elevada preparação e sólidos conhecimentos técnicos por parte da equipa de apoio à decisão, tanto para a facilitação das reuniões de grupo como para a aplicação dos métodos e técnicas da análise de decisão e utilização de *software* especializado.

É importante destacar a forma como o zero da escala de valor de cada objetivo foi definido (Golabi et al., 1981; Bana e Costa et al., 2006; Clemen e Smith, 2009). Na literatura sobre multicritério é usual estabelecer que o valor segundo um determinado objetivo deve ser representado sobre uma escala de intervalos, sendo esta escala invariante a transformações lineares afins (Stevens, 1946). Mas esta transformação não é válida em problemas de afectação multicritério de recursos, uma vez que esta transformação não garante a preservação do ponto zero da escala. Por exemplo, considerem-se três atividades (*A*, *B* e *C*) com, respetivamente, 30, 20 e 10 unidades de benefício e 20, 16 e 10 unidades de custo, as suas razões benefício/custo são respetivamente, $r_A = 1,5$ (30/20), $r_B = 1,25$ (20/16) e $r_C = 1$ (10/10). Se adicionarmos 50 unidades de benefício a cada uma das atividades, que é uma operação perfeitamente lícita para escalas de intervalos, as razões benefício/custo alteram-se, ficando $r_A = 4$ (80/20), $r_B = 4,375$ (70/16) e $r_C = 6$ (60/10). Observe-se que a ordem das razões era originalmente $r_A > r_B > r_C$ e que com a adição das 50 unidades de benefício passou a ser $r_C > r_B > r_A$, o que altera completamente a ordem de seleção de atividades a incluir no portefólio. A solução adotada para resolver este problema, no caso da REFER aqui apresentado (Secção 6), passou por considerar que qualquer atividade que não acrescentasse valor nem retirasse valor segundo um determinado objetivo (ou seja, se mantivesse o *status quo* nesse objetivo) teria uma pontuação de zero unidades de benefício nesse objetivo.

Em último lugar há que realçar que a investigação na área da afetação multicritério de recursos – ou *Portfolio Decision Analysis* (Salo et al., 2011a e 2011b) – não se esgota na abordagem que foi aplicada na REFER. Efetivamente, a investigação produzida nesta área tem tido uma dinâmica crescente nos últimos tempos, como se pode comprovar pelo elevado número de

publicações recentes (ver, por exemplo, a revisão da literatura apresentada em Salo et al., 2011c). Entre outros desenvolvimentos, destaca-se a existência de novas ferramentas e abordagens que permitem modelar, de forma direta, interdependências entre atividades, tais como relações de dependência, de sequencialidade (como as descritas na Secção 3) ou de mútua exclusividade entre atividades, sinergias ou antagonismos entre atividades e acomodar restrições de recursos em múltiplos períodos temporais, o que não é possível fazer-se com a metodologia que foi aplicada na REFER. Em problemas em que existe incerteza associada às estimativas de custos e benefícios das atividades bem como, eventualmente, hesitações relativamente aos pesos dos objetivos é sensato analisar a robustez do portefólio de atividades selecionado face à variação simultânea de todos os parâmetros incertos. Refira-se a propósito deste último aspeto que a metodologia proposta inicialmente à REFER incluía a realização de análises de sensibilidade aos parâmetros do modelo, principal razão pela qual se admitiu desde o início que as estimativas dos custos e impactos pudessem ser modeladas através de intervalos de valores (Secção 7).

Para lidar com as situações antes descritas podemos agora recorrer à abordagem e sistema de apoio à decisão denominados *Portfolio Robustness Evaluation* (PROBE) (Lourenço et al., 2012), que utiliza a programação matemática e algoritmos específicos para formar portefólios eficientes de atividades e para analisar a robustez desses portefólios. É interessante referir que este *software* também permite considerar custos associados a não financiar atividades, o que poderá ser útil quando a não seleção de uma atividade possa dar lugar, por exemplo, ao pagamento de uma indemnização (Lourenço et al., 2010). *Robust Portfolio Modeling* (RPM) (Liesjö et al., 2008) foi a primeira abordagem criada para analisar a robustez de portefólios de atividades. RPM considera logo à partida a existência de incerteza em todos os parâmetros do modelo para depois procurar todos os portefólios eficientes de atividades, o que resulta, tipicamente, num número muito elevado de portefólios. Posteriormente, utiliza um processo iterativo, através da aplicação de regras de exclusão e de redução progressiva dos domínios de incerteza dos parâmetros para reduzir o número de portefólios eficientes previamente encontrados até ao

ponto em que o decisor se sente em condições para selecionar um único portefólio de atividades.

Agradecimentos

Os autores agradecem à REFER por ter autorizado esta publicação e aos decisores e técnicos que colaboraram no trabalho aqui apresentado, em especial ao Eng.º Carrasquinho de Freitas (Diretor-Geral da DGEN). A redação final do artigo beneficiou significativamente dos comentários produzidos por revisores anónimos, aos quais deixamos o nosso agradecimento. Adicionalmente, os autores expressam publicamente o seu agradecimento ao Professor Carlos Bana e Costa pelos ensinamentos ministrados nas desafiantes áreas da análise de decisão e da afetação multicritério de recursos. Por último, agradecem à Fundação para a Ciência e a Tecnologia pelos meios disponibilizados para a sua formação científica e para a elaboração deste trabalho.

BIBLIOGRAFIA

BANA e COSTA, Carlos António, "Absolute and Relative Evaluation Problematiques: The Concept of Neutral Level and the MCDA Robot Technique", *Multicriteria Decision Making: Methods, Algorithms, Applications – Proceedings of the International Workshop*, Liblice, Czechoslovakia, 1992a, pp. 7-15.

BANA e COSTA, Carlos António, *Structuration, construction et exploitation d'un modèle multicritère d'aide à la décision*, Tese de Doutoramento, Universidade Técnica de Lisboa, 1992b.

BANA e COSTA, Carlos António, *Processo de apoio à decisão: Actores e acções; estruturação e avaliação*, Publicação CESUR, IST, Volume 618, 1993a.

BANA e COSTA, Carlos António, "Três convicções fundamentais na prática de apoio à decisão", *Pesquisa Operacional*, 13(1), 1993b, pp. 9-20.

BANA e COSTA, C. A. e Outros, "Metodologia multicritério de apoio à avaliação de propostas em concursos públicos", *Casos de Aplicação da Investigação Operacional*, McGraw-Hill, 2000, pp. 337-363.

BANA e COSTA, C. A. e Outros, "Multicriteria approach for strategic town planning: the case of Barcelos", *Aiding Decisions with Multiple Criteria: Essays in Honour of Bernard Roy*, Kluwer Academic Publishers, Dordrecht, 2002, pp. 429-456.

BANA e COSTA, C. A. e Outros, *M-MACBETH Versão 1.1 Guia do utilizador*, <http://www.m-macbeth.com>, 2005.

BANA e COSTA, C. A. e Outros, "Prioritisation of public investments in social infrastructures using multicriteria value analysis and decision conferencing: A case study", *International Transactions in Operational Research*, 13(4), 2006, pp. 279-297.

- BANA e COSTA, C. A. e Outros, "Development of reusable bid evaluation models for the Portuguese Electric Transmission Company", *Decision Analysis*, 5(1), 2008, pp. 22-42.
- BANA e COSTA, C. A. e Outros, "MACBETH", *International Journal of Information Technology & Decision Making*, 11(2), 2012, pp. 359-387.
- BANA e COSTA, C. A. e Outros, "A socio-technical approach for group decision support in public strategic planning: The Pernambuco PPA case", *Group Decision and Negotiation*, 23(1), 2014, pp. 5-29.
- BANA e COSTA, C. A. e VANSNICK, J. C., "MACBETH: an interactive path towards the construction of cardinal value functions", *International Transactions in Operational Research*, 1(4), 1994, pp. 489-500.
- BANA e COSTA, C. A. e VANSNICK, J. C., "The MACBETH approach: Basic ideas, software, and an application", *Advances in Decision Analysis*, Kluwer Academic Publishers, Dordrecht, 1999, pp. 131-157.
- BANXIA, *Decision Explorer - User's Guide Version 3.2. Banxia Software Limited*, Kendal, <http://www.banxia.com>, 2002.
- BELTON, V. e STEWART, T. J., *Multiple Criteria Decision Analysis: An Integrated Approach*, Kluwer, Boston, MA, 2002.
- CATALYZE, *Equity (version 3.4)*, Catalyze Ltd., Hurley, <http://www.catalyze.co.uk>, 2008.
- CLEMEN, Robert Taylor, *Making Hard Decisions: An Introduction to Decision Analysis*, Duxbury Press, Belmont, CA, 1996.
- CLEMEN, R. T. e SMITH, J. E., "On the choice of baselines in multiattribute portfolio analysis: A cautionary note", *Decision Analysis*, 6(4), 2009, pp. 256-262.
- DIOGO, Alberto de Almeida, "Plano de Actividades e Orçamentos: Realizar mais actividades com os mesmos recursos", *Notícias REFER*, Ano 11, N.º 57, 2007, p. 4.
- DOGSON, J. e Outros, *DTLR Multi-Criteria Analysis Manual*, Office of the Deputy Prime Minister, UK, 2000.
- DYER, J. S. e SARIN, R. K., "Measurable multiattribute value functions", *Operations Research*, 27(4), 1979, pp. 810-822.
- EDEN, Colin, "Cognitive mapping", *European Journal of Operational Research*, 36(1), 1998, pp. 1-13.
- EDEN, C. e SIMPSON, P., "SODA and cognitive mapping in practice", *Rational Analysis for a Problematic World: Problem Structuring Methods for Complexity, Uncertainty and Conflict*, John Wiley & Sons, New York, 1989, pp. 43-70.
- EDWARDS, Ward, "How to use multiattribute utility measurement for social decisionmaking", *IEEE Transactions on Systems, Man and Cybernetics*, 7(5), 1977, pp. 326-340.
- EDWARDS, W. e BARRON, F. H., "SMARTS and SMARTER: Improved simple methods for multiattribute utility measurement", *Organizational Behavior and Human Decision Processes*, 60(3), 1994, pp. 306-325.
- GOLABI, K. e Outros, "Selecting a portfolio of solar energy projects using multiattribute preference theory", *Management Science*, 27(2), 1981, pp. 174-189.
- GOODWIN, P. e WRIGHT, G., *Decision Analysis for Management Judgement*, John Wiley & Sons, Chichester, 2004.
- HAMMOND, J. S. e Outros, "The hidden traps in decision making", *Harvard Business Review*, 84(1), 2006, pp. 118-126.

- KEENEY, Ralph L., *Value-focused thinking: A path to creative decision making*, Harvard University Press, Cambridge, MA, 1992.
- KEENEY, Ralph L., "Developing objectives and attributes", *Advances in Decision Analysis: From Foundations to Applications*, Cambridge University Press, Cambridge, 2007, pp. 104-128.
- KEENEY, R. L. e RAIFFA, H., *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*, John Wiley & Sons, New York, 1976.
- LIESIÖ, J. e Outros, "Robust portfolio modeling with incomplete cost information and project interdependencies", *European Journal of Operational Research*, 190(3), 2008, pp. 679-695.
- LOURENÇO, J. C. e Outros, "Software packages for multi-criteria resource allocation", *IEEE International Engineering Management Conference*, Estoril, Portugal, 2008.
- LOURENÇO, J. C. e Outros, "Portfolio decision analysis with PROBE: Addressing costs of not financing projects", *12th WSEAS International Conference on Mathematical and Computational Methods in Science and Engineering*, University of Algarve, Faro, Portugal, 2010.
- LOURENÇO, J. C. e Outros, "PROBE - A multicriteria decision support system for portfolio robustness evaluation", *Decision Support Systems*, 54(1), 2012, pp. 534-550.
- MONTEIRO, José Alves, "O Controlo das Actividades na REFER: Uma importante ajuda para melhorar o nosso desempenho", *Notícias REFER*, Ano 10, N.º 47, 2006, p.4.
- MATEUS, R. e Outros, "Multicriteria decision analysis (MCDA): Central Porto high-speed railway station", *European Journal of Operational Research*, 187(1), 2008, pp. 1-18.
- MATEUS, R. e Outros, "Full disclosure of tender evaluation models: Background and application in Portuguese public procurement", *Journal of Purchasing and Supply Management*, 16(3), 2010, pp. 206 - 215.
- NEVES, A. R. e Outros, "A multi-criteria approach to local energy planning: The case of Barreiro municipality", *1st International Conference on Operations Research and Enterprise Systems*, Vilamoura, Portugal, 2012.
- OLIVEIRA, R. C. e LOURENÇO, J. C., "A multicriteria model for assigning new orders to service suppliers", *European Journal of Operational Research*, 139(2), 2002, pp. 390-399.
- PHILLIPS, Lawrence D., "People-centred group decision support", *Knowledge-based Management Support Systems*, Ellis Horwood, Chichester, 1989, pp. 208-224.
- PHILLIPS, Lawrence D., "Decision analysis for group decision support", *Tackling strategic problems: The role of group decision support*, Sage, London, 1990, pp. 142-153.
- PHILLIPS, Lawrence D., "Decision conferencing", *Advances in Decision Analysis: From Foundations to Applications*, Cambridge University Press, Cambridge, 2007, pp. 375-399.
- PHILLIPS, L. D. e BANA e COSTA, C. A., "Transparent prioritisation, budgeting and resource allocation with multi-criteria decision analysis and decision conferencing", *Annals of Operations Research*, 154(1), 2007, pp. 51-68.
- REFER, Empresa Pública, *Manual de Procedimentos da Direcção Geral de Engenharia*, 2004.
- REFER, Empresa Pública, *Relatório de Sustentabilidade*, 2005a.
- REFER, Empresa Pública, *Relatório e Contas*, 2005b.
- RUSSO, J. E. e SCHOEMAKER, P. J. H., *Decision Traps: The Ten Barriers to Brilliant Decision-Making and How to Overcome Them*, Doubleday, New York, 1989.
- SALO, A. e Outros, "Portfolio decision analysis: Improve resource allocation", *Analytics (November/December)*, 2011a, pp. 23-26.
- SALO, A. e Outros, *Portfolio Decision Analysis: Improved Methods for Resource Allocation*, Springer, New York, 2011b.

SALO, A. e Outros, "An invitation to portfolio decision analysis", *Portfolio Decision Analysis: Improved Methods for Resource Allocation*, Springer, New York, 2011c, pp. 3-27.

STEVENS, Stanley Smith, "On the theory of scales of measurement", *Science*, 103(2684), 1946, pp. 677-680.

VON WINTERFELDT, D. e EDWARDS, W., *Decision Analysis and Behavioural Research*, Cambridge University Press, Cambridge, MA, 1986.

(Página deixada propositadamente em branco)

Parte IV

SETOR dos
SERVIÇOS

(Página deixada propositadamente em branco)

CASO 15

DESENHO DE PROMOÇÕES DIFERENCIADAS EM EMPRESAS DE RETAILHO RECORRENDO À SEGMENTAÇÃO DE CLIENTES

Vera L. Miguéis

Faculdade de Engenharia da Universidade do Porto
vera.migueis@fe.up.pt

Ana S. Camanho

Faculdade de Engenharia da Universidade do Porto
acamanho@fe.up.pt

João Falcão e Cunha

Faculdade de Engenharia da Universidade do Porto
jfcunha@fe.up.pt

RESUMO

Uma boa relação entre as empresas e os seus clientes é um factor essencial para a competitividade das empresas. De facto, a fidelização dos clientes é um requisito para o sucesso sustentado de uma empresa. Este artigo propõe um método de apoio ao desenho de promoções diferenciadas para empresas de retalho, que visa reforçar a relação das empresas com os seus clientes. Este método baseia-se no conhecimento extraído de dados transacionais recolhidos através de cartões de fidelização. Inicialmente, os clientes são segmentados usando o algoritmo de agrupamento k-médias e posteriormente o perfil dos clientes integrados em cada grupo é caracterizado através de uma árvore de decisão. De seguida, identificam-se os produtos que são comprados em conjunto pelos clientes de cada segmento, e que constituem a base das promoções diferenciadas propostas. O estudo apresentado é feito em colaboração com uma empresa europeia que opera no domínio do retalho de base alimentar.

PALAVRAS-CHAVE

Retalho, Promoções, *Data Mining*, Agrupamento, Classificação, Associação

1 Introdução

As recentes alterações económicas e sociais que têm ocorrido na Europa transformaram o sector de retalho. No passado, as empresas centravam-se na venda dos produtos e serviços sem procurarem conhecer os clientes que compravam os referidos produtos ou serviços. Com a proliferação da concorrência, tornou-se mais difícil atrair novos clientes e conseqüentemente as empresas viram-se obrigadas a concentrar esforços para manter os seus clientes. Para além disso, a evolução da sociedade e das condições económicas fez com que o estilo de vida dos clientes se alterasse, e como resultado estes passaram a ser menos influenciáveis pelas atividades de marketing das empresas. Neste contexto, numerosas empresas abandonaram estratégias centradas no produto e passaram a reger-se por estratégias centradas no cliente. A criação de relações de fidelidade com os clientes tem sido considerada um elemento crítico da estratégia de negócio das empresas. De facto, as empresas que se pretendem manter competitivas apostam na melhoria da sua relação com os seus clientes.

Algumas empresas investiram na criação de bases de dados capazes de armazenar os dados relativos a transações dos clientes que possuem cartões de fidelização. Por cada cliente, milhões de registos são recolhidos, permitindo a análise do histórico completo das compras. Contudo, o conhecimento que potencialmente se pode extrair destes dados é raramente integrado no processo de tomada de decisão, nomeadamente nas campanhas de marketing. A grande quantidade de dados armazenados resulta muitas vezes num problema de sobrecarga de informação acompanhado de escassez de conhecimento. Em geral, os analistas não têm sido capazes de estudar os dados e transformá-los em conhecimento útil para fins de aplicação em campanhas de marketing.

As técnicas de *data mining* têm surgido como ferramentas de grande potencial na análise dos dados resultantes da atividade transacional dos clientes. Estas técnicas podem ser usadas para detectar padrões e regras subjacentes ao comportamento dos clientes. Contudo, o uso das ferramentas de *data mining* no contexto do marketing é ainda incipiente. A segmentação de clientes ou a identificação de grupos de clientes com padrões de comportamento

semelhantes é feita, na maior parte dos casos, de uma forma ad-hoc. Para além disto, a maior parte das empresas ainda usa estratégias de comunicação em massa para chegar aos seus clientes.

Neste contexto, este estudo tem como objectivo segmentar os clientes de uma empresa europeia de retalho e propor políticas promocionais ajustadas aos clientes de cada segmento, com vista a aumentar o nível de serviço percebido pelos clientes e a sua consequente fidelização. Isto é feito com base no conhecimento extraído a partir de uma grande base de dados transaccional, com recurso a ferramentas de *data mining*. A referida base de dados resulta do uso do cartão de fidelização por parte dos clientes da empresa.

Este capítulo tem a estrutura que se segue. A Secção 2 inclui uma breve revisão da literatura, com vista à contextualização do estudo. A Secção 3 inclui a apresentação das técnicas usadas: técnicas de agrupamento, classificação e associação. A Secção 4 apresenta a empresa usada como caso de estudo, bem como os resultados da análise efetuada. Finalmente, apresentam-se as conclusões e algumas ideias para trabalho futuro.

2 Revisão da literatura

A fidelização de clientes por parte das empresas tem ganho proeminência na literatura de marketing e gestão. De acordo com Keaveney (1995), a fidelização dos clientes actua frequentemente como factor de retenção, de onde resulta um papel crucial na competitividade das empresas.

Bettencourt and Gwinner (1996) considera que o tratamento personalizado dos clientes é um dos pilares da sua satisfação. Desta forma, é possível fomentar a fidelização dos clientes às empresas através de promoções dirigidas. Contrariamente às políticas promocionais tradicionais que tratam os clientes indistintamente, o desenho de promoções dirigidas requer um conhecimento profundo do comportamento dos clientes. Isto possibilita a recomendação de produtos ou serviços que respondem às necessidades específicas dos clientes. De acordo com Ngai et al. (2009), a caracterização dos clientes é usualmente feita através da segmentação do mercado e a recomendação de produtos é feita com recurso à análise dos cabazes de compra.

As primeiras abordagens ao problema de segmentação basearam-se no uso de critérios geográficos. As empresas agrupavam os seus clientes de acordo com a sua área de residência ou de trabalho. Posteriormente, surgiram abordagens baseadas em critérios demográficos, através das quais os clientes eram agrupados de acordo com a idade, sexo, rendimento, ocupação, entre outros. Nos anos 60 a segmentação do mercado sofreu um crescendo de popularidade. Twedt (1964) sugeriu o uso de modelos de segmentação baseados no volume de vendas. Assim, as atividades de marketing deviam centrar-se nos clientes com um número considerável de transações. Esta teoria, denominada «heavy half theory», sugeria que metade dos clientes poderia representar 80% das vendas de uma empresa. Frank et al. (1967) criticou este modelo de segmentação, afirmando que este assume que os clientes mais representativos em termos de vendas tinham características que os diferenciavam dos restantes, o que não foi comprovado pelo modelo de regressão que desenvolveu. Posteriormente, Haley (1968) introduziu um modelo de segmentação que diferencia os clientes de acordo com os benefícios que estes esperavam do produto ou serviço. Assim, o mercado poderia ser segmentado de acordo com as valias esperadas pelos clientes: qualidade, desempenho, imagem, serviço, entre outras. Estes modelos promoveram mais investigação, o que permitiu o desenvolvimento de modelos complexos de segmentação baseados no estilo de vida dos clientes. O conceito de estilo de vida, associado ao marketing por Lazer (1964), é baseado no facto de cada indivíduo possuir padrões de vida próprios, que podem influenciar a sua motivação para a compra de determinados produtos e marcas. Durante os anos 70, a validade destes modelos de segmentação multivariada foi questionada (Green and Wind, 1973), o que levou à modelação teórica do comportamento dos clientes (por exemplo, Blattberg et al. (1978)). Uma década mais tarde, Mitchell (1983) desenvolveu um modelo de segmentação psicográfica que divide o mercado em grupos de acordo com a classe social, estilo de vida e características da personalidade. Contudo, as dificuldades de implementação deste modelo de segmentação mais complexo viriam a revelar-se críticas (Piercy and Morgan, 1993) e (Dibb and Simkin, 1997).

Mais recentemente, a literatura de marketing tem dado ênfase ao abandono de padrões previsíveis de comportamento e consumo por parte dos clientes. A diversidade de necessidades e desejos, influenciada pelo estilo de vida, rendimento e idade, têm tornado menos efetivas as abordagens de segmentação propostas no passado. Assim, os modelos actuais de segmentação de mercado são baseados no comportamento dos clientes inferido a partir dos registos das transações e de inquéritos. Os dados recolhidos são explorados com técnicas de *data mining*, nomeadamente técnicas de agrupamento. Exemplos da aplicação de técnicas de *data mining* na segmentação de clientes, com recurso a inquéritos, incluem Kiang et al. (2006). No contexto dos serviços de comunicação a longa distância, os clientes foram segmentados usando variáveis psicográficas, recolhidas a partir de um inquérito constituído por 68 questões relacionadas com o seu comportamento. Min and Han (2005) agruparam os clientes com gostos semelhantes em relação a filmes, usando a informação fornecida por cada cliente. A classificação relativa dos filmes feita pelos clientes permitiu deduzir o valor percebido por cada cliente em relação a cada filme. Helsén and Green (1991) também identificaram os segmentos de mercado para um novo sistema informático com base em técnicas de agrupamento, usando dados obtidos a partir de um inquérito. A segmentação baseou-se na importância dada às características do produto.

No que se refere às abordagens de segmentação baseadas nos dados transacionais armazenados em bases de dados, o modelo de segmentação «recency, frequency and monetary value» (RFM), introduzido nos anos 20 (Roel, 1988), é um modelo de segmentação muito utilizado. Este modelo considera a data da última compra (recency), a frequência com que o cliente faz compras (frequency) e o valor gasto (monetary value), extraídos a partir de uma base de dados transacional. Liu and Shih (2005) é um exemplo de um estudo recente de segmentação que usa o modelo RFM e que tem como objectivo especificar os segmentos no mercado de retalho de *hardware*.

Agrupados os clientes com características semelhantes, é muitas vezes necessário inferir o seu comportamento ou estilo de vida a partir dos dados transacionais. Para isso poderá ser interessante encontrar padrões comuns nas suas compras, como por exemplo conjuntos de produtos que são normalmente

adquiridos em conjunto. As técnicas de associação de *data mining* são frequentemente usadas para a análise de cabazes de compra, com o objectivo de extrair regras de associação de produtos. O uso destas técnicas para suportar estratégias promocionais diferenciadas pode ser considerado relativamente novo na literatura de marketing. Por exemplo, Van den Poel et al. (2004) usa este tipo de técnicas para definir vendas cruzadas e Brijs et al. (2004) usa-as para definir a disposição dos produtos nas lojas. A literatura que aborda as técnicas de associação centra-se principalmente no desenvolvimento de algoritmos, sendo a sua aplicação em casos reais e a integração com as políticas de marketing menos frequente.

Neste contexto, este estudo propõe a aplicação de técnicas de *data mining*, incluindo técnicas de agrupamento e de associação, para desenhar promoções diferenciadas para uma empresa de retalho europeia.

3 Metodologia

A metodologia usada neste estudo tem como objectivo apoiar o desenho de promoções por forma a aumentar os níveis de fidelização dos clientes. Numa primeira fase, os clientes da empresa usada como caso de estudo são segmentados usando uma técnica de agrupamento. A isto segue-se a caracterização do perfil dos clientes integrados em cada segmento, com recurso a uma árvore de decisão. Posteriormente, identificam-se os produtos que são frequentemente comprados em conjunto pelos clientes de cada segmento, usando uma técnica de associação. Através deste procedimento, é possível definir descontos promocionais para os clientes seleccionados, tendo em conta o histórico de compras e as associações de produtos identificadas para os clientes do mesmo segmento. Seguidamente descrevem-se as técnicas subjacentes à aplicação da metodologia descrita.

3.1 Análise de agrupamento

As técnicas de agrupamento são técnicas de *data mining* muito populares que agrupam os dados em conjuntos previamente desconhecidos, com base na sua semelhança. Existe uma grande variedade de algoritmos de agrupamento (ver Jain et al. (1999) para uma revisão). Estes algoritmos podem ser

classificados como sendo do tipo particional ou do tipo hierárquico. Os algoritmos particionais permitem dividir os dados em grupos que não se sobrepõem, ou seja, cada objeto pertence unicamente a um grupo. As técnicas particionais requerem a definição à priori do número de grupos. Apesar desta limitação, estas técnicas têm a vantagem de permitir a optimização da função de erro. Os algoritmos hierárquicos podem ser do tipo aglomerativo ou do tipo divisivo. Os algoritmos aglomerativos iniciam-se com cada objeto num grupo e, de seguida, juntam os grupos até que todos os dados estejam no mesmo grupo. Em cada iteração os dois grupos mais semelhantes são agrupados. O algoritmo divisivo inicia-se com um grupo contendo todos os objetos e iterativamente divide este grupo em grupos mais pequenos. Quer o algoritmo aglomerativo quer o algoritmo divisivo produzem uma sequência de grupos, em que o do topo inclui todos os objetos, e em que os do fundo incluem um único objeto. O diagrama resultante da aglomeração ou divisão dos grupos, denominado de dendrograma, permite definir facilmente o número apropriado de grupos para atingir um determinado objectivo. Contudo, o algoritmo hierárquico não permite a realocação de objetos que tenham sido «incorretamente» agrupados ou separados em fases anteriores do procedimento.

O algoritmo k-médias, introduzido por Forgy (1965) e posteriormente desenvolvido por MacQueen (1967), é um dos algoritmos de agrupamento mais populares. O k-médias, que pertence à classe dos algoritmos particionais, tem duas vantagens principais: é fácil de implementar e não exige grande esforço computacional, o que o torna apropriado para grandes bases de dados. Neste estudo aplica-se o algoritmo k-médias para segmentar os clientes, tendo em conta os dados transacionais. Este algoritmo visa distribuir um conjunto de n objetos de dados por k grupos, por forma a obter-se uma semelhança grande intra-grupos e uma semelhança reduzida entre os diferentes grupos. O algoritmo k-médias tem a desvantagem de requerer a definição à priori do número de grupos (tal como todos os algoritmos particionais), e depende das sementes iniciais (objetos de dados definidos como centroides iniciais dos grupos). De facto, este algoritmo implica a definição das sementes para a primeira iteração. A selecção de sementes diferentes pode conduzir à obtenção de grupos diferentes, especialmente quando os dados em análise contêm

valores extremos. Este algoritmo não possui qualquer meio para determinar as sementes iniciais mais apropriadas. Contudo, o procedimento standard é correr o algoritmo com diferentes sementes e escolher aquelas que permitem obter o valor mais baixo para a função de erro, normalmente o erro quadrado. Esta medida de erro avalia a distância entre os objetos de dados e os centroides dos respetivos grupos. Neste estudo usou-se o software RapidMiner, que usa um conjunto aleatório de objetos de dados como centroides iniciais para o agrupamento. Depois de correr o algoritmo k-médias com várias sementes aleatórias, selecionaram-se aquelas que permitiram obter o menor erro quadrado.

Depois da selecção das sementes iniciais, cada objeto é alocado ao grupo mais próximo, de acordo com a distância (tipicamente a distância euclidiana) entre o objeto e o centroide dos grupos. No final de cada iteração, os centroides são atualizados com base na média dos objetos incluídos no grupo, para que se possa iniciar uma nova iteração. Este processo termina quando a função de erro converge para um valor próximo do mínimo.

No que diz respeito à definição do número de grupos, são referidas na literatura várias heurísticas, já que teoricamente não é possível determinar o número óptimo de grupos (ver Tibshirani et al. (2001) para uma revisão). Neste estudo usaram-se dois critérios: o índice de Davies-Bouldin e o erro quadrático médio. O índice de Davies-Bouldin, desenvolvido por Davies and Bouldin (1979), é baseado na razão entre a soma da dispersão interna dos grupos e a distância entre grupos. O número apropriado de grupos é aquele que corresponde a um valor mais baixo deste índice. O erro quadrático médio, usado neste contexto para construir a denominada curva do cotovelo (Aldenderfer and Blashfield, 1984), representa a dispersão dentro dos grupos, definida tipicamente como a soma dos quadrados das distâncias entre todos os objetos e o centroide do grupo correspondente, dividida pelo número de grupos. À medida que o número de grupos aumenta, a medida de erro diminui monotonicamente e, a partir de determinado valor de k, a diminuição deixa de ser significativa. Este «cotovelo» é normalmente usado para definir o número adequado de grupos.

3.2 Regras de classificação

A extração das regras de classificação é um processo que pode ser usado para identificar as características que distinguem os objetos de diferentes grupos. Uma reconhecida técnica de classificação é a árvore de decisão, que permite facilmente obter as regras que caracterizam os grupos. Isto é feito através da identificação dos atributos, ou seja, as variáveis de agrupamento, consideradas relevantes para a descrição dos grupos. Uma árvore de decisão tem a estrutura de uma árvore com nós e ramos. Os nós podem ser de dois tipos: folha e nó de decisão. Uma folha representa uma classificação, ou seja, o nome do grupo, enquanto um nó de decisão está associado a uma escolha entre dois ramos, correspondendo a diferentes valores de um determinado atributo. Os ramos com origem num determinado nó de decisão representam as alternativas possíveis para o intervalo de valores dos atributos.

Existem na literatura inúmeros algoritmos de construção de árvores de decisão. O CHAID (Kass, 1980), o ID3 (Quinlan, 1986), o CART (Breiman et al., 1984), o C4.5 (Quinlan, 1992) e o AID (Morgan and Sonquist, 1963) são alguns exemplos. Por forma a remover os ramos da árvore que possuem um reduzido poder de discriminação entre os objetos, a árvore de decisão obtida é frequentemente podada. A principal razão para a execução deste procedimento é a obtenção de árvores genéricas, facilmente interpretáveis e que classificam de uma forma precisa a maior parte dos objetos.

O algoritmo usado neste estudo para a construção da árvore de decisão foi o algoritmo C4.5. Os detalhes sobre a parametrização deste algoritmo podem ser obtidos em Rapid-I (2009). Os valores dos parâmetros usados foram os seguintes: o critério de selecção do atributo foi a razão de ganho, a dimensão mínima para a divisão foi 40.000, a dimensão mínima das folhas foi 20.000, o ganho mínimo foi 0.19, o comprimento máximo foi 20 e o erro pessimista foi 0.5. Esta parametrização foi especificada a partir da comparação da percentagem de clientes classificados corretamente para diferentes configurações. Para avaliar a precisão da árvore de decisão, a base de dados foi dividida em 80% para efeitos de treino e 20% para testes. Esta divisão foi estratificada, de forma que a percentagem de clientes pertencentes aos

diferentes grupos nas amostras de treino e teste fosse aproximadamente idêntica à da base de dados inicial.

A árvore construída pode ser usada para caracterizar cada segmento de mercado. Isto requer que se inicie a caracterização a partir da raiz da árvore e que se faça um movimento ao longo dos ramos, contendo os valores mutuamente exclusivos dos atributos, até se atingir uma folha (contendo o nome do grupo). Assim, para se extrair as regras referentes a um grupo é necessário considerar todos os caminhos desde a raiz até às folhas contendo o nome do grupo.

3.3 Análise do cabaz de compras

A análise do cabaz de compras com recurso a técnicas de associação visa identificar grupos de produtos ou serviços comprados em simultâneo. Uma regra de associação pode ser representada na forma $X \Rightarrow Y$, o que significa que quando X é comprado Y é também comprado. X é denominado de antecedente e Y de consequente, de tal forma que o antecedente desencadeia a compra do consequente. O problema de associação envolve duas fases. A primeira fase diz respeito à descoberta dos produtos comprados mais frequentemente. Os algoritmos mais usados nesta fase são o algoritmo Apriori (Agrawal and Srikant, 1994), o algoritmo Frequent-pattern growth (Han et al., 2000) e o algoritmo Eclat (Zaki et al., 1997). O algoritmo usado neste estudo foi o algoritmo Apriori. A segunda fase da determinação das regras de associação diz respeito ao processo de geração de regras de associação, ou seja, à definição de quais os produtos antecedentes (X) e quais os consequentes (Y).

Os algoritmos de descoberta dos produtos comprados mais frequentemente envolvem várias fases. Na primeira fase, são identificados os produtos que possuem um suporte mínimo definido. A medida de suporte, representada por $s(X)$, corresponde à percentagem de cabazes de compra analisados que contêm o produto X . Em seguida, os produtos com um suporte maior do que o limite especificado são combinados dois a dois, e é calculado o suporte desses conjuntos de produtos, isto é, estima-se o valor de $s(X,Y)$ (ver expressão (1)). Os grupos de dois produtos com um suporte maior do que o mínimo definido pelo analista são considerados no passo seguinte. A cada um deles é

adicionado mais um produto, seleccionado de entre os produtos frequentes, identificados no primeiro passo. Este processo iterativo continua até que não seja possível definir mais conjuntos de produtos com um suporte acima do mínimo especificado.

$$s(X \Rightarrow Y) = P(X \cup Y) \quad (1)$$

Uma vez concluída a identificação dos produtos comprados mais frequentemente, é necessário identificar, dentro dos conjuntos de produtos frequentes, os produtos antecedente e consequente. Isto requer o cálculo da medida de confiança. Considerando-se a regra de associação $X \Rightarrow Y$, a confiança $c(X \Rightarrow Y)$ é a razão entre o número de cabazes que contêm ambos os produtos X e Y e o número de cabazes que contêm apenas X (ver expressão (2)). No caso da medida de confiança ser superior ao limite mínimo definido, a regra é considerada uma regra de associação.

$$c(X \Rightarrow Y) = P(Y | X) = s(X \cup Y) / s(X) \quad (2)$$

O *lift* é outra medida comum na análise das regras de associação, que avalia o nível de dependência entre os produtos que integram a regra de associação. É obtido dividindo o suporte de X e Y, $s(X, Y)$ (representando a percentagem de cabazes que contêm X e Y em relação ao conjunto total de cabazes) pelo produto do suporte de X e Y, considerados em separado (ver expressão 3).

$$lift(X, Y) = s(X, Y) / (s(X) \cdot s(Y)) \quad (3)$$

Se o *lift* for igual a 1, existe independência entre a ocorrência de vendas dos produtos X e Y. Se o *lift* for maior do que 1, os produtos tendem a ser comprados em conjunto, e se for menor do que 1, os produtos tendem a ser comprados separadamente. As regras que apresentam um *lift* inferior a 1 são geralmente ignoradas, já que apenas as regras com *lift* superior a 1 são interessantes para suportar políticas de marketing.

4 Desenho de promoções dirigidas

4.1 Apresentação da empresa usada como caso de estudo

Este estudo desenvolve um método de desenho de promoções no sector de retalho, considerando as associações de produtos observadas nos grupos homogêneos de clientes. Usou-se como caso de estudo uma empresa europeia de retalho. Esta empresa de retalho é constituída por uma cadeia de hipermercados e duas cadeias de supermercados (grandes e pequenas lojas). Os formatos das lojas diferem essencialmente na diversidade e preço dos produtos oferecidos, na área de venda e no tamanho da cidade onde se localizam.

O estabelecimento de relações de fidelidade com os clientes tornou-se um elemento essencial da estratégia desta empresa. O desenvolvimento dos sistemas de informação da empresa e a implementação de um programa de fidelização têm permitido à empresa recolher os dados relativos ao perfil de cada cliente (ex: nome do cliente, morada, data de nascimento, sexo, número de pessoas no agregado familiar, número de telefone e número de um documento identificativo) e às suas transações (data, hora, loja, produtos e preços). Este programa é suportado essencialmente por um cartão de fidelização, e atualmente cerca de 80% do número de transações são registradas usando o cartão de fidelização.

Atualmente, os clientes da empresa são segmentados de duas formas. Uma delas consiste no agrupamento de clientes com base nos seus hábitos de consumo. Este modelo de segmentação é uma versão simplificada do modelo RFM, e é denominado internamente como modelo de Frequência e Valor Monetário (FM). De acordo com os valores destas duas variáveis, a empresa especifica 8 grupos de clientes. Cada cliente integra um destes grupos, com base no número médio de compras efetuadas num período de 8 semanas e no valor médio gasto por compra. As alterações na percentagem de clientes pertencentes a cada grupo são usadas para sinalizar uma eventual necessidade de promover ações para melhorar o relacionamento com o cliente. Por exemplo, se o número de clientes nos grupos com mais visitas às lojas sofrer uma redução, a empresa é alertada para lançar campanhas de marketing, a fim

de motivar os clientes a ir às lojas com mais frequência. O outro método de segmentação baseia-se nas necessidades e preferências dos clientes. Neste caso, os clientes são agrupados em 7 segmentos de acordo com o balanço entre as categorias de produtos que compram. Para isto, calculam-se as percentagens de produtos que cada cliente compra pertencentes a cada categoria de produtos predefinida pela empresa. Os clientes que apresentam percentagens semelhantes são agrupados, usando um algoritmo de agrupamento.

Quanto à estratégia promocional da empresa, existem essencialmente 3 tipos de políticas:

- Descontos em produtos específicos anunciados nas prateleiras das lojas e folhetos, que podem ser usados por todos os clientes com um cartão de fidelização;
- Descontos nas compras feitas em determinados dias (desconto percentual ou absoluto sobre o valor total das compras). Estes descontos podem ser usufruídos por clientes que apresentem o cupão de desconto enviado pelo correio no ponto de venda (POS);
- Descontos em produtos específicos em dias selecionados. Estes podem ser enviados pelo correio ou emitidos no POS.

Os dois primeiros tipos de promoções não fazem distinção entre os clientes de diferentes segmentos. O terceiro tipo, em vez de utilizar os modelos de segmentação descritos anteriormente, utiliza um modelo baseado no histórico das compras do produto incluído na promoção. Os descontos só são emitidos para os compradores mais frequentes do produto, ou para aqueles clientes que normalmente não compram o produto, para os incentivar a tornarem-se compradores.

A análise descrita neste capítulo baseia-se nos dados transacionais de clientes com o cartão de fidelização e é independente dos métodos usados atualmente pela empresa. A metodologia proposta apoia-se essencialmente em técnicas de *data mining*, ao contrário da metodologia usada atualmente pela empresa. A base de dados utilizada inclui os registos do último trimestre de 2009. Cada transação inclui o número de identidade do cliente, a data e a hora da transação, o produto transacionado e o preço do produto. Além das informações transacionais, a empresa forneceu informações demográficas sobre

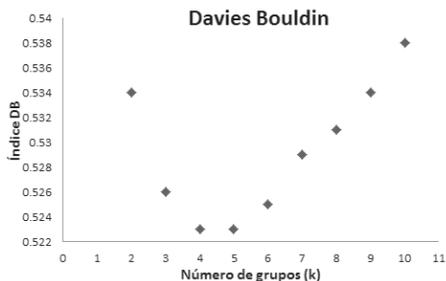
cada cliente: código postal da residência, cidade, data de nascimento, sexo e número de pessoas no agregado familiar. A preparação da base de dados para a análise exploratória envolveu a integração dos dados provenientes de diferentes fontes, a eliminação dos extremos e a seleção dos dados relevantes para a análise. Os clientes cujo valor médio gasto por compra ou cujo número médio de compras por mês estão fora do intervalo definido pela média mais três desvios padrão foram excluídos da análise. Estes extremos representavam 1,96% dos clientes incluídos na base de dados original. Dado que se está interessado no desenho de promoções para famílias, selecionou-se para a análise apenas os clientes cuja média do valor gasto por compra era igual ou inferior a €500. Estes clientes representavam 97,28% dos clientes incluídos na base de dados original. Habitualmente as compras superiores a este valor são feitas por pequenos retalhistas que revendem os produtos em lojas concorrentes, pelo que não devem ser alvo de programas promocionais. Após o processo de preparação dos dados a analisar, a base de dados passou a conter 2.142.439 clientes.

4.2 Segmentação

Neste estudo, a segmentação de clientes baseou-se na frequência e valor monetário das compras dos clientes. Estes indicadores representam os hábitos de compra dos clientes. À semelhança de Marcus (1998), definiram-se como variáveis exploratórias de agrupamento as seguintes variáveis: o número médio de compras feitas por mês e o valor médio gasto por compra. Embora a literatura sugira o uso das variáveis RFM, neste estudo não se incluiu a variável referente à data da última compra, já que se considerou que o período de análise não era suficientemente longo para permitir a diferenciação de clientes nesta dimensão.

A segmentação dos clientes por meio do algoritmo k-médias requer a definição à priori do número de grupos (k). A fim de se definir o número de segmentos, foi analisada a curva do cotovelo e calculado o índice de Davies-Bouldin para diferentes valores de k, como representado na Figura 1. De acordo com o índice de Davies-Bouldin, o número mais adequado de segmentos seria quatro ou cinco, uma vez que estes correspondem ao menor

valor do índice. A partir da curva do cotovelo, podemos concluir que cinco grupos parece ser a opção mais adequada. Por este motivo, os clientes analisados foram agrupados em cinco grupos.



(a) Índice de Davies-Bouldin.



(b) Curva do erro quadrático médio em torno dos centroides.

Figura 1 - Medidas de erro para diferentes valores de k.

A percentagem de clientes incluídos em cada grupo especificado pelo algoritmo k-médias foi a seguinte: 37% no *Cluster 4*, 27 % no *Cluster 2*, 20 % no *Cluster 3*, 8 % no *Cluster 0* e 8% no *Cluster 1*.

A fim de caracterizar o perfil de clientes pertencentes a cada segmento, extraíram-se as regras subjacentes a esta classificação utilizando uma árvore de decisão. Os perfis dos segmentos resultantes da árvore de decisão são ilustrados na Figura 2, e podem ser descritos como se segue. O *Cluster 0* inclui os clientes que fazem compras mais do que 6,2 vezes por mês. O *Cluster 3* corresponde aos clientes que vão às compras entre 3,2 e 6,2 vezes por mês. O *Cluster 1* inclui os clientes fazem compras menos do que 3,2 vezes por mês e gastam mais do que €135,9 por visita. O *Cluster 2* inclui os clientes que fazem compras entre 1,5 e 3,2 vezes por mês e gastam menos do que €135,9 por visita. O *Cluster 4* corresponde aos clientes que fazem compras menos do que 1,5 vezes por mês e gastam menos de €135,9 por visita. Importa salientar que a percentagem de clientes classificados corretamente com recurso à árvore de decisão é de 97.1%, o que significa que as referidas regras podem ser um bom suporte à segmentação de novos clientes.

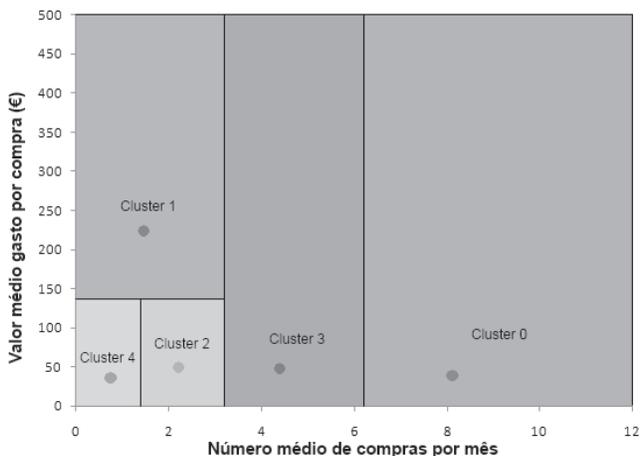


Figura 2 - Caracterização dos grupos.

4.3 Desenho de promoções diferenciadas

A análise dos cabazes de compra, destinada à identificação de associações de produtos dentro dos grupos, pode apoiar o desenho de promoções diferenciadas. Segundo Morales (2005) e Palmatier et al. (2009), estas promoções podem promover a fidelização dos clientes pois fazem com que os clientes sintam benefícios em se manter fiéis à empresa e permitem criar um sentimento de gratidão para com a empresa. As promoções diferenciadas visam premiar o relacionamento do cliente com a empresa e sugerir a futura aquisição de produtos que podem suscitar o interesse do cliente.

Para o propósito desta análise, um cabaz é o conjunto de todos os produtos que foram comprados por um cliente nos três meses analisados. Note-se que este estudo não se foca na análise dos produtos que foram comprados na mesma transação. A análise dos cabazes de compra é feita ao nível da subcategoria do produto, e não ao nível do produto, já que se pretende descobrir o tipo de produtos que pode ser potencialmente interessante para os clientes de um determinado segmento.

Vamos supor que, para fins ilustrativos, a empresa está interessada em promover ações de fidelização dirigidas aos clientes que apresentam uma baixa frequência de visitas à loja e baixo valor das compras. Estes clientes

provavelmente fazem a maior parte das suas compras em lojas da concorrência, de modo que a empresa pode estar interessada em motivá-los a visitar as lojas com maior frequência e ampliar a diversidade de produtos adquiridos. O segmento alvo que se enquadra neste perfil é o *Cluster 4*. Assim, procedeu-se à aplicação do algoritmo de associação considerando os cabazes de compra deste grupo (785.679 cabazes).

As subcategorias foram consideradas associadas quando se verificaram as seguintes condições: um *lift* maior que 1, uma confiança maior ou igual a 50% e um suporte maior ou igual a 2%. Isto significa que tem de haver pelo menos 15.713 cabazes que incluam as subcategorias consideradas associadas. Considerando estas condições, identificaram-se 29 regras de associação de subcategorias.

A maior parte das regras de associação obtidas incluem produtos pertencentes à mesma categoria, tais como peras e maçãs. Um grande número de regras incluem os produtos comprados mais regularmente, como por exemplo o arroz e o leite, como poderia ser antecipado. Para além destas, foram também identificadas algumas relações entre produtos pertencentes a diferentes categorias. A título de exemplo, a Tabela 1 apresenta algumas das regras de associação de subcategorias identificadas, ordenadas de acordo com a confiança.

Tabela 1: Regras de associação para o Cluster 4

Antecedente (x)	Consequente (y)	Conf.	Lift	s(x,y)	s(x)	s(y)
Condicionador de cabelo	Champô	64%	5,50	3%	5%	12%
Tomates	Vegetais para a salada	60%	4,55	5%	9%	14%
Couve	Vegetais para sopa	58%	3,68	6%	10%	16%
Arroz	Massa	58%	3,08	9%	16%	19%
Fiambre em fatias	Queijo flamengo	57%	3,89	7%	12%	15%
Sal	Arroz	53%	3,40	4%	8%	16%
Peras	Maçãs	51%	3,68	6%	12%	14%
Carne picada	Miudezas de porco	51%	3,52	4%	8%	15%
Óleo	Arroz	51%	3,26	6%	12%	16%
Vegetais embalados	Vegetais para a sopa	51%	3,22	3%	6%	16%

Considere-se que a empresa quer motivar os clientes para a compra de produtos que podem ser interessantes para estes, embora não tenham vindo a fazer parte das suas listas de compras recentes. Para isto, a empresa pode emitir um cupão de desconto no balcão de pagamento que vise o produto

consequente da regra de associação que não tem sido comprado, apesar do produto antecedente ter vindo a ser comprado. Por exemplo, a empresa pode sugerir um desconto na próxima compra de champô aos clientes que têm comprado condicionador de cabelo, mas que não têm comprado champô. Através da análise da base de dados da empresa concluiu-se que esta ação promocional pode ser relevante e motivar cerca de 15 mil clientes a comprar champô, dado que recentemente apenas compraram condicionador. Este tipo de promoções não só permite motivar os clientes a visitar mais vezes as lojas, mas também a aumentar a diversidade de produtos comprados. Para além disto, este tipo de ações tendem a fidelizar o cliente, na medida em que este se pode sentir parte integrante da estratégia da empresa.

A fim de verificar se as ações promoções seriam diferentes se a empresa definisse como alvo outro grupo de clientes, apresenta-se no Apêndice algumas das regras de associação de produtos que resultam da análise dos cabazes de compra dos clientes dos restantes segmentos. A primeira regra apresentada para cada grupo de clientes é aquela que apresenta maior confiança. Impondo critérios semelhantes aos definidos para o *Cluster 4*, *lift* (>1), confiança ($\geq 50\%$) e suporte ($\geq 2\%$), o número total de regras identificadas foi de 18.866 para o *Cluster 0*, 4.911 para o *Cluster 1*, 1.761 para o *Cluster 2* e 10.408 para o *Cluster 3*. É interessante verificar que, para os pequenos segmentos, tais como o *Cluster 0*, utilizando os mesmos critérios, é possível obter mais regras de associação de subcategorias do que para os segmentos maiores como o *Cluster 4*. Dado que os clientes do *Cluster 4* fazem compras esporádicas, é mais difícil encontrar padrões de compra. Embora algumas das regras identificadas para os diferentes grupos de clientes sejam idênticas, tais como o facto de a compra de condicionador de cabelo despoletar a compra de champô, pode verificar-se pela análise da Tabela 1 que os clientes têm hábitos de compra diferentes. A regra com maior confiança identificada para cada grupo de clientes é diferente para todos os grupos. Portanto, acredita-se que o procedimento sugerido neste estudo, que consiste no desenvolvimento da segmentação antes do processo de descoberta das regras de associação, pode contribuir para a melhoria da relação entre a empresa e os seus clientes.

5 Conclusão

Este estudo propõe um modelo para a segmentação dos clientes de uma empresa europeia de retalho e propõe políticas promocionais dirigidas aos clientes de cada segmento, com o objetivo de reforçar as suas relações de fidelização.

A aplicação de técnicas de *data mining* permitiu encontrar grupos naturais de clientes, com base nos dados transacionais armazenados na base de dados relativa ao cartão de fidelização da empresa. A segmentação baseou-se na frequência (número médio de compras feitas por mês) e no valor monetário (valor médio gasto por compra) das compras dos clientes. Usando uma técnica particional de agrupamento, os clientes foram agrupados em cinco grupos de acordo com os seus hábitos de compras. A análise também envolveu a construção de uma árvore de decisão, a fim de extrair as regras subjacentes à segmentação dos clientes. Assim, foi possível definir um perfil dos clientes de cada segmento, que pode ser usado para classificações futuras de clientes com elevada precisão.

O estudo descrito também identificou regras de associação de produtos dentro de cada segmento, tendo em conta os cabazes de compra dos clientes. Estas regras permitiram o desenho de promoções diferenciadas, que podem ser fundamentais para motivar os clientes a aumentar as suas compras e a manter-se fieis à empresa.

Como trabalho futuro espera-se que seja feita a implementação da metodologia proposta na empresa usada como caso de estudo. Após a implementação, seria importante entrevistar clientes pertencentes a cada grupo, a fim de verificar se estão satisfeitos com as promoções que lhes são dirigidas. Seria também interessante verificar se o comportamento dos clientes revela uma intensificação da sua relação com a empresa, como resultado das acções promocionais diferenciadas baseadas na metodologia proposta neste trabalho. Deverão ser considerados indicadores de fidelização, como por exemplo a taxa de aumento do consumo e a relação entre os descontos oferecidos e usados.

REFERÊNCIAS

- Agrawal, R. and Srikant, R. (1994). Fast algorithms for mining association rules. pages 487-499.
- Aldenderfer, M. and Blashfield, R. (1984). *Cluster Analysis*. Number 07-044. Sage Publications, NewburyPark, California.
- Bettencourt, L. A. and Gwinner, K. (1996). Customization of the service experience: the role of the frontline employee. *International Journal of Service Industry Management*, 7(2):3-20.
- Blattberg, R., Buesing, T., Peacock, P., and Sen, S. (1978). Identifying deal prone segment. *Journal of Marketing Research (JMR)*, 15(3):369-377.
- Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. (1984). *Classification and Regression Trees*, volume 1. Chapman and Hall/CRC, New York.
- Brijs, T., Swinnen, G., Vanhoof, K., and Wets, G. (2004). Building an association rules framework to improve product assortment decisions. *Data Mining and Knowledge Discovery*, 8(1):7-23.
- Davies, D. and Bouldin, D. (1979). Cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence*, 1(2):224-227.
- Dibb, S. and Simkin, L. (1997). A program for implementing market segmentation. *Journal of Business & Industrial Marketing*, 12(1):51 - 65.
- Forgy, E. W. (1965). Cluster analysis of multivariate data: efficiency vs interpretability of classifications. *Biometrics*, 21:768-769.
- Frank, R. E., Massy, W. F., and Boyd, H. W. (1967). Correlates of grocery product consumption rates. *Journal of Marketing Research (JMR)*, 4(2):184-190.
- Green, P. E. and Wind, Y. (1973). *Multiaattribute Decisions in Marketing*. Dryden Press.
- Haley, R. I. (1968). Benefit segmentation: A decision-oriented research tool. *Journal of Marketing*, 32(3):30-35.
- Han, J., Pei, J., and Yin, Y. (2000). Mining frequent patterns without candidate generation. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, pages 1-12, Dallas, Texas, United States. ACM.
- Helsen, K. and Green, P. E. (1991). A computational study of replicated clustering with an application to market-segmentation. *Decision Sciences*, 22(5):1124-1141.
- Jain, A. K., Murty, M. N., and Flynn, P. J. (1999). Data clustering: A review.
- Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, 29(2):119.
- Keaveney, S. M. (1995). Customer switching behavior in service industries: An exploratory study. *The Journal of Marketing*, 59(2):71-82.
- Kiang, M. Y., Hu, M. Y., and Fisher, D. M. (2006). An extended self-organizing map network for market segmentation—a telecommunication example. *Decision Support Systems*, 42(1):36-47.
- Lazer, W. (1964). Lifestyle concepts and marketing. In *Toward scientific marketing*. American Marketing Association, Chicago.
- Liu, D. and Shih, Y. (2005). Integrating AHP and data mining for product recommendation based on customer lifetime value. *Information & Management*, 42(3):387-400.

- MacQueen, J. B. (1967). Some methods for classification and analysis of MultiVariate observations. In *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297. University of California Press.
- Marcus, C. (1998). A practical yet meaningful approach to customer segmentation. *Journal of Consumer Marketing*, 15(5):494–504.
- Min, S. and Han, I. (2005). Detection of the customer time-variant pattern for improving recommender systems. *Expert Systems with Applications*, 28(2):189–199.
- Mitchell, A. (1983). *The nine American lifestyles: Who we are and where we're going*. Warner, New York.
- Morales, A. C. (2005). Giving firms an “E” for effort: Consumer responses to high-effort firms. *Journal of Consumer Research*, 31(4):806–812.
- Morgan, J. N. and Sonquist, J. (1963). Problems in the analysis of survey data, and a proposal. *Journal of the American Statistical Association*, 58:415–434.
- Ngai, E., Xiu, L., and Chau, D. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, 36(2, Part 2):2592–2602.
- Palmatier, R. W., Jarvis, C. B., Bechhoff, J. R., and Kardes, F. R. (2009). The role of customer gratitude in relationship marketing. *Journal of Marketing*, 73(5):1–18.
- Piercy, N. and Morgan, N. (1993). Strategic and operational market segmentation: a managerial analysis. *Journal of Strategic Marketing*, 1:123–140.
- Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1):81–106.
- Quinlan, J. R. (1992). *C4.5: programs for machine learning*. Morgan Kaufmann Publishers Inc.
- Rapid-I (2009). *RapidMiner 4.6 User Guide*. Rapid-I GmbH, Germany.
- Roel, R. (1988). Direct marketing's 50 big ideas. *Direct Marketing*, 50:45–52.
- Tibshirani, R., Walther, G., and Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 63:411–423.
- Twedt, D. W. (1964). How important to marketing strategy is the “Heavy user”? *The Journal of Marketing*, 28(1):71–72.
- Van den Poel, D., Schamphelaere, J. D., and Wets, G. (2004). Direct and indirect effects of retail promotions on sales and profits in the do-it-yourself market. *Expert Systems with Applications*, 27(1):53–62.
- Zaki, M. J., Parthasarathy, S., Ogihara, M., and Li, W. (1997). New algorithms for fast discovery of association rules. Technical report, University of Rochester.

APÊNDICE

Tabela A1: Regras de associação.

<i>Cluster 0</i>						
Antecedente (x)	Consequente (y)	Conf.	<i>Lift</i>	s(x,y)	s(x)	s(y)
Doces sazonais	Chocolates	92%	1,19	26%	28%	73%
Tomates	Vegetais para a salada	86%	1,27	45%	53%	68%
Peras	Maças	84%	1,21	50%	61%	70%
Farinha	Açúcar	81%	1,22	41%	51%	67%
Leites infantis	Fraldas	79%	5,65	2%	3%	14%
Fiambre fatiado	Queijo flamengo	79%	1,26	42%	54%	63%
Maças	Peras	73%	1,21	50%	70%	61%
Papel higiênico	Guardanapos	73%	1,20	43%	60%	61%
Guardanapos	Papel higiênico	72%	1,20	43%	61%	60%
Sumos/refrigerados	Legumes transformados	63%	1,55	2%	3%	40%
<i>Cluster 1</i>						
Antecedente (x)	Consequente (y)	Conf.	<i>Lift</i>	s(x,y)	s(x)	s(y)
Legumes secos	Massas	86%	1,48	6%	7%	58%
Corn Flakes	Leite UHT	83%	1,31	5%	6%	64%
Pão ralado	Especiarias secas	72%	1,94	3%	5%	28%
Refeições pré-preparadas	Vegetais em conserva	72%	1,83	2%	3%	39%
Cebolas	Ovos	64%	1,80	13%	20%	35%
Carne de coelho embalada	Carne de frango embalada	61%	2,59	2%	4%	23%
Barras de cereais	Bolachas doces	59%	1,59	4%	7%	37%
Boião de fruta	Iogurtes infantis	58%	3,23	3%	5%	18%
Insecticida	Acessórios de limpeza	54%	1,79	4%	7%	30%
Pastelaria/padaria	Natas/Chantilly	53%	2,06	2%	4%	25%
<i>Cluster 2</i>						
Antecedente (x)	Consequente (y)	Conf.	<i>Lift</i>	s(x,y)	s(x)	s(y)
Couves	Vegetais para a sopa	74%	1,79	21%	28%	41%
Condicionador de cabelo	Champô	72%	2,40	10%	13%	30%
Tomates	Vegetais para a salada	72%	2,04	18%	24%	35%
Polpas de tomate	Vegetais em conserva	62%	1,72	14%	22%	36%
Frutas tropicais	Bananas	61%	1,50	18%	29%	40%
Rolos de cozinha	Papel higiênico	58%	1,80	12%	21%	32%
Artigos de criança	Bolachas doces	57%	1,80	10%	18%	32%
Cebolas	Batatas	55%	2,25	11%	17%	24%
Farináceos	Açúcar	54%	1,70	14%	25%	36%
Fraldas	Toalhetes bebé	53%	4,76	4%	8%	11%
<i>Cluster 3</i>						
Antecedente (x)	Consequente (y)	Conf.	<i>Lift</i>	s(x,y)	s(x)	s(y)
Achocolatados	Leite UHT	85%	1,15	16%	19%	74%
Tortas	Bolos	78%	1,46	3%	4%	53%
Condicionador de cabelo	Champô	79%	1,76	17%	21%	45%
Vermutes	Cervejas com álcool	62%	1,66	2%	3%	38%
Frutas congeladas	Natas/cantilly	60%	1,51	3%	4%	40%
Fraldas	Iogurtes infantis	58%	2,18	7%	12%	27%
Esparregado	Legumes transformados	55%	1,78	2%	4%	31%
Carne de coelho embalada	Carne de bovino embalada	54%	1,87	3%	5%	29%
Frutos vermelhos	Legumes embalados	53%	1,88	3%	5%	29%
Frango do campo	Legumes aromáticos	53%	1,48	3%	6%	36%

(Página deixada propositadamente em branco)

CASO 16

OTIMIZAÇÃO DA OPERAÇÃO DE REDES DE DISTRIBUIÇÃO: CONSIDERAÇÃO SIMULTÂNEA DE VARIÁVEIS TOPOLÓGICAS, TOMADAS DE TRANSFORMADORES E BATERIAS DE CONDENSADORES UTILIZANDO EPSO E HEURÍSTICAS

Jorge Pereira

INESC Porto, Faculdade de Economia
Universidade do Porto
jpereira@inescporto.pt

Jorge Alves

INESC Porto, Faculdade de Engenharia
Universidade do Porto
pds11007@fe.up.pt

Manuel Matos

INESC Porto, Faculdade de Engenharia
Universidade do Porto
mmatos@inescporto.pt

RESUMO

A identificação de estratégias ótimas de exploração da rede tendo em conta a minimização da potência ativa de perdas assume uma relevância crescente na operação de distribuição de energia elétrica. Um novo método para esta identificação foi desenvolvido através da resolução do problema da determinação do ponto de operação ótimo, tendo como possibilidade a alteração da configuração da topologia de rede conjuntamente com a alteração das posições das tomadas dos transformadores e baterias de condensadores. O método de otimização EPSO modificado foi utilizado para resolver este problema.

PALAVRAS-CHAVE

Reconfiguração topológica, *Voltage Var Control*, OPF, EPSO, DMS

Introdução

A operação das redes de distribuição de energia elétrica tem vindo a assumir uma importância cada vez maior, dado que é neste tipo de redes que se verifica a ocorrência de situações problemáticas tanto ao nível da qualidade de serviço, assim como em aspetos relacionados com a operação ótima do ponto de vista económico. Desta forma, a identificação de estratégias ótimas de exploração da rede tendo em conta a minimização da potência ativa de perdas assume uma relevância crescente. Um novo modelo para esta identificação foi desenvolvido e integrado num pacote de *software* comercial para gestão de redes da EFACEC, que está a ser instalado no novo Centro de Despacho da Região de Ática na Grécia.

O modo como as redes de distribuição são exploradas é fortemente determinado pela configuração da topologia em operação (em particular quando é imposta uma estrutura radial), pelas posições das tomadas nos transformadores e pelos escalões das baterias de condensadores colocados em serviço. Estes aspetos refletem-se sob diversas formas nas condições de exploração das redes e na qualidade de serviço oferecida aos consumidores. O problema relacionado com a determinação da configuração da topologia de rede ótima é geralmente considerado separadamente do problema da determinação das posições ótimas das tomadas dos transformadores e baterias de condensadores. O método de otimização proposto representa uma unificação destes dois problemas, uma vez que são tratados em simultâneo através da implementação de um algoritmo EPSO (*“Evolutionary Particle Swarm Optimization”*) modificado, como é descrito nas secções seguintes.

A existência de produção distribuída de potência reativa, em diversos locais da rede, é um recurso que permite diminuir os trânsitos de potência reativa nos ramos das redes. Os ramos da rede poderão assim ser utilizados de forma mais completa para transmissão de potência ativa, evitando-se ou adiando-se assim novos investimentos. Por outro lado, a identificação da topologia em operação, assim como das tomadas de transformadores e dos escalões de baterias de condensadores, pode ser realizada especificando um conjunto alargado de restrições que deverão ser verificadas. Nestas restrições poderão ser consideradas: restrições associadas aos limites de intensidade de corrente em

ramos; e restrições impondo gamas admissíveis para o módulo da tensão nos nós. A consideração das restrições para o módulo das tensões pode ser interpretada como a possibilidade de se permitir a especificação de variações máximas admissíveis para a tensão e, desta forma, inclui-se a especificação de quedas de tensão máximas admissíveis. Assim, há uma contribuição para garantir uma qualidade de serviço acrescida.

No caso de existirem geradores ligados à rede de distribuição (e não apenas ligações a redes de níveis de tensão mais elevados), decisões de despacho dessas máquinas podem também influenciar o resultado económico da exploração. A otimização global pode ser feita tendo em consideração um ou mais cenários de carga, podendo ainda incluir-se a otimização da transição entre cenários, com minimização do critério do número de manobras.

Se no problema de otimização for considerado que se podem ter avarias que conduzam a reconfigurações de emergência para restabelecimento de serviço, pode ainda ser acrescentado na formulação do problema um critério de minimização de potência cortada (mesmo sem reconfiguração) ou de potência não recuperada depois da reconfiguração. Esta hipótese é especialmente interessante se existirem cargas com diferentes graus de prioridade, correspondendo de certo modo às restrições de segurança do OPF (*Optimal Power Flow* - Trânsito de Potências Ótimo) para um EMS (*Energy Management System* – Sistema de Gestão das Redes de Transmissão). A reconfiguração para restabelecimento de serviço corresponde a um módulo de cálculo elétrico que é oferecida dentro de um DMS (*Distribution Management System* – Sistema de Gestão das Redes de Distribuição).

Nas próximas secções descreve-se a resolução do problema da determinação do ponto de operação ótimo, tendo como possibilidade a alteração da configuração da topologia de rede ótima conjuntamente com a alteração das posições das tomadas dos transformadores e baterias de condensadores. Assim, nas secções seguintes descrevem-se mais em detalhe os principais problemas existentes na otimização da operação de redes de distribuição, desenvolvendo-se uma formulação matemática para o problema, e dando-se saliência à incorporação de técnicas de Investigação Operacional, em particular algoritmos de otimização. Na secção 2 é descrito o problema da

reconfiguração topológica, na secção 3 é apresentado o problema de controlo de tensão e potência reativa e finalmente na secção 4 é apresentada a metodologia proposta para a resolução conjunta dos dois problemas anteriores. Finalmente, na secção 5, são apresentados os resultados obtidos com a metodologia proposta para uma rede de teste.

Reconfiguração Topológica

Reconfiguração de redes é um problema de otimização em que se pretende otimizar o ponto de operação da rede em termos da minimização das perdas técnicas ou da melhoria dos perfis de tensão, dando como resultado sugestões de alterações na topologia da rede (também designada por configuração da rede).

As redes elétricas de distribuição são normalmente exploradas radialmente, mas têm uma estrutura que inclui malhas ou anéis (com alguns ramos normalmente abertos), que permite diferentes configurações de operação e aumenta a fiabilidade do sistema. A reconfiguração destas redes, através da alteração de estado dos aparelhos de corte, permite otimizar o funcionamento em face de um regime de carga, minimizando as perdas de energia e garantindo a satisfação das restrições técnicas (limites de carga dos ramos e queda de tensão nos nós).

Diversos autores dedicaram-se a estudar este problema, desenvolvendo e implementando diversas abordagens. Em 1975, MERLIN, A. e BACK, H. propuseram um método de otimização que utiliza uma técnica denominada por *branch-and-bound* para encontrar uma nova configuração admissível. CINVALAR, S. et al. (1988) por seu lado apresentam uma heurística baseada na estimação eficiente da redução das perdas ativas na rede, pela transferência de carga entre duas saídas de uma subestação de distribuição.

Mais recentemente têm sido aplicadas técnicas de otimização baseadas em meta-heurísticas, como algoritmos genéticos GUPTA, N. et al. (2010), *simulated annealing* EL-HAWARY, M. (1998), e métodos baseados em populações de soluções como o *Particle Swarm Optimization* GUPTA, N., SWARNKAR, A., NIAZI, K.R. (2011).

Em termos de reconfiguração topológica, para além do problema de otimização da operação, há ainda um problema adicional relacionado com o restauro de serviço na sequência de uma avaria, onde a reconfiguração visa permitir realimentar pelo menos parte dos consumidores que ficaram privados de eletricidade, através de percursos alternativos. Em ambos os problemas descritos, a minimização do número de manobras (ações de abertura/fecho de aparelhos de corte) pode ser também um critério adicional a considerar.

Um algoritmo desenvolvido e integrado num DMS, por MATOS, M.A., MELO, P. (1999), para lidar com estes problemas de reconfiguração baseia-se numa heurística fundada num processo de substituição de ramos (*branch exchange*) onde a seleção de ramos a fechar e abrir é realizada a partir de valores locais de tensão nos nós e correntes nos ramos adjacentes. Este algoritmo utilizou uma meta-heurística de otimização, o *simulated annealing*, sendo as vizinhanças definidas pela abertura/fecho de ramos (de modo a manter a rede radial).

Na referência KHODR, H.M. et al. (2009), o objetivo e as restrições deste problema foram incluídos no problema de controlo de tensão e potência reativa (descrito na secção seguinte), tendo sido assim reunidos num problema único, e que foi designado por DOPF (*Distribution Optimal Power Flow*). Nesta referência, foi feita uma pesquisa bibliográfica detalhada sobre o problema de reconfiguração topológica, e concluiu-se que o problema de reconfiguração ótima, não foi apresentado ou estudado ao mesmo tempo e no mesmo modelo que o problema controlo de tensão e potência reativa. Uma abordagem a este problema integrado é descrita na secção 4.

Controlo de tensão e potência reativa

A otimização do funcionamento de sistemas elétricos de energia também está relacionada com a posição das tomadas de baterias de condensadores e a posição de tomadas de transformadores com regulação em carga. Com a alteração da quantidade de potência reativa injetada pela bateria de condensadores pode-se fazer subir ou baixar a tensão numa determinada zona da rede onde a bateria está localizada. Com a alteração da razão de transformação de um transformador pode-se fazer subir ou baixar a tensão na

zona da rede ligada ao secundário do transformador. No caso de existirem geradores síncronos ligados à rede, que permitem controlar a tensão dos nós a que estão ligados, também há influência no resultado económico da exploração por causa da tensão em que é colocado o respetivo barramento.

Várias abordagens a este problema têm sido sugeridas ao longo dos anos por diferentes autores. Em LIU, Y., ZHANG, P., QIU, X. (2002), foi utilizada programação dinâmica em conjunto com um sistema de lógica difusa para controlar a tensão à saída da subestação. Em HSU, Y. (1998), foi considerado o problema do controlo de tensão multitemporal, para as 24 horas seguintes. Foi utilizada uma rede neuronal e programação dinâmica difusa para resolver este problema multitemporal, tendo por base heurísticas definidas com base na experiência passada dos operadores de rede.

Outros trabalhos sugerem a ideia de utilizar plataformas de inferência difusa para corrigir desvios de tensão e controlo de potência reativa JUNGES, M. (2000) e introduzem conceitos baseados em lógica difusa para o controlo de sistemas e subsistemas de energia elétrica EKEL, P. et al. (1998). Exemplos de aplicação de controladores difusos em problemas de estabilidade podem ser encontrados em LIU, C.W., CHANG, C.S., JIANG, J.A. (2001). LIANG, R. (2003) apresenta uma metodologia baseada em *simulated annealing* que, em conjunto com uma estratégia difusa de controlo de potência reativa, é aplicada ao problema de determinar o escalonamento ótimo das tomadas de baterias de condensadores e transformadores, para um sistema de distribuição.

A utilização de conceitos baseados em lógica difusa para o controlo de sistemas de energia elétrica surge como alternativa ou complemento ao uso de métodos como o *simulated annealing* PEREIRA, J., SARAIVA, J.T., PONCE DE LEÃO, M.T. (1999), algoritmos evolucionários MIRANDA, V., FONSECA, N. (2002) ou outras meta-heurísticas MIRANDA, V., CALISTO, P. (2002) cujo principal objetivo é o de derivar ações de controlo para controlar tensões e reduzir perdas, visto que apresenta uma eficiência muito mais elevada em termos de esforço computacional.

A resolução deste problema é um dos módulos integrantes de um DMS, cujo objetivo é controlar tensões em redes que possuam diversos tipos de equipamentos que permitam a regulação em carga. Este módulo é

habitualmente designado por *Voltage Var Control* - Controlo de tensão e potência reativa.

A função objetivo deste problema de otimização representa as perdas técnicas da rede (1), e resolve-se como um problema de minimização. As restrições do problema estão relacionadas com: a satisfação das equações do problema do balanço de potências em cada nó (2) – primeira lei de Kirchhoff – e respetivo cálculo dos trânsitos de potência nas linhas – segunda lei de Kirchhoff -; a manutenção das tensões em todos os nós da rede dentro de uma banda admissível (3); e a garantia que o trânsito de potência em todas as linhas da rede não exceda os seus limites térmicos (4). Este problema é um problema de otimização não linear misto, com as variáveis inteiras representadas em (5) e (6).

$$\text{Min } Z = \sum_{ij} -G_{ij}(V_i^2 + V_j^2 - 2V_iV_j \cos(\theta_i - \theta_j)) \quad (1)$$

sujeito a:

$$g(V_i, \theta_i, t_f, e_c) = 0 \quad (2)$$

$$V^{\min} \leq V_i \leq V^{\max} \quad \text{para cada nó} \quad (3)$$

$$|I_{ij}(V_i, \theta_i, V_j, \theta_j)| \leq |I_{ij}(V_i, \theta_i, V_j, \theta_j)|^{\max} \quad \text{para cada ramo } ij \quad (4)$$

$$t_f \in \{t_f^1, t_f^2, \dots, t_f^{k_f}\} \quad \text{para cada transformador } f \text{ com } k_f \text{ tomadas} \quad (5)$$

$$e_c \in \{e_c^1, e_c^2, \dots, e_c^{k_c}\} \quad (6)$$

para cada bateria de condensadores c com k_c secções

$$i, j \in \{1, 2, \dots, nb\} \quad (7)$$

Nesta formulação:

- $g()$ representa as equações do problema de trânsito de potências;
- t_f representa um dos valores disponíveis de tomadas de transformadores;
- e_c representa um dos valores disponíveis de secções de baterias de condensadores;
- V_i e θ_i representam o módulo e a fase da tensão em cada nó da rede;
- G_{ij} corresponde à parte real da matriz das admitâncias nodais (característica física dos ramos da rede);

- $|I_{ij}(V_i, \theta_i, V_j, \theta_j)|$ representa o módulo da intensidade de corrente no ramo que liga o nó i ao nó j ;
- nb representa o número de barramentos da rede;

Este problema pode ser resolvido através de um sistema de controlo de tensões como o que é apresentado em MIRANDA, V., CALISTO, P. (2002) e MIRANDA, V., MOREIRA, A., PEREIRA, J. (2007), sendo um procedimento eficiente em termos de tempo de execução necessário para encontrar uma solução viável. Este controlador demonstra grande superioridade face a outros tipos de controlo que utilizam algoritmos computacionalmente mais pesados. É frequente que estes tipos de algoritmos de controlo possam necessitar de centenas a milhares de iterações, enquanto o controlador desenvolvido necessita apenas de um reduzido número de iterações para atingir um resultado viável.

MIRANDA, V., MOREIRA, A., PEREIRA, J. (2007) descrevem uma arquitetura para o módulo *Voltage Var Control* que tem por base uma cascata de controladores difusos do tipo Mamdani, que constituem um sistema de inferência baseado em lógica difusa (ver Figura 1), e uma rotina de trânsitos de potência que avalia, periodicamente, o resultado de todas as ações propostas. As regras difusas apresentadas estão relacionadas com a eficiência de uma ação de controlo, bem como com a posição atual de cada um dos diferentes equipamentos utilizados no controlo.

Além do principal objetivo de manter as tensões em todos os nós da rede dentro de uma banda admissível, é da responsabilidade do controlador difuso garantir que o trânsito de potências em todas as linhas da rede não exceda os limites térmicos. Este controlador difuso será utilizado no final do processo de otimização apresentado na secção seguinte para dar mais garantias ao algoritmo da solução encontrada ser mais próxima da ótima.

Reconfiguração com controlo de tensão e potência reativa

O algoritmo para resolver o problema do DOPF deverá ser capaz de encontrar a configuração ótima de operação de um sistema de distribuição de energia elétrica, para um cenário de operação, avaliando a carga distribuída

pelas subestações e pelas saídas destas. A solução final deverá ser uma solução admissível, onde todas as cargas deverão estar alimentadas e os limites técnicos dos componentes do sistema deverão ser cumpridos. Este é um problema com um grande número de soluções alternativas, sendo que o algoritmo deverá escolher aquela que otimiza a função objetivo escolhida.

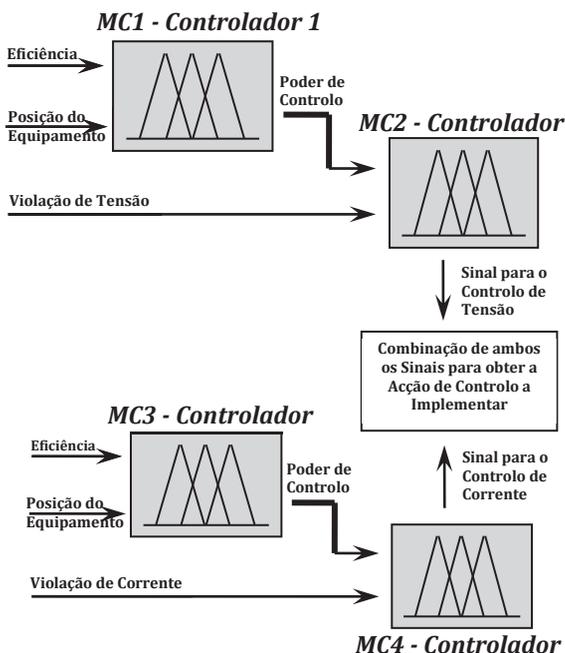


Figura 1: Esquema do sistema de inferência baseado em lógica difusa para controlo de tensão e potência reativa

Na secção anterior apresentou-se a formulação do problema de controlo de tensão e potência reativa, onde a função objetivo é a minimização das perdas na rede (1), com restrições relativas ao balanço de potência nos nós – primeira lei de Kirchhoff -; cálculo de tensões nos nós da rede – segunda lei de Kirchhoff -; limites de trânsito de potência nas linhas; limites de máximo e mínimo de tensão; e limites de tomadas de transformadores e baterias de condensadores associados a variáveis inteiras.

A formulação completa do problema integrado de reconfiguração com controlo de tensão e potência reativa, com a função objetivo e as restrições, é apresentada nos seguintes parágrafos.

Função objetivo de minimização das perdas técnicas da rede:

$$\text{Min } P_{\text{perdas}}^{\text{total}} = \sum_{ij} -G_{ij}(y_{ij})(V_i^2 + V_j^2 - 2V_iV_j \cos(\theta_i - \theta_j)) \quad (8)$$

Sujeito a:

Limites técnicos dos trânsitos de potência nos ramos (linhas e transformadores):

$$S_{ij}^{\text{min.}} \leq S_{ij}(V_i, \theta_i, V_j, \theta_j) \leq S_{ij}^{\text{max.}} \quad (9)$$

Limites técnicos de tensão nos nós da rede:

$$V_i^{\text{min.}} \leq V_i \leq V_i^{\text{max.}} \quad (10)$$

Condição de radialidade, que consiste em verificar que o número de linhas fora de serviço no início deve ser igual no final do processo, com todos os consumidores abastecidos:

$$\sum_{ij}(1 - y_{ij}) = N_{\text{abertos}}^{\text{inic.}} \quad (11)$$

$$S_L^{\text{total}} = S_L^{\text{inicial}} \quad (12)$$

$$\text{real}(S_L^{\text{total}}) = \sum_i P_i^L \quad (13)$$

$$\text{imag}(S_L^{\text{total}}) = \sum_i Q_i^L \quad (14)$$

Limites máximos e mínimos de potência produzida nos geradores:

$$S_{Gi}^{\text{min.}} \leq S_{Gi} \leq S_{Gi}^{\text{max.}} \quad (15)$$

Limites de tomadas nos transformadores:

$$t_f^1 \leq t_f \leq t_f^{\text{kf}} \quad (16)$$

Limites de tomadas nas baterias de condensadores:

$$e_c^1 \leq e_c \leq e_c^{\text{kc}} \quad (17)$$

Balço de potência na rede, a produção tem de ser igual ao consumo:

$$S_L^{\text{total}} + S_{\text{perdas}}^{\text{total}} = S_G^{\text{total}} \quad (18)$$

$$\text{real}(S_G^{\text{total}}) = \sum_i P_i^G \quad (19)$$

$$\text{imag}(S_G^{\text{total}}) = \sum_i Q_i^G \quad (20)$$

$$\text{real}(S_{\text{perdas}}^{\text{total}}) = P_{\text{perdas}}^{\text{total}} \quad (21)$$

$$\text{imag}(S_{\text{perdas}}^{\text{total}}) = \sum_{ij} B_{ij}(y_{ij})(V_i^2 + V_j^2 - 2V_i V_j \cos(\theta_i - \theta_j))$$

Equações do trânsito de potência ativa nos ramos e balanço nulo num nó:

$$P_i^G - P_i^L = V_i \sum_j V_j (G_{ij}(y_{ij}) \cos(\theta_i - \theta_j) + B_{ij}(y_{ij}) \sin(\theta_i - \theta_j)) \quad (22)$$

Equações do trânsito de potência reativa nos ramos e balanço nulo num nó:

$$Q_i^G - Q_i^L = V_i \sum_j V_j (G_{ij}(y_{ij}) \sin(\theta_i - \theta_j) - B_{ij}(y_{ij}) \cos(\theta_i - \theta_j)) \quad (23)$$

$$G_{ij}(y_{ij}) = \begin{cases} G_{ij} & , \text{ se } y_{ij} = 1 \\ 0 & , \text{ se } y_{ij} = 0 \end{cases} \quad (24)$$

$$B_{ij}(y_{ij}) = \begin{cases} B_{ij} & , \text{ se } y_{ij} = 1 \\ 0 & , \text{ se } y_{ij} = 0 \end{cases} \quad (25)$$

As variáveis de decisão são:

- t_f – variável inteira de decisão associada à tomada do transformador f
- e_c – variável inteira de decisão associada à tomada da bateria de condensadores c
- y_{ij} – variável binária de decisão relativa à configuração topológica do ramo ij (1 – ramo ij em serviço, 0 – ramo ij fora de serviço)

Nesta formulação:

- S_G^{total} – total de potência aparente gerada
- S_L^{total} – total de potência aparente consumida
- $S_{ij}^{\text{min.}}$ – mínimo de potência aparente no ramo ij
- $S_{ij}^{\text{max.}}$ – máximo de potência aparente no ramo ij
- $V_i^{\text{min.}}$ – valor mínimo para o módulo da tensão no nó i
- $V_i^{\text{max.}}$ – valor máximo para o módulo da tensão no nó i
- $N_{\text{abertos}}^{\text{inic.}}$ – número de linhas fora de serviço na configuração inicial
- $S_{Gi}^{\text{min.}}$ – potência aparente mínima no gerador i

- S_{Gi}^{\max} – potência aparente máxima no gerador i
- t_f^1 e t_f^{kf} – tomada 1 (menor valor das tomadas) e tomada kf (maior valor das tomadas) do transformador f com kf tomadas
- e_c^1 e e_c^{kc} – tomada 1 (menor valor das tomadas) e tomada kc (maior valor das tomadas) da bateria de condensadores c com kc secções
- V_i – módulo da tensão no nó i
- θ_i – fase da tensão no nó i
- G_{ij} – elemento da parte real da matriz das admitâncias nodais
- B_{ij} – elemento da parte imaginária da matriz das admitâncias nodais
- P_i^G – potência ativa gerada no nó i
- Q_i^G – potência reativa gerada no nó i
- P_i^L – potência ativa consumida no nó i
- Q_i^L – potência reativa consumida no nó i
- S_L^{inicial} – valor da potência aparente consumida, no estado inicial da rede

A principal diferença entre a função objetivo (1) e (8) reside na variável de decisão relacionada com a topologia, y_{ij} , que terá apenas dois valores possíveis: 0, se o ramo ij está fora de serviço; e 1, se o ramo ij está em serviço. Desta forma, a parte real da matriz das admitâncias nodais, G_{ij} , será igual a zero no caso do ramo estar fora de serviço e sendo, por isso, eliminada do cálculo das perdas globais. Por outro lado, se o ramo estiver em serviço, ele será considerado para o cálculo das perdas, pois nesta situação G_{ij} tem o valor característico do ramo ij .

O método de resolução do problema proposto admite a possibilidade de existência de anéis fechados na rede, por exemplo entre duas saídas de uma mesma subestação, ou entre duas subestações diferentes. Caso existam estes anéis fechados na rede, o algoritmo trata-os como zonas não otimizáveis, garantindo a operação radial nas restantes zonas da rede, sujeitas a otimização.

O algoritmo implementado para resolver este problema parte de uma configuração topológica inicial, e a partir desta faz uma pesquisa para encontrar a configuração ótima, gerando um conjunto de configurações alternativas que se movem no espaço de soluções por meio de um operador. Todas as configurações que não pertençam ao espaço admissível de soluções

por violação de limites técnicos são penalizadas de forma a não serem escolhidas.

Foi utilizado o algoritmo EPSO MIRANDA, V., FONSECA, N. (2002) para resolver este problema, conjuntamente com algumas heurísticas locais que têm em conta o conhecimento que existe sobre as relações existentes entre as variáveis de decisão e os seus efeitos nas restrições. Estas heurísticas têm a função de garantir a admissibilidade topológica das soluções encontradas e ajustes de variáveis relacionadas com o controlo de tensão na rede, fazendo com que configurações, que à partida poderiam ser penalizadas por não satisfação dos critérios de tensão ou dos limites nos ramos, possam ser consideradas como admissíveis por alteração de algumas variáveis de controlo relativas ao módulo da tensão.

O algoritmo EPSO foi proposto por Miranda e Fonseca em 2002 como um novo algoritmo de otimização que explora as fraquezas do algoritmo PSO (*Particle Swarm Optimization*) já existente KENNEDY, J., EBERHART, R. (1995). O EPSO utiliza os mesmos operadores que o PSO, mas no primeiro os pesos dos operadores internos do método são também otimizados, utilizando conceitos evolucionários. Ambas as estratégias de otimização são estratégias populacionais, ou seja, criam um conjunto de soluções, denominadas de partículas, que se movem pelo espaço de soluções de acordo com as respetivas posições passadas e a melhor posição encontrada até ao momento, do conjunto de partículas (enxame).

O operador de movimento que faz mover cada partícula possui três termos, cada um com o respetivo peso: inércia - relacionado com as posições passadas da partícula; memória - relacionado com a melhor posição passada da partícula; cooperação - relacionado com a melhor posição encontrada pelo enxame. São os pesos de cada um destes termos, que no algoritmo EPSO, são otimizados utilizando o conceito de mutação dos algoritmos evolucionários.

As modificações introduzidas no algoritmo EPSO para o adaptar ao problema aqui descrito baseiam-se no facto da convergência do método ser acelerada através da utilização de heurísticas simples, que permitem gerar, na maioria dos casos, soluções admissíveis. Desta forma, no início do processo é criado um grafo da rede, que a simplifica e permite a identificação dos ramos

que pertencem a cada anel fundamental da rede. As heurísticas implementadas relacionadas com configuração da rede não permitem que mais do que um ramo se encontre fora de serviço em cada anel, e que cada ramo a abrir pertença a um anel da rede.

Por outro lado, no que diz respeito ao controlo de tensão, as tomadas dos transformadores e das baterias de condensadores são ajustadas caso haja violação dos limites de tensão ou dos limites dos ramos. Nas redes de distribuição de energia elétrica o controlo do módulo da tensão é efetuado através de transformadores e de baterias de condensadores. Estes equipamentos possuem tomadas que podem ser ajustadas de modo a manter o módulo da tensão, no nó de controlo, num determinado valor especificado, sendo portanto um controlo discreto da tensão. Ao controlar e equilibrar os módulos da tensão nos nós da rede, os fluxos de energia na rede são diminuídos, baixando-se assim os trânsitos de potência nos ramos.

Após a execução do EPSO para reconfigurar a rede de forma integrada é executado um algoritmo de controlo de tensão e potência reativa [18] que faz os ajustes finais das tomadas de controlo, obtendo-se assim uma solução bastante boa para o problema.

Exemplo

Para testar o algoritmo descrito na secção anterior, foi considerada a rede de teste apresentada na Figura 2, que corresponde a uma versão simplificada de uma rede de distribuição de energia elétrica. A rede de teste apresentada é uma versão modificada da rede utilizada em BHOOMESH, R., AH KING, R.T.F., RUGHHOOPUTH, H.C.S. (2003), sendo constituída por 3 *feeders* (saídas de subestações), 3 transformadores (T1, T2 e T3), 19 nós (F1, F2, F3, 1 a 16), 20 linhas (L1 a L20), 13 cargas elétricas (correspondentes a Postos de Transformação onde se liga a rede de Baixa Tensão ou correspondentes a Consumidores de Média Tensão) e 7 baterias de condensadores (C1 a C7). As linhas possuem um aparelho de corte que através da sua abertura inibe a passagem de corrente elétrica pela linha. Cada uma das referidas saídas pode corresponder a três saídas de uma mesma subestação ou três saídas de três subestações diferentes.

A montante dos nós F1, F2 e F3 está a rede de transporte e os geradores de um sistema elétrico convencional. As cargas estão representadas por setas nos nós de consumo e o seu valor total é 28,6 MW. Refira-se que o cenário de operação da rede é caracterizado pelo consumo em cada um dos nós de consumo da rede, sendo a operação da rede otimizada para este cenário de consumo.

A configuração inicial da rede é a que está apresentada na Figura 2, onde os ramos L6, L8, L12, L17, L18, L19 e L20 estão abertos. Para alterar a configuração da rede existem aparelhos de corte em cada ramo, que permitem ligar ou desligar o respetivo ramo. De notar que os ramos que se encontram abertos fecham caminhos entre as diferentes saídas, o que permite transferir carga de uma saída para outra. Obviamente, esta transferência de carga está limitada pela capacidade de cada uma das saídas.

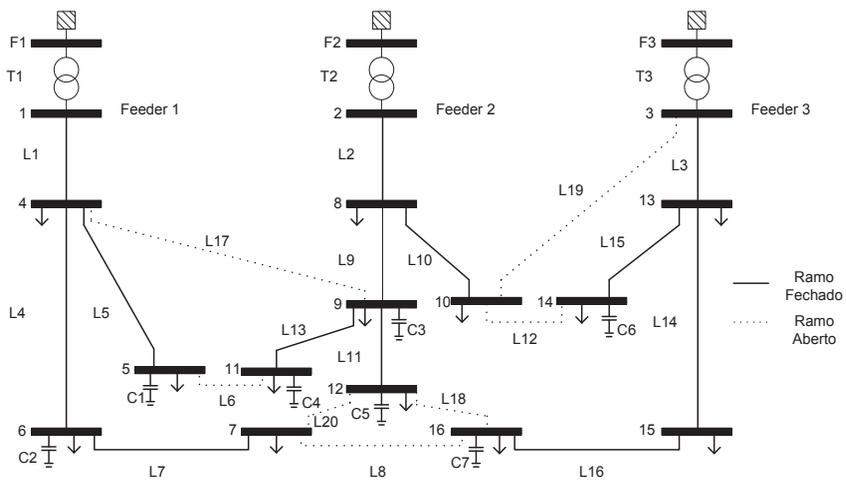


Figura 2: Rede de distribuição com possibilidade de reconfiguração.

O algoritmo EPSO proposto para a reconfiguração ótima da rede foi configurado com os seguintes parâmetros: uma população com 45 indivíduos; a probabilidade de comunicação é de 0,2; o parâmetro de mutação igual a 0,8; e o critério de paragem utilizado foi definido como o número máximo de iterações e igual a 35.

A otimização da rede, relativamente às perdas de energia elétrica, foi obtida através de quatro modos de otimização diferentes: reconfiguração e controlo de tensão (CT) integrados (Reconfiguração e CT); reconfiguração seguida do controlo de tensão (Reconfiguração + CT, resolução separada e tradicional); só reconfiguração (Reconfiguração); só controlo de tensão (CT).

Esta otimização foi realizada para um cenário de operação da rede, caracterizado pelo nível de carga em cada um dos nós. Os resultados estão apresentados na Tabela 1, e mostram que a utilização combinada dos módulos de reconfiguração e controlo de tensão é preferível pois conduz ao menor valor de perdas finais na rede (uma redução de 31% relativamente às perdas iniciais), para o mesmo cenário de operação. Por outro lado quando o modo de operação considera a execução separada das operações de reconfiguração e das operações de controlo de tensão apenas conduz a uma redução das perdas de cerca de 29%. Estes resultados mostram a vantagem do método proposto relativamente ao método tradicional no qual as operações de reconfiguração e controlo de tensão são consideradas separadamente.

Modo do algoritmo	Perdas iniciais (kW)	Perdas finais (kW)	Equipamentos alterados	Tempo de execução (s)
Reconfiguração e CT	335,5	230,1	F1, F2, F3, C1, C2, C3, C4, C5, C6, C7, T2, T3, swit10, swit11, swit13, swit15, swit06, swit12, swit18, swit19	0,41
Reconfiguração + CT		237,7	F1, F2, F3, C1, C2, C3, C4, C5, C6, C7, T1, T2, T3, swit10, swit11, swit19, swit20	0,7
Reconfiguração		313,7	swit10, swit11, swit19, swit20	0,41
CT		246,9	F1, F2, F3, T1, T2, T3, C1, C2, C3, C4, C5, C6, C7	0,01

Tabela 1: Resultados da execução dos módulos de reconfiguração e controlo de tensão integrados, sequencialmente, e isoladamente cada um.

As operações topológicas efetuadas pelo algoritmo sugerem um alívio de carga do ramo central da rede para os periféricos. Isto é devido ao facto de as cargas situadas nos nós 9 e 12 serem cargas com valores mais elevados quando comparadas com as restantes cargas da rede, o que faz com que os ramos

centrais se encontrem mais carregados e por isso com maiores perdas. Assim, o algoritmo procura aliviar estes ramos de forma a reduzir as perdas globais do sistema.

Notar que apesar de todos os aparelhos com capacidade de controlo de tensão, baterias de condensadores, equivalentes de rede e transformadores, serem todos operados nos quatro diferentes modos de execução do algoritmo de otimização, os estados finais de cada um destes equipamentos não é o mesmo.

Os tempos de execução da Tabela 1 e Tabela 2 foram obtidos num computador com um processador Intel® Core i7 CPU 3,40GHz e com 8 GB de memória RAM. Analisando os tempos de execução da Tabela 1, o modo integrado apresenta um tempo de execução de 0,41 segundos, que é um tempo inferior ao modo sequencial e semelhante ao modo de apenas reconfiguração, e conseguindo uma solução melhor.

Modo do algoritmo	Perdas iniciais (kW)	Perdas finais (kW)	Tempo de execução (s)
Reconfiguração e CT	42561,8	35007,0	223,8
Reconfiguração + CT		35878,5	412,55
Reconfiguração		42471,8	184,4
CT		36033,4	44,0

Tabela 2 – Resultados da execução dos módulos de reconfiguração e controlo de tensão integrados, sequencialmente, e isoladamente para uma rede real.

Para validar o bom comportamento do algoritmo desenvolvido em redes reais e de maior dimensão, na Tabela apresentam-se os resultados para uma rede de distribuição real, que é constituída por 3909 nós e 3914 ramos. Nesta rede continua a verificar-se um melhor desempenho em termos de solução encontrada por parte do modo integrado, conseguindo a melhor solução em termos do mínimo de perdas. Neste caso, o tempo de execução do modo integrado é bastante inferior ao modo de execução sequencial, obtendo a solução em tempo útil.

O algoritmo proposto para a reconfiguração ótima de rede com base na estratégia de otimização EPSO, com o objetivo de minimização das perdas

ativas na rede, provou ser uma metodologia válida para lidar com este problema de uma forma mais eficiente.

Conclusões

O modelo desenvolvido revela-se um procedimento extremamente eficiente que resolve de forma integrada dois problemas que habitualmente são tratados de forma independente.

O exemplo apresentado neste artigo confirma os resultados obtidos por outros investigadores que aplicaram outros conceitos para a reconfiguração da rede e para o controlo de tensões e minimização de perdas em sistemas de energia elétrica de forma separada. No entanto, quando a resolução é feita de forma integrada, conseguem-se obter melhores resultados e de uma forma eficiente.

A utilização do EPSO para resolver este problema, conjuntamente com algumas heurísticas locais que têm em conta as relações existentes entre as variáveis, permitiu definir um algoritmo que resolve em tempo útil este problema. Quando comparado com outros algoritmos, que fazem apenas reconfiguração, o tempo total é da mesma ordem de grandeza.

Uma versão comercial deste algoritmo foi integrada num DMS comercial da EFACEC, estando atualmente a ser instalado em alguns dos seus clientes.

Agradecimentos

Este trabalho é parcialmente financiado por Fundos FEDER através do Programa Operacional Fatores de Competitividade – COMPETE e por Fundos Nacionais através da FCT – Fundação para a Ciência e a Tecnologia no âmbito do projeto GEMS PTDC/EEA-EEL/105261/2008.

BIBLIOGRAFIA

- MERLIN, A., BACK, H. (1975) “Search for a minimum-loss operating spanning tree configuration in an urban power distribution”, *Proceedings of Fifth Power System Computation Conference*, Cambridge, UK, pp. 1 – 18.
- CINVALAR, S., GRAINGER, J.J., YIN, H., LEE, S.S.H. (1988) “Distribution feeder reconfiguration for loss reduction”, *IEEE Transactions on Power Delivery*, vol. 3, no. 3, pp. 1217 – 1223.

GUPTA, N., SWARNKAR, A., NIAZI, K.R., BANSAL, R.C. (2010) "Multi-objective reconfiguration of distribution systems using adaptive genetic algorithm in fuzzy framework", *IET Generation, Transmission & Distribution*, vol. 4, no. 12, pp. 1288-1298, 2010.

EL-HAWARY, M. (editor) (1998) *Electric Power Applications of Fuzzy Systems*, IEEE Press Series on Power Engineering.

GUPTA, N., SWARNKAR, A., NIAZI, K.R. (2011) "Reconfiguration of Distribution Systems for Real Power Loss Minimization Using Adaptive Particle Swarm Optimization", *Electric Power Components and Systems*, vol. 39, no. 4, pp. 317 – 330.

MATOS, M.A., MELO, P. (1999) "Multi-objective Reconfiguration for Loss Reduction and Service Restoration Using Simulated Annealing", *Proceedings of IEEE Power Tech '99*, Budapest, Hungary.

KHODR, H.M., MARTÍNEZ-CRESPO, J., MATOS, M.A., PEREIRA, J. (2009) "Distribution Systems Reconfiguration Based on OPF Using Benders Decomposition", *IEEE Transactions on Power Delivery*, vol. 24, no. 4, pp. 2166-2176.

LIU, Y., ZHANG, P., QIU, X. (2002) "Optimal volt/var control in distribution systems," *International Journal of Electrical Power & Energy Systems*, vol. 24, no. 4, pp. 271-276.

HSU, Y. (1998) "A combined artificial neural network-fuzzy dynamic programming approach to reactive power/voltage control in a distribution substation", *IEEE Transactions on Power Systems*, vol. 13, no. 4, pp. 1265-1271.

JUNGES, M. (2000) *Lógica fuzzy em controle de tensão e potência reativa em sistemas de potência*, M.Sc. Thesis, Pontifícia Universidade Católica de Minas Gerais, no. 621.31J95|T - PucMinas Library, Belo Horizonte, Brazil.

EKEL, P., TERRA, L., JUNGES, M., POPOV, V. (1998) "Fuzzy technology in the design, planning and control of power systems and subsystems", *Proceedings of EUFIT'98 - European Congress on Intelligent Techniques and Soft Computing*, vol. 2, pp. 1126-1130, Aachen, Alemanha.

LIU, C.W., CHANG, C.S., JIANG, J.A. (2001) "Genetic algorithms as a reactive power source dispatching aid for voltage security enhancement," *Proceedings of National Sci Council ROC(A)*, vol. 25, no. 1, pp. 53–62.

LIANG, R. (2003) "Fuzzy-based reactive power and voltage control in a distribution system," *IEEE Transactions on Power Delivery*, vol. 18, no. 2, pp. 610-618.

PEREIRA, J., SARAIVA, J.T., PONCE DE LEÃO, M.T. (1999) "Identification of operation strategies of distribution networks using a simulated annealing approach", *Proceedings of IEEE Power Tech '99*, Budapest, Hungary.

MIRANDA, V., FONSECA, N. (2002) "New Evolutionary Particle Swarm Algorithm (EPSO) Applied to Voltage /VAR control", *Proceedings 14th PSCC – Power Systems Computation Conference*, Sevilha, Espanha.

MIRANDA, V., CALISTO, P. (2002) "A fuzzy inference system to voltage/VAR control in DMS - distribution management systems", *Proceedings 14th PSCC – Power Systems Computation Conference*, Sevilha, Espanha.

MIRANDA, V., MOREIRA, A., PEREIRA, J. (2007) "An Improved Fuzzy Inference System for Voltage/VAR Control", *IEEE Transactions on Power Systems*, vol. 22, no. 4, pp. 2013-2020.

KENNEDY, J., EBERHART, R. (1995) "Particle swarm optimization", *Proceedings of IEEE International Conference on Neural Networks*, vol.4, pp.1942-1948.

BHOOMESH, R., AH KING, R.T.F., RUGHOOPUTH, H.C.S. (2003) "A modified genetic algorithm for optimal electrical distribution network reconfiguration", in *CEC '03 - The 2003 Congress on Evolutionary Computation*, vol. 2, pp. 1472-1479.

CASO 17

OTIMIZAÇÃO DO PLANEAMENTO INVERSO DE TRATAMENTOS EM RADIOTERAPIA

H. Rocha

INESC-Coimbra, Coimbra, Portugal
hrocha@mat.uc.pt

J.M. Dias

INESC-Coimbra, Coimbra, Portugal
Faculdade de Economia, Universidade de Coimbra, Coimbra
joana@fe.uc.pt

B.C. Ferreira

I3N, Departamento de Física, Universidade de Aveiro, Aveiro, Portugal
Serviço de Física Médica, IPOC-FG, EPE, Coimbra, Portugal
brigida@ua.pt

M.C. Lopes

I3N, Departamento de Física, Universidade de Aveiro, Aveiro, Portugal
Serviço de Física Médica, IPOC-FG, EPE, Coimbra, Portugal
mclopes@ipocoimbra.min-saude.pt

RESUMO

O número de casos de cancro continua a crescer à escala planetária. Existem vários tipos de tratamento usados habitualmente, dependendo do tipo e fase do cancro, incluindo cirurgia, radioterapia, quimioterapia, imunoterapia, etc. Neste trabalho, vamos focar a otimização do planeamento de tratamentos em radioterapia, em particular de tratamentos em radioterapia com intensidade modulada (IMRT), onde a otimização tem um papel importante na melhoria da qualidade dos tratamentos. O contínuo desenvolvimento de novas máquinas de tratamento contribui para a melhoria da precisão e maior controlo da irradiação dos pacientes. A investigação operacional tem acompanhado a evolução destas máquinas e tecnologias, e tem feito contribuições significativas na melhoria do planeamento dos tratamentos em radioterapia, sendo os tratamentos em IMRT um bom exemplo disso. Devido à complexidade dos problemas de otimização do planeamento de tratamentos em IMRT, estes são usualmente divididos em três sub-problemas de menor dimensão que podem ser resolvidos sequencialmente: problema de otimização angular, problema de otimização das intensidades e problema de sequenciação das folhas. O primeiro destes sub-problemas consiste na escolha das melhores direções de incidência da radiação, encontrando-se ainda por resolver de forma satisfatória. Na prática clínica, a maior parte das vezes, as direções de incidência da radiação continuam a ser escolhidas manualmente, com base essencialmente na experiência do físico-médico, sem garantia da obtenção dos planos de tratamento de melhor qualidade. A abordagem que usamos considera os problemas de otimização angular e otimização das intensidades de forma integrada. Vamos demonstrar que a investigação operacional tem um papel importante na obtenção de melhores planos de tratamento, focando-nos principalmente na aproximação por nós sugerida para o problema de otimização angular.

PALAVRAS-CHAVE

Radioterapia; Otimização; Planeamento inverso.

1 Introdução

O número de casos de cancro continua a crescer à escala planetária. Um estudo da American Cancer Society (2010) estima que mais de 33 mil novos casos de cancro surjam diariamente. O mesmo estudo prevê que, em 2050, existam mais de 27 milhões de pessoas com cancro. Em 2012, o número estimado de novos casos nos EUA atinge os 1.5 milhões, sendo esperado que mais de meio milhão de norte-americanos morram de cancro, mais de 1500 por dia. Na Europa, com quase 500 milhões de habitantes, estima-se que sejam diagnosticados, por ano, cerca de 2 milhões de novos casos. Em Portugal, não existem números oficiais, devido à ausência de um registo sistematizado de ocorrências. No entanto, a *Liga Portuguesa Contra o Cancro* estima que, por ano, morram 42 mil pessoas com cancro em Portugal e deixa ainda o alerta relativo à duplicação de novos casos até 2030.

A prevenção e o rápido diagnóstico são as melhores armas contra o cancro. Para os cancros diagnosticados nos EUA entre 1996 e 2004, a taxa de sobrevivência a 5 anos é de 66%, melhor do que os 50% verificados entre 1975 e 1977 (American Cancer Society, 2010). Esta melhoria reflete o esforço em prevenção e rápido diagnóstico, mas também a melhoria contínua da qualidade dos tratamentos oferecidos para os diferentes tipos de cancro. Existem várias possibilidades de tratamento usados habitualmente, dependendo do tipo e fase do cancro, e incluem cirurgia, radioterapia, quimioterapia, imunoterapia, etc. Uma combinação destes tratamentos é frequentemente usada para obter os melhores resultados. Neste trabalho, vamos focar-nos no planeamento de tratamentos em radioterapia, em particular com intensidade modulada (IMRT), onde a otimização tem um papel importante na melhoria da qualidade dos tratamentos. Os termos radioterapia e otimização estão relacionados entre si desde 1959 (Carlson & Morkovin, 1959). A interação entre a investigação operacional e a física médica teve o seu primeiro passo em 1968 com o uso do primeiro modelo de programação linear para ajudar ao desenho de modelos em radioterapia (Bahr *et al*, 1968). Apareceu deste modo uma área multidisciplinar florescente e de importância crescente. Muitos artigos de revisão foram já produzidos nesta área multidisciplinar (e.g., Bortfeld, 2006;

Censor, 2003; Ehrgott *et al*, 2008; Holder & Salter, 2004; Rocha & Dias, 2009) o que é mais uma prova do interesse que este tema tem suscitado.



Figura 1: Acelerador linear (linac) irradiando em torno do paciente (RadiologyInfo, 2012).

O objetivo da radioterapia é administrar a dose de radiação prescrita pelo médico ao volume delineado (o volume do tumor mais uma margem de segurança), designado por PTV, de forma a matar as células cancerosas minimizando os danos nos tecidos e órgãos circundantes. A radioterapia baseia-se no facto de as células cancerosas se reproduzirem muito mais rapidamente do que as células saudáveis e serem incapazes de sobreviver quando danificadas por radiação. Como tal, o objetivo do tratamento é emitir radiação suficiente de forma a matar as células cancerosas mas não demasiada que comprometa a capacidade das células saudáveis em sobreviver. Tipicamente, a radiação é gerada por um acelerador linear (linac) que roda em torno de um eixo central, e é emitida com o paciente imobilizado numa mesa de tratamentos que também pode rodar (Figura 1). A combinação destes dois movimentos de rotação permite uma irradiação a partir de qualquer ângulo em torno do PTV. Contudo, como forma de reduzir um problema já complexo, o

uso de ângulos coplanares continua a ser predominante, pertencendo os ângulos habitualmente usados ao plano de rotação do linac em torno do paciente. Na prática clínica, a maior parte das vezes, o número de ângulos é definido a priori e as direções de incidência da radiação são selecionadas manualmente pelo físico-médico, que confia essencialmente na sua experiência. A literatura apresenta, no entanto, evidência de que direções de incidência da radiação apropriadas podem levar a melhorias na qualidade dos tratamentos (Das & Marks, 1997; Liu *et al*, 2006; Rowbottom *et al*, 1998). O contínuo desenvolvimento de novas máquinas contribui para a melhoria da precisão e maior controlo da irradiação. A investigação operacional tem acompanhado a evolução das máquinas e tecnologia e tem feito contribuições significativas na melhoria do planeamento dos tratamentos em radioterapia (Bahr *et al*, 1968; Ferris *et al*, 2003a; Ferris *et al*, 2003b; Lim *et al*, 2007; Lim & Lee, 2007; Romeijn *et al*, 2006), sendo os tratamentos em IMRT um bom exemplo disso.

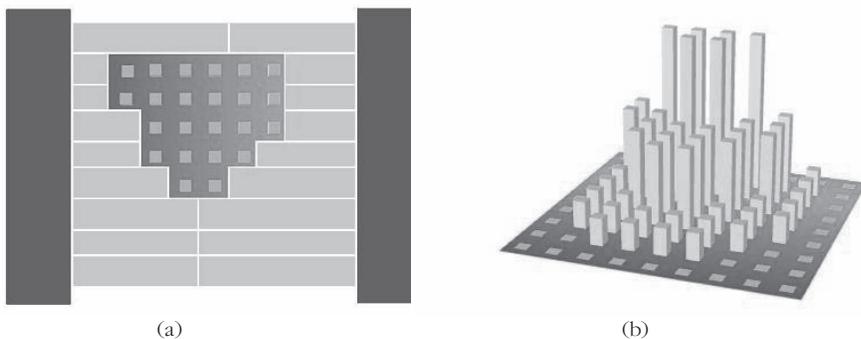


Figura 2: Ilustração de um MLC (com 9 pares de folhas) – 2(a) e ilustração de um mapa de intensidades (9X9) – 2(b).

A otimização do planeamento de um tratamento pode ser interpretada como a seleção ótima de uma dada configuração dos parâmetros do acelerador linear, entre um conjunto de possíveis configurações (soluções admissíveis). Dada a complexidade do problema, muitas vezes o planeamento é feito através de um processo de tentativa e erro, designado por planeamento direto. No planeamento direto, uma dada configuração é definida e as doses de

radioterapia que vão ser absorvidas pelo organismo do doente são calculadas. Se estas doses estiverem de acordo com a prescrição médica, o planeamento termina. Em caso contrário, o processo continua através da alteração manual de alguns dos parâmetros da configuração. O planeamento inverso consiste em calcular a configuração ótima, dadas as doses prescritas, usando modelos e programas de otimização. O planeamento inverso permite modelar problemas de planeamento de tratamentos altamente complexos e a otimização tem um papel fundamental no sucesso deste procedimento. Um tipo importante de tratamentos com planeamento inverso são os tratamentos em IMRT, onde o feixe de radiação é modulado por um colimador multifolhas (MLC) (Figura 2(a)), o que permite transformar o feixe de radiação em pequenos sub-feixes de intensidades independentes (Figura 2(b)). Apesar da ilustração na Figura 2(b), os sub-feixes não existem fisicamente, sendo gerados pelo movimento das folhas do MLC que bloqueiam parte do feixe durante porções do tempo de irradiação. Uma forma comum de resolver os problemas de planeamento inverso na otimização de tratamentos em IMRT é usar uma aproximação baseada nos sub-feixes de radiação, o que origina problemas de programação de grande dimensão com centenas de variáveis e dezenas de milhar de restrições. A qualidade do plano de tratamento, dependendo dos modelos de programação e correspondentes métodos de resolução, determina a eficácia do tratamento. Devido à complexidade de tais problemas de otimização, estes são usualmente divididos em três sub-problemas de menor dimensão, que podem ser resolvidos sequencialmente: problema de otimização angular, problema de otimização das intensidades e problema de sequenciação das folhas. A abordagem que usamos considera os problemas de otimização angular e otimização das intensidades de forma integrada.

O problema de otimização angular tenta encontrar o número ótimo de ângulos (direções) a utilizar no tratamento e a sua localização. Muitas vezes considera-se o número de ângulos fixo a priori, limitando-se a otimização à procura dos melhores ângulos dentro de todas as combinações possíveis. O problema de otimização das intensidades considera os ângulos já definidos, e tenta otimizar as intensidades dos feixes de radiação para cada direção de

incidência. O problema de sequenciação das folhas determina de que forma será efetuado o movimento das folhas do MLC, tendo em conta as restrições físicas do equipamento, por forma a se obterem as intensidades de radiação desejadas. A maior parte dos esforços dos investigadores em otimização de tratamentos em IMRT foram dedicados à otimização das intensidades (Craft, 2007), sendo os ângulos de irradiação determinados pelo físico-médico. Comparativamente menos esforço de investigação tem sido dedicado ao problema de otimização angular (Ehrgott & Johnston, 2003). Este é um problema de otimização muito difícil, ainda em aberto, visto ser um problema altamente não convexo com muitos mínimos locais. Neste trabalho, vamos demonstrar que a investigação operacional tem um papel importante na obtenção de planos de tratamento de maior qualidade, focando a aproximação por nós sugerida para o problema de otimização angular, e descrevendo muito sucintamente os outros sub-problemas. Para tal, vamos usar dois exemplos clínicos de casos de tumores de cabeça e pescoço tratados retrospectivamente no *Instituto Português de Oncologia Francisco Gentil de Coimbra* (IPOC), instituição com quem temos colaborado. É importante salientar que as melhorias obtidas nos planos de tratamento otimizados se refletem numa efetiva melhoria da qualidade de vida dos pacientes. Na próxima secção vamos descrever o problema de otimização das intensidades, necessário para a nossa formulação do problema de geometria angular. Na secção 3 apresentamos o problema de otimização angular, a formulação e método de resolução por nós sugeridos. Os resultados numéricos para os casos de tumores de cabeça e pescoço descritos na secção 4 são apresentados na secção 5. Na última secção temos a conclusão.

2 Problema de otimização das intensidades

Considerando que os ângulos (ou direções) de irradiação estão já definidos, o paciente é tratado usando um plano ótimo obtido após a resolução do problema de otimização das intensidades - o problema de determinar a intensidade ótima de cada um dos sub-feixes para cada um dos feixes fixos correspondentes às direções de irradiação já definidas. Muitos modelos

matemáticos e algoritmos têm sido propostos para este problema, incluindo modelos lineares (e.g., Romeijn *et al*, 2005; Romeijn *et al*, 2003), modelos lineares inteiros mistos (e.g., Lee *et al*, 2003; Preciado-Walters *et al*, 2006), modelos não lineares (e.g., Aleman *et al*, 2008; Spirou & Chui, 1998) e modelos multiobjetivo (e.g., Craft *et al*, 2006; Romeijn *et al*, 2004).

A distribuição de dose depositada no paciente, medida em Grays (Gy), precisa de ser aferida com precisão de forma a resolver o problema de otimização das intensidades. O volume de cada estrutura é discretizado em voxéis (pequenos elementos de volume) e a dose é calculada para cada voxel usando o princípio da superposição, i.e., considerando a contribuição de cada sub-feixe. Tipicamente, uma matriz de dose D é construída a partir do conjunto de todas as intensidades dos sub-feixes, indexando as linhas de D a cada voxel e as colunas de D a cada sub-feixe, i.e., o número de linhas de D é igual ao número total de voxéis (N_v) e o número de colunas é igual ao número total de sub-feixes (N_b) de todos os feixes de radiação considerados. O cálculo preciso da matriz D é alvo contínuo de investigação, existindo diferentes algoritmos de dose, sendo o Monte Carlo a técnica padrão. Por razões de economia de tempo, utilizamos um algoritmo menos preciso mas mais rápido (pencil beam) para calcular a matriz D , que se mantém constante ao longo do processo de otimização para cada conjunto de direções de irradiação. Portanto, usando formato matricial, podemos dizer que a dose total recebida pelo voxel i é dada por $\sum_{j=1}^{N_b} D_{ij} w_j$, com w_j a intensidade do sub-feixe j . Usualmente, o número de voxéis considerados atinge as dezenas de milhar, portanto o número de linhas da matriz de dose é dessa magnitude. O tamanho da matriz D origina problemas de grande dimensão, sendo uma das principais razões para a dificuldade em resolver o problema de otimização das intensidades. Uma das estratégias para reduzir a dimensão dos problemas de otimização e simultaneamente o tempo total de cálculo é usar sampling (Rocha *et al*, 2011). A agregação de voxéis em certas estruturas (sobretudo nas células do tecido normal), origina uma redução da dimensão do problema, o que conduz a uma resolução mais rápida do mesmo.

As primeiras tentativas de abordagem ao problema de otimização das

intensidades fizeram uso de modelos de programação linear. Algumas das razões para o uso destes modelos incluem o facto de a dose ser depositada de forma linear, os modelos lineares serem fáceis de implementar, usados extensivamente, rápidos a obter soluções ótimas e resolúveis mesmo para modelos de muito grande dimensão. Dada uma prescrição com limites inferior e superior para a dose do PTV, limites superiores para as doses dos órgãos de risco (OARs) em torno do tumor, limite superior para a dose do restante tecido de células normais (NT) e um valor máximo para as intensidades dos sub-feixes de radiação (M), a maioria das formulações dos modelos de programação linear pertencem a uma classe de modelos de otimização com restrições, tal que uma função objetivo é minimizada enquanto as doses permanecem dentro dos limites impostos:

$$\begin{aligned}
 \min_w \quad & f(D) \\
 \text{t.q.} \quad & LB_{PTV} \leq \sum_{j=1}^{N_b} D_{ij} w_j \leq UB_{PTV}, \forall i \in PTV, \\
 & \sum_{j=1}^{N_b} D_{ij} w_j \leq UB_{OAR}, \forall i \in OAR, \\
 & \sum_{j=1}^{N_b} D_{ij} w_j \leq UB_{NT}, \forall i \in NT, \\
 & 0 \leq w_j \leq M, \quad j = 1, \dots, N_b.
 \end{aligned}$$

Outras restrições poderiam ser incluídas dependendo do que for pretendido pelo físico-médico. Uma variedade de diferentes modelos pode ser formulada combinando diferentes restrições com diferentes objetivos. Por forma a reduzir o impacto dos diferentes tamanhos das estruturas, uma função objetivo usada usualmente considera a média dos desvios da dose prescrita em cada estrutura considerada (Lim *et al*, 2007a):

$$f(D) = \alpha_{PTV} \frac{\|D_{PTV} - TG_{PTV}\|_p}{card(PTV)} + \alpha_{OAR} \frac{\|(D_{OAR} - UB_{OAR})_+\|_p}{card(OAR)} + \alpha_{NT} \frac{\|(D_{NT} - UB_{NT})_+\|_p}{card(NT)}, \quad p = 1, 2, \infty, \quad (1)$$

onde D_{PTV} e TG_{PTV} correspondem à dose e objetivo de dose no PTV respetivamente, D_{OAR} e D_{NT} correspondem à dose nos órgãos de risco e tecidos de células normais respetivamente, $card(\cdot)$ denota o número total de voxels na estrutura considerada, p a norma considerada e $\alpha_{(\cdot)}$ os pesos de cada estrutura que têm de ser ajustados pelo físico-médico. Diferentes parcelas, correspondendo a diferentes estruturas, podem ser considerados usando diferentes normas na mesma função objetivo. A norma L_1 penaliza o desvio da dose prescrita, em valor absoluto, em cada voxel, ponderado de acordo com a estrutura ao qual o voxel pertence. A norma L_2 penaliza a soma dos quadrados dos desvios da dose prescrita, ponderados pelos pesos definidos. A norma L_∞ penaliza “pontos quentes” em estruturas sensíveis (OARs e NT) e “pontos frios” no PTV. Note-se que o uso da norma L_2 origina modelos não lineares quadráticos.

Um revés importante no uso destes modelos de programação linear reside no facto de as soluções ótimas calculadas serem pontos extremos da região admissível, havendo restrições que são satisfeitas como igualdade. Quer isto dizer que, na maior parte das vezes, a solução ótima atribui a dose máxima permitida aos OARs e/ou o PTV recebe a dose mínima permitida. Caso as restrições sejam demasiado apertadas, por forma a minorar o revés anterior, temos um revés ainda maior: encontrar uma solução admissível pode tornar-se muito difícil ou até mesmo impossível.

Para além dos modelos de programação linear, é também possível considerar modelos não lineares que apresentam como vantagem o facto de garantirem a existência de soluções admissíveis. Considere-se que cada voxel é penalizado de acordo com o quadrado da diferença entre a dose recebida pelo voxel e a dose desejada/permitida para aquele voxel. Esta formulação origina um problema de programação quadrático apenas com restrições de não

negatividade dos valores das intensidades dos sub-feixes (Romeijn *et al*, 2003):

$$\min_w \sum_{i=1}^{N_v} \frac{1}{V_S} \left[\underline{\lambda}_i \left(T_i - \sum_{j=1}^{N_b} D_{ij} w_j \right)_+^2 + \bar{\lambda}_i \left(\sum_{j=1}^{N_b} D_{ij} w_j - T_i \right)_+^2 \right]$$

$$t.q. \quad w_j \geq 0, \quad j = 1, K, N_b,$$

onde T_i é a dose desejada para o voxel i , $\underline{\lambda}_i$ e $\bar{\lambda}_i$ são as penalidades de subdosagem e sobredosagem do voxel i e $(\cdot)_+ = \max\{0, \cdot\}$. Apesar desta formulação permitir pesos diferentes para cada voxel, analogamente à implementação em Aleman *et al* (2008) podemos atribuir pesos por estrutura por forma a que todos os voxéis de uma dada estrutura tenham os mesmos pesos (dividido pelo número de voxéis da estrutura). Esta formulação não linear implica que uma pequena quantidade de subdosagem ou sobredosagem possa ser aceitável num processo de decisão clínica, mas desvios maiores relativamente à dose desejada/prescrita são decrescentemente tolerados (Aleman *et al*, 2008).

Do problema de otimização de intensidades irá surgir uma solução que consiste na definição das intensidades de cada sub-feixe para cada direção de incidência. É depois necessário reproduzir de forma o mais fidedigna possível, estas intensidades (contínuas) em intensidades que possam ser realizadas pelo MLC. Tipicamente, as intensidades dos sub-feixes são discretizadas num conjunto de valores (0 a 10, e.g.) e uma das muitas técnicas existentes (Alber & Nusslin, 2001; Men *et al*, 2007; Romeijn *et al*, 2005; Shepard *et al*, 2002; Tervo *et al*, 2003) é usada para construir as sequências de posições das folhas do linac. Note-se que a discretização sem critério das intensidades ótimas leva a uma deterioração da qualidade do plano otimizado, que pode e deve ser minorada (Rocha *et al*, 2014a; Rocha *et al*, 2014b).

3 Problema de otimização angular

Iremos considerar o problema de otimização angular em que pretendemos determinar as direções ótimas de incidência dos feixes de radiação, considerando o número total de ângulos definido a priori. Sendo assim, queremos escolher uma determinada combinação de ângulos, dentro de todas

as combinações possíveis. A formulação do problema de otimização angular como um problema de programação matemática requer uma medida quantitativa para comparar a qualidade de diferentes combinações de direções de incidência da radiação. A maneira mais correta de se quantificar a qualidade de uma combinação de ângulos terá de ter em conta a solução do problema de otimização de intensidades que se encontra imediatamente a jusante deste problema. Qualquer outra medida de qualidade não oferece qualquer garantia de optimalidade ou fiabilidade, pois tem sido extensamente reportado que direções de irradiação ótimas de tratamentos em IMRT são muitas vezes não intuitivas (Stein *et al.*, 1997). Como tal, a nossa formulação do problema de otimização angular é baseada nas soluções ótimas do problema de otimização das intensidades. Posteriormente, descrevemos o método de resolução por nós sugerido.

3.1 Formulação do problema de otimização angular

Consideremos n como sendo o número de direções de irradiação coplanares fixas, i.e., são escolhidos n ângulos num círculo em torno do tumor. Tipicamente, o problema de otimização angular é formulado como sendo um problema de otimização combinatória no qual os n ângulos são selecionados entre um conjunto de ângulos candidatos. Geralmente, o intervalo de ângulos possíveis, $[0^\circ, 360^\circ]$, é discretizado em direções equidistantes, com um dado incremento de ângulo (e.g., 5 ou 10 graus). Podemos pensar numa pesquisa global exaustiva de todas as possíveis combinações dos n ângulos. Contudo, isso requer um tempo de cálculo e comparação de todas as diferentes distribuições de dose incompatível com uma resposta num espaço temporal clinicamente aceitável. Por exemplo, escolhendo $n = 5$ ângulos entre 72 candidatos, $\{0^\circ, 5^\circ, \dots, 355^\circ\}$, existem $C_5^{72} = 13,991,544$ combinações possíveis. Diminuindo o número de candidatos para 36, $\{0^\circ, 10^\circ, \dots, 350^\circ\}$, o número de combinações possíveis é ainda de $C_5^{36} = 376,992$, o que requer um tempo de cálculo de grande magnitude, independentemente da medida usada para comparar os planos de tratamento resultantes. Portanto, uma pesquisa exaustiva num problema combinatório de grande dimensão é considerada

demasiadamente lenta e inapropriada para uma resposta clínica atempada. Muitas heurísticas e meta-heurísticas têm sido apresentadas como tentativas para reduzir o número de comparações a efetuar. Contudo, a maioria requer um número proibitivo de avaliações da função objetivo, quando a medida usada é o valor ótimo do problema de otimização das intensidades.

A nossa opção consiste em considerar uma aproximação diferente para a formulação do problema de otimização angular. Todos os ângulos contínuos no intervalo $[0^\circ, 360^\circ]$ são considerados em vez de uma amostra discreta. Como o ângulo -5° é equivalente ao ângulo 355° , e o ângulo 365° é o mesmo que o ângulo 5° , podemos evitar uma formulação com restrições. Uma formulação para o problema de otimização angular é obtida selecionando uma função objetivo tal que o melhor conjunto de direções ou ângulos de irradiação é obtido para o mínimo da função:

$$\begin{aligned} \min \quad & f(\theta_1, \dots, \theta_n) \\ \text{t. q.} \quad & (\theta_1, \dots, \theta_n) \in \mathbb{R}^n \end{aligned} \tag{2}$$

Aqui, a função objetivo $f(\theta_1, \dots, \theta_n)$ que mede a qualidade do conjunto de direções de irradiação $\theta_1, \dots, \theta_n$ é o valor ótimo do problema de otimização das intensidades para o conjunto de direções de irradiação fixo $\theta_1, \dots, \theta_n$. Cada conjunto de direções de irradiação fixo $\theta_1, \dots, \theta_n$ origina uma matriz de dose D diferente e, conseqüentemente, um problema de otimização das intensidades com soluções (w) e valores ótimos da função objetivo $f(\theta_1, \dots, \theta_n)$ diferentes. Estas funções têm inúmeros mínimos locais, o que exponencia a dificuldade em obter uma boa solução global. Portanto, a escolha do método de resolução é um aspeto crítico para obter uma boa solução. O uso de métodos com derivadas é bastante comum pelo facto de convergirem muito rapidamente para um ótimo local. A inclusão destes métodos nos softwares dos sistemas de tratamentos para resolução do problema de otimização angular é usual, tal como acontece no software do sistema de tratamento usado na instituição com que colaboramos (IPO Coimbra), mas origina resultados muitos fracos, sendo essa uma das principais razões pelas quais nos dedicamos a este

problema em particular. A nossa formulação foi essencialmente motivada pela possibilidade de usar uma classe de métodos que, ao contrário dos métodos que usam derivadas, consegue evitar o aprisionamento em mínimos locais, característica que achamos fundamental para atacar com sucesso o problema de otimização angular: métodos de procura em padrão.

3.2 Métodos de procura em padrão

Os métodos de procura em padrão são métodos de procura direcional que pertencem à classe mais abrangente de métodos sem derivadas (ver Conn *et al*, 2009) para uma perspetiva detalhada de métodos de otimização sem derivadas). Os métodos de procura em padrão são métodos iterativos que geram uma sucessão de pontos $\{x_k\}$ usando bases positivas (ou conjuntos geradores positivos) e movem-se numa direção que pode produzir uma redução no valor da função objetivo. Uma base positiva para \mathbb{R}^n pode ser definida como um conjunto de vetores não nulos de \mathbb{R}^n , cuja combinação positiva gera \mathbb{R}^n , mas nenhum subconjunto próprio o faz.

Uma das principais características das bases positivas, que constitui a principal motivação para o seu uso nos métodos de procura direcional, é que, caso o ponto corrente não seja um ponto estacionário, existe sempre um vetor v_i numa base positiva que é uma direção de descida (Davis, 1954), i.e., existe um $\Delta > 0$ tal que $f(x_k + \Delta v_i) < f(x_k)$. Isto é o cerne dos métodos de procura direcional e em particular dos métodos de procura em padrão.

Os métodos de procura em padrão são métodos iterativos que geram uma sucessão de pontos $\{x_k\}$ não crescente (em termos de valor da função objetivo). Dado o ponto corrente x_k , em cada iteração k , o próximo ponto x_{k+1} é escolhido entre um conjunto finito de candidatos numa dada grelha, definida usando os vetores que formam um conjunto gerador positivo, tentando obter um decréscimo no valor da função objetivo: $f(x_{k+1}) < f(x_k)$. Estes métodos consideram dois passos em cada iteração. O primeiro passo consiste numa procura finita na grelha, com o objetivo de encontrar um novo ponto que decresça o valor da função objetivo no ponto atual. Este passo, chamado passo de procura, tem a flexibilidade de usar qualquer estratégia, método ou

heurística, ou tirar partido de particularidades ou conhecimento a priori do problema em mãos, desde que apenas um número finito de pontos da grelha seja testado. Este passo fornece a flexibilidade para uma procura global pois permite pesquisas longe da vizinhança do ponto atual, e influencia a qualidade do minimizante local ou ponto estacionário encontrado pelo método. Se o passo de procura não tiver sucesso, é realizado um segundo passo, chamado passo de sondagem, em torno do ponto atual, com o propósito de melhorar o valor da função objetivo. O passo de sondagem segue regras mais estritas e apela aos conceitos de bases positivas. Neste passo, faz-se uma pesquisa local numa vizinhança da grelha do ponto atual que, para um parâmetro de tamanho de grelha, Δ_k , suficientemente pequeno, garante uma redução no valor da função objetivo, caso o ponto atual não seja um ponto estacionário (Alberto *et al.*, 2004). Portanto, se o passo de sondagem falhar, o parâmetro Δ_k deve ser reduzido. A escolha mais comum para atualizar este parâmetro é mantê-lo ou duplicá-lo em iterações com sucesso, e dividir por dois, em caso contrário. Note-se que se o parâmetro de tamanho de grelha inicial for uma potência de 2, ($\Delta_0 = 2^p, p \in \mathbb{N}$), e o ponto inicial for um vetor de inteiros, usando esta atualização usual do parâmetro do tamanho da grelha, todos os pontos seriam um vetor de inteiros até o parâmetro do tamanho da grelha ser inferior a 1. Esta possibilidade é particularmente interessante para o nosso problema de otimização angular.

Na abordagem ao problema de otimização angular, é de extrema importância a eficiência em termos de número de avaliações da função objetivo. Portanto, o número de pontos pesquisado no passo de procura deve ser minimalista, e guiado por algum significado físico ou biológico. Por outro lado, quando o passo de procura falha em obter um decréscimo no valor da função objetivo, o passo de sondagem deve ser feito com as direções de procura criteriosamente ordenadas de forma a reduzir ainda mais o número de avaliações da função objetivo (pelo menos nas iterações com sucesso). Recentemente, a eficiência dos métodos de procura em padrão foi melhorada significativamente, reordenando as direções de pesquisa de acordo com indicadores de descida construídos a partir de gradientes simpléticos (Custódio

& Vicente, 2007). Posteriormente, no passo de procura, foram incluídos modelos quadráticos de norma de Frobenius mínima, para serem minimizados numa região de confiança, o que pode levar a melhorias significativas (Custódio *et al*, 2010). Tirando partido da existência de uma implementação dos métodos de procura em padrão, que incorpora as melhorias referidas, usamos uma adaptação da última versão do SID-PSM (Custódio *et al*, 2010; Custódio & Vicente, 2007) para a resolução do problema de otimização angular.

Os benefícios do uso dos métodos de procura em padrão na resolução do problema altamente não convexo de otimização angular são ilustrados usando um conjunto de exemplos clínicos de tumores de cabeça e pescoço que são apresentados seguidamente.

4 Exemplos clínicos de tumores de cabeça e pescoço

Dois exemplos clínicos de casos de tumores de cabeça e pescoço tratados retrospectivamente no IPOC foram usados para testar a nossa aproximação usando métodos de procura em padrão. Os exemplos selecionados foram sinalizados no IPOC como casos complexos, onde uma irradiação apropriada do tumor e a poupança dos OARs, em particular a poupança das parótidas, demonstrou ser particularmente difícil de obter com os planos de tratamento usuais com 7 ângulos de irradiação equidistantes. As tomografias axiais computadorizadas (TACs) dos pacientes com as estruturas delineadas foram exportadas via Dicom RT para um ambiente computacional para investigação em radioterapia de acesso livre (ver Figura 3). Como a região da cabeça e pescoço é uma área complexa onde, e.g., as glândulas parótidas estão usualmente muito próximas ou até sobrepostas ao PTV, uma seleção cuidadosa das direções de incidência da radiação pode ser determinante para obter um plano de tratamentos satisfatório.

A medula e o tronco encefálico são dois dos OARs mais críticos nos casos de tumores de cabeça e pescoço. Trata-se de órgãos em série, i.e., órgãos cuja funcionalidade total fica comprometida mesmo quando apenas uma pequena porção do órgão é danificada. Portanto, se a dose máxima tolerada for excedida, tal pode resultar na perda funcional de todo o órgão. Logo, é de vital

importância não exceder a dose tolerada prescrita para esse tipo de órgãos. Além da medula e do tronco encefálico, as glândulas parótidas são também OARs importantes. As glândulas parótidas são as maiores das três glândulas salivares. Uma complicação comum devido à irradiação das parótidas é a xerostomia (o termo médico para a boca seca devido à falta de saliva). Isto provoca uma perda acentuada da qualidade de vida dos pacientes submetidos a radioterapia para tratamento de tumores de cabeça e pescoço, causando dificuldade em engolir, perda de apetite, etc. As parótidas são órgãos paralelos, i.e., se apenas uma pequena porção do órgão for danificada, a funcionalidade do resto do órgão pode não ser afetada. A sua dose de tolerância depende fortemente da fração do volume que foi irradiada. Portanto, se apenas uma pequena fração do órgão foi irradiada, a dose de tolerância é muito maior do que se uma fração maior fosse irradiada. Logo, para essas estruturas paralelas, a dose média é geralmente considerada em vez da dose máxima na otimização do planeamento inverso de tratamentos em IMRT.

Em geral, a região da cabeça e pescoço é uma área complexa a tratar em radioterapia devido ao grande número de órgãos sensíveis presentes nesta região (e.g., olhos, mandíbula, laringe, cavidade oral, etc.). Para simplificar, neste estudo, os OARs usados na otimização do planeamento foram limitados à medula, tronco encefálico e glândulas parótidas.

Para estes casos de cabeça e pescoço em estudo, o volume tumoral delineado foi separado em dois com diferentes doses prescritas: PTV1 e PTV2. As doses prescritas para os tumores e as doses de tolerância para os OARs considerados na otimização são apresentadas na Tabela 1.

As glândulas parótidas estão muito próximas ou até mesmo sobrepostas aos PTVs, o que ajuda a explicar a dificuldade em poupar as parótidas. Direções de incidência da radiação adequadas podem ajudar todo o processo de otimização, em particular diminuindo a dose de radiação absorvida pelas parótidas.

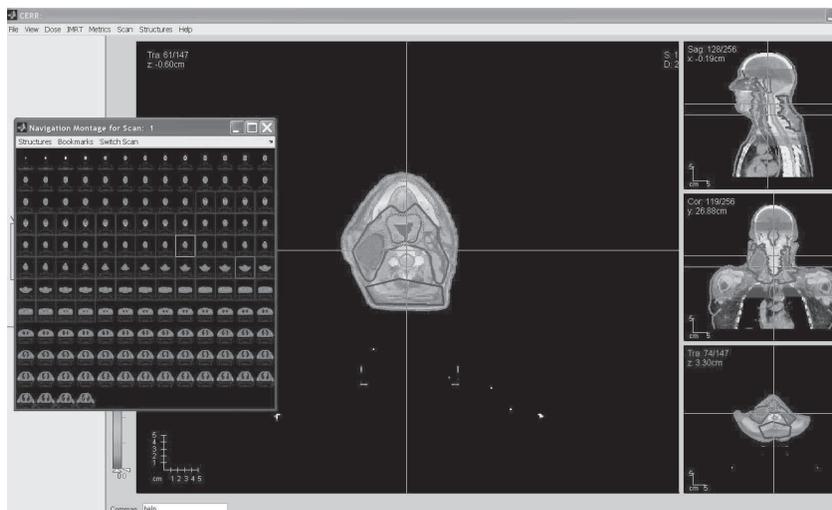


Figura 3: Estruturas consideradas na otimização do planejamento de tratamentos em IMRT.

	Dose média	Dose máxima	Dose prescrita
Medula	–	45 Gy	–
Tronco encefálico	–	54 Gy	–
Parótida esquerda	26 Gy	–	–
Parótida direita	26 Gy	–	–
PTV1	–	–	70.0 Gy
PTV2	–	–	59.4 Gy
Corpo	–	80 Gy	–

Tabela 1: Doses prescritas e doses de tolerância para todas as estruturas consideradas na otimização do planejamento de tratamentos em IMRT.

5 Resultados

Os métodos de procura em padrão aplicados ao problema de otimização angular foram testados usando dois exemplos clínicos de casos de tumores de cabeça e pescoço tratados retrospectivamente no IPOC. A nossa aproximação demonstrou vantagem clara quando comparada com aproximações usando derivadas e estratégias de *multistart*. Por limitação de espaço, clareza e simplificação da apresentação dos resultados não serão aqui apresentadas essas comparações. Como tarefa futura imediata do projeto de investigação que engloba estes trabalhos de investigação em radioterapia temos a implementação de algoritmos genéticos talhados para o problema de

otimização angular. A comparação com os resultados obtidos por outros algoritmos que não usam derivadas é também de grande interesse pois, tal como os métodos de procura em padrão, não ficam aprisionados em mínimos locais.

Um plano de tratamento típico, para casos de tumores de cabeça e pescoço, consiste em irradiar o paciente usando entre 5 e 9 direções de incidência da radiação coplanares equidistantes em torno do paciente. Planos de tratamento com 7 direções de incidência da radiação coplanares equidistantes foram usados no IPOC para tratar os pacientes e são usuais na prática clínica para tratamento deste tipo de casos (Aleman *et al*, 2008). Portanto, foram obtidos planos de tratamento com 7 direções de incidência da radiação coplanares usando métodos de procura em padrão, denotados por *7 PSM*, e comparados com os tratamentos usuais com 7 direções de incidência da radiação coplanares equidistantes, denotados por *7 equi*. O objetivo desta comparação é demonstrar que a qualidade dos planos de tratamentos podem ser melhoradas usando configurações angulares ótimas obtidas pela nossa aproximação para o problema de otimização angular.

O software de acesso livre CERR – computational environment for radiotherapy research (Deasy *et al*, 2003) – é usado amplamente para investigação do planeamento de tratamentos em IMRT por facilitar um acesso conveniente aos dados dos planos de tratamento dos pacientes, a sua visualização e análise, bem como permitir a obtenção de dados de dosimetria essenciais para a otimização de planos de tratamento. Este software desenvolvido em MATLAB foi eleito como a plataforma de suporte da nossa otimização do problema inverso do planeamento de tratamentos em radioterapia e serve-nos essencialmente para importar as TACs dos pacientes com as estruturas delineadas para posterior visualização e análise. Tal como referido na secção 2, apesar de o CERR também disponibilizar um algoritmo de Monte Carlo para o cálculo da matriz de dose D , por razões de economia de tempo, utilizamos o algoritmo QIB do CERR, menos preciso mas mais rápido (pencil beam). Os nossos testes foram feitos num PC Intel Core Duo com 2.66Ghz de velocidade, usando o MATLAB 7.4.0 (R2007a). Para cada um dos

dois casos de tumores de cabeça e pescoço, a taxa de sampling usada para o Corpo foi 32 e para as restantes estruturas foi 4, resultando em 21607 (23251) voxels e 933 (1226) sub-feixes para o plano de tratamentos com 7 direções equidistantes do primeiro (segundo) caso. Foi implementado o cálculo automatizado de dose para cada conjunto de direções de irradiação, em vez do tradicional cálculo de dose usando o módulo IMRTP acessível a partir da barra de opções do CERR. Esta automatização do cálculo da dose foi essencial à integração do nosso algoritmo de otimização angular no CERR.

Na resolução da nossa formulação não linear convexa do problema de otimização das intensidades usamos um algoritmo de regiões de confiança (*fmincon*) implementado na Optimization Toolbox do MATLAB 7.4.0. A última versão do SID-PSM foi usada como implementação base dos nossos métodos de procura em padrão. A escolha de pontos iniciais feita de forma aleatória é um procedimento comum, assim como o uso de estratégias de *multistart*. Como queríamos melhorar a qualidade dos planos de tratamento obtidos usando 7 direções de irradiação equidistantes, consideramos como pontos iniciais no processo de otimização angular exatamente estas configurações angulares equidistantes. A escolha destes pontos iniciais conjuntamente com a propriedade não decrescente da sucessão de pontos gerados pelo SID-PSM implica que cada iteração com sucesso corresponde a uma melhoria efetiva relativamente à configuração equidistante usual. O conjunto gerador positivo usado foi o conjunto $([e - e \ I - I])$, sendo I a matriz identidade e $e = [1 \dots 1]^T$. Cada uma destas direções corresponde à rotação de todas as direções de incidência da radiação no sentido dos ponteiros do relógio, à rotação de todas as direções de incidência da radiação no sentido contrário ao dos ponteiros do relógio, à rotação individual de cada direção de incidência da radiação no sentido dos ponteiros do relógio e à rotação individual de cada direção de incidência da radiação no sentido contrário à dos ponteiros do relógio, respetivamente. O tamanho da grelha inicial considerado foi 4. Tamanhos de grelha maiores foram testados, nomeadamente 8, 16 e 32, sem melhoria dos resultados, e com um aumento significativo do número de avaliações da função objetivo. Como os pontos considerados são vetores inteiros, todos os pontos da

sucessão gerada pelos métodos de procura em padrão terão valores inteiros enquanto o tamanho da grelha não for inferior a 1. Portanto, o critério de paragem adotado foi o tamanho da grelha tornar-se inferior a 1.

Uma das principais vantagens dos métodos de procura em padrão é a flexibilidade fornecida pelo passo de procura, onde qualquer estratégia pode ser aplicada desde que apenas um número finito de pontos seja testado. Isto permite a inserção de estratégias/heurísticas previamente usadas com sucesso na resolução do problema de otimização angular, fazendo uma pesquisa global e influenciando a qualidade do mínimo local encontrado pelo método. Na última versão do SID-PSM, o passo de procura calcula um único ponto usando modelos quadráticos com norma de Frobenius mínima a serem minimizados numa região de confiança, o que permitiu uma melhoria significativa na procura direta em funções “black-box” (Custódio *et al*, 2010) semelhantes às do problema de otimização angular em mãos.

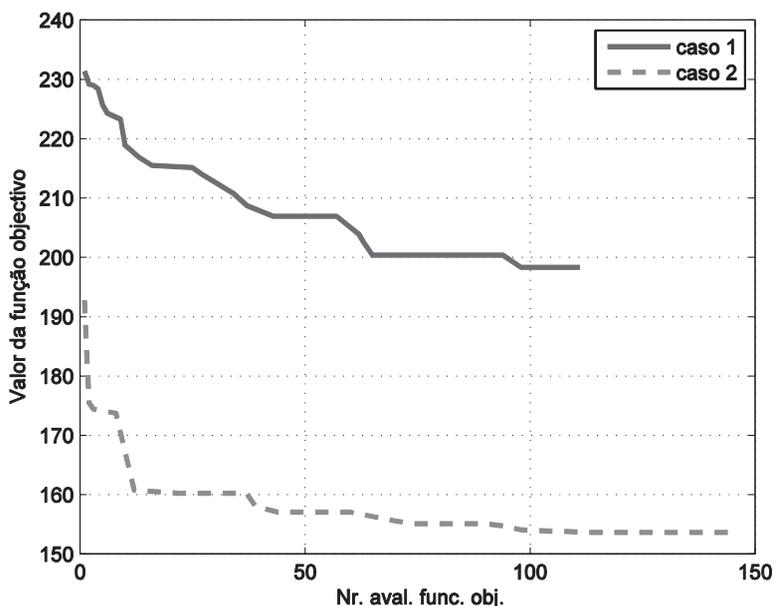


Figura 4: História do processo de otimização angular para cada caso de tumor de cabeça e pescoço em termos do valor da função objetivo.

A história do processo de otimização angular para cada caso de tumor de

cabeça e pescoço, em termos do valor da função objetivo, usando métodos de procura em padrão, é apresentada na Figura 4. O decréscimo do valor da função objetivo foi em média de 18% com um tempo computacional médio de 2.6 horas. O cálculo da dose usando o algoritmo QIB do CERR consumiu a maior parte do tempo computacional total. Para diferentes algoritmos de dose, diferentes métodos de otimização das intensidades e diferentes estratégias para a função objetivo, o tempo computacional total pode ter magnitudes significativamente diferentes. Portanto, tão importante quanto mencionar que em média demorou 2.6 horas ao método para convergir, é importante realçar o reduzido número de avaliações da função objetivo requerido pelos métodos de procura em padrão.

Apesar das melhorias do valor ótimo do problema de otimização das intensidades serem um bom indicativo, a qualidade dos resultados pode ser aferida considerando uma variedade de outras métricas. Tipicamente, a qualidade dos resultados é julgada pelos seus histogramas de dose-volume (DVH). O DVH mostra a fração de volume de uma estrutura que recebe pelo menos uma certa dose. Outra métrica usualmente usada para a avaliação de um plano é o volume do PTV que recebe pelo menos 95% da dose prescrita. Tipicamente, 95% do volume do PTV é requerido. As doses médias e/ou máximas dos OARs são também usualmente representadas para se verificar se há poupança dos OARs.

Na Figura 5, apresentam-se os DVHs onde é feita a comparação dos planos de tratamento *7 equi* com os planos de tratamento *7 PSM*. Como as parótidas são os órgãos mais difíceis de poupar, e como os limites das doses máximas da medula e do tronco encefálico foram sempre respeitados, de forma a ilustrar os resultados obtidos o mais nitidamente possível, os DVHs apenas incluem as parótidas e os PTVs. Os asteriscos indicam 95% do volume dos PTVs versus 95% das doses prescritas e os círculos indicam a dose média de tolerância para as parótidas. Podemos observar que a irradiação dos tumores em ambos os casos é satisfatória, com as curvas dos PTVs a passarem acima dos asteriscos. Isto significa que, em ambos os casos, os dois planos de tratamento cumprem o objetivo de irradiar pelo menos 95% do volume dos PTVs com 95% das doses

prescritas. Apesar de as irradiações serem muito semelhantes em ambos os planos de tratamento para os dois casos, pode-se verificar uma ligeira vantagem para as curvas obtidas pelo plano de tratamentos *7 PSM*, nomeadamente para o segundo caso. Como esperado, as principais diferenças residem na poupança das parótidas. Os planos de tratamento *7 equi* não conseguem satisfazer os limites das doses médias das parótidas ao contrário dos planos de tratamento *7 equi*. Em média, os planos de tratamento *7 PSM* reduzem a dose média nas parótidas em 3 Gy comparativamente aos planos de tratamento *7 equi*, facto que é bem ilustrado na Figura 5, o que demonstra a vantagem em usar planos de tratamento com direções de incidência da radiação obtidas pela nossa aproximação em comparação com planos que usam direções equidistantes usadas habitualmente na prática clínica.

É importante salientar que as melhorias obtidas nos planos de tratamento otimizados se refletem numa efetiva melhoria da qualidade de vida dos pacientes. Consideramos que a qualidade dos resultados obtidos pela nossa aproximação é satisfatória quer em termos da melhoria do valor da função objetivo quer em termos da consequente melhoria da qualidade dos planos de tratamento, aferida, e.g., pela poupança das parótidas com uma redução de 3 Gy da dose média comparativamente aos planos de tratamento com direções equidistantes. Contudo, o físico-médico pode não considerar relevante as melhorias obtidas por achar que não refletem uma efetiva melhoria da qualidade dos planos de tratamento oferecido aos pacientes. Isto constituiu uma diferença significativa entre exercícios académicos e este problema real em concreto onde a solução encontrada tem que ser validada. A obtenção de uma solução ótima em termos de valor da função objetivo é por norma o objetivo final de um exercício académico mas, neste problema real, a obtenção de soluções ótimas não leva necessariamente à aceitação do plano de tratamentos resultantes como melhor por parte do físico-médico. Por último, resta realçar que as soluções e correspondentes planos de tratamento obtidos pela nossa aproximação são consideradas pelos físicos-médicos, coautores deste trabalho, como refletindo uma efetiva melhoria relativamente aos planos de tratamento usados habitualmente na prática clínica.

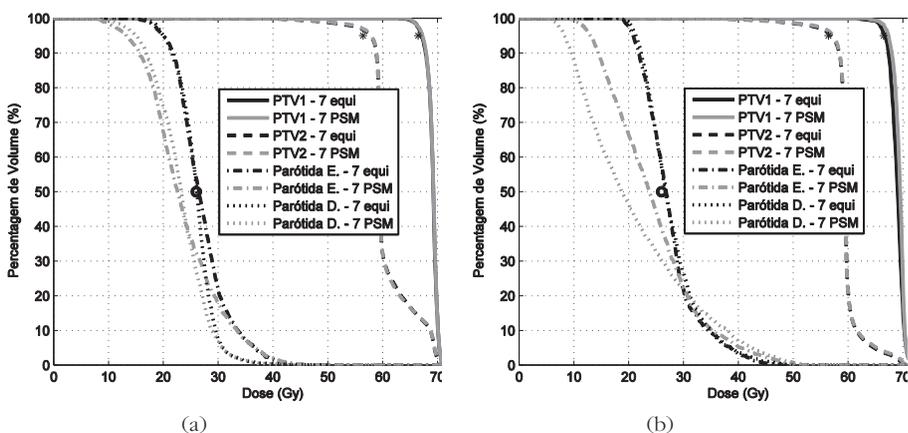


Figura 5: Histogramas de dose-volume comparando os resultados obtidos por 7 PSM e 7 equi para os casos 1 e 2 , (a) e (b) respetivamente.

6 Conclusão

O planeamento de tratamentos em radioterapia, nomeadamente o planeamento inverso, tem tido uma evolução fantástica, desde o uso dos primeiros modelos de otimização usados em radioterapia em 1968, até à atual aplicação clínica generalizada em todo o mundo. A investigação operacional seguiu (e contribuiu para) a evolução das novas máquinas e tecnologias e tem feito contribuições muito relevantes na contínua melhoria do planeamento dos tratamentos em radioterapia. Contudo, novos (e antigos) desafios continuam a necessitar de ser abordados, e este campo de investigação multidisciplinar continua a necessitar da contribuição da investigação operacional para oferecer aos pacientes planos de tratamento com cada vez mais qualidade.

Um dos principais desafios, ainda por resolver de forma satisfatória, é a escolha das melhores direções de incidência da radiação. Na prática clínica, a maioria das vezes, as direções de incidência da radiação continuam a ser escolhidas manualmente pelo físico-médico que confia essencialmente na sua experiência. Tipicamente são necessárias muitas iterações de tentativa e erro entre selecionar as direções de incidência da radiação e calcular as intensidades ótimas dos feixes correspondentes até que se obtenha um plano de tratamentos satisfatório. Este processo é fastidioso, moroso e sem garantias de produzir planos de tratamento de qualidade. Processos automatizados de confiança

deveriam substituir o físico-médico nesta tarefa e libertá-lo para outras tarefas importantes no processo de planeamento, nomeadamente assegurar a qualidade do mesmo. A investigação operacional tem aqui um papel muito importante e planos de tratamento com qualidade superior aos planos de tratamento equidistantes usualmente usados podem ser obtidos. Demonstramos isso na secção 5 onde os modelos e métodos por nós propostos para a resolução do problema de otimização angular originaram planos de tratamento com maior qualidade relativamente aos planos de tratamento usuais na prática clínica.

Resta ainda muito trabalho a fazer nesta área de investigação aplicada e multidisciplinar e a investigação operacional pode e deve continuar a fazer contributos para a melhoria clínica dos tratamentos em radioterapia. Um dos muitos exemplos ainda por explorar é a otimização das direcções de incidência da radiação não coplanares de extrema utilidade em casos de tumores de cabeça.

BIBLIOGRAFIA

Alber, M., Nusslin, F. (2001). Optimization of intensity modulated radiotherapy under constraints for static and dynamic MLC delivery. *Phys. Med. Biol.* **46**, 3229–3239.

Alberto, P., Nogueira, F., Rocha, H., Vicente, L.N. (2004). Pattern search methods for user-provided points: Application to molecular geometry problems. *SIAM J. Optim.* **14**, 1216–1236.

Aleman, D.M., Kumar, A., Ahuja, R.K., Romeijn, H.E., Dempsey, J.F. (2008). Neighborhood search approaches to beam orientation optimization in intensity modulated radiation therapy treatment planning. *J. Global Optim.* **42**, 587–607.

American Cancer Society. Cancer facts and figures (2010). <http://www.cancer.org/acs/groups/content/@epidemiologysurveillance/documents/document/acspc-026238.pdf>. Consultado online em 2012-01-11.

Bahr, G.K., Kereiakes, J.G., Horwitz, H., Finney, R., Galvin, J., Goode, K. (1968). The method of linear programming applied to radiation treatment planning. *Radiology* **91** 686–693.

Bortfeld, T. (2006). IMRT: a review and preview. *Phys. Med. Biol.* **51** 363–379.

Carlson, W.D., Morkovin, D. (1959). Time-dose relations for optimal radiation treatment of malignancy with minimal damage to normal tissues. *Miss. Valley Med. J.* **81** 287–288.

Censor, Y. (2003). Mathematical optimization for the inverse problem of intensity-modulated radiation therapy. In: J.R. Palta and T.R. Mackie (Eds.), *Intensity-Modulated Radiation Therapy: The State of The Art*, American Association of Physicists in Medicine (AAPM), Medical Physics Monograph No. 29, Medical Physics Publishing, Madison, Wisconsin, USA, 25–49.

Conn, A.R., Scheinberg, K., Vicente, L.N. (2009). *Introduction to Derivative-Free Optimization*. SIAM, Philadelphia.

- Craft, D. (2007). Local beam angle optimization with linear programming and gradient search. *Phys. Med. Biol.* **52**, 127–135.
- Craft, D., Halabi, T., Shih, H., Bortfeld, T. (2006). Approximating convex Pareto surfaces in multiobjective radiotherapy planning. *Med. Phys.* **33**, 3399–3407.
- Custódio, A.L., Rocha, H., Vicente, L.N. (2010). Incorporating minimum Frobenius norm models in direct search. *Comput. Optim. Appl.* **46**, 265–278.
- Custódio, A.L., Vicente, L.N. (2007). Using sampling and simplex derivatives in pattern search methods. *SIAM J. Optim.* **18**, 537–555.
- Das, S.K., Marks, L.B. (1997). Selection of coplanar or non coplanar beams using three-dimensional optimization based on maximum beam separation and minimized nontarget irradiation. *Int. J. Radiat. Oncol. Biol. Phys.* **38**, 643–655.
- Davis, C. (1954). Theory of positive linear dependence. *Am. J. Math.* **76**, 733–746 .
- Deasy, J.O., Blanco, A.I., Clark, V.H. (2003). CERR: A Computational Environment for Radiotherapy Research. *Med. Phys.* **30**, 979–985.
- Deasy, J.O., Lee, E.K., Bortfeld, T., Langer, M., Zakarian, K., Alaly, J., Zhang, Y., Liu, H., Mohan, R., Ahuja, R., Pollack, A., Purdy, J., Rardin, R. (2006). A collaboratory for radiation therapy planning optimization research. *Ann. Oper. Res.* **148**, 55–63 .
- Ehrgott, M., Guler, C., Hammacher, H.W., Shao, L. (2008). Mathematical optimization in intensity modulated radiation therapy. *4OR* **6**, 199–262.
- Ehrgott, M., Johnston, R. (2003). Optimisation of beam directions in intensity modulated radiation therapy planning. *OR Spectrum* **25**, 251–264.
- Ferris, M.C., Lim, J.-H., Shepard, D.M. (2003a). Optimization approaches for treatment planning on a Gamma Knife. *SIAM J. Optim.* **13**, 921–937.
- Ferris, M.C., Lim, J.-H., Shepard, D.M. (2003b). Radiosurgery treatment planning via nonlinear programming. *Ann. Oper. Res.* **119**, 247–260.
- Holder, A., Salter, B. (2004). A tutorial on radiation oncology and optimization. In: H. Greenber (Eds.), *Emerging Methodologies and Applications in Operations Research*, Kluwer Academic Press, Boston, USA.
- Lee, E.K., Fox, T., Crocker, I. (2003). Integer programming applied to intensity-modulated radiation therapy treatment planning. *Ann. Oper. Res.* **119**, 165–181.
- Lim, G.J., Ferris, M.C., Wright, S.J., Shepard, D.M., Earl, M.A. (2007a). An optimization framework for conformal radiation treatment planning. *INFORMS J. Comput.* **19**, 366–380.
- Lim, G.J., Lee, E.K. (2007b). *Optimization in Medicine and Biology*. Taylor and Francis, Auerbach Publications, USA.
- Liu, H.H., Jauregui, M., Zhang, X., Wang, X., Dongand, L., Mohan, R. (2006). Beam angle optimization and reduction for intensity-modulated radiation therapy of non-small-cell lung cancers. *Int. J. Radiat. Oncol. Biol. Phys.* **65**, 561–572.
- Men, C., Romeijn, H.E., Taskin, Z.C., Dempsey, J. F. (2007). An exact approach to direct aperture optimization in IMRT treatment planning. *Phys. Med. Biol.* **52**, 7333–7352.
- Preciado-Walters, F., Langer, M.P., Rardin, R.L., Thai, V. (2006). Column generation for IMRT cancer therapy optimization with implementable segments. *Ann. Oper. Res.* **148**, 65–79.
- RadiologyInfo. <http://www.radiologyinfo.org/en/info.cfm?pg=linac>. Consultado online em 2012-01-11.
- Rocha, H., Dias, J.M. (2009). *On the Optimization of Radiation Therapy Planning*. Inesc

Research Reports, nr. 15.

Rocha, H., Dias, J.M., Ferreira, B.C., Lopes, M.C. (2011). Influence of Sampling in Radiation Therapy Treatment Design. In: B. Murgante et al. (Eds.): ICCSA 2011, Lecture Notes in Computer Science, Springer-Verlag, Berlin, **6784** 215-230.

Rocha, H., Dias, J.M., Ferreira, B.C., Lopes, M.C. (2014a). Combinatorial optimization for an improved transition from fluence optimization to fluence delivery in IMRT treatment planning. To appear in Optimization. DOI:10.1080/02331934.2011.607498.

Rocha, H., Dias, J.M., Ferreira, B.C., Lopes, M.C. (2014b). Discretization of optimal beamlet intensities in IMRT: a binary integer programming approach. To appear in Mathematical and Computer Modelling. DOI:10.1016/j.mcm.2011.11.056.

Romeijn, H.E., Ahuja, R.K., Dempsey, J.F., Kumar, A. (2006). A new linear programming approach to radiation therapy planning problems. *Oper. Res.* **54**, 201–216.

Romeijn, H.E., Ahuja, R.K., Dempsey, J.F., Kumar, A. (2005). A column generation approach to radiation therapy treatment planning using aperture modulation. *SIAM J. Optim.* **15**, 838–862.

Romeijn, H.E., Ahuja, R.K., Dempsey, J.F., Kumar, A., Li, J. (2003). A novel linear programming approach to fluence map optimization for intensity modulated radiation therapy treatment planing. *Phys. Med. Biol.* **48**, 3521–3542.

Romeijn, H.E., Dempsey, J.F., Li, J. (2004). A unifying framework for multi-criteria fluence map optimization models. *Phys. Med. Biol.* **49**, 1991–2013.

Rowbottom, C.G., Webb, S., Oldham, M. (1998). Improvements in prostate radiotherapy from the customization of beam directions. *Med. Phys.* **25**, 1171–1179.

Shepard, D.M., Earl, M.A., Li, X.A., Naqvi, S., Yu, C. (2002). Direct aperture optimization: a turnkey solution for the step-and-shoot IMRT. *Med. Phys.* **29**, 1007–1018.

Spirou, S., Chui, C.-S. (1998). A gradient inverse planning algorithm with dose-volume constraints. *Med. Phys.* **25**, 321–333.

Stein, J., Mohan, R., Wang, X.H., Bortfeld, T., Wu, Q., Preiser, K., Ling, C.C., Schlegel, W. (1997). Number and orientation of beams in intensity-modulated radiation treatments. *Med. Phys.* **24**, 149–160.

Tervo, J., Kolmonen, P., Lyyra-Laitinen, T., Pintér, J.D., Lahtinen, T. (2003). An optimization-based approach to the multiple static delivery technique in radiation therapy. *Ann. Oper. Res.* **119**, 205–227.

Vaz, A.I.F., Vicente, L.N. (2007). A particle swarm pattern search method for bound constrained global optimization. *J. Global Optim.* **39**, 197–219.

(Página deixada propositadamente em branco)

CASO 18

AVALIAÇÃO DE DESEMPENHO DAS UNIDADES DE SAÚDE PARA A MELHORIA DOS CUIDADOS DE SAÚDE PRIMÁRIOS

M.C. Gouveia

IPC - ISCA Coimbra
INESC Coimbra
mgouveia@iscac.pt

L.C. Dias

Faculdade de Economia, Universidade de Coimbra
INESC Coimbra
lmc dias@fe.uc.pt

C.H. Antunes

DEEC-FCT Universidade de Coimbra
INESC Coimbra
cantunes@inescc.pt

M.A. Mota

ACES Baixo Mondego I
mariaaugustamota@gmail.com

E.M. Duarte

ACES Baixo Mondego I
emrduarte@gmail.com

E.M. Tenreiro

ACES Baixo Mondego I
Eduardo.tenreiro@gmail.com

RESUMO

Este artigo relata um estudo de avaliação de desempenho de unidades de saúde prestadoras de cuidados de saúde primários realizado com o envolvimento dos decisores. Foram utilizadas metodologias, algumas desenvolvidas para este caso em particular, que exploram as ligações entre a Data Envelopment Analysis (DEA) e Análise de Decisão Multicritério (Multi-Criteria Decision Analysis - MCDA), para incorporar preferências de gestão na avaliação de desempenho de doze unidades com dados relativos ao ano de 2010. Neste trabalho são apresentados dois modelos distintos, resultantes de discussões com um grupo de decisores, que combinam aspectos de avaliação como o acesso, os recursos utilizados, os serviços prestados e os resultados alcançados nos Cuidados de Saúde Primários. É utilizada uma variante do modelo DEA aditivo em duas fases para a avaliação de eficiência. Esta análise contribui para a identificação das melhores práticas, das fontes de ineficiência, bem como das diferenças em relação às melhores práticas. O interesse da abordagem reside não só na contribuição para conceber medidas corretivas adequadas baseadas nesses resultados, mas também na vertente de proporcionar informação relevante para a tomada de decisões mais fundamentadas sobre as metas dos indicadores de desempenho no futuro.

PALAVRAS-CHAVE

Data Envelopment Analysis; Análise de eficiência; Análise Multicritério; Cuidados de saúde primários.

1. Introdução

Os Cuidados de Saúde Primários (CSP) são considerados de crucial importância no sistema de saúde (SNS) português. A experiência internacional demonstra que os modelos de organização e gestão dos CSP, bem como as estratégias de financiamento adotadas e a forma de articulação com os restantes níveis de cuidados, podem potenciar ou, pelo contrário, comprometer o cumprimento dos princípios basilares dos sistemas de saúde, bem como das metas definidas pela política de saúde ao nível da equidade, acesso, eficiência e qualidade dos cuidados.

A reforma dos CSP, iniciada em 2005, tem como propósito melhorar o acesso aos cuidados de saúde e a qualidade dos serviços prestados. Ao mesmo tempo, visa ganhos de eficiência, porquanto a dimensão do SNS é de tal ordem que práticas inadequadas podem causar enormes desperdícios. O impacto desta reforma está agora a ser testemunhado, o que justifica a necessidade de avaliar os seus resultados de modo a identificar as melhores práticas nos CSP.

O objetivo do presente trabalho é avaliar a eficiência das unidades de CSP monitorizadas pelo Agrupamento de Centros de Saúde (ACES) Baixo Mondego I. Esta avaliação pretende conciliar algumas das prioridades estabelecidas pelo Ministério da Saúde em 2011, em particular a secção sobre o "custo-efetividade, custo-benefício e custo-utilidade" que constam do mapa geral das necessidades de novos conhecimentos elaborado pela Coordenação Estratégica da Reforma dos CSP. Este trabalho envolve decisores reais participantes na avaliação de desempenho das unidades de CSP que monitorizam, recorrendo a uma nova abordagem. São apresentados neste texto apenas dois dos modelos analisados, que visam estabelecer uma ligação entre aspectos de avaliação como o acesso, os recursos utilizados, os serviços prestados e os resultados obtidos. A escolha dos fatores incluídos nos modelos foi resultado de discussões com o grupo de decisores (GD), constituído por profissionais do ACES Baixo Mondego I, com base no Plano de Desempenho (Coordenação Estratégica dos CSP [CECSP], 2010). A análise realizada utilizou dados relativos ao ano de 2010, compreendendo todas as 12 unidades prestadoras de cuidados de saúde que o ACES Baixo Mondego I monitoriza e acompanha.

A contribuição deste trabalho para o ACES Baixo Mondego I reside, essencialmente, na capacidade de usar múltiplos fatores num modelo integrado, dando a possibilidade de identificar as melhores práticas, as fontes de ineficiência, bem como as diferenças para as melhores práticas. Esta abordagem proporciona informações no sentido de serem introduzidas medidas corretivas e ajuda a estabelecer as metas de desempenho para o futuro.

O estudo da avaliação de eficiência na prestação de CSP tem sido alvo da atenção de vários autores, especialmente no Reino Unido. Na maioria dos trabalhos publicados para avaliar a prestação de serviços de saúde e qualidade de atendimento constata-se que são utilizados indicadores de *performance* (IPs) (Donabedian, 1980; Baker, 1992; Campbell et al., 1998; McColl et al., 1998; Van den Hombergh et al., 1998; Proctor e Campbell, 1999). Contudo, os modelos de avaliação de desempenho que são baseados num conjunto de IPs para comparar as entidades conduzem a análises pobres, uma vez que os diferentes IPs são analisados individualmente. Thanassoulis et al. (1996) e Giuffrida et al. (2000) apontaram as desvantagens do uso de IPs na avaliação de desempenho e propuseram técnicas alternativas, tais como a *Data Envelopment Analysis* (DEA). A DEA é uma abordagem não paramétrica, baseada em programação linear (PL), para avaliar a eficiência relativa de unidades organizacionais (*Decision Making Unit* - DMUs) semelhantes, gerando um indicador global de eficiência para cada unidade em avaliação, com base nos *inputs* que consome e nos *outputs* que produz. Desde meados de 1980 que esta metodologia é muitas vezes escolhida para avaliar a eficiência dos serviços de saúde (para revisões mais abrangentes ver Hollingworth, 2003; Amado & Dyson, 2008). Existem alguns estudos realizados sobre a eficiência na prestação de CSP que recorrem à técnica DEA (e.g., Huang & McLaughlin, 1989; Linna et al., 2003; Kirigia et al., 2004; Rosenman & Friesner, 2004; Rouse et al., 2011; Thanassoulis et al., 2012) e relativamente poucos estudos (Salinas-Jimenez & Smith, 1996; Wagner et al. 2003, Amado & Santos, 2009; Milliken et al., 2011; Nuti et al., 2011) incluem medidas de qualidade da assistência/atendimento na medição da eficiência.

Apesar de, nos últimos anos, se terem realizado muitos estudos de avaliação de eficiência em diferentes áreas utilizando DEA, poucos incorporam na análise as preferências dos gestores/decisores (ver Thanassoulis et al., 2004). A estrutura matemática dos modelos clássicos DEA permite total liberdade na escolha dos coeficientes de ponderação (pesos) dos *inputs* e dos *outputs*. Logo, a solução pode gerar coeficientes de ponderação nulos para variáveis importantes e eventualmente relações entre as variáveis, resultantes do próprio contexto do estudo, podem não ser consideradas. Como consequência, os resultados atingidos podem ser criticados pelo decisor, ficando comprometida a sua confiança na metodologia. Para ultrapassar este problema utilizam-se restrições aos coeficientes de ponderação com o objetivo de incorporar a opinião do decisor (Allen et al., 1997). As restrições mais comuns são as baseadas em restrições de razões entre pesos, na qual são definidos os limites inferiores e superiores atribuídos às razões entre os coeficientes de ponderação de duas quaisquer variáveis. Estas restrições definem uma região preferida da fronteira, onde são observadas as melhores práticas do modelo de produção em avaliação.

Este trabalho utiliza o método de duas fases desenvolvido por Gouveia et al. (2008), que é uma variante do modelo DEA aditivo com projeções orientadas (Ali et al., 1995), recorrendo a conceitos desenvolvidos no âmbito da MCDA com informação imprecisa (Athanassopoulos & Podinovski, 1997; Dias & Clímaco, 2000). Com as DMUs a assumir o papel de alternativas a serem comparadas, neste método os fatores de *input* e *output* são convertidos em funções de utilidade de acordo com informação de preferências fornecida pelos decisores e as funções de utilidade aditiva são utilizadas para agregar as utilidades (marginais) associadas a cada alternativa, com base na Teoria da Utilidade Multi-Atributo (*Multi-Attribute Utility Theory* - MAUT) (ver Keeney & Raiffa, 1976). Ao invés de deixar cada DMU escolher livremente os coeficientes de ponderação associados a estas funções, foram incluídas no modelo as preferências dos decisores no sentido de construir restrições aos coeficientes de ponderação seguindo alguns dos protocolos de diálogo recomendados (ver von Winterfeldt & Edwards, 1986). Neste método de duas fases, a medida de eficiência atribuída a cada DMU tem um significado intuitivo, correspondendo

a uma medida "*min-max regret*" (perda de utilidade). O modelo de Gouveia et al. (2008) foi neste estudo adaptado para incluir o conceito de super-eficiência (Gouveia et al., 2013), a fim de assinalar, entre as DMUs eficientes, quais o são com maior margem.

O método das duas fases de Gouveia et al. (2008), com as modificações para incluir o conceito de super-eficiência, é apresentado na secção 2. A secção 3 descreve brevemente o sistema de cuidados de saúde português. Na secção 4 são apresentados os dados, os modelos discutidos com o GD e os protocolos de diálogo utilizados para obter as suas preferências. Alguns resultados dos modelos propostos são exibidos e analisados na secção 5. A secção 6 destaca as propostas de melhoria para as unidades ineficientes e os novos desenvolvimentos do método apresentado na secção 2. As considerações finais são apresentadas na secção 7.

2. O método das duas fases baseado no modelo DEA aditivo com incorporação de preferências

As origens da DEA remontam aos trabalhos de Farrell (1957) sobre métodos para avaliação da produtividade, em que era reconhecida a necessidade de dispor de indicadores de eficiência global, e não de mera produtividade de fatores, que resultassem da combinação de múltiplos *inputs* e *outputs*. O primeiro modelo de DEA, assumindo rendimentos constantes à escala (*Constant Returns to Scale* - CRS), resultou do trabalho de Charnes, Cooper e Rhodes (Charnes et al., 1978), motivado pela avaliação de programas educacionais para estudantes desfavorecidos em escolas públicas dos Estados Unidos da América. A característica fundamental deste modelo (geralmente designado na literatura científica por modelo CCR ou *Engineering Ratio*) é, para cada entidade sob avaliação, a transformação da situação de múltiplos *outputs* / múltiplos *inputs* na situação de *output* virtual / *input* virtual. Para cada DMU, a razão entre o *output* virtual e o *input* virtual fornece uma medida de eficiência, a qual é função de um conjunto de multiplicadores que constituem as variáveis de um modelo de programação linear fraccionária (depois reduzido a um modelo de PL). As DMUs devem possuir um grau de controlo sobre os *inputs* consumidos e os *outputs* produzidos, ou seja,

autonomia de gestão, operando num ambiente relativamente homogêneo. Neste modelo espera-se uma variação proporcional dos produtos a partir da alteração de recursos em todos os níveis de escala.

A extensão do modelo CCR para o caso de rendimentos variáveis à escala (*Variable Returns to Scale - VRS*), foi proposta em Banker, Charnes e Cooper (Banker et al., 1984), modelo designado por BCC, onde basicamente se introduz uma restrição que garante que cada DMU em estudo será comparada com uma combinação convexa das restantes DMUs. De acordo com Ozcan (1993), pode dizer-se que o modelo VRS é a escolha mais adequada quando se pretende comparar unidades de saúde de tamanhos muito diferentes entre si, de modo que as unidades tenham como referência as outras com perfil semelhante.

Os diferentes modelos DEA procuram determinar quais das n DMUs formam a fronteira eficiente no sentido da eficiência de Pareto-Koopmans. Uma DMU será eficiente se, comparativamente às demais, tiver maior produção para quantidades fixas de recursos (orientação a *outputs*) e/ou utilizar menos recursos para gerar uma quantidade fixa de produtos (orientação a *inputs*). A DEA constrói uma fronteira empírica eficiente, ou "superfície envolvente" de máxima eficiência (Ali & Seiford, 1993). Isto permite a identificação de unidades de referência (*benchmarks*) cujos valores de eficiência servem como modelo para as restantes unidades envolvidas pela fronteira de eficiência.

Charnes et al. (1985) propuseram o modelo DEA aditivo como uma alternativa para o modelo BCC, que também considera que os rendimentos podem ser variáveis à escala, mas não requer uma escolha entre modelo orientado a *inputs* e modelo orientado a *outputs*.

A distinção entre modelo DEA aditivo e os modelos orientados reside no facto destes últimos terem um mecanismo de projecção em duas etapas. Nos modelos orientados, numa primeira etapa todos os fatores são reduzidos ou aumentados na mesma proporção até à superfície envolvente, e numa segunda etapa são calculadas as folgas que dão o ponto projetado na fronteira eficiente. Por sua vez, o modelo DEA aditivo utiliza somente a segunda etapa. O valor obtido na primeira etapa de modelos orientados fornece uma medida da

ineficiência, entre 0 e 1, para as DMUs ineficientes de acordo com a orientação utilizada.

Seja $\{DMU_j: (j = 1, \dots, n)\}$ o conjunto das n DMUs a ser avaliadas. Cada DMU j consome a quantidade $x_{ij} (i = 1, \dots, m)$ do *input* i e produz a quantidade $y_{rj} (r = 1, \dots, p)$ do *output* r . O modelo aditivo a resolver para cada DMU tem a formulação primal (1):

$$\begin{aligned}
 \min_{\lambda, s^-, s^+} z_k &= -\left(\sum_{i=1}^m s_i^- + \sum_{r=1}^p s_r^+\right) \\
 \text{s. a: } -\sum_{j=1}^n \lambda_j x_{ij} - s_i^- &= -x_{ik}, \quad i = 1, \dots, m \\
 \sum_{j=1}^n \lambda_j y_{rj} - s_r^+ &= y_{rk}, \quad r = 1, \dots, p \\
 \sum_{j=1}^n \lambda_j &= 1 \\
 \lambda_j, s_i^-, s_r^+ &\geq 0, \quad j = 1, \dots, n; i = 1, \dots, m; r = 1, \dots, p
 \end{aligned} \tag{1}$$

O modelo DEA aditivo não devolve uma medida para as DMUs ineficientes. Este modelo apenas mede o excesso de *inputs* (s^-) e o déficit de *outputs* (s^+), em que uma DMU k está a operar quando comparada com as DMUs que operam na fronteira eficiente. Se o valor obtido for negativo, então a DMU em análise está a operar de forma ineficiente em alguns fatores. Este valor é simétrico da soma das distâncias em cada dimensão à superfície envolvente (distância L_1).

O método de duas fases desenvolvido por Gouveia et al. (2008) é uma variante do modelo DEA aditivo com projeções orientadas (Ali et al., 1995), que incorpora as preferências do decisor utilizando conceitos da MAUT com informação imprecisa (Dias & Clímaco, 2000).

O objetivo da MAUT é encontrar uma única expressão que represente a atratividade global de cada alternativa, de acordo com as preferências do decisor, tendo em consideração os múltiplos critérios. É também propósito desta teoria simplificar a construção de funções de utilidade multicritério deixando que o decisor foque a sua atenção num critério de cada vez, e posteriormente na agregação dos vários critérios, em vez de avaliar diretamente a utilidade global.

A eliciação de preferências é realizada interativamente, em pequenos passos, de forma construtiva. von Winterfeldt e Edwards (1986) apresentam detalhadamente várias técnicas de questionar o decisor no sentido de construir as funções de utilidade marginais (parciais) consistentes com as respostas obtidas.

Seja $u_c(a_j)$ a utilidade de uma alternativa a_j de acordo com o critério c ($c = 1, \dots, q$). A função de utilidade aditiva agrega estas funções através de uma soma ponderada na qual são usados q coeficientes de escala, w_1, \dots, w_q : $U(a_j) = \sum_{c=1}^q w_c u_c(a_j)$ (ver Keeney & Raiffa, 1976). A quantidade $u_c(a_j)$ pode ser interpretada como uma medida de desempenho da alternativa a_j no critério c . Estas funções são definidas para cada critério convertendo a escala de desempenhos originais na escala de utilidades $[0,1]$. Por sua vez, w_c é o coeficiente de ponderação que reflete a importância da função u_c ($c = 1, \dots, q$). Os coeficientes de ponderação são normalizados de forma a que $\sum_{c=1}^q w_c = 1$, com $w_c \geq 0, \forall c = 1, \dots, q$, podendo ainda ser acrescentadas pelo decisor outras restrições, caso se pretenda impedir combinações de valores para estes coeficientes que não estejam alinhados com os objetivos da organização.

Este procedimento ultrapassa o problema das escalas associado ao modelo DEA aditivo, uma vez que todas as medidas de *input* e *output* são convertidas em utilidade. Além disso, os coeficientes de ponderação usados na agregação adquirem um significado específico: são os coeficientes de escala das funções de utilidade. Os coeficientes de ponderação são valores escolhidos por cada DMU, a fim de se tornar a melhor DMU (se possível) ou então no sentido de minimizar a diferença de utilidade para a melhor DMU, ao invés de serem fixados previamente como no modelo de Ali et al. (1995).

A teoria de decisão clássica oferece algumas regras (tais como *max-min* e *min-max regret*) que podem ser aplicadas a situações onde as funções de utilidade são incertas.

No desenvolvimento do método das duas fases utilizou-se a regra *min-max regret*, referente ao conceito de perda de utilidade de escolher uma alternativa face a outra. A ideia é determinar para cada alternativa o potencial “arrependimento” associado à sua escolha, observando quão melhor poderia outra alternativa ser. Para um dado vetor de coeficientes de ponderação, a

perda de utilidade de escolher a alternativa a_j em vez de outra alternativa $a_j \neq a_k$ é dada por (2):

$$R(a_k, a_j) = \max_{j \neq k} U(a_j) - U(a_k) \quad (2)$$

A perda de utilidade máxima associada a uma dada alternativa a_k , $R_{max}(a_k)$, é dada pela máxima perda de oportunidade associada à escolha dessa alternativa (3):

$$R_{max}(a_k) = \max_w \{(\max_{j \neq k} U(a_j)) - U(a_k)\} \quad (3)$$

A decisão ótima *min-max regret* minimiza a perda no pior caso de acordo com possíveis realizações do vetor de coeficientes w , i.e., alternativas com $R_{max}(a_k)$ mínimo são preferidas. Assim, a medida de eficiência atribuída a cada DMU ganha um significado intuitivo: corresponde a uma medida "*min-max regret*" (perda de utilidade).

Considerando que as alternativas a avaliar de acordo com q ($q = m+p$) critérios são as DMUs, definem-se q funções de utilidade de modo a que o pior nível da escala tenha o valor 0 e o melhor nível da escala tenha o valor 1. Por outras palavras, quanto melhor (neste caso menor) for o desempenho associado a um fator de *input* maior será a sua utilidade (mais próxima estará de 1); para o caso dos *outputs*, quanto maior for o desempenho maior será a sua utilidade. Assim, depois de todos os fatores terem sido convertidos em utilidades, passam a ser considerados como *outputs* a maximizar. Para cada alternativa (DMU), atendendo ao modelo aditivo MAUT, a utilidade obtida é dada por: $U(DMU_j) = \sum_{c=1}^q w_c u_c(DMU_j)$, onde $w_c \geq 0, \forall c = 1, \dots, q$ e $\sum_{c=1}^q w_c = 1$. Os coeficientes de escala w_1, \dots, w_q são os coeficientes de ponderação das funções de utilidade e traduzem as taxas de compensação (*trade-offs*) da utilidade do decisor entre os vários fatores.

Para atender ao conceito de super-eficiência, introduzido em DEA por Andersen e Petersen (1993), o método de duas fases de Gouveia et al. (2008) foi modificado no sentido de dar a possibilidade de discriminar as unidades eficientes (Gouveia et al., 2013). Assim, o problema (4) é resolvido (Fase 1):

$$\begin{aligned}
& \min_{d_k, w} d_k \\
& \text{s. a: } \sum_{c=1}^q w_c u_c(DMU_j) - \sum_{c=1}^q w_c u_c(DMU_k) \leq d_k, \quad j = 1, \dots, n; j \neq k \\
& \quad \sum_{c=1}^q w_c = 1 \\
& \quad w_c \geq 0, c = 1, \dots, q
\end{aligned} \tag{4}$$

O valor ótimo d_k^* indica a distância definida pela diferença de utilidade para a melhor de todas as DMUs (excluindo a que está em avaliação).

O propósito desta abordagem é, para a DMU k , calcular o vetor w de ponderação que minimiza a distância (a diferença de utilidade) da unidade em causa para a melhor (note-se que a melhor alternativa dependerá também de w), excluindo-se ela própria do conjunto de referência. Depois, no caso de ser uma DMU ineficiente ou fracamente eficiente, encontra um alvo (par) eficiente (Fase 2). Os detalhes deste processo são os seguintes:

Fase 1: Converter os fatores de *input* e *output* em escalas de utilidade. Calcular a medida de eficiência, d_k^* , de cada DMU, $k = 1, \dots, n$, e o vetor de ponderação correspondente.

Fase 2: Se $d_k^* \geq 0$, então resolver o problema "aditivo ponderado" (5), usando o vetor de ponderação ótimo resultante da Fase 1, w_k^* , e determinar o correspondente ponto projetado da DMU em análise.

$$\begin{aligned}
& \min_{\lambda, s} z_k = - \sum_{c=1}^q w_c^* s_c \\
& \text{s. a: } \sum_{j=1, j \neq k}^n \lambda_j u_c(DMU_j) - s_c = u_c(DMU_k), \quad c = 1, \dots, q \\
& \quad \sum_{j=1, j \neq k}^n \lambda_j = 1 \\
& \quad \lambda_j \geq 0, s_c \geq 0, \quad j = 1, \dots, k-1, k+1, \dots, n; c = 1, \dots, q
\end{aligned} \tag{5}$$

Se o valor ótimo d_k^* da função objetivo em (4) é não positivo, então a DMU k em avaliação é eficiente (se o valor ótimo for nulo a DMU poderá ser fracamente eficiente). Caso contrário, é ineficiente e d_k^* é a diferença mínima de utilidade para a melhor DMU (ou seja, a DMU com maior utilidade global). Contudo, com a modificação feita no método original, é possível discriminar as unidades eficientes. As variáveis $\lambda_1, \dots, \lambda_j, j = 1, \dots, k-1, k+1, \dots, n$, definem uma combinação convexa das n DMUs. As DMUs eficientes (podendo ser apenas uma) que definem a combinação convexa com $\lambda_j > 0$ são chamadas de "pares" da DMU k em avaliação. A combinação convexa corresponde a um

ponto na fronteira eficiente, melhor do que a DMU k com uma diferença de valor de s_c (folga) em cada critério c .

O modelo (5) admite alvos ótimos alternativos, cada um dos quais correspondendo a uma forma diferente de anular o valor da diferença d_k^* . Esses alvos correspondem a diferentes projeções sobre a fronteira eficiente. Existe a possibilidade de escolher quais os fatores que se permite alterar e limitar a melhoria máxima que se pode exigir em cada fator para atingir o valor global de utilidade do alvo, através da inserção de restrições às folgas s_c . O objetivo é evitar a fixação de alvos irrealistas, permitindo que as propostas de melhoria das unidades ineficientes sejam concordantes com os requisitos do decisor.

3. O sistema de saúde português – Cuidados de saúde primários

Os CSP são a primeira e principal forma de contacto da população com o sistema de saúde. Em 1979, o SNS foi criado com o objectivo de defender o direito à saúde e o dever de defesa e protecção da saúde da população. O SNS forneceria um serviço nacional, universal, integral e gratuito, assegurado pelo Estado. A especialidade de medicina geral e familiar foi também criada e implementada. Apesar da maioria da população ser utilizadora do SNS, há vários subsistemas de seguros de saúde, mas a população coberta por estes também tem acesso aos serviços do SNS, traduzindo-se, muitas vezes, numa dupla cobertura.

Sucessivas tentativas de reforma do SNS conduziram ao entendimento de que não foi dada a devida atenção ao acesso aos cuidados de saúde, sendo necessário projetar políticas de saúde que se traduzam em soluções técnicas que realmente funcionem. É igualmente importante que a base política de qualquer reforma seja alicerçada numa análise rigorosa das questões económicas, sociais, culturais e políticas (Observatório Português dos Sistemas de Saúde [OPSS], 2008).

Em 2005 começou uma nova fase no sector da saúde, especialmente em relação aos CSP, focando em particular a reorganização dos centros de saúde, com a criação de Unidades de Saúde Familiar (USF), constituídas por equipas multidisciplinares com autonomia técnica e funcional, e de outras soluções

mais adequadas para a prestação de cuidados de saúde primários. O impacto desta reforma está a ser acompanhado, o que justifica a necessidade de avaliar os seus resultados, em particular no que diz respeito às melhores práticas em CSP.

Do ponto de vista organizacional dos CSP, pode referir-se a existência, em 2005, de 351 centros de saúde, 1823 extensões de saúde e 254 Serviços de Atendimento Permanente (SAPs) em Portugal Continental, de acordo com o Instituto de Gestão Informática e Financeira da Saúde (2006). Em termos de modelo de contratualização, constatava-se a ausência de um papel claro e bem definido de um agente que desse resposta às necessidades dos diversos atores com responsabilidades no sector da saúde, especialmente no que diz respeito à negociação de objetivos e metas. De acordo com um estudo da Organização para a Cooperação e Desenvolvimento Económico (2006), os gastos com a saúde aumentariam no futuro, estimando-se representar 13.1% do PIB em 2050, se não fossem implementadas reformas para contrariar esta tendência.

A prestação de serviços de saúde é caracterizada por um grande número de utilizadores sem acesso a médico de família, especialmente na área geográfica do litoral e nas grandes áreas urbanas. Os cidadãos podem escolher livremente o centro de saúde onde se inscrevem, bem como o seu profissional de medicina geral e familiar. Este médico fica designado como médico de família do utente para todas as consultas. No entanto, em 2005, cerca de 10.6% dos utentes do SNS não tinham médico de família (OPSS, 2007).

Durante todo o ano de 2007 houve um compromisso do Ministério da Saúde para a criação de mais USF e, a 31 de dezembro de 2007, já se encontravam 104 USF em funcionamento.¹

Foram estabelecidos modelos de acompanhamento, avaliação e contratualização, com padrões de exigência elevados. Estes conduziram à criação de um clima de confiança entre os profissionais de saúde envolvidos e as Administrações Regionais de Saúde (ARS). O ambiente criado mobilizou cerca de 1300 médicos, 1300 enfermeiros e 950 funcionários administrativos, que aceitaram o desafio e se disponibilizaram para trabalhar em USF. Para além destas mudanças organizacionais, a lista de utentes inscritos, em média,

¹ Dados retirados do sítio www.mcsp.min-saude.pt, consultado em 15 de janeiro de 2012.

aumentou para cerca de 1750 utentes por profissional de medicina geral e familiar (OPSS, 2008).

A 22 de fevereiro de 2008 foi publicado o Decreto-Lei n.º 28/2008, que regulamenta a criação dos ACES. Estes são definidos como “*serviços públicos de saúde com autonomia administrativa, constituídos por várias unidades funcionais, que agrupam um ou mais centros de saúde, e que têm por missão garantir a prestação de cuidados de saúde primários à população de determinada área geográfica*”. A constituição dos ACES pressupõe a sua organização em unidades funcionais, nomeadamente em USF, unidades de cuidados de saúde personalizados (UCSP - integram os restantes médicos de família e outros profissionais, mas não estão organizados em USF), unidades de cuidados na comunidade, unidades de saúde pública e unidades de recursos assistenciais partilhados. O ACES também prevê a existência de unidades de apoio à sua governação, como é o caso do Conselho da Comunidade, o Conselho Clínico, a Unidade de Apoio à Gestão e o Gabinete do Cidadão. Importa realçar que a autonomia do ACES é administrativa e está intimamente relacionada com a existência de um contrato-programa, a ser estabelecido entre o ACES e a ARS, Instituto Público.

De acordo com o Decreto-Lei n.º 222/2007, publicado em 29 de maio, as cinco ARS existentes “têm por missão garantir à população da respectiva área geográfica de intervenção o acesso à prestação de cuidados de saúde de qualidade, adequando os recursos disponíveis às necessidades em saúde e cumprir e fazer cumprir o Plano Nacional de Saúde na sua área de intervenção”.

Em 2010, pela primeira vez, os Diretores Executivos/Conselhos Clínicos, com o apoio dos Departamentos de Contratualização das ARS, desencadearam a contratualização interna com as USF, já iniciada anteriormente, e o mesmo modelo foi adaptado para as UCSP.

Conforme consta no contrato-programa de 2010, para um futuro a curto prazo “*espera-se que o processo de contratualização e a sofisticação da alocação dos recursos permita uma efetiva autonomização e responsabilização dos ACES pela efetividade dos cuidados de saúde prestados à população*”.

4. Dados e modelos desenvolvidos atendendo às preferências do Grupo de Decisores

Esta secção descreve a avaliação de desempenho das 12 unidades prestadoras de cuidados de saúde monitorizadas pelo ACES Baixo Mondego I, com dados relativos a 2010. As unidades a avaliar são USF e UCSP, uma vez que todas elas estão no mesmo contrato-programa estabelecido entre o ACES e a ARS e prosseguem os mesmos objetivos. A avaliação de desempenho destas unidades é realizada periodicamente pelo ACES, tendo como base os indicadores que constam no Plano de Desempenho, sem atender a diferenças existentes a nível organizacional. Portanto, e a pedido do GD, manteve-se o mesmo critério na avaliação proposta neste estudo, podendo-se considerar que as unidades são homogéneas.

A monitorização e acompanhamento destas unidades assenta num conjunto de 15 indicadores agrupados em quatro grandes áreas: acesso, desempenho assistencial, qualidade percecionada e desempenho económico, cujas metas são negociadas entre a unidade e o Diretores Executivos/Conselhos Clínicos, tendo em consideração os objetivos de saúde estabelecidos pelo ACES e pela ARS. Um bom desempenho por parte das unidades poderá traduzir-se na atribuição de recompensas de vária ordem.

Os dados analisados foram extraídos de diversos sistemas de informação, desde os que são utilizados para monitorizar as atividades assistenciais até aos do departamento de contabilidade e gestão. A fim de incluir na análise um fator que reflete a qualidade percecionada de serviço, foi utilizada informação sobre a satisfação dos utentes com a qualidade do serviço, proveniente de um inquérito (*Avaliação sobre o funcionamento das Unidades de Saúde do ACES*, 2010). Este abrange 19 perguntas sobre acessibilidade aos serviços, informação e atendimento, instalações das unidades de saúde e dados sociodemográficos. O inquérito não considerou um indicador de satisfação global, mas foi entendido pelo GD que a pergunta "*quando adoecer de repente, consegue consulta no médico de família no próprio dia?*" (resposta sim/não) seria uma boa aproximação da satisfação dos utentes com a qualidade do serviço.

Face ao objetivo de avaliação do desempenho das 12 DMUs (USF e UCSP) monitorizadas pelo ACES Baixo Mondego I, foi explicada ao GD a metodologia

proposta por Gouveia et al. (2008). O GD aceitou esta abordagem que considerou muito útil, dada a possibilidade de usar múltiplos fatores num modelo integrado e a capacidade de identificar as melhores práticas, as fontes de ineficiência, as diferenças relativamente às melhores práticas, bem como as oportunidades de melhoria.

4.1. Definição dos modelos

Depois de traçados os objetivos do estudo, foram definidos os fatores de *input* e *output* para cada modelo, atendendo aos indicadores das principais áreas do Plano de Desempenho, que eram essencialmente rácios. Como resultado de reuniões surgiram vários modelos diferentes, pois vários entendimentos possíveis decorriam das diferentes perspectivas de avaliação utilizadas. No entanto, neste trabalho apenas serão apresentados dois desses modelos.

Com o objetivo de combinar desempenho assistencial, qualidade percebida e desempenho económico, foi desenvolvido um modelo de eficiência face a custos. Neste, os custos associados à prestação de cuidados de saúde primários foram considerados como *inputs*; como *outputs* foram incluídos os fatores considerados como a principal atividade dos centros de saúde, juntamente com o fator relativo à satisfação dos utentes com a qualidade do serviço de saúde prestado. Os *outputs* (à exceção do último referido) são separados por tipo de consulta e por grupo etário dos utentes, a fim de controlar o “*case-mix*” (Tabela 1). Este modelo foi chamado de “eficiência dos serviços de saúde prestados atendendo às características do utente”, a partir de agora referido como modelo 1.

Uma vez que no modelo 1 não está refletido o trabalho de enfermagem, houve a necessidade por parte do GD de criar um segundo modelo de eficiência (modelo 2). Este modelo, considerado mais abrangente, alia os vários eixos de avaliação mencionados no Plano de Desempenho, expressando em simultâneo todo o trabalho médico, de enfermagem e administrativo. Concordou-se com a inclusão de nove fatores para a construção deste modelo (ver Tabela 2), cujos *inputs* são os indicadores económicos do Plano de Desempenho, os primeiros quatro *outputs* estão relacionados com o acesso

(número de utentes inscritos com pelo menos uma consulta no ano de referência) e desempenho assistencial e o último *output* é relativo à qualidade percebida. Este modelo fornece ao GD uma perspectiva diferente das unidades em avaliação e foi considerado muito importante, uma vez que as recompensas institucionais são dependentes do desempenho das unidades em alguns destes fatores, nomeadamente o número de domicílios médicos e de enfermagem.

Tabela 1. Fatores do modelo 1.

<i>Inputs</i>	<i>Outputs</i>
X_{MCDT} : custo total com meios complementares de diagnóstico e tratamento faturados ao SNS	y_{CPF} : número de consultas médicas de planeamento familiar
X_{MED} : custo total com medicamentos faturados ao SNS	y_{CSM} : número de consultas de saúde materna
X_{RH} : custo total com recursos humanos	y_{CSI} : número de consultas de saúde infantil (idade 0-18)
X_{OC} : custos totais com medicamentos, materiais de consumo clínico e outros	y_{CSA} : número de consultas de saúde a adultos (idade 19-64)
	y_{CSS} : número de consultas de saúde a seniores (idade superior ou igual a 65)
	y_{SQ} : satisfação com a qualidade dos serviços

Amplamente discutidos com o GD e em concordância com os indicadores estabelecidos no Plano de Desempenho, estes *inputs* e *outputs* são típicos deste tipo de aplicação DEA (e.g. Amado & Santos, 2009).

4.2. Construção das funções de utilidade dos fatores

Existem várias técnicas para obter informações de preferência de um Grupo de Decisores/Decisor com o propósito de construir funções de utilidade compatíveis com as suas respostas (ver von Winterfeldt & Edwards, 1986; Goodwin & Wright 1998); no entanto, as perguntas para esse fim devem ser

enquadradas no contexto de cada estudo. No método de Gouveia et al. (2008) o objetivo de converter os fatores numa escala de utilidade (funções de utilidade não-lineares) é o de espelhar as preferências do GD. Para construir as funções de utilidade seguiu-se um protocolo de diálogo já utilizado por Almeida e Dias (2012).

Tabela 2. Fatores do modelo 2.

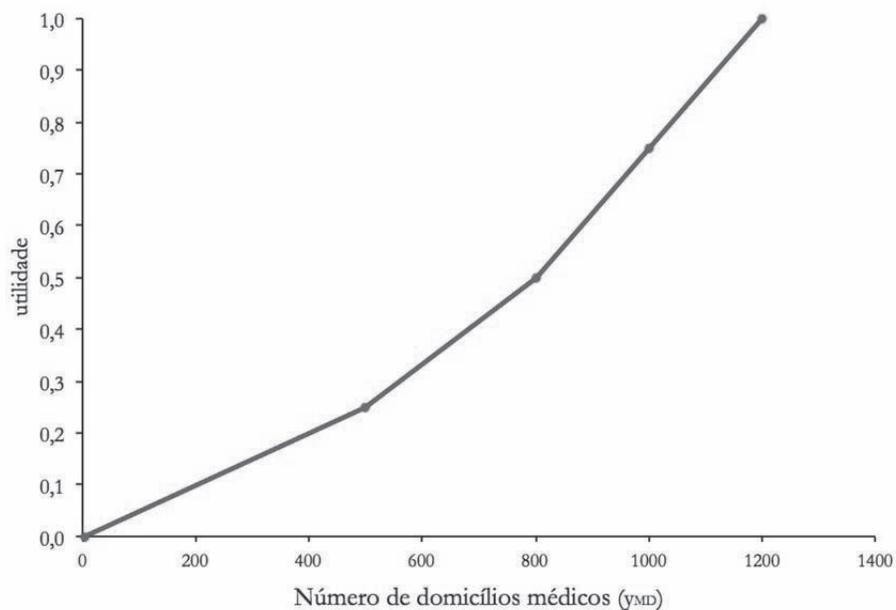
<i>Inputs</i>	<i>Outputs</i>
x_{MCDT} : custo total com meios complementares de diagnóstico e tratamento faturados ao SNS	y_{DM} : número de domicílios médicos y_{CM} : número de consultas médicas na unidade
x_{MED} : custo total com medicamentos faturados ao SNS	y_{DE} : número de domicílios de enfermagem
x_{RH} : custo total com recursos humanos	y_{CE} : número de consultas de enfermagem na unidade
x_{OC} : custos totais com medicamentos, materiais de consumo clínico e outros	y_{SQ} : satisfação com a qualidade dos serviços

O protocolo de diálogo baseou-se na comparação entre o mérito de aumentar um *output* (ou diminuir um *input*) de a para b contra o aumento do mesmo *output* (ou diminuição do mesmo *input*) de a' para b' , mantendo todos os outros níveis de desempenho iguais, e pedindo ao GD para ajustar um destes quatro valores de tal forma que o aumento de mérito seja aproximadamente igual. Para fazer esta conversão será apenas assumido que as funções são contínuas.

Por exemplo, considerando o *output* y_{DM} a questão colocada ao GD foi: “Tem mais mérito aumentar o número de domicílios médicos de 3 para 600 ou de 600 para 1200, mantendo todos os outros desempenhos iguais?” A resposta foi: “tem mais mérito aumentar de 600 para 1200”. Então o nível foi ajustado e a questão foi reformulada como: “Tem mais mérito aumentar o número de domicílios médicos de 3 para 800 ou de 800 para 1200, mantendo todos os outros desempenhos iguais?” A resposta foi: “tem igual mérito”. Um procedimento análogo foi utilizado para dicotomizar os intervalos de mérito [3,800] e [800,1200].

As funções de utilidade, para a maioria dos fatores, foram representadas por aproximação a funções lineares definidas por ramos. Quando as respostas do GD se conseguiram ajustar a curvas pré-definidas (como funções logarítmicas) as funções utilidade foram representadas por essas funções conhecidas.

Os intervalos encontrados foram escolhidos com o intuito de incluir as gamas de desempenho observadas para cada fator, mais (ou menos) uma margem de tolerância máxima (neste caso de 20%).



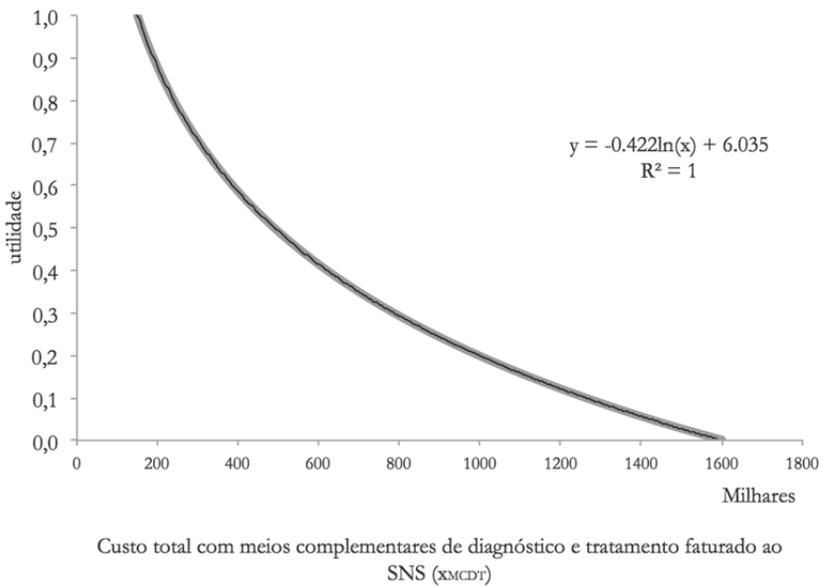
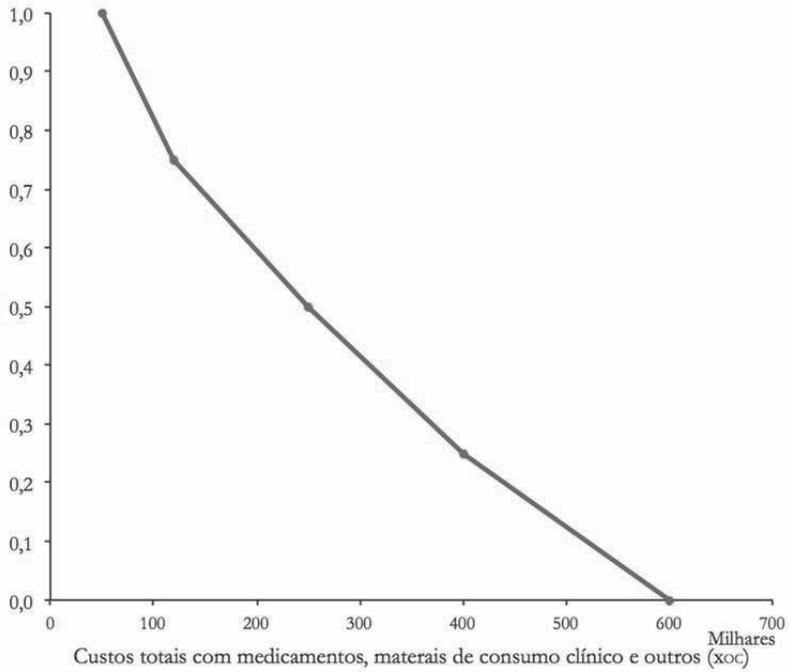


Figura 1. Três das funções de utilidade eliciadas.

As funções de utilidade lineares definidas por ramos para alguns dos fatores estão exibidas na Figura 1. Por exemplo, para o número de domicílios médicos:

$u_{DM}(1200) - u_{DM}(1000) = u_{DM}(1000) - u_{DM}(800) = u_{DM}(800) - u_{DM}(500) = u_{DM}(500) - u_{DM}(3)$, mantendo-se todos os outros níveis de desempenho iguais. As outras funções de utilidade para os restantes fatores *output* dos modelos (y_{CPF} , y_{CSM} , y_{CSI} , y_{CSA} , y_{CSS} , y_{CM} , y_{DE} , y_{CE} , e y_{SQ}) são semelhantes à função de utilidade y_{DM} .

Para os custos x_{MCDT} , x_{MED} , e x_{RH} , as funções de utilidade foram obtidas realizando o ajustamento de uma função conhecida às preferências reveladas pelo GD.

4.3. Restrições aos coeficientes de ponderação

Para os modelos apresentados houve a necessidade de incorporar restrições aos coeficientes de ponderação porque, caso contrário, não era possível alcançar uma classificação das unidades com discriminação aceitável (veja-se Dyson et al., 2001). Isto foi explicado ao GD, acrescentando que a introdução de restrições aos coeficientes de ponderação nos modelos traduz melhor os objetivos da organização e, portanto, garante resultados significativamente mais próximos daquilo que o GD considera como sendo as melhores práticas.

Existem na literatura várias abordagens para a definição de restrições aos coeficientes de ponderação, podendo a especificação destes de forma adequada revelar-se um desafio (Salo & Hämäläinen, 2001; Podinovski, 2004).

A técnica de *swings* começa pela construção de duas hipóteses extremas, B e G; a primeira diz respeito ao pior desempenho (utilidade 0) em todos os critérios e a segunda corresponde ao melhor desempenho (utilidade 1). O protocolo de diálogo consiste em consultar o decisor no sentido deste olhar para potenciais ganhos ao passar de B para G em cada critério e, em seguida, decidir quais os critérios que prefere mudar para a hipótese G. Supõe-se que a transição da hipótese B para a hipótese G, num determinado critério, vale 100 unidades numa escala hipotética. Em seguida, é solicitada ao decisor a atribuição de um valor (<100) ao segundo critério movido para G, depois ao

terceiro critério e assim sucessivamente, até que o último critério é movido para G.

O procedimento utilizado neste trabalho foi: primeiro obter uma ordenação de coeficientes de ponderação (baseado na técnica de *swings*) e, depois, estabelecer um limite para a razão entre os coeficientes de ponderação que ficaram em primeiro e em último lugar (pergunta de *trade-off*), por forma a evitar pesos nulos.

Ao incluir os coeficientes de ponderação, a Fase 1 e a Fase 2 (ver secção 2) passam a assumir outras formulações: às restrições do programa linear da Fase 1 acrescentam-se as restrições aos pesos, ao passo que no programa linear da Fase 2 é necessário permitir às folgas s_c ter valores negativos, retirando-se a restrição $s_c \geq 0$ (ver detalhes em Almeida & Dias, 2012).

5. Resultados

A implementação computacional do método, bem como das suas extensões, foram realizadas recorrendo a uma folha de cálculo.

O analista em conjunto com o GD construíram dois modelos (1 e 2) para avaliar a eficiência das unidades de CSP. A Tabela 3 apresenta os dados originais e as utilidades correspondentes para o modelo 1.

Tabela 3. Desempenhos das DMUs nas escalas originais e na escala de utilidades para o modelo 1.

Fatores nas escalas originais										
DMU	<i>X_{MCDT}</i>	<i>X_{MED}</i>	<i>X_{RH}</i>	<i>X_{OC}</i>	<i>Y_{CPF}</i>	<i>Y_{CSM}</i>	<i>Y_{CSI}</i>	<i>Y_{CSA}</i>	<i>Y_{CSS}</i>	<i>Y_{SQ}</i>
1	362900	1114935	1398033	123382	845	323	5133	25380	17149	92
2	463549	1192330	1050558	128606	449	311	3197	19983	14783	74
3	466561	1113734	1589818	101750	85	76	691	5136	2434	81
4	277140	1321779	1001297	152974	1628	605	4934	19283	13811	83
5	232565	1147295	1101200	141864	732	363	1668	17775	12108	62
6	613535	2491923	1724028	218890	2534	782	5068	29422	18854	75
7	1332317	4365605	2368401	360011	3726	1081	8073	54200	40307	77
8	701121	2289334	2225321	491140	2641	1142	4530	41247	27858	80
9	314090	1196353	1054564	65416	666	502	4513	21909	13476	91
10	606148	2954054	1348651	203098	3353	662	5068	29422	18854	54
11	1007909	3350321	1933791	371468	4763	1084	6587	42838	30037	78
12	708936	2793686	2227392	278936	3087	1052	5328	35475	25593	74
Fatores em escalas de utilidade										
DMU	<i>U_{MCDT}</i>	<i>U_{MED}</i>	<i>U_{RH}</i>	<i>U_{OC}</i>	<i>U_{CPF}</i>	<i>U_{CSM}</i>	<i>U_{CSI}</i>	<i>U_{CSA}</i>	<i>U_{CSS}</i>	<i>U_{SQ}</i>
1	0.633	0.824	0.572	0.743	0.084	0.123	0.392	0.256	0.210	0.750
2	0.529	0.789	0.788	0.733	0.042	0.117	0.193	0.190	0.178	0.400
3	0.527	0.824	0.475	0.815	0.003	0.010	0.014	0.014	0.006	0.521
4	0.746	0.735	0.825	0.687	0.168	0.253	0.367	0.182	0.164	0.563
5	0.820	0.809	0.753	0.708	0.072	0.141	0.083	0.164	0.140	0.220
6	0.411	0.406	0.413	0.560	0.271	0.364	0.384	0.324	0.234	0.417
7	0.084	0.115	0.173	0.317	0.457	0.568	0.759	0.737	0.633	0.450
8	0.355	0.450	0.220	0.136	0.288	0.618	0.316	0.521	0.381	0.500
9	0.694	0.787	0.785	0.945	0.065	0.205	0.314	0.213	0.159	0.729
10	0.416	0.318	0.599	0.590	0.399	0.289	0.384	0.324	0.234	0.140
11	0.202	0.253	0.326	0.298	0.659	0.570	0.573	0.547	0.417	0.467
12	0.350	0.347	0.219	0.452	0.357	0.543	0.416	0.425	0.343	0.400

Como anteriormente foi referido, para estes dois modelos, dada a relação entre o número de fatores e o número de unidades em avaliação, foi necessário incorporar restrições aos coeficientes de ponderação.

De acordo com o procedimento explicado na sub-secção 4.3, as restrições aos coeficientes de ponderação foram obtidas pedindo ao GD para comparar os "swings" da utilidade de 0 para 1 para o modelo 1, como descrito na Tabela 4.

O GD foi convidado a considerar uma unidade com o desempenho no nível 0 para todos os fatores e perguntou-se: "se pudesse melhorar um e apenas um fator para o nível 1, qual seria?". A resposta do GD foi unânime: "custo total com medicamentos faturados ao SNS". Isto permite a ilação de que w_{MED} é o maior coeficiente de escala. Ao repetir esta pergunta sucessivamente para os demais fatores, e depois de alguma discussão, o GD confluía na seguinte ordenação para os coeficientes de ponderação:

$$w_{MED} \geq w_{MCDT} \geq w_{RH} \geq w_{SQ} \geq w_{CPF} \geq w_{CSI} \geq w_{CSM} \geq w_{OC} \geq w_{CSA} \geq w_{CSS}.$$

Tabela 4. Desempenhos extremos associados com os níveis de utilidade 0 e 1 (modelo 1).

Nível de utilidade	X_{MCDT}	X_{MED}	X_{RH}	X_{OC}	Y_{CPF}	Y_{CSM}	Y_{CSI}	Y_{CSA}	Y_{CSS}	Y_{SQ}
$u(.)=0$	1600000	5500000	3000000	600000	60	55	500	4000	2000	40
$u(.)=1$	150000	800000	800000	50000	6000	1500	10000	65000	50000	100

Com o objetivo de evitar valores nulos nos coeficientes de ponderação, foi feita ao GD uma pergunta de *trade-off* para limitar a relação dos coeficientes de ponderação classificados na primeira (w_{MED}) e última posição (w_{CSS}): "Qual seria o maior valor z em euros correspondente ao custo total com medicamentos faturados ao SNS que permitiria a uma unidade com 50000 consultas de saúde a seniores realizadas e custo total com medicamentos faturados ao SNS associado de 5500000 euros ser considerada como tendo mais mérito do que uma unidade com 2000 consultas de saúde a seniores realizadas e com um custo total com medicamentos faturados ao SNS de z euros?". A resposta permite construir a seguinte desigualdade:

$$w_{CSS} u(50000) + w_{MED} u(5500000) \geq w_{CSS} u(2000) + w_{MED} u(z).$$

O limite obtido foi $z = 2500000$. Com o aumento deste valor o GD começava a hesitar se a desigualdade se manteria. Substituindo esse valor de z na expressão anterior: $w_{MED} \leq 2.47 w_{CSS}$.

Os resultados para o modelo 2 estão representados nas Tabela 5 e 6.

Tabela 5. Resultados da Fase 1, com restrições aos coeficientes de ponderação (modelo 1).

Fase 1											
DMU	d^*	w^*_{MCDT}	w^*_{MED}	w^*_{RH}	w^*_{OC}	w^*_{CPF}	w^*_{CSM}	w^*_{CSI}	w^*_{CSA}	w^*_{CSS}	w^*_{SQ}
1	0.0223	0.087	0.215	0.087	0.087	0.087	0.087	0.087	0.087	0.087	0.087
2	0.0813	0.171	0.171	0.171	0.069	0.069	0.069	0.069	0.069	0.069	0.069
3	0.1426	0.087	0.215	0.087	0.087	0.087	0.087	0.087	0.087	0.087	0.087
4	0.0047	0.122	0.122	0.122	0.049	0.122	0.122	0.122	0.049	0.049	0.122
5	0.0566	0.171	0.171	0.171	0.069	0.069	0.069	0.069	0.069	0.069	0.069
6	0.1114	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100
7	0.0605	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100
8	0.1112	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100
9	-0.0248	0.156	0.156	0.156	0.063	0.063	0.063	0.063	0.063	0.063	0.156
10	0.1204	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100
11	0.0585	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100
12	0.1044	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100

Tabela 6. Resultados da Fase 2, com restrições aos coeficientes de ponderação e folgas livres (modelo 1).

Fase 2										
DMU	\hat{s}^*_{MCDT}	\hat{s}^*_{MED}	\hat{s}^*_{RH}	\hat{s}^*_{OC}	\hat{s}^*_{CPF}	\hat{s}^*_{CSM}	\hat{s}^*_{CSI}	\hat{s}^*_{CSA}	\hat{s}^*_{CSS}	\hat{s}^*_{SQ}
1	0.061	-0.037	0.213	0.201	-0.019	0.082	-0.078	-0.043	-0.051	-0.021
2	0.164	-0.002	-0.003	0.211	0.023	0.088	0.121	0.023	-0.018	0.329
3	0.167	-0.037	0.311	0.130	0.062	0.195	0.300	0.200	0.153	0.208
4	-0.053	0.052	-0.039	0.258	-0.103	-0.048	-0.053	0.031	-0.005	0.167
5	-0.127	-0.022	0.033	0.237	-0.007	0.064	0.231	0.049	0.019	0.509
6	0.283	0.381	0.372	0.385	-0.206	-0.159	-0.069	-0.110	-0.075	0.313
7	0.610	0.672	0.612	0.628	-0.392	-0.362	-0.445	-0.523	-0.473	0.279
8	0.339	0.337	0.565	0.809	-0.223	-0.413	-0.002	-0.308	-0.222	0.229
9										
10	0.277	0.469	0.186	0.355	-0.334	-0.084	-0.069	-0.110	-0.075	0.589
11	0.492	0.534	0.459	0.647	-0.594	-0.365	-0.259	-0.334	-0.258	0.263
12	0.344	0.440	0.566	0.493	-0.293	-0.338	-0.102	-0.211	-0.184	0.329

Neste caso há apenas uma unidade eficiente (DMU 9), que é a referência para todas as outras (Tabela 5). Esta unidade tem as melhores práticas de acordo com as preferências do GD.

Obteve-se a seguinte ordenação:

DMU 9 > DMU 4 > DMU 1 > DMU 5 > DMU 11 > DMU 7 > DMU 2 > DMU 12 > DMU 8 > DMU 6 > DMU 10 > DMU 3.

Ao GD agradou que a DMU 9 ficasse em primeiro lugar mas ficou um pouco surpreendido, pois esperava que outra unidade assumisse o primeiro lugar. Este resultado suscitou o debate no GD, tendo surgido como possível explicação o facto de os fatores de *output* traduzirem pouco o trabalho de enfermagem. A DMU que ficou em último lugar foi realmente surpreendente, mas o GD conjecturou que existe uma forte possibilidade de não terem sido registadas todas as consultas médicas naquela unidade. No entanto, seria necessário averiguar melhor a causa desta última posição neste modelo.

O modelo 2 é aquele que o GD considera como sendo o mais abrangente, dado que combina os vários eixos de avaliação correspondentes ao acesso, desempenho assistencial, qualidade percebida e desempenho económico e reflete todo o trabalho médico, de enfermagem e administrativo. Os dados originais e as utilidades correspondentes para o modelo 2 são expostos na Tabela 7

As restrições aos coeficientes de ponderação foram obtidas de forma semelhante às do modelo 1 e ao GD foi pedido para comparar os "swings" da utilidade de 0 para 1 apresentados na Tabela 8 .

A ordenação dos coeficientes de escala obtida foi:

$$w_{MED} \geq w_{MCDT} \geq w_{RH} \geq w_{DM} \geq w_{OC} \geq w_{SQ} \geq w_{CE} \geq w_{CM} \geq w_{DE}.$$

De acordo com a resposta conseguida à questão de julgamento de indiferença, foi estabelecida a seguinte desigualdade: $w_{MED} \leq 2.47 w_{DE}$.

A ordenação das unidades é (d^* mostrado entre parênteses):

DMU 1 (-0.0562) > DMU 9 (-0.0013) > DMU 4 (0.0229) > DMU 5 (0.0503) > DMU 2 (0.0725) > DMU 3 (0.1395) > DMU 6 (0.1668) > DMU 10 (0.1810) > DMU 11 (0.1886) > DMU 12 (0.2003) > DMU 7 (0.2429) > DMU 8 (0.2619).

Embora a DMU 9 não esteja na primeira posição, continua a ser eficiente neste modelo.

Tabela 7. Desempenhos das DMUs nas escalas originais e na escala de utilidades para o modelo 2.

Fatores nas escalas originais									
DMU	<i>X_{MCDT}</i>	<i>X_{MED}</i>	<i>X_{RH}</i>	<i>X_{OC}</i>	<i>Y_{DM}</i>	<i>Y_{CM}</i>	<i>Y_{DE}</i>	<i>Y_{CE}</i>	<i>Y_{SQ}</i>
1	362900	1114935	1398033	123382	985	48038	2006	24273	92
2	463549	1192330	1050558	128606	241	37977	1424	20580	74
3	466561	1113734	1589818	101750	96	8265	844	11677	81
4	277140	1321779	1001297	152974	308	38102	1979	17150	83
5	232565	1147295	1101200	141864	4	21665	792	27328	62
6	613535	2491923	1724028	218890	195	50768	3645	34610	75
7	1332317	4365605	2368401	360011	570	96530	4166	23781	77
8	701121	2289334	2225321	491140	182	68599	987	34052	80
9	314090	1196353	1054564	65416	302	40283	1756	17435	91
10	606148	2954054	1348651	203098	244	45819	2249	45119	54
11	1007909	3350321	1933791	371468	365	75904	5031	35040	78
12	70936	2793686	2227392	278936	246	64165	3878	34831	74
Fatores em escalas de utilidade									
DMU	<i>U_{MCDT}</i>	<i>U_{MED}</i>	<i>U_{RH}</i>	<i>U_{OC}</i>	<i>U_{DM}</i>	<i>U_{CM}</i>	<i>U_{DE}</i>	<i>U_{CE}</i>	<i>U_{SQ}</i>
1	0.633	0.824	0.572	0.743	0.731	0.194	0.185	0.225	0.750
2	0.529	0.789	0.788	0.733	0.120	0.147	0.108	0.170	0.400
3	0.527	0.824	0.475	0.815	0.047	0.008	0.032	0.039	0.521
4	0.746	0.735	0.825	0.687	0.153	0.148	0.181	0.120	0.563
5	0.820	0.809	0.753	0.708	0.001	0.071	0.025	0.274	0.220
6	0.411	0.406	0.413	0.560	0.097	0.207	0.441	0.404	0.417
7	0.084	0.115	0.173	0.317	0.308	0.478	0.528	0.217	0.450
8	0.355	0.450	0.220	0.136	0.090	0.304	0.051	0.394	0.500
9	0.694	0.787	0.785	0.945	0.150	0.158	0.152	0.124	0.729
10	0.416	0.318	0.599	0.590	0.121	0.184	0.217	0.642	0.140
11	0.202	0.253	0.326	0.298	0.182	0.349	0.672	0.411	0.467
12	0.350	0.347	0.219	0.452	0.122	0.276	0.480	0.408	0.400

Tabela 8. Desempenhos extremos associados com os níveis de utilidade 0 e 1 (modelo 2).

Nível de utilidade	X_{MCDT}	X_{MED}	X_{RH}	X_{OC}	Y_{DM}	Y_{CM}	Y_{DE}	Y_{CE}	Y_{SQ}
$u(.)=0$	1600000	5500000	3000000	600000	3	6500	600	9000	40
$u(.)=1$	150000	800000	800000	50000	1200	150000	6500	55000	100

Os resultados alcançados corresponderam às expectativas do GD, em vários aspectos: a DMU 3 recupera, subindo várias posições no *ranking*, com a inclusão do trabalho de enfermagem; a DMU 5 está na 2.^a posição entre as ineficientes, mantendo a 4.^a posição no *ranking* geral; a DMU 1 é, como esperado, a melhor de todas as unidades, sendo uma referência para a grande maioria das unidades ineficientes (exceto para uma); a DMU 8 é classificada como a mais ineficiente, facto nada surpreendente para o GD.

6. Melhorias das unidades ineficientes e outros desenvolvimentos

Considere-se que W denota o conjunto de vetores de coeficientes de ponderação compatíveis com o *ranking* e limite obtidos (ver 4.3.). A Fase 2 do método usa o vetor de ponderação ótimo em W^* para resolver o modelo de PL da Fase 2 com as variáveis folga a assumir qualquer valor real. A solução é uma proposta de um alvo de eficiência (projeção) para cada DMU ineficiente. Para atingir o estado de eficiência, essas DMUs ineficientes devem mudar a sua utilidade em cada fator pelo valor indicado por s^* , considerando os dois modelos.

À partida, não havia restrições aos alvos: qualquer alvo era admitido desde que permitisse à DMU anular o seu défice de valor global. No entanto, tal não impedia que o alvo escolhido fosse pior nalguns fatores, desde que estas variações fossem suficientemente compensadas por melhorias nos restantes fatores. Logo, nem todas as mudanças propostas correspondem a melhorias em todos os fatores, uma vez que algumas folgas podiam ser negativas. Note-se, porém, que uma folga negativa para um fator de *input* corresponde a uma sugestão de aumentar o consumo desse *input*, e uma folga para um fator de *output* corresponde a uma sugestão de reduzir a produção desse *output*. O GD não concordou com estas possibilidades, uma vez que não tinha vantagem

alguma em aumentar recursos, nomeadamente custos, nem em diminuir serviços prestados, pelo que houve a necessidade de propor uma outra solução.

Considera-se que u_k^* é a utilidade da melhor DMU usando o vetor de ponderação no conjunto ótimo W^* , ou seja, o valor da utilidade que a DMU k deve alcançar:

$$u_k^* = \sum_{c=1}^q w_c^* u_c(DMU_k) + d_k^* \quad (6)$$

Sabe-se que o modelo de PL resolvido na Fase 2 do método de Gouveia et al. (2008) admite alvos ótimos alternativos, cada um dos quais corresponde a uma forma diferente de anular o valor da diferença d_k^* . Esses alvos correspondem a diferentes projeções sobre a fronteira eficiente. No presente caso, o GD tem a intenção de restringir os alvos de eficiência que se propõem alcançar u_k^* , não só para evitar os alvos que implicam um aumento de *inputs* ou uma redução dos *outputs*, mas também para escolher, dadas as características de cada unidade, quais os fatores que podem ser alterados.

O conjunto $S_< = \{c \in \{1, \dots, q\}: s_c^* < 0 \text{ na Fase 2}\}$ denota as folgas negativas no ótimo de (5); essas folgas vão passar agora a ser constantes nulas. O conjunto $S_> = \{c \in \{1, \dots, q\}: s_c^* \geq 0 \text{ na Fase 2}\}$ denota as restantes folgas, que serão consideradas como variáveis não-negativas. Portanto, a formulação que conduz a um alvo alternativo que mantém ou melhora o desempenho de todos os fatores pode ser obtida resolvendo o modelo de PL (7), no qual a folga máxima (em termos de valor) deve ser minimizada para atingir o valor global de utilidade do alvo. Não são permitidas folgas negativas, mas o alvo já não será uma combinação convexa das DMUs observadas (para mais detalhes ver Almeida & Dias, 2012).

A formulação abaixo difere da utilizada por Almeida e Dias (2012) na medida em que introduz um novo parâmetro σ_c (um valor positivo) para limitar o valor que uma folga pode ter. O objetivo é evitar a fixação de alvos irrealistas, permitindo que as propostas de melhoria sejam mais ajustadas às exigências do GD. Os alvos nunca excedem o valor 1 em qualquer fator, devido à restrição $u_c(DMU_k) + s_c \leq 1$. Tal garante que a função de utilidade não ultrapassa os intervalos dos desempenhos obtidos.

$$\begin{aligned}
& \min_{\rho_k, s} \rho_k \\
& \text{s. a: } \sum_{c \in S_{\geq}} w_c^* s_c = d_k^* \\
& \quad u_c(DMU_k) + s_c \leq 1, \quad c = 1, \dots, q \\
& \quad s_c - \sigma_c \leq 0, \quad c \in S_{\geq} \\
& \quad \rho_c \geq s_c \geq 0, \quad c = 1, \dots, q
\end{aligned} \tag{7}$$

No que diz respeito ao modelo 1, todas as unidades ineficientes escolheram como par a DMU 9 e o GD começou por analisar os resultados em termos de propostas de melhoria das unidades ineficientes, bloqueando alterações nos fatores que tinham folgas negativas na Fase 2 (ver proposta 1 na Tabela 9), não permitindo que, nestes fatores, o alvo tivesse pior desempenho que o desempenho original das unidades.

Para algumas unidades a proposta 1 tem desempenhos atingíveis, mas para outras tais seriam impossíveis de alcançar. Esta realidade obrigou à construção de uma nova formulação, (8), na qual foi feita uma distribuição por todos os fatores (*inputs* e *outputs*) do valor da diferença relativamente ao par, de uma forma equilibrada, através do seguinte modelo de PL:

$$\begin{aligned}
& \min_{\rho_k, s} \rho_k \\
& \text{s. a: } \sum_{c=1}^q w_c^* s_c = d_k^* \\
& \quad u_c(DMU_k) + s_c \leq 1, \quad c = 1, \dots, q \\
& \quad s_c - \sigma_c \leq 0, \quad c = 1, \dots, q \\
& \quad \rho_c \geq s_c \geq 0, \quad c = 1, \dots, q
\end{aligned} \tag{8}$$

Mesmo assim, o GD não entendeu como praticáveis, dadas as características de certas unidades, algumas das reduções propostas, nomeadamente no custo total com recursos humanos proposto para a DMU 3, e nos aumentos de algumas consultas propostas para as DMUs 6 e 10 (ver proposta 2 na Tabela 9). Neste último caso a justificação está relacionada com estatísticas nacionais associadas à taxa de natalidade, bem como ao facto de alguns utentes (apesar de inscritos no serviço de saúde público) preferirem ir ao serviço de saúde privado, possivelmente por motivos de desempenho percebido. Assim, surgiu a proposta de limitar a redução ou aumento de alguns fatores de *input* ou *output*, para as unidades 3, 6 e 10 no modelo 1. Para este efeito, a diferença entre a unidade em avaliação e a DMU 9 deveria ser suprimida pela redução ou aumento de outros fatores de *input* ou *output* (ver proposta 3 na Tabela 9).

Tabela 9. Propostas de melhoria para as unidades ineficientes nas escalas originais (modelo 1)

Proposta 1										
DMU	X _{MCDT}	X _{MED}	X _{RH}	X _{OC}	Y _{CPF}	Y _{CSM}	Y _{CSI}	Y _{CSA}	Y _{CSS}	Y _{SQ}
1	311190	1114935	1285500	10415	845	462	5133	25380	17149	92
2	350315	1192330	1050558	85950	1741	610	4644	29710	14783	81.8
3	303800	1113734	1250900	50750	1780	472	3230	20340	15500	89.7
4	277140	1311200	1001297	145745	1628	605	4934	20422	13811	83.6
5	232565	1147295	970000	105000	732	573	3013	25570	19028	69
6	362000	1622900	1285000	110900	2534	782	5068	29422	18854	86.6
7	1000840	3462876	2009990	288090	3726	1081	8073	54200	40307	83.4
8	414500	1490900	1659300	335300	2641	1142	4530	41247	27858	90.6
9										
10	342900	1855000	981100	97400	3353	662	5068	29422	18854	72.8
11	763900	2672000	1658890	301300	4763	1084	6587	42838	30037	84.0
12	432500	1869000	1689900	166100	3087	1052	5328	35475	25593	85.2
Proposta 2										
DMU	X _{MCDT}	X _{MED}	X _{RH}	X _{OC}	Y _{CPF}	Y _{CSM}	Y _{CSI}	Y _{CSA}	Y _{CSS}	Y _{SQ}
1	344500	1068900	1358900	115650	1048	371	5308	26710	18740	92.7
2	381560	1020700	942490	101690	1207	488	4190	26300	14780	78.8
3	332600	846100	1316900	61750	1420	386	2680	17100	12700	87.8
4	274440	1288720	996200	150430	1666	612	4964	19667	14132	83.2
5	203699	1030610	1022900	115990	1254	486	2453	22522	18854	66.6
6	472200	2010000	1487990	160900	3242	960	5957	36100	25080	81.3
7	1153760	3887255	2187990	324100	4085	1153	8556	56760	43700	80.5
8	539200	1850900	1923300	402800	3351	1275	5415	47890	34490	85.3
9										
10	455000	2342000	1149100	140100	4090	855	6030	36610	26280	65.6
11	878540	2996000	1790990	336700	5040	1153	7052	46330	33580	81.2
12	553490	2287190	1939300	221100	3757	1177	6160	41710	31880	80.2
Proposta 3										
DMU	X _{MCDT}	X _{MED}	X _{RH}	X _{OC}	Y _{CPF}	Y _{CSM}	Y _{CSI}	Y _{CSA}	Y _{CSS}	Y _{SQ}
3	380500	1000900	1000000	101690	12175	489	4210	21340	10789	86.3
6	471200	2000000	1488010	160990	3249	800	5500	36190	25120	82.3
10	454890	2341090	1149090	140000	4150	800	5500	36623	26340	70

As propostas de melhoria para o modelo 1 em todos os cenários referidos são apresentadas na Tabela 9, com os fatores que não foram alterados ou estão limitados (proposta 3) mostrados em negrito.

O GD ficou muito satisfeito com a possibilidade de escolher, para cada caso, a proposta de melhoria que entende ser a mais adequada, uma vez que conhece bem as características específicas de cada unidade.

Para o modelo 2, e atendendo às unidades em avaliação e aos fatores em causa, o GD entendeu suficiente a apresentação apenas das propostas 1 e 2 (Tabela 10).

Tabela 10. Propostas de melhoria para as unidades ineficientes nas escalas originais (modelo 2).

Proposta 1									
DMU	X_{MCDT}	X_{MED}	X_{RH}	X_{OC}	Y_{DM}	Y_{CM}	Y_{DE}	Y_{CE}	Y_{SQ}
1									
2	376000	103900	1050558	99900	418	57000	2100	26480	79.3
3	298200	1113734	1240300	101750	470	48580	2280	24500	90
4	277140	1232500	1001297	134300	380	45000	2260	19600	84.7
5	232565	1147295	976900	106700	183	40850	1480	27328	68.6
6	339500	1535900	1238000	103300	616	50768	3645	34610	80
7	636500	2737000	1569100	183100	897	96530	4166	41090	92.3
8	316100	1195900	1426900	266500	712	68599	3328	34052	94.8
9									
10	319300	1753900	1348651	88700	672	92720	2249	45119	74.7
11	516690	19339800	1332000	208600	758	75904	5031	35040	89.6
12	347800	1566900	1498400	119580	707	64165	3878	34831	89.6
Proposta 2									
DMU	X_{MCDT}	X_{MED}	X_{RH}	X_{OC}	Y_{DM}	Y_{CM}	Y_{DE}	Y_{CE}	Y_{SQ}
1									
2	389790	1037990	953800	104500	384	53480	1972	25490	78.4
3	334900	851090	1323400	62470	372	38080	1910	21200	87.7
4	262985	126600	971500	141580	352	43000	2150	18690	84.1
5	206100	1042500	1031000	117900	105	32290	1180	30120	66.2
6	412800	1809000	1384000	132000	515	79760	4645	42560	84
7	749500	2392000	1719100	219000	841	126500	5585	37750	89.2
8	377000	13839000	1575400	311500	622	107860	2874	45598	92.4
9									
10	394800	2083000	1061700	114200	562	78320	3385	50740	69.2
11	644190	2332800	1507000	258600	645	104500	5938	43600	87.4
12	441590	1899900	1709000	171190	587	96100	5080	43880	84.8

7. Considerações Finais

Este estudo relata um processo interativo em que o analista e especialistas em saúde trabalharam juntos na avaliação de unidades de CSP, com o intuito de identificar as melhores práticas e contribuir para uma tomada de decisão informada. A abordagem DEA/MCDA proposta anteriormente pelos autores foi utilizada como base de modelação e posteriormente adaptada aos requisitos desta aplicação, ou seja, para definir alvos realistas (com a introdução de um novo parâmetro no modelo) no sentido de possibilitar a ordenação total das DMUs.

A intervenção para avaliar o desempenho das unidades de saúde começa com a escolha pelo GD dos fatores a serem incluídos no modelo, abarcando quatro grandes áreas: acesso, desempenho assistencial, qualidade percebida e desempenho económico. Fruto das discussões, resultaram dois modelos para avaliar as unidades de CSP. A incorporação das preferências do GD nesses modelos foi possível graças à conversão dos fatores *inputs* e *outputs* em funções de utilidade, o que exigiu interpretar estas funções de utilidade como dispositivos para comparar a variação de mérito que corresponde a diferenças de desempenho. Os modelos tornam-se mais completos e credíveis para o GD apenas com a introdução de restrições de coeficientes de ponderação que foram obtidas recorrendo à técnica dos *swings* para estabelecer uma ordenação dos coeficientes de ponderação. Além disso, foi usada uma questão baseada em *trade-offs* para limitar a relação entre os coeficientes de ponderação posicionados no primeiro e último lugares e, desta forma, evitar coeficientes de ponderação nulos.

Para cada modelo obteve-se o *ranking* completo das unidades e foram identificados os pares que lhes serviriam de referência. As propostas de melhoria para as unidades ineficientes surgiram aproveitando o conhecimento do GD sobre as unidades, a fim de obter alvos mais realistas. Para este efeito, um novo modelo foi desenvolvido onde é introduzido um novo parâmetro que limita o valor que uma folga pode ter, possibilitando propostas de melhoria mais ajustadas às preferências do GD.

Este trabalho permitiu facultar informação ao GD no sentido da identificação das melhores práticas, fontes de ineficiência e diferenças em

relação às melhores práticas, bem como possibilidades e oportunidades de melhoria. Os resultados foram úteis não só para a conceção de medidas corretivas, mas também para informar as decisões sobre os objetivos dos IPs no futuro. Todo o processo de avaliação das unidades de CSP desenrolou-se alternando reuniões de discussão e apresentação de resultados com desenvolvimento metodológicos e testes computacionais, podendo este não ser um processo replicável em todos os casos. Os desenvolvimentos futuros deste trabalho incluem a implementação de um sistema de apoio à decisão que responda em tempo real aos requisitos dos utilizadores.

BIBLIOGRAFIA

- ALI, A.I., & Seiford, L. (1993) The mathematical programming approach to efficiency analysis. In Fried, Lovell & Schmidt (Orgs.). *The measurement of productive efficiency: techniques and applications* (pp.120-113). New York: Oxford University.
- ALI, A.I., Lerne, C.S., & Seiford, L. (1995). Components of efficiency evaluation in Data Envelopment Analysis. *European Journal of Operational Research*, 80, 462-473.
- ALLEN, R., Athanassopoulos, A., Dyson, R.J., & Thanassoulis, E. (1997). Weights restrictions and value judgments in Data Envelopment Analysis: evolution, development and future directions. *Annals of Operations Research*, 73,13-34.
- ALMEIDA, P.N., & Dias, L.C. (2012). Value-based DEA models: application-driven developments. *Journal of the Operational Research Society*, 63, 16-27.
- AMADO, C.A.F., & Dyson, R.G. (2008). On comparing the performance of primary care providers. *European Journal of Operational Research*, 185, 915-932.
- AMADO, C.A.F., & Santos, S.P. (2009). Challenges for performance assessment and improvement in primary health care: The case of the Portuguese health centres. *Health Policy*, 91, 43-56.
- ANDERSEN, P., & Petersen, N.C. (1993). A procedure for ranking efficient units in data envelopment analysis. *Management Science*, 39, 1261-1264.
- ATHANASSOPOULOS, A.D., & Podinovski, V.V. (1997). Dominance and potential optimality multiple criteria models decision analysis with imprecise information. *Journal of the Operational Research Society*, 48 (1), 142-150.
- BAKER, R. (1992). General practice in Gloucestershire, Avon and Somerset: Explaining variations in standards. *British Journal of General Practice*, 42, 415-418.
- BANKER, R.D., Charnes, A., & Cooper, W.W. (1994). Some models estimating technical and scale inefficiencies in Data Envelopment Analysis. *Management Science*, 3 (9), 1078-1092.
- CAMPBELL, S.M., Roland, M.O., Quayle, J.A., Buetow, S.A., & Shekelle, P.G. (1998). Quality indicators for general practice: Which ones can general practitioners and health authority managers agree are important and how useful are they?. *Journal of Public Health Medicine*, 20, 414-421.
- CHARNES, A., Cooper, W.W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2 (6), 429-444.

CHARNES, A., Cooper W.W., Golany, B., Seiford, L., & Stutz, J. (1995). Foundations of Data Envelopment Analysis for Pareto-Koopmans efficient empirical production functions. *Journal of Econometrics*, 30, 91-107.

Avaliação sobre o funcionamento das Unidades de Saúde do ACES – BM1. (2010). Coimbra.

Coodenação Estratégica dos CSP. (2010). *Metodologia de Contratualização*, Administração Central do Sistema de Saúde.

DIAS, L.C., & Clímaco, J.N. (2000). Additive Aggregation with Variable Interdependent Parameters: the VIP Analysis Software. *Journal of the Operational Research Society*, 51 (9), 1070-10822.

DONABEDIAN, A. (1980). *The Definition of Quality and Approaches to its Assessment*. Ann Arbor, MI, Health Administration Press.

DULÁ, J.H., & Hickman, B.L. (1997). Effects for excluding the column being scored from DEA envelopment LP technology matrix. *Journal of the Operational Research Society*, 48, 1001-1012.

DYSON, R.G., Allen, R., Camanho, A.S., Podinovski, V.V., Sarrico, C.S. & Shale, E.A. (2001). Pitfalls and protocols in DEA. *European Journal of Operational Research*, 132, 245-259.

FARRELL, M.J. (1997). The Measurement of productive efficiency. *Journal of Royal Statistical Society*, 120 (A), 353-290.

GIUFFRIDA, A., Gravelle, H., & Sutton, M. (2000). Efficiency and administrative costs in primary care. *Journal of Health Economics*, 19 (6), 983-1006.

GOODWIN, P., & Wright, G. (1998). *Decision analysis for management judgment*. Wiley: Chichester, 1998.

GOUVEIA, M.C., Dias, L.C., & Antunes, C.H. (2008). Additive DEA based on MCDA with imprecise information. *Journal of the Operational Research Society*, 59, 54-63.

GOUVEIA, M.C., Dias, L.C., & Antunes, C.H. (2013). Super-efficiency and stability intervals in additive DEA. *Journal of the Operational Research Society*, 64, 86-96.

HOLLINGSWORTH, B. (2003). Non-Parametric and Parametric Applications Measuring Efficiency in Health Care. *Health Care Management Science*, 6, 203-218.

HUANG, Y.G., & McLaughlin, C.P. (1989). Relative efficiency in rural primary health care: an application of data envelopment analysis. *Health Services Research*, 24, 143-158.

Instituto de Gestão Informática e Financeira da Saúde. (2006). *Estatística do Movimento Assistencial dos Centros de Saúde em 2005*.

KEENEY, R.L., Raiffa, H. (1976). *Decisions with multiple objectives: preferences and value tradeoff*. Wiley: New York.

KIRIGIA, J.M., Emrouznejad, A., Sambo, L.G., Munguti, N., & Liambila, W. (2004). Using data envelopment analysis to measure the technical efficiency of public health centers in Kenya. *Journal of Medical Systems*, 28, 155-166.

LINNA, M., Nordblad, A., & Koivu, M. (2003). Technical and cost efficiency of oral health care provision in Finnish health centres. *Social Science & Medicine*, 56, 343-353.

MILLIKEN, O., Devlin, R.A., Barham, V., Hogg, W., Dahrouge, S., & Russell, G. (2011). Comparative Efficiency Assessment of Primary Care Service Delivery Models Using Data Envelopment Analysis. *Canadian Public Policy*, 37, 85-109.

MCCOLL, A., Roderick, P., Gabbay, J., Smith, H., Moore, M. (1998). Performance indicators for primary care groups: An evidence based approach. *British Medical Journal*, 317, 1354-1360.

NUTI, S., Daraio, C., Speroni, C., & Vainieri, M. (2011). Relationships between technical efficiency and the quality and costs of health care in Italy. *International Journal for Quality in Health Care*, 23, 324-330.

Organização para a Cooperação e Desenvolvimento Económico. (2006). *Projecting OECD health and long term care expenditures: what are the main drivers?* Economics Department. Working Papers 477.

Observatório Português dos Sistemas de Saúde. (2007). *Relatório de Primavera 2007. Luzes e sombras: a governação da saúde*, Coimbra: Mar da Palavra.

Observatório Português dos Sistemas de Saúde. (2008). *Relatório de Primavera 2008. Sistema de Saúde Português – Riscos e Incertezas*, Coimbra: CEISUC.

OZCAN, Y.A. (1993). Sensitivity Analysis of Hospital Efficiency under Alternative Output/Input Combinations and Peer Groupings. *International Journal of Knowledge and Policy*, 4, 1-31.

PROCTOR, S., & Campbell, J. (1999). A developmental performance framework for primary care. *International Journal of Health Quality Assurance*, 12 (7), 279-286.

PODINOVSKI, V.V. (2004). Production trade-offs and weight restrictions in data envelopment analysis. *Journal of the Operational Research Society*, 55 (12), 1311-22.

ROSEMAN, R., & Friesner, D. (2004). Scope and scale inefficiencies in physician practices. *Health Economics*, 13, 1091-1116.

ROUSE, P., Harrison, J., & Turner, N. (2011). Cost and Performance: Complements for Improvement. *Journal of Medical Systems*, 35, 1063-1074.

SALINAS-JIMENEZ, J., & Smith, P. (1996). Data envelopment analysis applied to quality in primary health care. *Annals of Operational Research*, 67, 141-161.

SALO, A.A., & Hämäläinen, R.P. (2001). Preference Ratios in Multiattribute Evaluation (PRIME) - elicitation and decision procedures under incomplete information, *IEEE Transactions on Systems Man and Cybernetics part A*, 31, 533-545.

THANASSOULIS, E., Boussofiane, A., & Dyson, R.G. (1996). A comparison of data envelopment analysis and ratio analysis as tools for performance assessment. *Omega*, 24, 229-244.

THANASSOULIS, E., Portela, M.C., Allen, R. (2004). Incorporating value judgments in DEA. In Cooper, W.W., Seiford, L.M., Zhu, J. (eds.), *Handbook on Data Envelopment Analysis* (pp.99-137). Boston: Kluwer.

THANASSOULIS, E., Portela, M.C., Graveney, M. (2012). Estimating the scope for savings in referrals and drug prescription costs in the General Practice units of a UK Primary Care Trust. *European Journal of Operational Research*, 221, 432-444.

VON WINTERFELDT, D., Edwards, W. (1986). *Decision analysis behavioural research*. New York: Cambridge University Press.

VAN DEN HOMBERGH, P., Grol, R., van den Hoogen, H.J.M., & van den Bosch, W.J.H.M. (1998). Assessment of management in general practice: Validation of a practice visit method. *British Journal of General Practice*, 48, 1743-1750.

WAGNER, J., Shimshak, D., Novak, M. (2003). Advances in physician profiling: the use of DEA. *Socio-Economic Planning Sciences*, 37, 141-163.

CASO 19

OTIMIZAÇÃO DE REDES HOSPITALARES: UM MODELO DE LOCALIZAÇÃO HIERÁRQUICO E MULTI-SERVIÇO APLICADO AO CASO PORTUGUÊS

Ana Maria Mestre

Centro de Estudos em Gestão do Instituto Superior Técnico
Universidade de Lisboa
anamestre@tecnico.ulisboa.pt

Mónica Duarte Oliveira

Centro de Estudos em Gestão do Instituto Superior Técnico
Universidade de Lisboa
monica.oliveira@tecnico.ulisboa.pt

Ana Paula Barbosa Póvoa

Centro de Estudos em Gestão do Instituto Superior Técnico
Universidade de Lisboa
apovoa@tecnico.ulisboa.pt

RESUMO

Sistemas hospitalares enquadrados num Serviço Nacional de Saúde (SNS) são sujeitos a um planeamento da rede hospitalar, com o intuito de promover equidade na distribuição dos serviços e acesso aos cuidados de saúde hospitalares, tendo em conta os recursos disponíveis. Este trabalho desenvolve um modelo de apoio à reorganização da rede hospitalar na zona sul de Portugal, com base num modelo de programação linear inteira mista. O modelo desenvolvido é uma extensão de modelos de localização hierárquicos que permite modelar vários aspetos relevantes no planeamento hospitalar, nomeadamente: as unidades hospitalares como prestadoras de múltiplos serviços, a cooperação de unidades em diferentes níveis com fluxos ascendentes e descendentes na hierarquia, e limites na capacidade instalada. Para exemplificar a utilidade do modelo são apresentados os resultados de dois cenários que capturam diferentes perspetivas relevantes para a reorganização hospitalar e incluem os projetos hospitalares em estudo e discussão em Portugal. O modelo fornece resultados relevantes para o planeamento da rede hospitalar, como a localização e dimensionamento das unidades hospitalares, a definição das áreas de influência, e indicadores de nível de serviço. Na secção final analisa-se a aplicação do modelo a outro caso de estudo, assim como se identificam as linhas de investigação a prosseguir.

PALAVRAS-CHAVE

Modelos de localização, Planeamento de redes, Programação matemática, Saúde, Rede hospitalar

1. Introdução

Em países com um sistema de saúde estruturado num Serviço Nacional de Saúde (SNS), um objetivo central do sistema consiste na prossecução da equidade na distribuição dos serviços e no acesso. No caso do SNS português, as leis fundamentais instituem cobertura universal da população, acesso tendencialmente gratuito a cuidados de saúde e financiamento do sistema público através de impostos e do orçamento geral do estado. Neste contexto, o planeamento estratégico de uma rede hospitalar num SNS envolve decisões de distribuição dos recursos de forma a promover a equidade geográfica e a considerar questões de eficiência e custos.

O presente trabalho descreve a resolução de um caso de estudo no SNS português onde se pretende analisar como reestruturar a atual rede hospitalar na zona sul do país de forma a melhorar o acesso das populações aos cuidados hospitalares. Esta melhoria no acesso pode ser alcançada com a reorganização da atual rede e principalmente com a construção de unidades de menor dimensão, face às existentes na rede atual, as quais se localizarão mais próximas da população. Todavia, estas soluções podem resultar como demasiado dispendiosas e levar a ineficiências na prestação de serviços, havendo pois a necessidade de estabelecer um compromisso entre dois fatores chave: acesso e custo. A fim de conseguir atingir estes objetivos, desenvolveu-se um modelo de otimização hierárquico e multi-serviço que permite analisar decisões relativas à (re)organização de redes hospitalares no contexto de um SNS. Nesta abordagem deu-se ênfase à melhoria do acesso estimando as respetivas implicações nos custos hospitalares (abrangendo custos de investimento e operacionais). Numa fase posterior do trabalho pretende-se estabelecer soluções de compromisso entre estes dois objetivos de extrema importância no planeamento hospitalar. O modelo pode pois ser usado como uma ferramenta de apoio à decisão no planeamento integrado de redes hospitalares. O uso do modelo gera informação chave para responder simultaneamente a várias questões relevantes para o planeamento hospitalar, nomeadamente: onde localizar novas unidades hospitalares e como redimensionar as unidades hospitalares existentes? Que combinação de serviços deve cada unidade hospitalar prestar? Como devem as unidades

hospitalares estar organizadas na rede de prestadores e no sistema de referenciação? Que área de influência deve ter cada unidade hospitalar, e que populações devem estar afetadas a cada hospital? Que fluxos de utentes, intra- e inter-hospitalares são expectáveis? Qual o custo de investimento e de prestação na rede para diferentes níveis de acesso?

O modelo desenvolvido é uma extensão do modelo clássico de localização da p-mediana, e considera como enquadramento um decisor que pretende maximizar o acesso no âmbito de um SNS com as seguintes características:

- ✓ A rede hospitalar é representada através de um sistema hierárquico com dois níveis que disponibilizam diferentes tipos de serviço. O nível inferior presta cuidados de proximidade às populações referenciando os utentes para o nível superior sempre que existe necessidade de cuidados mais diferenciados. Por outro lado, o nível superior presta ambos os cuidados de saúde à sua área de influência direta;
- ✓ As unidades hospitalares são consideradas prestadoras de múltiplos serviços: internamento, urgência e consulta externa;
- ✓ Existem fluxos de utentes entre serviços hospitalares de um mesmo hospital e entre hospitais de diferentes níveis hierárquicos;
- ✓ São impostos limites na capacidade hospitalar, o que indiretamente considera questões de economia de escala e eficiência na prestação dos serviços;
- ✓ Como regra de planeamento, é considerado que a afetação de uma população é feita para a unidade mais próxima, regra pertinente para a definição de áreas de influência compactas;
- ✓ O modelo considera custos de investimento relacionados com a construção e redimensionamento dos hospitais, assim como custos operacionais decorrentes da atividade diária de prestação de serviços hospitalares – a análise de custos é central para a reorganização de uma rede hospitalar.

O caso de estudo, como referido anteriormente, envolve a reorganização da rede hospitalar da zona sul de Portugal. Esta região é independente e autossuficiente na prestação de cuidados de saúde e integra as áreas cobertas pelas Administrações Regionais de Saúde de Lisboa e Vale do Tejo, Alentejo e

Algarve. Estas regiões apresentam características e populações distintas, assim como redes hospitalares com diferentes níveis de especialização e tecnologia. Enquanto Lisboa e Vale do Tejo é uma zona urbana com boas acessibilidades físicas, onde se concentram grande parte dos recursos hospitalares para tratamento de casos mais complexos, o Alentejo é uma região rural que tem fracas acessibilidades e uma população envelhecida e dispersa pelo território; a contrastar, o Algarve é uma região com zonas urbanas e rurais e que apresenta uma grande flutuação da população devido à sazonalidade da sua maior atividade económica, o turismo.

Atualmente existem vários projetos para a reorganização desta rede hospitalar que definem a construção de novos hospitais e a substituição de unidades existentes. Estando estes projetos em fase de estudo, discussão e construção, o modelo de otimização desenvolvido foi usado para analisar se estes projetos são os que mais contribuem para melhorias no acesso.

Os dados usados no caso de estudo resultam de informação institucional regularmente disponibilizada, e os resultados obtidos ilustram de que forma o modelo pode ser usado para apoiar os decisores no planeamento de redes hospitalares. Para tal, estudaram-se vários cenários que permitiram testar diferentes políticas de reorganização da rede e implicações no acesso e custo. Uma versão mais alargada do trabalho aqui descrito pode ser consultada em Mestre et al. (2012).

O presente trabalho prossegue com uma fundamentação teórica onde se caracterizam os modelos de localização e se analisam alguns dos trabalhos mais relevantes realizados na literatura da área. Em seguida, é feito o enquadramento do modelo desenvolvido para a otimização de redes hospitalares, são descritos os principais pressupostos assumidos na elaboração do modelo e a sua formulação matemática. Posteriormente, o caso de estudo é analisado em detalhe, sendo descrita a informação de base usada e apresentados exemplos de alguns resultados obtidos com a aplicação do modelo. Na secção final é feito o balanço do trabalho realizado e discutem-se futuros desenvolvimentos.

2. Modelos de localização

2.1. Aspectos genéricos em modelos de localização

Os modelos de localização têm sido frequentemente usados como ferramentas de apoio à decisão de planeamento de sistemas de saúde, tal como descrevem Daskin e Dean (2004). A literatura na área de localização é vasta, pelo que a presente revisão bibliográfica não pretende ser exaustiva, identificando-se apenas os trabalhos mais relevantes para o enquadramento do trabalho desenvolvido. Informação mais detalhada sobre modelos de localização pode ser encontrada no livro de Daskin (1995), ou nos artigos de revisão de Owen e Daskin (1998), Marianov e Serra (2004), ReVelle e Eiselt (2005) e Melo et al. (2009). No contexto de sistemas de saúde, diversas aplicações têm surgido em países em desenvolvimento, tal como revisto por Rahman e Smith (2000) e descrito por Smith et al. (2009).

Daskin (2008) classificou os modelos de localização discretos em dois grandes grupos: os modelos de cobertura e os modelos p-mediana. Nos modelos p-mediana o objetivo é minimizar o tempo médio (ou total) de percurso normalmente ponderado pela procura. Por outro lado, os modelos de cobertura são caracterizados pela existência de uma distância crítica que não pode ser excedida para a procura ser considerada satisfeita, e surgem na literatura com diferentes objetivos: minimizar o número de unidades necessárias para cobrir toda a procura ('set covering models'), maximizar a procura coberta com um número pré-definido de unidades ('maximal covering models'), ou minimizar a distância necessária para cobrir toda procura com um número pré-definido de unidades ('p-center models'). Marianov e Serra (2004) analisam aplicações no sector público de ambos os tipos de modelos de localização. Em alternativa, alguns autores tentaram conjugar as vantagens dos modelos de cobertura e dos modelos p-mediana, introduzindo uma restrição que estabelece uma distância máxima (ou tempo máximo de percurso) que não deve ser ultrapassada. Este conceito foi introduzido por Toregas et al. (1971) e usado no contexto de sistemas de saúde por vários autores, incluindo Ruth (1981) e Meherez et al. (1996).

O modelo construído no presente estudo segue este último caso: tem uma estrutura p-mediana, minimizando o tempo de acesso ponderado pelo número

de utentes, assim como considera restrições com os tempos máximos de percurso aceitáveis.

2.2. Objetivos para planeamento de sistemas de saúde

Os modelos de localização desenvolvidos para o contexto de um sistema de saúde frequentemente consideram objetivos de custo, equidade, eficiência e acesso. Dentro destes objetivos a prossecução de equidade tem tido especial ênfase e sido um tópico investigado por diversos estudos. Contudo, múltiplas definições alternativas de equidade são citadas na literatura. Marsh e Schilling (1994) apresentam uma revisão das diversas medidas propostas para atingir equidade, sendo que este conceito depende da aplicação em causa e das preferências do decisor. No presente estudo, o modelo proposto maximiza a acessibilidade, considerando indiretamente questões de custo e de eficiência.

2.3. Modelos de localização para planeamento de sistemas de saúde

Tendo em conta as especificidades dos sistemas de saúde, vários modelos hierárquicos têm sido usados no planeamento desses sistemas (ReVelle e Eiselt, 2005). Estes modelos permitem modelar o sistema como um todo e evitam soluções sub-ótimas que se obtêm quando cada nível de serviço é otimizado individualmente. Narula (1986) propôs uma classificação para os modelos de localização hierárquicos que os divide em modelos hierárquico inclusivos, quando a unidade disponibiliza todos os níveis inferiores de serviço, e hierárquico exclusivos, quando uma unidade presta apenas um nível de serviço. O autor também reviu os artigos aplicados na área da saúde até à data, e concluiu que é mais comum usar modelos hierárquicos inclusivos. Uma revisão mais atual de modelos hierárquicos pode ser encontrada em Sahin e Süral (2007).

Alguns exemplos de modelos de localização hierárquicos aplicados ao sistema de saúde são da autoria de Dökmeci (1973), Ruth (1981), Moore e ReVelle (1982), Galvão et al. (2002), Smith et al. (2009) e Smith et al. (2010). Dökmeci (1973) introduziu um modelo hierárquico para determinar a localização e capacidade de serviços de saúde de forma a minimizar os custos do sistema. Ruth (1981) propôs um modelo para o planeamento hospitalar

considerando três níveis hierárquicos e restrições de capacidade mínima que indiretamente modelam questões relacionadas com economias de escala e eficiência. Moore e ReVelle (1982) desenvolveram uma extensão hierárquica de um modelo de cobertura para otimizar os serviços de saúde nas Honduras. Galvão et al. (2002) desenvolveram um modelo hierárquico de três níveis para a organização de serviços de assistência materna e perinatal no Município do Rio de Janeiro. Smith et al. (2009), dentro do contexto de países em desenvolvimento, propuseram um conjunto de modelos hierárquicos do tipo de cobertura e p-mediana. Smith et al. (2010) analisaram os modelos hierárquicos numa perspectiva bi-objetivo equilibrando equidade e eficiência.

Embora esta revisão bibliográfica seja orientada para os modelos hierárquicos devido à sua relevância no planeamento de sistemas de saúde, e em particular no trabalho desenvolvido para redes hospitalares, importa referir que existem na literatura outras abordagens importantes, que todavia não serão aqui analisadas.

Acresce que o planeamento de redes hospitalares envolve decisões que são difíceis de reverter, que têm impactos políticos e que influenciam a qualidade de vida das populações. O desenvolvimento de métodos que permitam uma análise quantitativa e global da rede é pois fundamental para o apoio à decisão e relevante para a sustentação das escolhas efetuadas. O modelo desenvolvido apresenta-se como uma ferramenta de apoio à decisão para o planeamento de redes hospitalares que permite informar os decisores sobre questões relativas à localização e dimensionamento de redes hospitalares. O modelo é do tipo inteiro misto, tem uma estrutura hierárquica, e contribui para a literatura ao considerar simultaneamente vários aspetos relevantes para o planeamento, i.e., retratando unidades hospitalares como prestadoras de vários serviços, fluxos de utentes entre níveis hierárquicos e serviços hospitalares, e limites na capacidade instalada.

3. Modelo de localização hierárquico

Tal como referido anteriormente, em sistemas de saúde baseados num SNS os principais objetivos a alcançar prendem-se com a promoção da igualdade de acesso e com a equidade na distribuição espacial dos recursos e na sua

utilização. Inerente a estes sistemas está a cobertura universal da população e o acesso tendencialmente gratuito aos serviços. Os decisores políticos em países com SNS têm de planear o sistema tendo em conta as necessidades de cuidados de saúde e questões orçamentais.

No presente trabalho apresenta-se um modelo de localização hierárquico de programação matemática que tem como objetivo auxiliar o planeamento integrado de redes hospitalares, informando simultaneamente sobre várias decisões como a localização de novas unidades hospitalares e o redimensionamento de unidades hospitalares existentes, a definição de redes referenciação entre hospitais, a definição das áreas de influência dos hospitais, a composição da oferta de serviços em cada unidade hospitalar, e os fluxos de utentes intra- e inter- hospitalar. As várias opções usadas são descritas em seguida e, posteriormente será apresentado o modelo desenvolvido.

3.1. Estrutura hierárquica e fluxos de utentes

O modelo proposto considera que as redes hospitalares estão organizadas em sistemas hierárquicos (inclusivos) com dois níveis de serviços. No nível inferior, unidades de menor dimensão prestam cuidados de proximidade às populações da sua área direta de influência. Estas unidades são denominadas de Hospitais Distritais (HD). No nível superior, unidades mais especializadas recebem utentes transferidos de unidades de nível inferior. Estas unidades são denominadas de Hospitais Centrais (HC) e prestam tantos serviços especializados a uma maior área de influência como cuidados de proximidade às populações locais dentro da sua área de influência direta.

O modelo apresentado neste trabalho considera que os hospitais são prestadores de três tipos de serviços: internamento, urgência e consulta externa. Considera ainda que os hospitais estão organizados em redes que cooperam entre si, existindo um fluxo de utentes entre serviços e níveis hierárquicos. A Figura 1 contextualiza o problema e evidencia as principais características da rede hospitalar que serão analisadas nas subsecções seguintes. Os fluxos de utentes na hierarquia devem ser interpretados da seguinte forma: recorrendo a taxas de utilização normalizadas, a população é convertida em procura para os três serviços de saúde que podem ser disponibilizados pelos HD ou HC; se um

utente é servido por uma unidade de nível inferior, pode ser transferido para uma unidade de nível superior para receber serviços mais especializados não disponíveis na unidade de nível inferior; o fluxo inverso (de HC para HD) pode ocorrer depois de um atendimento; se um utente dá entrada no serviço de urgência pode ser internado. Os fluxos representados na Figura 1 são os mais relevantes dentro de um sistema hospitalar, mesmo que outros fluxos possam ser considerados no modelo.

Atendendo às características do sistema e à necessidade de modelar em detalhe os fluxos na rede, optou-se por desenvolver uma extensão de um modelo *p-mediana*. Esta tipologia de modelos é apropriada para retratar situações em que o tempo médio (ou total) de percurso é relevante (Daskin e Dean, 2004). O modelo inclui ainda restrições que limitam o tempo máximo de percurso (ou distância máxima). De acordo com o descrito na revisão bibliográfica, a inclusão desta restrição permite conjugar os modelos de cobertura com os modelos *p-mediana*.

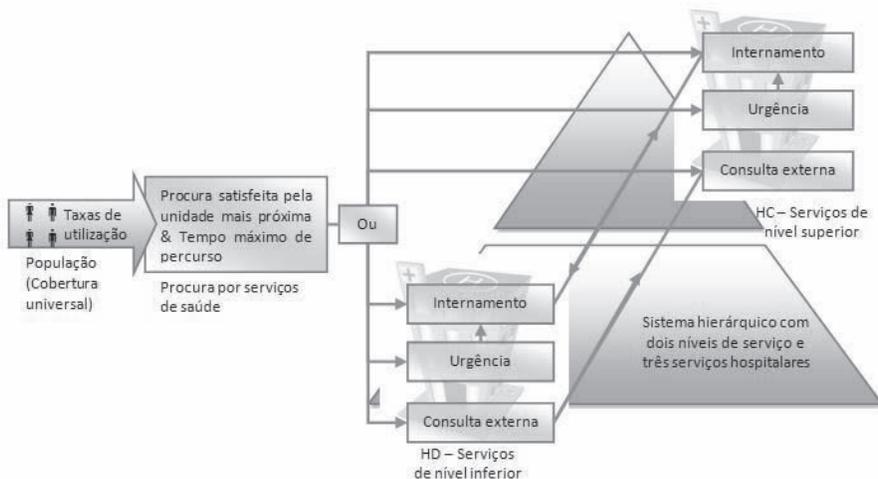


Figura 1: Esquematização do sistema e dos fluxos hospitalares

3.2. Critérios de afectação

Outro aspeto relevante para o planeamento de redes hospitalares é a escolha do critério de afetação. Dependendo do sistema em questão, os utentes podem estar afetos a uma unidade ou serem livres de escolher onde querem

ser atendidos. No caso de os utentes serem livres de escolher o prestador de cuidados de saúde, é necessário compreender os fatores que influenciam as suas escolhas e modelar o seu comportamento (Mayhew e Leonardi, 1982).

Compreendendo as vantagens e desvantagens dos diferentes critérios, optou-se no presente trabalho por definir a afetação das populações a uma única unidade, devido ao facto de no SNS português estar definido na prática um sistema de referenciação e de não promover mecanismos de livre escolha por parte dos utentes. Por outro lado, a necessidade de criação de áreas de influência compactas levou à escolha de uma regra de planeamento em que a afetação é feita à unidade mais próxima (Gerrard e Church, 1996).

3.3. Limites de capacidades e seleção de potenciais localizações

O modelo desenvolvido considera limites na capacidade instalada, dado que uma dimensão crítica de serviços está relacionada com qualidade na prestação de serviços e indiretamente permite modelar economias de escala na prestação dos serviços.

Em modelos de planeamento, a escolha do conjunto de potenciais novas localizações reveste-se de grande relevância. O estabelecimento de um requisito mínimo de população para as potenciais localizações de novas unidades é importante para assegurar que existem recursos físicos e humanos para prestar os serviços numa dada localização. Os decisores e utilizadores dificilmente aceitariam unidades em localizações de pequena dimensão, pelo que esta situação é considerada no modelo. Este critério é também compatível com a inclusão no modelo de limites de capacidade mínima, dado que por questões de eficiência, custo e qualidade não devem existir hospitais de muito pequena (ou de muito grande) dimensão. Para os casos de hospitais previamente existentes, a reorganização da rede hospitalar deve ter em conta essa superestrutura e as opções políticas de alteração dessa rede, tal como é o caso da rede hospitalar analisada neste trabalho (ver secção 4).

3.4. Síntese da estrutura do modelo: inputs, outputs e objetivo

Para além das questões atrás analisadas, o modelo foi construído de forma a incluir outras questões importantes para o planeamento, fornecendo

informação relevante para o planeamento de redes hospitalares, nomeadamente sobre: o acesso geográfico, a localização das unidades hospitalares, a afetação das populações, as áreas de influência, os fluxos de utentes e indicadores do nível de serviço. A estrutura global do modelo desenvolvido está sintetizada na Tabela 1.

Tabela 1: Esquematização do modelo: inputs, outputs e objetivo.

Dada a(o): (inputs)	Procura por serviços hospitalares Percentagens de transferências entre hospitais e entre serviços Capacidades hospitalares máximas e mínimas Tempo médio de percurso entre pontos da rede Tempo médio de internamento Custos operacionais e de investimento Superestrutura existente da rede e possíveis alterações
Determinar (outputs)	A estrutura da rede: localização das unidades hospitalares, afetação das populações, capacidade hospitalar, fluxos de utentes e indicadores de nível de serviço
De forma a	Melhorar o acesso através da minimização do tempo de percurso ponderado pela utilização

3.5. Síntese da estrutura do modelo: inputs, outputs e objetivo

O modelo de programação matemática desenvolvido faz uso da notação apresentada na Tabela 2.

Tabela 2: Notação do modelo hierárquico: índices, variáveis e parâmetros.

	Notação	Descrição
Índices	$i \in I$	Conjunto dos pontos de procura
	$j, j' \in J$	Conjunto das potenciais localizações para Hospitais Distritais (HD)
	$k, k' \in K$	Conjunto das potenciais localizações para Hospitais Centrais (HC)
	$w, v, a \in W$	Conjunto dos serviços hospitalares ($w=$ int para serviço de internamento, $w=$ urg para serviço de urgência e $w=$ ce para serviço de consulta externa)
Variáveis	X_{jw}^D X_{kw}^C	$=1$ se o HD/HC é localizado no ponto j/K e disponibiliza o serviço w ; $=0$ caso contrário
	Y_{ijw}^D Y_{ikw}^C	Fluxos de utentes entre o ponto de procura i e o serviço w no HD/HC j/k
	Y_{jkvw}^{DC} $Y_{k'jvw}^{CD}$	Fluxos de utentes entre o serviço w no HD j e o serviço v no HC k , e vice-versa
	cap_{jw}^D cap_{kw}^C	Capacidade do serviço w no HD j /HC k (variáveis auxiliares em camas para o serviço de internamento e número de utentes servidos nos restantes serviços)
Parâmetros	d_{ij}^1 d_{ik}^2	Tempo médio de percurso entre os pontos de procura e HD j / HC k
	d_{jk}^3 d_{kj}^4	Tempo médio de percurso entre o HD j e HC k / HC k e HD j
	α	Ponderador que permite diferenciar uma primeira entrada de uma transferência

	Notação	Descrição
Parâmetros	D_{iw}	Procura pelo serviço w no ponto i
	S_{aw}	Percentagem de utentes transferidos do serviço a para o serviço w dentro do mesmo hospital
	S_{ww}^{DC}	Percentagem de utentes transferidos do serviço w num HD para o serviço v num HC
	S_{vw}^{CD}	Percentagem de utentes transferidos do serviço v num HC para o serviço w num HD
	$alos_w^D$ $alos_v^C$	Demora média da estadia no serviço w/v num HD/HC
	$capmin_{jw}^D$	Capacidade mínima necessária no serviço w no HD j / HC k
	$capmin_{kw}^C$	Capacidade mínima necessária no serviço w no HD j / HC k
	$capmax_{jw}^D$	Capacidade máxima necessária no serviço w no HD j / HC k
	$capmax_{kw}^C$	Capacidade máxima necessária no serviço w no HD j / HC k
	b^w	Assiste a conversão de fluxos em capacidade instalada (em camas no internamento)
	d^D d^C d^{\max}	Tempo máximo de percurso para aceder a um HD / HC
$icap_{jint}^D$	Capacidade atualmente instalada no serviço de internamento do HD j	

De acordo com a notação descrita, o modelo de programação matemática pode ser definido através da função objetivo [1] e das restrições [2] a [19] que de seguida se apresentam.

$$\begin{aligned} \text{Min} \quad & \sum_{i \in I} \sum_{j \in J} \sum_{w \in W} d_{ij}^1 Y_{ijw}^D + \sum_{i \in I} \sum_{k \in K} \sum_{v \in W} d_{ik}^2 Y_{ikv}^C \\ & + \sum_{j \in J} \sum_{k \in K} \sum_{w \in W} \sum_{v \in W} \alpha \left(d_{jk}^3 Y_{jkwv}^{DC} + d_{kj}^4 Y_{kvw}^{CD} \right) \end{aligned} \quad [1]$$

Sujeito a:

$$\sum_{j \in J} Y_{ijw}^D + \sum_{k \in K} Y_{ikw}^C = D_{iw} \quad \forall_{i \in I} w \in W \quad [2]$$

$$\left[\sum_{i \in I} Y_{ijw}^D + \sum_{i \in I} \sum_{a \in W} Y_{ija}^D \cdot S_{aw} \right] S_{ww}^{DC} = \sum_{k \in K} Y_{jkwv}^{DC} \quad \forall_{j \in J} w \in W v \in W \quad [3]$$

$$Y_{kvw}^{CD} = S_{vw}^{CD} \cdot Y_{jkwv}^{DC} \quad \forall_{j \in J} k \in K w \in W v \in W \quad [4]$$

$$cap_{jw}^D = \left[\sum_{i \in I} Y_{ijw}^D + \sum_{i \in I} \sum_{a \in W} Y_{ija}^D \cdot S_{aw} + \sum_{k \in K} \sum_{v \in W} Y_{kvw}^{CD} \right] \times \frac{alos_w^D}{b^w} \quad \forall_{j \in J} w \in W \quad [5]$$

$$cap_{kv}^C = \left[\sum_{i \in I} Y_{ikv}^C + \sum_{i \in I} \sum_{a \in W} Y_{ika}^C \cdot S_{av} + \sum_{j \in J} \sum_{w \in W} Y_{jkwv}^{DC} \right] \times \frac{alos_v^C}{b^v} \quad \forall_{k \in K} v \in W \quad [6]$$

$$cap \min_{jw}^D X_{jw}^D \leq cap_{jw}^D \leq cap \max_{jw}^D X_{jw}^D \quad \forall_{j \in J \ w \in W} \quad [7]$$

$$cap \min_{kw}^C X_{kw}^C \leq cap_{kw}^C \leq cap \max_{kw}^C X_{kw}^C \quad \forall_{k \in K \ w \in W} \quad [8]$$

$$Y_{ijw}^D + D_{iw} X_{j'w}^D \leq D_{iw} \quad \forall_{i \in I \ j \in J \ j' \in \{j \mid d_{ij}^1 < d_{ij}^1\} \ w \in W} \quad [9]$$

$$Y_{ikw}^C + D_{iw} X_{k'w}^C \leq D_{iw} \quad \forall_{i \in I \ k \in K \ k' \in \{k \mid d_{ik}^2 < d_{ik}^2\} \ w \in W} \quad [10]$$

$$Y_{ijw}^D + D_{iw} X_{kw}^C \leq D_{iw} \quad \forall_{i \in I \ j \in J \ k \in \{k \mid d_{ik}^2 < d_{ij}^1\} \ w \in W} \quad [11]$$

$$Y_{ikw}^C + D_{iw} X_{jw}^D \leq D_{iw} \quad \forall_{i \in I \ j \in \{j \mid d_{ij}^1 < d_{ik}^2\} \ k \in K \ w \in W} \quad [12]$$

$$Y_{jkwv}^{DC} + D_{iw} X_{k'w}^C \leq D_{iw} \quad \forall_{j \in J \ k \in K \ k' \in \{k \mid d_{jk}^3 < d_{jk}^3\} \ w \in W \ v \in W} \quad [13]$$

$$Y_{ijw}^D \leq D_{iw} X_{jw}^D \quad \forall_{i \in I \ j \in J \ w \in W} \quad [14]$$

$$Y_{ikw}^C \leq D_{iw} X_{kw}^C \quad \forall_{i \in I \ k \in K \ w \in W} \quad [15]$$

$$Y_{jkwv}^{DC} \leq S_{wv}^{DC} \sum_{i \in I} D_{iw} X_{kw}^C \quad \forall_{j \in J \ k \in K \ w \in W \ v \in W} \quad [16]$$

$$Y_{ijw}^D = 0 \quad \forall_{i \in I \ j \in J \ w \in W \ j \in \{j \mid d_{ij}^1 > d^{Dmax}\}} \quad [17]$$

$$Y_{ikw}^C = 0 \quad \forall_{i \in I \ k \in K \ w \in W \ k \in \{k \mid d_{ik}^2 > d^{Cmax}\}} \quad [18]$$

$$X_{jw}^D X_{kw}^C \in \{0, 1\}, \quad Y_{ijw}^D Y_{ikw}^C Y_{jkwv}^{DC} Y_{kjwv}^{CD} \geq 0 \quad \forall_{i \in I \ j \in J \ k \in K \ w \in W \ v \in W} \quad [19]$$

A função objetivo representada em [1] pretende melhorar o acesso geográfico, através da minimização do tempo total de percurso ponderado pela utilização. A restrição [2] é relativa à procura de serviços hospitalares, determinando que toda a procura deve ser satisfeita por um HD ou por um HC. As restrições [3] e [4] definem respetivamente os fluxos ascendentes e descendentes na hierarquia. As restrições [5] e [6] são auxiliares na formulação matemática e determinam a capacidade hospitalar com base nos fluxos de entrada direta nas unidades hospitalares, nas transferências de outros serviços dentro da mesma unidade e de transferências de unidades de nível diferente. No caso do serviço de internamento, a capacidade é convertida em camas. As restrições [7] e [8] estabelecem que a capacidade hospitalar é limitada por valores máximos e mínimos que indiretamente modelam questões de economias de escala. As equações [9] a [13] modelam a regra de afetação

escolhida, estabelecendo que a procura deve ser satisfeita pela unidade mais próxima. As restrições [14] a [16] asseguram que a afetação e referenciação são feitas para unidades abertas. Estas restrições são redundantes na formulação matemática pois as restrições de capacidade máxima ([7] e [8]) igualmente regulam a afetação, sendo porém usadas para fortalecer a relaxação linear e melhorar o desempenho do modelo. As restrições [17] e [18] impõem um tempo máximo de percurso que não pode ser excedido para a procura ser considerada satisfeita. As restrições [19] são relativas à natureza das variáveis e definem condições de integralidade e não negatividade.

O modelo de programação matemática anteriormente descrito permite modelar de forma genérica questões relevantes no planeamento de redes hospitalares. No entanto, o modelo pode ser expandido para contemplar outras questões específicas relacionadas com a sua aplicação em casos particulares. As restrições [20] a [22] são exemplos de questões adicionais que podem surgir – as restrições estão definidas para o caso de HD, mas restrições semelhantes podem ser definidas para HC.

$$\sum_{j \in J} X_{jw}^D = p \quad \forall_{w \in W} \quad [20]$$

$$X_{jw}^D = 0 \quad \forall_{j \in \{j | \text{pop}_j < \text{pop}^{\text{Dmin}}\}} \quad \forall_{w \in W} \quad [21]$$

$$X_{j\text{int}}^D \geq X_{j\text{urg}}^D \quad \wedge \quad X_{j\text{int}}^D \geq X_{j\text{ce}}^D \quad \forall_{j \in J} \quad [22]$$

A restrição [20] determina o número de unidades em funcionamento/ou a abrir, definido por p. Na restrição [21] as potenciais localizações para novas unidades são avaliadas em função da sua população. Desta forma a restrição determina que a população de uma potencial localidade (pop_j) deve ter uma dimensão mínima (pop^{Dmin}), caso contrário essa abertura não é permitida ($X_{jw}^D=0$). Apesar de ser possível a prestação dos diferentes serviços hospitalares em separado, existem sinergias na sua prestação conjunta dentro da mesma unidade, e economias de gama (Samuelson e Nordhaus, 2010) uma vez que a prestação de diferentes serviços pode partilhar os mesmos recursos físicos e humanos, gerando ganhos a nível da produção hospitalar e no atendimento ao

utente. Desta forma a restrição [22] estabelece que, por exemplo, o serviço de urgência ($w=urg$) e de consulta externa ($w=ce$) só pode ser prestado por unidades que disponham do serviço de internamento ($w=int$).

O modelo apresentado considera que o principal objetivo do decisor é a melhoria da acessibilidade geográfica, o que justifica a escolha da função objetivo [1]. Contudo, na reorganização da rede, questões relativas ao custo são indispensáveis para viabilizar os projetos. Neste estudo os custos operacionais e de investimento em novas unidades são determinados após a otimização, através das equações [23] e [24], produzindo informação para o decisor sobre o custo de cada configuração da rede.

$$TOC^D = \sum_{w \in W} \sum_{j \in J} cap_{jw}^D CO_{jw}^D \quad [23]$$

$$TIC^D = \sum_{j \in J_E} (cap_{jint}^D - icap_{jint}^D) CE_{jint}^D + \sum_{J \in J_N} cap_{jint}^D CI_{jint}^D \quad [24]$$

A equação [23] permite determinar os custos operacionais relativos aos hospitais distritais (TOCD) através da capacidade hospitalar instalada (cap_{jw}^D) e dos custos operacionais de prestação de cada serviço (CO_{jw}^D). Através da equação [24] é possível determinar os custos de investimento em hospitais distritais (TIC^D) que se dividem em custos de expansão (CE_{jint}^D) de camas em serviços de internamento ($w=int$) de hospitais existentes (J_E), e custos de investimento (CI_{jint}^D) para construção de novas unidades (J_N). Para os HC foram introduzidas restrições semelhantes.

4. Resolução do Caso de estudo

O modelo desenvolvido foi utilizado para estudar a reconfiguração da rede hospitalar da zona sul de Portugal que, como foi anteriormente descrito, é constituída por três Regiões de Saúde com características diferentes. Vários projetos de reorganização desta rede hospitalar têm sido discutidos publicamente, tendo sido alguns aprovados (estando atualmente em fase de construção) e outros encontram-se em fase de discussão e análise, dada a atual situação económica do país. Estes projetos que envolvem a reorganização da

rede estão sumariamente descritos na Tabela 3 e serão explorados ao longo do caso de estudo.

Tabela 3: Projetos para a reorganização da rede Hospitalar da zona sul de Portugal.

Região de Saúde de Lisboa	<ul style="list-style-type: none"> - Unidade de substituição com 600 camas em Lisboa. - Novas unidades em Sintra e Loures (unidade que está atualmente em construção sendo esperada a sua abertura para breve) - Unidades de substituição em Cascais e Vila Franca de Xira - Nova unidade com cerca de 150 camas no Seixal - Expansão da unidade de Almada
Região de Saúde do Alentejo	<ul style="list-style-type: none"> - Unidade de substituição de carácter central em Évora com cerca de 384 camas
Região de Saúde do Algarve	<ul style="list-style-type: none"> - Manutenção do Hospital de Faro (com cerca de 486 camas) - Construção de uma nova unidade de carácter central (550 camas) no eixo Faro-Loulé - Construção de um novo HC e manter parte da capacidade da atual unidade localizada em Faro

Na secção seguinte (secção 4.1) são apresentados e discutidos os dados usados na resolução do caso de estudo e posteriormente são analisados os resultados obtidos (secção 4.2).

4.1. Dados

A aplicação do modelo hierárquico multi-serviço desenvolvido necessitou de informação relacionada com a procura e oferta dos serviços hospitalares, dados que estão sintetizados na Tabela 4. A procura por serviços de saúde foi determinada com base em estimativas da população por concelho (INE, 2006) e em índices de utilização dos serviços. Para estimar os índices de utilização do serviço de internamento recorreu-se à base de dados que contém informação sobre atendimentos no serviço de internamento classificados de acordo com os Grupos de Diagnóstico Homogéneo (GDH). Os índices de utilização para o serviço de internamento foram determinados tendo em conta a faixa etária e o género dos utentes. Esta base de dados permitiu ainda quantificar as transferências entre níveis hierárquicos no serviço de internamento. Para o caso do serviço de urgência e consulta externa foi usada informação reportada pela Direção Geral de Saúde (Ministério da Saúde, 2003), a qual permitiu ainda

estimar a percentagem de utentes que são internados após uma urgência e que necessitam de uma segunda consulta externa num HC.

Tabela 4: Parâmetros do modelo hierárquico.

	Notação	Valor	Descrição
Parâmetros Gerais	D_{iw}	$D_{i\text{int}}$	Índices de utilização determinados com base na utilização observada, de acordo com a faixa etária e género
		579	Número de consultas externas esperadas por 1 000 de habitantes
		584	Número de urgências esperadas por 1 000 de habitantes
	S_{aw}	9%	Percentagem de utentes transferidos da urgência para o internamento
	S_{wv}^{DC}	13%	Percentagem de utentes transferidos do internamento em HD para HC
		19,6%	Percentagem de utentes que necessita de uma consulta externa em HC depois de uma consulta em HD
	S_{vw}^{CD}	1.24%	Percentagem de utentes transferidos do internamento em HC para HD
	$alos_w^D$ $alos_v^C$	7 / 8.9	Demora média no internamento em HD/HC
	$capmin_{jw}^D$	150	Capacidade mínima no internamento para novos HD
	$capmin_{kw}^C$	300	Capacidade mínima no internamento para novos HC
	$pop^{D\text{min}}$	15 000	População mínima para a abertura de um HD
	$pop^{C\text{min}}$	50 000	População mínima para a abertura de um HC
	α	0.5	Uma transferência recebe metade do peso de uma entrada direta
Parâmetros de custos	CO_{jw}^D CO_{kw}^C	€ 380 / € 544	Custo médio de um dia de internamento em HD/HC
		€ 143 / € 119	Custo médio de uma urgência em HD/HC
		€ 94 / € 118	Custo médio de uma consulta externa em HD/HC
	$CI_{j\text{int}}^D$	€ 200 000	Custo de uma nova cama num novo HD
	$CI_{k\text{int}}^C$	€ 224 500	Custo de uma nova cama num novo HC
	$CE_{j\text{int}}^D$	€ 100 000	Custo de expansão de uma cama em HD existente
$CE_{k\text{int}}^C$	€ 124 750	Custo de expansão de uma cama em HC existente	

Relativamente à produção hospitalar foi necessário estimar a demora média no serviço de internamento, onde mais uma vez se recorreu à base de dados de GDH. O modelo considera a demora média diferente nos dois níveis hierárquicos para capturar as diferenças existentes na complexidade das situações tratadas em cada nível hierárquico. Relativamente à capacidade mínima a instalar em novas unidades, os dados usados têm por base o estudo para a reorganização da capacidade de Lisboa: (Ministério da Saúde, 2006a).

No presente estudo optou-se por não introduzir capacidade máxima nas unidades uma vez que, nas instâncias geradas com o caso de estudo da zona sul, existe uma rede hospitalar pré-existente que, ao ser considerada simultaneamente com limites na capacidade instalada e restrições de afetação à unidade mais próxima, resultaria numa estrutura pouco flexível e levaria a soluções não viáveis. A relaxação desta restrição de capacidade máxima nesta primeira abordagem permite dar ênfase à definição de áreas de influência.

As estimativas de custos operacionais foram determinadas com base em informação disponibilizada pela Administração Central do Sistema de Saúde que reporta a contabilidade analítica das unidades hospitalares (Ministério da Saúde, 2006b). Os dados relativos aos custos de investimento e de expansão são de um estudo realizado na Região de Saúde do Algarve (Região Saúde do Algarve, 2004).

O caso de estudo baseado na zona sul de Portugal compreende regiões urbanas e rurais com acessibilidades muito baixas, nomeadamente no Alentejo e no interior do Algarve. Neste contexto, a introdução de distâncias críticas relevantes para definir um nível de serviço adequado não é compatível com a definição de capacidade mínima hospitalar, tal como ilustrado em Mestre et al. (2012). Por este motivo optou-se por introduzir uma cobertura menos restritiva. Dado que o modelo considera a superestrutura inicial da rede e que esta contempla um HD em cada distrito, as restrições [17] e [18] foram adaptadas para impedir a afetação das populações a unidades fora do distrito (com exceção dos concelhos limítrofes onde já se observa a afetação para uma unidade fora do distrito). Esta definição de afetação é relevante em termos de planeamento regional e torna o modelo mais eficiente do ponto de vista computacional, uma vez que a eliminação de possíveis afetações diminui a complexidade do modelo. A adaptação das restrições [17] e [18] é feita alterando o domínio das restrições que em vez de ser limitado por uma distância ($d^D \max, d^C \max$), passa a ser controlado por um parâmetro que define se o ponto de procura $i \in I$ e a unidade $j \in J$ pertencem ao mesmo distrito.

A qualidade dos resultados depende em grande parte do rigor dos dados. A informação acima descrita, apesar de ser a melhor disponível, apresenta algumas lacunas, nomeadamente no que respeita à qualidade da informação

sobre índices de utilização, transferências e custos. Por exemplo, os índices de utilização retratam a procura observada no passado, a qual pode estar condicionada pela oferta disponível; adicionalmente, espera-se que alterações devido à evolução tecnológica ou ao desenvolvimento de novas técnicas em saúde alterem o padrão futuro de prestação de cuidados de saúde. Face a estas limitações, foi realizada uma análise de sensibilidade aos parâmetros que têm maior incerteza associada.

4.2. Resultados

Na aplicação do modelo são estudados dois cenários (I e II) que capturam diferentes opções de reconfiguração da rede hospitalar. Estes cenários não têm como objetivo apresentar uma única resposta para a otimização da rede, mas ilustrar como o modelo pode ser usado para apoiar o processo de decisão e fornecer informação ao decisor sobre as consequências de diferentes alternativas.

O modelo foi implementado na linguagem *General Algebraic Modelling System* (GAMS) e resolvido usando a ferramenta comercial de otimização CPLEX 11 da IBM ILOG. Os resultados foram obtidos com um Intel® Core™ i3 CPU M350@2.27 GHz utilizando as definições de origem do GAMS/CPLEX e estabelecendo como critério de paragem a otimalidade.

4.2.1.1. Cenário I

Este cenário considera a reconfiguração da rede hospitalar mantendo as unidades atuais, ou seja, não permitindo o encerramento de unidades. Desta forma a reorganização da rede é obtida através de ajustes na capacidade instalada ou através de novas unidades que podem abrir em localidades que respeitem o requisito de população anteriormente definido. De forma a apresentar soluções viáveis para ajustes na capacidade instalada, considera-se possível uma redução na ordem dos 20% na capacidade de hospitais em regiões urbanas (com exceção de Lisboa onde se estabelece um mínimo de 2 000 camas) e de 30% para hospitais em regiões rurais. Esta redução diferenciada possibilita que as unidades localizadas em zonas rurais disponham de maiores reduções, o que é razoável considerando que estas zonas têm mais baixa acessibilidade e níveis mais baixos de procura.

Os resultados obtidos com a aplicação do modelo estão apresentados na Figura 2 onde se pode observar: a localização das unidades atualmente existentes e das unidades a abrir; as áreas de influência de cada unidade, definidas pelas cores; a referenciação entre os níveis hierárquicos representados através das setas; a capacidade a instalar em cada serviço recomendada pelo modelo (definida em termos camas para o serviço de internamento e do número de utentes servidos nos restantes serviços); e informação adicional sobre custos e sobre o desempenho do modelo.

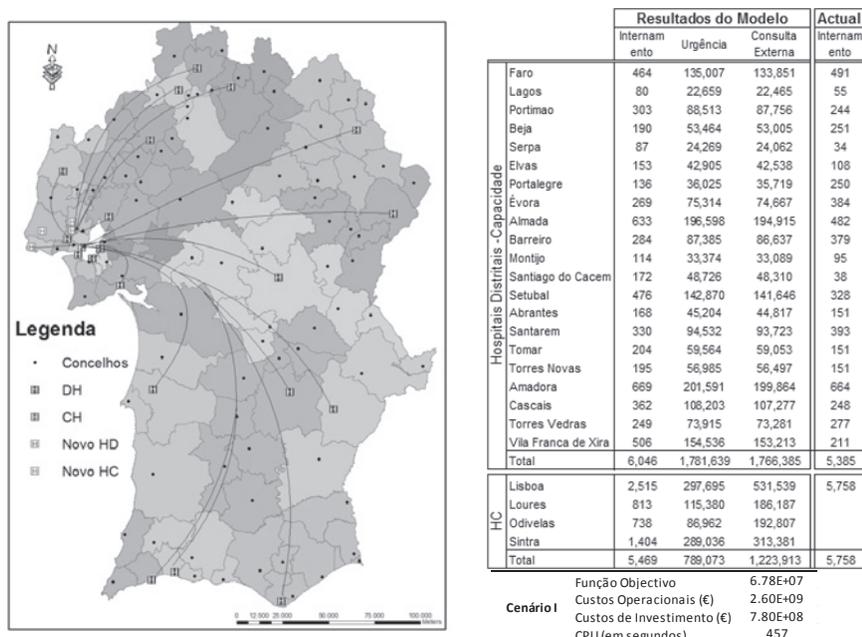


Figura 2: Rede hospitalar para o cenário I: localização, áreas de influência e capacidade.

Os resultados sugerem alterações na localização e capacidade da rede hospitalar, recomendando a dispersão da capacidade instalada na área metropolitana de Lisboa através da abertura de novas unidades (Loures, Odivelas e Sintra), sendo que duas das unidades propostas são projetos em construção (Loures) e em discussão (Sintra). O modelo propõe ainda alterações significativas na capacidade instalada para responder à procura observada, nomeadamente nos hospitais localizados nas regiões mais rurais. Os resultados apresentados na Figura 2 relativos ao número de utentes esperados no serviço

de urgência mostram que a meta de 150 atendimentos por dia no serviço de urgência (que corresponderia a cerca de 54 750 atendimentos por ano) (Ministério da Saúde, 2007) não é alcançada em todas as unidades, como sejam, por exemplo, os casos de Elvas e Serpa. Estes casos, para além de serem unidades de pequena dimensão, estão localizadas em zonas rurais com baixa densidade populacional. Nestas circunstâncias a necessidade deste serviço carece de uma análise mais detalhada para determinar a forma mais adequada de prestar um serviço de qualidade às populações. Este resultado ilustra que em regiões rurais é necessário estabelecer um compromisso entre equidade e eficiência, e que eventualmente as unidades hospitalares têm em certos casos de funcionar com dimensão/produção abaixo dos níveis de produção desejáveis para poderem prestar um serviço de proximidade. Alternativamente, uma menor acessibilidade hospitalar pode ser compensada por maior disponibilidade e acesso a cuidados de saúde primários.

A Figura 3 apresenta indicadores do tempo médio e máximo de percurso para populações que vivem em diferentes sub-regiões de saúde. Apesar do tempo médio de percurso ser inferior a quarenta e cinco minutos em todas as sub-regiões, existem valores preocupantes no que respeita ao tempo máximo de percurso, nomeadamente nas regiões rurais do Algarve e Alentejo, onde os valores máximos estão muito acima dos valores de referência que apontam para sessenta minutos (Ministério da Saúde, 2006a). Uma análise detalhada revela que estes tempos de percurso são observados para uma pequena proporção da população, sendo contudo necessário criar alternativas que podem passar eventualmente pelo reforço dos cuidados primários e/ou dos cuidados pré-hospitalares ao nível da urgência.

Dado que existe incerteza associada a alguns parâmetros, principalmente devido a limitações na informação recolhida (ver secção 4.1), foi realizada uma análise de sensibilidade para perceber quais as implicações nos resultados de variações nos parâmetros do modelo. Desta forma realizou-se uma análise aos parâmetros que se consideram mais críticos, como sejam os parâmetros que determinam as transferências e estabelecem os fluxos de utentes, a demora média no serviço de internamento, e a procura por serviços hospitalares, de acordo com:

- ✓ Análise de sensibilidade às transferências: testam-se aumentos nas taxas de transferência entre HD e HC de 2% (cenário HD-HC1) e 4% (cenário HD-HC2). No fluxo inverso HC-HD também são testados aumentos da mesma ordem de grandeza (resultado nos cenários HC-HD1 e HC-HD2).
- ✓ Análise de sensibilidade na demora média: analisa-se o impacto de diminuir a demora média no serviço de internamento em meio-dia (DM1) e um dia (DM2), dado que a redução da demora média é uma meta definida para o serviço hospitalar.
- ✓ Análise de sensibilidade à procura: testam-se variações na procura de serviços hospitalares. Considera-se uma redução de 2% (Proc 1) e um aumento de 2% (Proc 2).

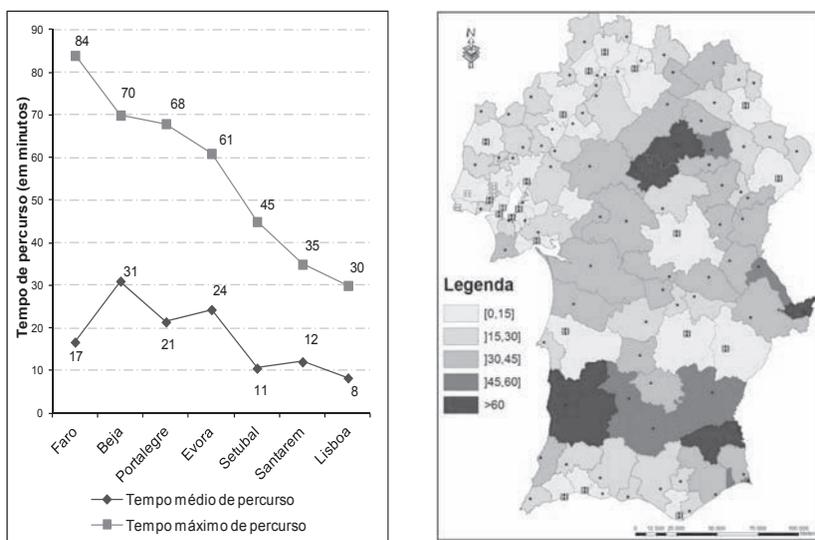


Figura 3: Tempo médio e máximo de percurso para o cenário I.

A Tabela 5 sintetiza os resultados obtidos na análise de sensibilidade e compara-os com os valores inicialmente obtidos para o cenário I (denominado na tabela por cenário base). Da análise de sensibilidade realizada pode-se concluir que os resultados são robustos no que respeita à localização. Porém, observam-se diferenças no que respeita ao dimensionamento das unidades: aumentos nas transferências entre HD e HC têm impacto direto na capacidade

dos HC e este aumento leva a acréscimos na capacidade instalada dos HD; aumentos nas transferências entre HC e HD apenas afetam as capacidades dos HD.

No que diz respeito aos resultados da demora média o impacto é muito mais significativo, podendo levar a uma diminuição significativa na capacidade total da rede. Por exemplo, uma redução da demora média em um dia justifica uma redução de 13% do total de capacidade instalada. Nesta análise de sensibilidade a função objetivo permanece inalterada, uma vez que a procura não sofre alterações. No entanto, os custos operacionais e de investimento sofrem uma redução que está associada à diminuição da capacidade instalada.

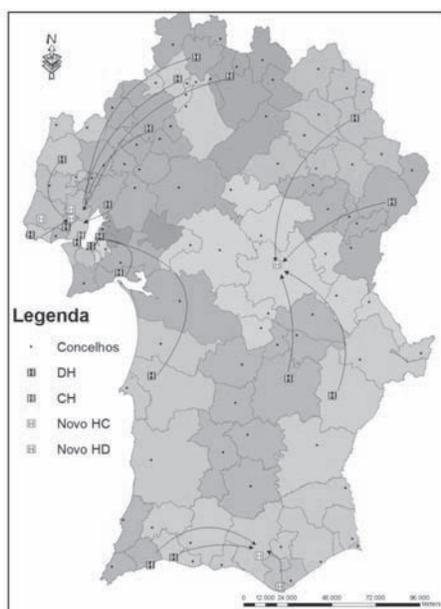
Relativamente a variações na procura observa-se que os resultados do modelo são robustos face à variação da procura testada ($\pm 2\%$). Porém, os resultados dos cenários de variação de procura (Proc 1 e Proc 2) apresentam desvios significativos na capacidade instalada quando comparados com o cenário base, o que indicia que as decisões de localização podem ser mais sensíveis a alterações na procura.

4.2.1.2. Cenário II

No cenário II são analisadas as possíveis alternativas para a reorganização da rede hospitalar atualmente em discussão. O cenário II parte dos mesmos pressupostos do cenário I, considerando igualmente informação adicional sobre a reorganização da rede. Contrariamente ao cenário I, o cenário II permite o encerramento de unidades que têm sido consideradas no âmbito da reorganização da rede (ver Tabela 3). Assim, no cenário II considera-se nomeadamente: o encerramento do HD de Évora de forma a testar a localização de uma unidade de substituição; a redução de capacidade ou o encerramento do HD de Faro de forma a testar a construção de uma nova unidade; e a redução da capacidade instalada em Lisboa de forma a testar a dispersão da capacidade hospitalar pela área metropolitana de Lisboa. Este cenário considera ainda as flutuações existentes na população da região do Algarve, com um conseqüente impacto na procura por serviços de saúde nessa região - assume-se para o efeito um aumento da procura de serviços de saúde de 10% para os habitantes desta região.

A Figura 4 apresenta os resultados do cenário II onde, em comparação com o cenário I, se observa a abertura de dois novos HC em Évora (unidade de substituição do atual HD) e Loulé que prestam, respetivamente, cuidados de nível superior para as populações do Alentejo e Algarve.⁷

No caso da Região de Saúde do Algarve o modelo sugere a manutenção da unidade hospitalar de Faro, de forma a aumentar o acesso dos utentes. No entanto, esta unidade passa a ter menos capacidade instalada e a servir apenas as populações de Faro e Olhão. Embora a solução apresentada no cenário II seja melhor (em relação á solução do cenário I) do ponto de vista da acessibilidade, ela representa uma solução muito mais dispendiosa (ver Figura 2 e Figura 4) pois requer a abertura de duas novas unidades de carácter central: uma em Évora e outra em Faro.



	Resultados do Modelo			Actual
	Internam ento	Urgência	Consulta Externa	Internam ento
Faro	219	65,529	64,969	491
Lagos	88	24,925	24,712	55
Portimão	256	73,439	72,811	244
Beja	190	53,464	53,005	251
Serpa	87	24,269	24,062	34
Élvás	153	42,905	42,538	108
Portalegre	136	36,025	35,719	250
Évora	633	196,598	194,915	384
Almada	284	87,385	86,637	482
Barreiro	114	33,374	33,089	379
Montijo	172	48,726	48,310	95
Santiago do Cacém	476	142,870	141,646	38
Setúbal	168	45,204	44,817	328
Abrantes	330	94,532	93,723	151
Santarem	204	59,564	59,053	393
Tomar	195	56,985	56,497	151
Torres Novas	669	201,591	199,864	151
Amadora	362	108,203	107,277	664
Cascais	249	73,915	73,281	248
Torres Vedras	249	73,915	73,281	277
Vila Franca de Xira	596	154,536	153,213	211
Total	5,742	1,697,954	1,683,419	5,385
Loulé	662	106,904	146,614	5,758
Évora	530	75,314	113,498	
Lisboa	1,946	297,695	413,023	
Loures	813	115,380	186,187	
Odivelas	738	86,962	192,807	
Sintra	1,404	289,036	313,381	
Total	6,092	971,291	1,365,510	5,758
Função Objectivo				5.88E+07
Cenário II Custos Operacionais (€)				2.70E+09
Custos de Investimento (€)				1.13E+09
CPU (em segundos)				335

Figura 4: Rede hospitalar para o cenário II: localização, áreas de influência e capacidade.

Tabela 5: Resultados da análise de sensibilidade às transferências e à demora média.

	Transferências HD-HC		Transferências HC-HD		Demora Média		Procura por serviços					
	Base	HD-HC 1	HD-HC 2	Base	HC-HD 1	HC-HD 2	Base	DM1	DM2	Proc 1	Base	Proc 2
Valor do Parâmetro	13,0%	15,0%	17,0%	1,2%	3,2%	5,2%	7,00	6,50	6,00	-2%	----	+ 2%
Função Objectivo	6,78E+07	6,81E+07	6,83E+07	6,78E+07	6,79E+07	6,79E+07	8,90	8,40	7,90	6,85E+07	6,78E+07	6,92E+07
Custos Operacionais (€)	2,60E+09	2,63E+09	2,66E+09	2,60E+09	2,61E+09	2,61E+09	2,60E+09	2,48E+09	2,36E+09	2,55E+09	2,60E+09	2,66E+09
Custos Investimento (€)	7,80E+08	7,96E+08	8,12E+08	7,80E+08	7,81E+08	7,82E+08	7,80E+08	7,17E+08	6,56E+08	7,58E+08	7,80E+08	8,01E+08
CPU (em segundos)	457	407	462	457	478	430	457	412	389	447	457	422
Faro	464	465	465	464	466	467	464	431	398	455	464	474
Lagos	80	80	80	80	81	81	80	75	69	79	80	82
Portimao	303	303	303	303	304	305	303	282	260	297	303	309
Beja	190	190	190	190	191	191	190	177	163	186	190	194
Serpa	87	87	87	87	88	88	87	81	75	86	87	89
Elvas	153	153	153	153	153	154	153	142	131	150	153	156
Portalegre	136	136	136	136	137	137	136	127	117	134	136	139
Évora	269	269	269	269	269	270	269	250	230	263	269	274
Almada	633	633	634	633	635	637	633	588	543	621	633	646
Barreiro	284	284	284	284	285	286	284	264	244	279	284	290
Montijo	114	114	114	114	114	114	114	106	98	112	114	116
Santiago do Cacém	172	172	172	172	173	173	172	160	147	169	172	176
Setúbal	476	476	476	476	477	478	476	442	408	467	476	486
Abraantes	168	168	168	168	168	169	168	156	144	164	168	171
Santarem	330	330	330	330	331	332	330	306	283	323	330	337
Tomar	204	204	204	204	205	205	204	190	175	200	204	208
Torres Novas	195	196	196	195	196	196	195	182	168	192	195	199
Amadora	669	669	669	669	671	672	669	621	573	656	669	682
Cascais	362	362	363	362	363	364	362	336	311	355	362	370
Torres Vedras	249	249	249	249	250	250	249	231	213	244	249	254
Vila Franca de Xira	506	506	506	506	507	508	506	470	433	496	506	516
Lisboa	2.515	2.597	2.680	2.515	2.515	2.515	2.515	2.374	2.232	2.465	2.515	2.565
Loures	813	838	864	813	813	813	813	767	721	796	813	829
Odivelas	738	757	776	738	615	738	848	684	644	711	738	740
Sintra	1.404	1.430	1.456	1.404	1.527	1.404	1.293	1.337	1.257	1.388	1.404	1.445

5. Discussão e trabalho futuro

Este trabalho propõe uma ferramenta para apoiar o planeamento de redes hospitalares num sistema que tem por base um Serviço Nacional de Saúde. O trabalho foi motivado pela necessidade de reorganização da rede hospitalar na região sul de Portugal. Para o efeito desenvolveu-se um modelo de localização hierárquico e multi-serviço cujo objetivo é a maximização do acesso das populações. Os cenários desenvolvidos introduzem diferentes pressupostos para a reorganização da rede hospitalar e permitem exemplificar a natureza dos resultados que podem ser recolhidos, como a localização e dimensionamento das unidades hospitalares, a definição das áreas de influência e indicadores de nível de serviço. Desta forma, o modelo provou ser uma ferramenta de apoio à decisão útil e flexível para modelar diversos cenários de interesse para o decisor.

O modelo considera a estrutura hierárquica das unidades hospitalares e modela em detalhe os fluxos de utentes entre unidades hospitalares a diferentes níveis e entre serviços hospitalares. Considera ainda outras questões relevantes para o planeamento de redes hospitalares, como sejam limites na capacidade instalada, critérios de afetação das populações que permitem definir áreas de influência e a existência de uma superestrutura hospitalar.

Este modelo foi também aplicado no âmbito de um projeto de reestruturação da rede hospitalar na Administração Regional da Saúde do Norte (ARSN). Do ponto de vista da decisão, a característica que se afigurou mais importante, tal como no caso acima descrito, foi a natureza integrada dos resultados que através de uma análise quantitativa abrangente disponibiliza informação sobre localização, dimensionamento, referenciação e áreas de influência. Por outro lado, a flexibilidade do modelo foi importante para testar diferentes cenários e realizar análise de sensibilidade a parâmetros incertos. Sendo uma rede hospitalar na ARSN constituída por várias unidades, as decisões mais relevantes neste caso de estudo prenderam-se com o dimensionamento das unidades existentes e conseqüente redefinição das áreas de influência das unidades, e a possibilidade de construir unidades de substituição. Sendo uma rede mais compacta que a analisada no caso de estudo da zona sul, foi necessário introduzir pequenas alterações no modelo. A

mais relevante foi a introdução de fluxos diretos para unidades de nível superior devido à existência de referenciação direta dos cuidados primários para estas unidades.

Apesar da generalidade e flexibilidade do modelo, tal como foi demonstrada neste caso de estudo, existem ainda aspetos que devem ser alvo de estudo. Nomeadamente, o modelo aqui proposto representa uma primeira abordagem onde o objetivo principal é a maximização da acessibilidade enquanto os custos são estimados após otimização. A integração de questões de acesso e custos numa análise bi-objetivo onde soluções de compromisso entre estes objetivos tipicamente conflituosos são definidas é um dos aspetos já em estudo pelas autoras. Os primeiros resultados desta abordagem podem ser encontrados em Mestre et al. (2011). A questão relativa ao planeamento futuro também deverá ser analisada através da introdução de um horizonte temporal que permita delinear a rede hospitalar tendo em conta a evolução do sistema. É de notar que embora as decisões mais importantes sejam aquelas que têm impacto imediato, é necessário acautelar que estas não limitem escolhas futuras e é necessário considerar projeções sobre possíveis estados do sistema (p.e. em parâmetros da procura e oferta ao longo do tempo).

Será igualmente importante analisar a robustez das localizações face a diferentes cenários, através de um modelo de programação estocástica, onde os potenciais cenários de incerteza possam ser explicitamente incluídos, e o modelo forneça soluções de compromisso. Os primeiros resultados deste trabalho podem ser encontrados em Mestre et al. (2010).

Bibliografia

- BARROS, P. P., S. R. MACHADO, e Outros., "Portugal: Health systems review", *Health Systems in Transition*, n.ºs 13(4), 2011, pp. 1-156.
- DASKIN, M., *Network and discrete location*, New York, Wiley Interscience, 1995.
- DASKIN, M., "What you should know about location modelling", *Naval Research Logistics*, n.ºs 55, 2008, pp. 283-294.
- DASKIN, M. S. and L. K. DEAN, *Location of health care facilities*, *Handbook of OR/MS in health care: a handbook of methods and applications*. F. Sainford, M. Brandeau and W. P. (eds). Boston, Kluwer, 2004, pp. 43-76.
- DÖKMECI, V. F., "An optimization model for a hierarchical spatial system", *Journal of Regional Science*, n.ºs 13(3), 1973, pp. 439.

GALVÃO, R. D., L. G. A. ESPEJO, e Outros., "A hierarchical model for the location of perinatal facilities in the municipality of Rio de Janeiro", *European Journal of Operational Research*, n°s 138(3), 2002, pp. 495-517.

GERRARD, R. and R. CHURCH, "Closest assignment constraints and location models: properties and structure", *Location Science*, n°s 4(4), 1996, pp. 251-270.

INE. "XIV census - Instituto Nacional de Estatística" Retrieved Info Service: www.ine.pt, 2006.

MARIANOV, V. and D. SERRA, *Location problems in the public sector Facility Location: applications and theory*. Z. Drezner and H. Hamacher. Berlin, Springer, 2004, pp. 119-143.

MARSH, M. T. and D. A. SCHILLING, "Equity measurement in facility location analysis: A review and framework", *European Journal of Operational Research*, n°s 74(1), 1994, pp. 1-17.

MAYHEW, L. D. and G. LEONARDI, "Equity, efficiency, and accessibility in urban and regional health-care systems", *Environment and Planning A*, n°s 14(11), 1982, pp. 1479-1507.

MEHEREZ, A., Z. SINUANY-STERN, e Outros., "On the implementation of quantitative facility location models: the case of hospital in a rural region", *Journal of the Operational Research Society*, n°s 47(5), 1996, pp. 612-625.

MELO, M. T., S. NICKEL, e Outros., "Facility location and supply chain management - A review", *European Journal of Operational Research*, n°s 196(2), 2009, pp. 401-412.

MESTRE, A., M. OLIVEIRA, e Outros. *Hospital network planning under uncertain demand*. Proceedings of the Uncertainty and Robustness in Planning and Decision Making conference, Coimbra - Portugal, 2010.

MESTRE, A., M. OLIVEIRA, e Outros. *Planning hospital networks: A dynamic and multiobjective approach*. Proceedings of the 37th Operational Research Applied to Health Services (ORAHs) conference: Operational Research Informing National Health Policy, Cardiff, United Kingdom, School of Mathematics, Cardiff University, 2011.

MESTRE, A., M. OLIVEIRA, e Outros., "Organizing hospitals into networks: a hierarchical and multiservice model to define location, supply and referrals in planned hospital systems", *OR Spectrum*, n°s 34(2), 2012, pp. 319-348.

MINISTÉRIO DA SAÚDE "Centros de saúde e hospitais: recursos e produção do SNS". Lisboa, (2003).

MINISTÉRIO DA SAÚDE "Reordenamento das capacidades hospitalares da cidade de Lisboa". Lisboa, (2006a).

MINISTÉRIO DA SAÚDE "Contabilidade Analítica - Hospitais do SNS". Lisboa, (2006b).

MINISTÉRIO DA SAÚDE "Proposta da rede de urgências - Relatório Final". Lisboa, (2007).

MOORE, G. and C. S. REVELLE, "The hierarchical service location problem", *Management Science*, n°s 28(7), 1982, pp. 775-780.

NARULA, S. C., "Minisum hierarchical location-allocation problems on a network: a survey", *Annals of Operations Research*, n°s 6, 1986, pp. 257-272.

OWEN, S. H. and M. S. DASKIN, "Strategic facility location: A review", *European Journal of Operational Research*, n°s 111(3), 1998, pp. 423-447.

RAHMAN, S. and D. K. SMITH, "Use of location-allocation models in health service development planning in developing nations", *European Journal of Operational Research*, n°s 123(3), 2000, pp. 437-452.

REGIÃO SAÚDE DO ALGARVE "Plano Director Regional da Saúde do Algarve." Faro, (2004).

REVELLE, C. S. and H. A. EISELT, "Location analysis: A synthesis and survey", *European Journal of Operational Research*, n°s 165(1), 2005, pp. 1-19.

RUTH, R. J., "A mixed integer programming model for regional planning of a hospital inpatient service", *Management Science*, n°s 27(5), 1981, pp. 521-533.

SAHIN, G. and H. SÜRAL, "A review of hierarchical facility location models", *Computers & Operations Research*, n°s 34(8), 2007, pp. 2310-2331.

SAMUELSON, P. A. and W. D. NORDHAUS, *Economics*, New York, McGraw-Hill 2010.

SMITH, H., P. HARPER, e Outros. "Bicriteria efficiency/equity hierarchical location models". Working Paper - 152075, University of Southampton, (2010).

SMITH, H. K., P. R. HARPER, e Outros., "Planning sustainable community health schemes in rural areas of developing countries", *European Journal of Operational Research*, n°s 193(3), 2009, pp. 768-777.

TOREGAS, C., R. SWAIN, e Outros., "The location of emergency service facilities", *Operations Research*, n°s 19(6), 1971, pp. 1363-1373.

CASO 20

BENCHMARKING DOS SERVIÇOS DOS HOSPITAIS PORTUGUESES: UMA APLICAÇÃO DE DATA ENVELOPMENT ANALYSIS

Ricardo A. S. Castro

Faculdade de Engenharia da Universidade do Porto
ricardo.alves.castro@fe.up.pt

Conceição Silva Portela

Universidade Católica Portuguesa no Porto
csilva@porto.ucp.pt

Ana S. Camanho

Faculdade de Engenharia da Universidade do Porto
acamanho@fe.up.pt

RESUMO

Neste artigo, apresentamos a aplicação de um modelo de Data Envelopment Analysis aos serviços de Medicina Interna da maioria dos hospitais do Serviço Nacional de Saúde Português, numa perspetiva de eficiência. O modelo de avaliação é apresentado para a generalidade dos serviços hospitalares, tendo os inputs e outputs sido escolhidos após análise dos seus impactos nos gastos globais dos diferentes serviços. Nesta aplicação impõem-se restrições aos pesos, por forma a evitar obter serviços com pesos pouco razoáveis para as variáveis do modelo. Observa-se que as maiores poupanças são possíveis nos medicamentos e material clínico. É feita uma comparação entre serviços eficientes e ineficientes, observando-se que os meios complementares de diagnóstico e terapêutica (recursos) e as variáveis de acesso aos cuidados (elementos da produção) são determinantes na definição do melhor desempenho dos serviços de Medicina Interna analisados.

PALAVRAS-CHAVE

Data Envelopment Analysis, Avaliação de hospitais, Benchmarking, Análise de eficiência

1. Introdução: Avaliação de Hospitais – Motivação e Desafios

Este artigo apresenta brevemente o contexto hospitalar nacional e os problemas inerentes à avaliação de unidades tão complexas quanto os hospitais. Pretende-se com este estudo apresentar uma abordagem à avaliação de serviços hospitalares usando *Data Envelopment Analysis* (DEA), com base em dados cedidos pela Administração Central do Sistema de Saúde (ACSS). A maioria dos dados, relativos a 2008, pode ser consultada na Base de Dados dos Elementos Analíticos (contabilidade de gestão) dos hospitais do Serviço Nacional de Saúde (SNS), disponível no sítio da ACSS. Os restantes dados referem-se aos recursos e à produção dos hospitais, também cedidos pela ACSS.

1.1 Os Gastos em Saúde na OCDE, UE e Portugal e a Agenda Política para o Setor

Em todos os países da Organização para a Cooperação e Desenvolvimento Económico (OCDE) tem-se assistido a um aumento constante das despesas em saúde. Existem diversas razões que levam a este aumento, destacando-se, do lado da procura, o envelhecimento da população, o alargamento do acesso e da cobertura dos diferentes seguros/sistemas nacionais de saúde e o crescimento do rendimento das populações. Enquanto as duas primeiras razões parecem óbvias, a terceira tem sido bastante estudada, existindo evidência que o que a sustenta é a correlação existente entre o aumento do rendimento e a maior exigência e consumo de serviços de saúde (Simões, 2004). Do lado da oferta, pode destacar-se o aumento da complexidade tecnológica das ferramentas de diagnóstico e tratamento disponíveis, assim como o aumento da formação do pessoal hospitalar. Em Portugal, desde a criação do SNS, em 1979, o aumento dos gastos com a saúde tem sido superior à média da OCDE e da União Europeia a 15 países (UE-15). Em 2004, estes gastos atingiram o valor de 10% do Produto Interno Bruto (PIB) (Oliveira et al., 2007), dos quais 7% eram gastos públicos. Dados do Instituto Nacional de Estatística (INE) revelam que, de 2005 a 2008, o valor total da despesa em Portugal se situou, em média, em torno dos 9,5% do PIB (Pordata). Desta despesa, neste último período, a percentagem de despesa pública correspondeu a um valor compreendido entre

os 65% e os 70% (Pordata, 2011), sendo a restante despesa proveniente das famílias, de seguros sociais, entre outras fontes privadas. Em 2004, os valores referidos fizeram Portugal aproximar o nível de despesa pública em saúde com o da OCDE e EU-15 (Oliveira et al., 2007), sem, no entanto, a ultrapassar. Portugal e a grande maioria dos países da OCDE apresentam mecanismos de proteção dos seus cidadãos na doença e, por isso, apresentam uma preponderância dos gastos públicos em saúde, face aos gastos privados. No nosso país, a maior parte da despesa pública incorrida pelo Serviço Nacional de Saúde (SNS) deve-se aos hospitais (mais de 50%).

Os aumentos de custos com saúde que se têm verificado no mundo ocidental, em especial no espaço europeu, têm exercido uma grande pressão sobre os défices nacionais, o que tem levado os diversos Estados a agir no sentido da sua contenção, tendo em vista a sustentabilidade dos sistemas de saúde no longo prazo. Muito em voga até ao fim da década de 80, a doutrina "mais mercado" na Saúde foi abandonada, mesmo pelo Banco Mundial, tendo o enfoque passado a ser a eficiência da intervenção estatal (Simões, 2004). No final da década de 90 e no princípio deste século, foram tomadas diversas medidas transversais a toda a União Europeia, que tinham como objetivo conter o aumento dos custos com a saúde, através de um aumento da eficiência (Simões, 2004). Em Portugal, a agenda política vigente, que tem conduzido as reformas efetuadas na última década, pode resumir-se numa delegação da responsabilidade de execução para níveis mais próximos do utilizador, numa responsabilização dos prestadores pelo desempenho e mesmo numa flexibilização da gestão dentro do setor público (Simões, 2004). Na prática, a execução desta agenda política tem levado à tentativa de mudança do paradigma hospitalocêntrico, sendo de destacar: a criação de unidades de internamento de longo prazo, inseridas na Rede de Cuidados Continuados (rede que promove a substituição do hospital como centro de tratamento de doentes crónicos), o desenvolvimento de serviços de cuidados paliativos (cuidados prestados no sentido de tratar e cuidar de doentes na fase final da vida), a reorganização do mapa hospitalar público e a transformação de hospitais do setor público administrativo em hospitais entidade pública empresarial, com independência de gestão e financiados pelo Ministério da

Saúde. Como conclui o relatório de Barros e Simões (2007) para o Observatório Europeu de Sistemas e Políticas de Saúde, o paradigma hospitalocêntrico vigente não se tinha alterado até à data (e não se alterou até ao ano de 2012), continuando-se as reformas para o alterar.

1.2 A Necessidade de Avaliação Hospitalar e Seus Problemas

A necessidade de controlar mais de perto a despesa pública tem levado à gradual aceitação da necessidade de se introduzirem mecanismos de avaliação mais profundos nos processos de tomada de decisão nas administrações públicas, nomeadamente ao nível da saúde. O princípio geral de prestar contas tem ganho espaço, devido ao aumento progressivo da informação, conhecimento e sentido crítico dos cidadãos. Esta avaliação dos serviços da Administração Pública é muitas vezes tida como uma exigência decorrente da responsabilização pela utilização de fundos estatais destinados a responder a necessidades públicas (Simões, 2004).

As políticas nacionais para o setor da saúde, nomeadamente a que conduziu à introdução de hospitais empresa e à separação do papel do Estado como prestador e financiador (Lei 27/2002 de 8 de Novembro), supõem, pela sua natureza, a criação de mecanismos de avaliação por parte das próprias instituições hospitalares e quase forçam uma avaliação aos hospitais, para se aferir o impacto da sua implementação. No entanto, a extensão de estudos de avaliação do desempenho hospitalar, em Portugal, é ainda reduzida. Existem, no entanto, exemplos de programas de certificação e acreditação promovidos principalmente por entidades anglófonas e mesmo alguns programas de *benchmarking*. É, contudo, de esperar uma evolução significativa neste campo, já que a cada vez maior penetração de *software* de gestão hospitalar poderá permitir a organização de cada vez maiores quantidades de informação com acesso mais fácil.

Apesar de necessária, a avaliação do desempenho de um hospital não é uma atividade trivial (Costa e Lopes, 2007). Das diferentes dificuldades, podem destacar-se a inexistência de objetivos claros e comuns para a atividade de todas as unidades, a existência de objetivos não mensuráveis (Simões, 2004) ou mesmo a existência de diferentes agentes (Costa e Lopes, 2007) no processo da

prestação de cuidados de saúde. Neste processo, o médico terá tendência a maximizar o estado de saúde dos indivíduos (maximizando a eficácia da atividade hospitalar), enquanto o gestor do hospital se preocupará mais com a minimização dos recursos usados para a prestação dos cuidados (maximização da eficiência da atividade hospitalar, que pode ser conflituante com a eficácia). No entanto, a única maneira de se praticar gestão e tomar decisões baseadas em factos é medindo esta atividade hospitalar, realizando um processo avaliativo que permita priorizar as oportunidades de melhoria e avaliar o desempenho do que se tem realizado. Para que a avaliação reflita da melhor forma a complexidade inerente ao setor da saúde, os indicadores escolhidos para medir a atividade de um prestador de cuidados deverão ter em conta os aspetos relacionados com a qualidade dos tratamentos, isto é, com o seu impacto na saúde das pessoas que os recebem (Marshall et al., 2004).

Num processo avaliativo, é conveniente não esquecer que a avaliação de um sistema tão complexo como o hospitalar pode trazer consigo algumas consequências indesejadas, fruto da usual utilização de um reduzido número de indicadores de desempenho (usualmente necessária para permitir a exequibilidade e interpretação da avaliação realizada, mas que pode não ser representativa da realidade medida). Entre as consequências conhecidas, podem destacar-se o aparecimento da visão de túnel, isto é, o grande enfoque no que é medido em detrimento de outras vertentes do desempenho hospitalar, a sub-otimização, que aparece quando alguns agentes se esforçam para a prossecução de objetivos setoriais, por vezes desalinhados dos organizacionais, ou a miopia, inerente à concentração em objetivos de curto-prazo, que são mais facilmente medíveis (Simões, 2004). Estes problemas podem e devem ser combatidos, nomeadamente recorrendo-se a processos que permitam envolver o pessoal avaliado no desenvolvimento e implementação dos indicadores do sistema de avaliação, flexibilizando a sua utilização e quantificando, sempre que possível, os objetivos a atingir.

1.3 A Motivação para um Benchmarking aos Hospitais Portugueses

Existem diversos modelos de avaliação para os hospitais. Dependendo do país e da existência ou não de um Serviço Nacional de Saúde, a forma como se

avalia e mesmo o tipo de indicadores utilizados pode variar. Os indicadores são usualmente agrupados em três dimensões da avaliação da qualidade da atividade hospitalar: estrutura, processos e resultados (Donabedian, 1980). No entanto, algo é comum às avaliações feitas pelos Estados (e pelas empresas de acreditação e *benchmarking* que lhes prestam serviços), no que toca à avaliação de resultados. Nestas avaliações não é usual utilizarem-se metodologias de fronteira para avaliar o desempenho dos hospitais, preferindo-se uma avaliação baseada numa longa lista de indicadores, sem se conseguir obter uma imagem única e direta da realidade hospitalar. Isto fica bem patente ao analisar-se os relatórios de retorno enviados pela ACSS aos hospitais do SNS, na longa lista pública de indicadores usados pelo *National Health System* britânico e pela proliferação de empresas que prestam serviços de acreditação e *benchmarking*: a IASIST, em Portugal, Espanha e Reino Unido, a *Caspe Health Knowledge System* (CHKS) e a *Dr. Foster*, no Reino Unido, ou a *Thomson Reuters*, que publica o *100 Top Hospitals*, nos Estados Unidos da América, entre tantas outras. As metodologias de fronteira são, no entanto, largamente utilizadas pela comunidade académica (Hollingsworth, 2008).

Tendo em conta que não existe uma referência absoluta de desempenho hospitalar, considera-se oportuna e valiosa qualquer ferramenta que permita avaliar os hospitais, comparando-os entre si. Um modelo de DEA pode ser um elemento fundamental na identificação das melhores práticas e na deteção das maiores fontes de ineficiência, agregando múltiplas variáveis que caracterizem a produção hospitalar, para que se obtenha uma imagem global e agregada de um hospital. Neste estudo apresenta-se um modelo para a avaliação da eficiência hospitalar ao nível dos serviços, com esta finalidade. Para uma avaliação completa, falta realizar a avaliação da eficácia, elemento fundamental numa avaliação hospitalar. Esta avaliação será alvo de trabalhos futuros.

Por fim, é de realçar a importância política que o *benchmarking* hospitalar começa a ganhar em Portugal, oficializada no memorando acordado entre o Estado português e a Troica, que negociou a sua ajuda externa. Neste memorando encontra-se um pacote de medidas que o Estado se comprometeu a executar, nomeadamente o estabelecimento de um sistema de *benchmarking* hospitalar (Banco Central Europeu et al., 2011, ponto 3.75).

2. Data Envelopment Analysis – Uma Metodologia para Avaliação de Eficiência e Realização de Benchmarking aos Hospitais

2.1 Definição de Benchmarking

Para avaliar o desempenho dos hospitais portugueses é necessário definir um padrão. Como organizações humanas complexas, os hospitais têm, usualmente, o seu padrão definido com base na análise comparativa (ou *benchmarking*) das unidades a avaliar. Para Camp (1989, 12), *benchmarking* significa “a procura das melhores práticas de uma indústria que levam a um desempenho superior”. Quando aplicado ao setor público, esta técnica tem em vista a melhoria da qualidade dos serviços prestados, assim como a sua eficiência. Desde a década de 80 que esta prática tem vindo a crescer, em especial no setor da saúde, impulsionada pelas reformas realizadas no setor público, nomeadamente no Reino Unido e nos Estados Unidos da América ou em países escandinavos (Triantafillou, 2007).

Apesar de alguma controvérsia e dificuldades inerentes ao processo de *benchmarking* (Triantafillou, 2007), são bastantes as virtudes que lhe são atribuídas. Em primeiro lugar, realizar *benchmarking* implica criar referências quantitativas comuns, que não se baseiam em opiniões ou valores, mas sim em medidas concretas de produtividade. A partir destas referências, a experiência sugere que um mecanismo quase automático é iniciado, já que a comparação insta, de facto, as organizações a agir (Triantafillou, 2007). Ao compararem-se com os seus pares, as organizações assumem o papel de sujeitos que buscam o seu próprio aperfeiçoamento e tentam implementar mudanças organizacionais, de procedimentos ou de gestão que as levem à superação de si mesmas. Outros autores referem que o *benchmarking* permite a melhoria do desempenho de uma organização, através da identificação direta de práticas e de objetivos a serem atingidos e de uma maior perceção das suas forças e fraquezas (Voss et al., 1997).

2.2 Definição de Eficiência

O desempenho de uma unidade está intimamente ligado à sua produtividade, que é definida como o rácio entre a produção por si obtida (o

output) e os recursos para tal utilizados (o *input*). Numa indústria, ou conjunto de unidades, a produtividade raramente é igual para todos os seus elementos. A variação resulta, usualmente, das diferenças na tecnologia de produção, na escala das operações, na eficiência operacional e no ambiente operacional no qual a produção ocorre (Fried et al., 2008). Enquanto o último fator é exógeno ao gestor de uma unidade de produção, os restantes encontram-se diretamente ligados à sua atuação. Ainda assim, a que será mais facilmente manipulável pelo gestor será mesmo a eficiência operacional, medida como o rácio entre o *output* obtido e o máximo teoricamente possível, para a tecnologia utilizada, dado um certo nível de *inputs* (perspetiva dos *outputs*), ou entre o *input* mínimo e o utilizado, para obter um dado nível de *outputs* (perspetiva dos *input*). No entanto, esta medição de eficiência esbarra num grande problema: como aferir o verdadeiro potencial produtivo de uma unidade. Não se tratando de uma reação química nem de uma máquina, mas sim de uma organização humana, inerentemente complexa e com uma imprevisibilidade associada, a única maneira de aferir este potencial reside no que é conhecido ou é observável na indústria, naquilo que pode ser denominado de fronteira das melhores práticas. É nesta definição de um padrão real que entra o conceito de *benchmarking* numa avaliação de eficiência, tida como uma medida de distância às melhores práticas observadas, ou como uma produtividade relativa. Por isto mesmo, é fundamental avaliar-se a eficiência de uma organização para se identificar o que é possível melhorar, com que extensão pode isto ser feito, e para se controlar a própria produção, tendo em vista melhores resultados.

Koopmans (1951) apresentou o conceito de unidade produtiva tecnicamente eficiente: uma unidade é tecnicamente eficiente se um aumento num qualquer *output* implicar a redução de um outro *output* ou o aumento de pelo menos um dos *inputs* (ou vice-versa) (Fried et al., 2008). Isto é, uma unidade é tecnicamente eficiente se se situar sobre a fronteira das melhores práticas da indústria. Para unidades ineficientes, a medida de eficiência é uma medida da distância do produtor à fronteira do conjunto de possibilidades produtivas (distância essa calculada, de acordo com Farrell (1957), seguindo a linha que liga a origem à fronteira, passando pela unidade ineficiente). É a medição da eficiência técnica, possível através das técnicas descritas de

seguida, que se tomará como aferidora do desempenho de uma unidade hospitalar.

2.3 Técnicas de Medição de Eficiência - DEA vs. SFA

Como resulta da definição da secção anterior, medir a eficiência envolve sempre a comparação do desempenho de uma unidade com o de uma outra, localizada na fronteira de produção da amostra considerada. Historicamente, estas técnicas classificam-se em paramétricas e não-paramétricas. As primeiras baseiam-se na criação de uma função de produção uni-produto (ou *output*) ou uni-recurso (ou *input*) e identificam duas possíveis causas de desvio da fronteira: a casualidade (ou ruído estatístico) e a ineficiência técnica. Para incorporarem estas duas fontes de desvio, os modelos utilizam dois termos estocásticos. Estes modelos pertencem à família denominada de *Stochastic Frontier Analysis* (SFA), que permite distinguir a magnitude dos dois efeitos, com a contrapartida da assunção de distribuições específicas para os erros e assumindo que estes mesmos erros são independentes entre si. Os métodos não-paramétricos, que incluem a técnica de *Data Envelopment Analysis* (DEA), baseiam-se em modelos de programação matemática que têm por objetivo modelar as fronteiras de produção como segmentos lineares, sem incorporarem formas de avaliar estatisticamente os erros inerentes a uma medição da realidade (Fried et al., 2008). Estes métodos permitem o estudo de unidades com múltiplos *inputs* e *outputs*.

Ambas as técnicas apresentam uma semelhança: baseiam-se na estimação da fronteira de produção e num exercício de *benchmarking* rigoroso, já que exploram funções de distância entre as diferentes unidades e as suas projeções sobre a fronteira eficiente. No entanto, as diferenças são várias. A formulação original do DEA não incorpora o erro estatístico impossibilitando a distinção do que é, de facto, falta de eficiência do que é casualidade ou ruído estatístico (possibilidade existente no SFA), sendo, por isso, esta técnica muito sensível a observações extremas (a sua maior fragilidade). Como é uma abordagem não-paramétrica, não é necessário fazer grandes assunções *a priori*, sendo de destacar a ausência de imposições relativas à forma da fronteira de produção e à distribuição dos erros. Desta forma, com o DEA não se corre o risco de

confundir a falta de eficiência com uma simples má especificação da forma funcional do modelo, ao contrário da utilização do SFA. Enquanto o SFA baseia a medição de eficiência numa comparação com uma fronteira teórica estimada, o DEA mede a eficiência relativamente a uma fronteira que envolve todas as unidades produtivas que foram observadas na prática. Para que isto aconteça, no DEA realiza-se uma otimização por cada unidade em estudo, enquanto no SFA se realizam apenas dois passos, estimando-se a fronteira no primeiro e calculando-se a distância da unidade a esta fronteira no segundo. Isto faz com que seja mais fácil analisar individualmente cada unidade utilizando o DEA, enquanto é mais fácil obter a informação relativa a toda a amostra usando o SFA, o que torna esta ferramenta mais útil para analisar políticas de saúde do que dar informação de gestão aos gestores hospitalares ou aos seus financiadores, como o faz o DEA (O'Neill et al., 2008).

Neste trabalho é escolhida a utilização da metodologia DEA, por ser a que melhor se adapta à multiplicidade de recursos e produtos existentes na atividade hospitalar, e por permitir uma exploração da fronteira de eficiência sem se fazer uma assunção da sua forma funcional.

2.4 Data Envelopment Analysis

O *Data Envelopment Analysis* (DEA) é uma metodologia que permite não só tratar a complexidade inerente à comparação entre unidades que apresentem multiplicidade de *inputs* e *outputs*, como permite agregar numa única medida os resultados desta comparação complexa. O resultado do DEA é uma medida de eficiência para cada unidade de decisão (em inglês *Decision Making Unit* ou DMU), calculada com base na comparação dessa mesma unidade com uma ou várias unidades de referência, com o melhor desempenho observado.

Como esta metodologia é baseada num cálculo de produtividade assente em múltiplos *inputs* e *outputs*, existe a necessidade de atribuir pesos distintos a cada um desses itens, de forma a evidenciar a importância relativa de cada um deles. Para eliminar arbitrariedades, o método consiste no cálculo da produtividade relativa (ou eficiência) de cada DMU, com os pesos que mais a beneficiam. Na prática, isto é realizado maximizando-se o rácio (1), sujeito às

restrições (2) e (3). Este programa fracional é o modelo que está na base desta metodologia e é denominado modelo CCR, após a apresentação por Charnes, Cooper e Rhodes (1978) do estudo considerado a origem do DEA. Neste programa, aplicado a cada uma das unidades em estudo, considera-se existir n unidades em avaliação, m *inputs* e s *outputs*, sendo o a unidade em estudo (Cooper et al., 2007, 23).

$$\max_{v,u} \theta = \frac{\sum_r u_r y_{ro}}{\sum_i v_i x_{io}} \quad (1)$$

$$\text{Sujeito a: } \frac{\sum_r u_r y_{rj}}{\sum_i v_i x_{ij}} \leq 1 \quad (j = 1, \dots, n) \quad (2)$$

$$v_i, u_r \geq 0 \quad (\forall i, r) \quad (3)$$

É a imposição das restrições em (2) que torna o resultado uma medida da eficiência de cada unidade, já que, ao limitar o resultado que cada uma pode obter, está a criar-se uma referência relativamente à qual todas as unidades se irão comparar. O lado esquerdo da inequação (2) tomará o valor de 1 se a unidade em causa for uma unidade de referência. As unidades de referência encontradas durante a otimização são aquelas cujos pontos representativos formam a fronteira da produção. Quando comparada com a fronteira de produção, cada DMU escolhe os pesos a atribuir a cada *input* e *output*, de forma a maximizar o seu resultado de eficiência. Assim sendo, quando uma unidade apresenta um resultado de eficiência inferior à unidade, isto significa que ela é, de facto, ineficiente. Por fim, há que referir que as unidades em estudo devem ser homogéneas, isto é, ter características e comportamentos semelhantes. A introdução no modelo de elementos não controláveis pela gestão (elementos não-discricionários), permite homogeneizar as unidades sujeitas a elementos exógenos não controláveis.

Na prática, a metodologia de DEA utiliza um modelo de programação linear para calcular as eficiências de cada uma das unidades em estudo. A linearização do modelo anterior, numa perspetiva orientada aos *inputs*, leva à alteração da função objetivo (1) para a (4), dando origem à restrição (5). A linearização transforma a restrição (2) na (6), mantendo-se a restrição de não-

negatividade dos pesos. Esta formulação do DEA é designada “modelo dos pesos”.

$$\max_{v,u} \theta = \sum_r u_r y_{ro} \quad (4)$$

Sujeito a: $\sum_i v_i x_{io} = 1 \quad (5)$

$$\sum_r u_r y_{rj} - \sum_i v_i x_{ij} \leq 0 \quad (j = 1, \dots, n) \quad (6)$$

$$v_i, u_r \geq 0 \quad (\forall i, r) \quad (7)$$

A nova função objetivo consiste na maximização do numerador do rácio de produtividade, continuando as variáveis de decisão a ser os pesos dos *inputs* e *outputs*. Como referido atrás, este é o modelo CCR e as suas principais características prendem-se com a assunção de rendimentos constantes à escala, isto é, a existência de uma variação proporcional dos *outputs* face a uma variação dos *inputs*. Se na situação em estudo os aumentos de *outputs* não forem proporcionais aos aumentos de *inputs* (rendimentos variáveis à escala), deve corrigir-se o modelo anterior, alterando-se a função objetivo e a restrição (6):

$$\max_{v,u} \theta = \sum_r u_r y_{ro} - w \quad (8)$$

Sujeito a: $\sum_i v_i x_{io} = 1 \quad (9)$

$$\sum_r u_r y_{rj} - \sum_i v_i x_{ij} - w \leq 0 \quad (j = 1, \dots, n) \quad (10)$$

$$v_i, u_r \geq 0 \quad (\forall i, r) \quad (11)$$

$$w \in \Re \quad (12)$$

Um avaliação de eficiência assumindo rendimentos variáveis à escala faz com que cada unidade seja avaliada relativamente a unidades de dimensão mais semelhante à sua, dando origem a uma fronteira que restringe mais o conjunto de possibilidades produtivas. O modelo que assume rendimentos variáveis à escala denomina-se BCC (do trabalho original de Banker, Charnes e

Cooper (1984)). O que há de comum entre este modelo e o CCR é a medição radial da eficiência que ambos realizam. Em ambos, a medida de eficiência obtida representa o fator pelo qual cada unidade deve multiplicar os seus *inputs*, de modo a situar-se sobre a fronteira de produção, isto é, sem alterar o seu *mix* (proporção) de *inputs* (ou de *outputs* para a orientação alternativa).

Existem outras variantes aos modelos básicos de DEA. Uma delas, que será usada neste estudo, consiste na introdução de restrições aos valores dos pesos. Esta variante é utilizada quando se pretende impor ao modelo características resultantes de um julgamento *a priori* do avaliador, que o torne mais realista, ou mesmo quando se pretende simplesmente analisar o impacto da escolha destes pesos de entre um determinado conjunto de valores (análise de sensibilidade). Uma das formas de impor este tipo de restrições consiste na utilização do método da região de confiança (traduzido do inglês *Assurance Region Method*, proposto por Thompson et al. (1986)), em que se restringe o valor dos rácios entre os pesos de *inputs* (ou *outputs*) distintos, de forma a enquadrar a relação entre ambos com um limite inferior (α) e superior (β) (Thanassoulis et al., 2004):

$$\alpha \leq \frac{v_2}{v_1} \leq \beta \tag{13}$$

Estas restrições poderiam ser também aplicadas aos pesos virtuais dos *inputs* e *outputs*, isto é, o produto do valor de cada variável com os respetivos pesos. De facto, são os pesos virtuais que representam a verdadeira importância de cada *input* e *output* no valor calculado da eficiência de uma unidade. Contudo, a sua utilização implica, de forma indireta, uma limitação absoluta dos pesos do modelo e não um enquadramento relativo, como o realizado com (13) (Thanassoulis et al., 2004), gerando por vezes problemas na otimização do modelo. A colocação da equação (13) no “modelo dos pesos”, sob a forma de duas restrições e consequente linearização, altera a fronteira eficiente, diminui o domínio do modelo de programação linear. Como consequência, as unidades veem a sua eficiência piorar ou manter-se, face ao modelo sem restrições de pesos. Também é possível que a referência com que

a DMU em avaliação é comparada, correspondente ao objetivo de desempenho proposto pelo modelo, tenha, para um ou mais *inputs*, valores superiores ao da DMU avaliada (e vice-versa na orientação para *outputs*).

2.5 A Utilização do DEA na Avaliação de Hospitais

2.5.1 Processo de Produção Hospitalar

Na avaliação de cuidados médicos, especialmente em hospitais ou unidades de cuidados primários, existem duas perspetivas de análise: uma ligada à gestão, que é responsável pela afetação dos recursos necessários à atividade clínica, e outra ligada ao tratamento dos pacientes, que se designa por perspetiva clínica, seguindo Chilingerian e Sherman (2004). Como pode ver-se na Figura 1, as perspetivas estão interligadas, já que os *outputs* da primeira coincidem com os *inputs* da segunda (elementos denominados de *outputs* intermédios). No entanto, enquanto na primeira perspetiva uma gestão eficiente terá como objetivo minimizar o *input*, para um dado nível de *output* intermédio (a capacidade instalada e disponibilizada ao médico), na segunda o tratamento clínico adequado dos doentes preocupa-se com a maximização dos *outcomes* positivos no estado da saúde dos indivíduos, que refletem os resultados finais sobre a sua saúde. Usando DEA, podemos calcular uma medida de desempenho em ambos os casos, sendo que no primeiro denominamos essa medida como eficiência e no segundo como eficácia. A responsabilidade pelo desempenho em cada um dos casos é diferente, já que a eficácia é controlada pelo médico, enquanto a eficiência é controlada pelo gestor.

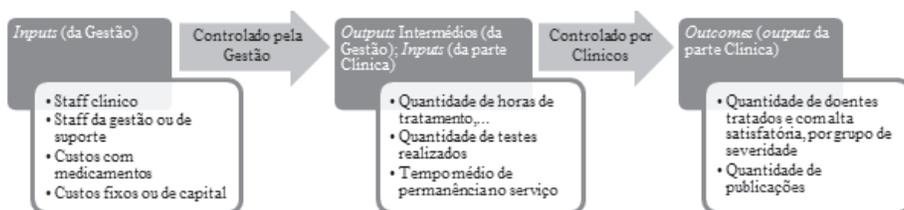


Figura 1 – Duas Perspetivas da Atividade Hospitalar (Chilingerian e Sherman, 2004, 496).

Para um estudo numa perspectiva de eficiência das unidades hospitalares usando DEA, não existe um modelo globalmente aceite relativamente aos *inputs* e *outputs* intermédios a considerar. A escolha deve ser feita de forma a adequar, o melhor possível, o modelo à realidade a estudar. Os *inputs* devem ser categorizados em grupos que façam sentido para os gestores que os controlam e que representem a maior parte dos gastos pelos quais são responsáveis. Quanto aos *outputs*, eles só estarão bem definidos se entrarem em linha de conta com a complexidade dos tratamentos realizados pelas unidades estudadas.

A perspectiva da eficácia clínica, não abordada neste trabalho, tem sido estudada limitando os resultados dos pacientes ao seu estado na alta, como vivo ou morto (Castro, 2011). Sendo esta uma visão simplista do assunto, o estudo do desempenho clínico de especialidades médicas concretas pode vir a alterar este paradigma. Esta é a perspectiva que apresenta maiores dificuldades de avaliação e que, por isso, apresenta o maior número de oportunidades para a aplicação de técnicas de fronteira para a avaliação do desempenho.

2.5.2 Utilização do DEA na Avaliação do Desempenho Hospitalar

O DEA é a metodologia mais usada em estudos de desempenho em saúde e são os hospitais, por larga margem, os alvos preferenciais destes mesmos estudos (Hollingsworth, 2008). Estes estudos tentam avaliar a eficiência destas organizações complexas, onde não existe um modelo de produção globalmente aceite e onde, por isso, as avaliações com DEA poderão ser muito sensíveis a erros de medição ou a uma fraca aferição da realidade, por omissão consciente ou inconsciente de variáveis (Hollingsworth et al., 1999). É por isto mesmo que se defende que a utilização de DEA numa avaliação de hospitais deveria cingir-se a aferir tendências gerais ou a avaliar unidades mais pequenas e concretas, como por exemplo serviços ou mesmo práticas médicas, onde os potenciais erros de medição serão menores. Os hospitais não são, por isso, considerados unidades suficientemente homogéneas para permitirem a utilização simples e direta desta técnica, apesar da literatura se centrar no seu estudo.

Ao contrário da literatura de avaliação de hospitais, que é vasta, a literatura de avaliação de unidades sub-hospitalares não o é. Os estudos de Puig-Junoy (1998) e Kontodimopoulos e Niakas (2005) são exemplos de análises feitas a este nível com recurso ao DEA. No primeiro caso, o estudo tenta discernir o que determina o desempenho de unidades de cuidados intensivos. Este estudo concluiu pela maior relevância do fator “procedimentos médicos utilizados” para a eficiência destas unidades, em detrimento da gravidade ou complexidade das doenças diagnosticadas, evidenciando uma ligação entre eficiência e decisões clínicas, o que sublinha a importância de se aferir o impacto dos cuidados sobre a saúde dos pacientes. O segundo trabalho estuda unidades de hemodiálise gregas, e discute as dificuldades de análise de unidades sub-hospitalares usando DEA. Neste estudo, a validade dos resultados é limitada pelo reduzido número de variáveis consideradas, o que se deve à parca informação disponibilizada por parte dos prestadores de cuidados e das autoridades governamentais.

Em Portugal, encontram-se estudos semelhantes a estes, como os que avaliam Grupos de Diagnósticos Homogêneos (GDH). Como exemplo destes estudos, pode destacar-se o de Dismuke e Sena (2001) que, tendo por base os dois mais frequentes GDH (o das desordens cerebrovasculares, exceto ataques isquémicos transitórios, e o das falhas e ataque cardíacos) estudaram a evolução da produtividade de 3 tecnologias de diagnóstico, a TAC (tomografia axial computadorizada), o eletrocardiograma e o ecocardiograma, ao longo de 3 anos (1992 a 1994). Estes estudos avaliaram se o aumento de produtividade nestas tecnologias trouxe consigo uma diminuição da qualidade, aferida pela mortalidade. Utilizando um modelo de DEA com *outputs* indesejáveis e os índices de Malmquist-Luenberger (usados para aferir a evolução da produtividade), verificou-se que não havia evidência de que o aumento da eficiência tenha provocado uma diminuição da qualidade, medida através da taxa de mortalidade.

A maioria dos estudos sobre hospitais não chega a aferir o impacto que os cuidados de saúde têm na melhoria real da condição dos doentes. É frequente usar-se, como medidas de *output*, variáveis que aferem apenas o resultado do processo produtivo hospitalar, como dias de permanência em internamento,

total de altas médicas ou número de consultas externas. Uma percentagem reduzida dos estudos utiliza medidas de resultado efetivo como a taxa de mortalidade ou a qualidade dos cuidados individuais prestados (Hollingsworth, 2008; O'Neill et al., 2008). Como exemplo destes estudos, pode citar-se o caso Nayar e Ozcan (2008). Avaliando hospitais do estado da Virgínia nos EUA, o estudo tenta descobrir qual o impacto sobre a medida de eficiência dos hospitais, de variáveis de qualidade nos *outputs*. Deste estudo surgiu evidência de que os hospitais tecnicamente eficientes eram também aqueles que apresentavam bons resultados em termos de qualidade de cuidados. No entanto, esta conclusão é enviesada, por ser uma consequência quase direta da metodologia usada, em que se consideraram as perspetivas de eficiência e eficácia simultaneamente.

A aplicação do DEA ao setor hospitalar usa tradicionalmente uma orientação para *inputs*, justificando-se este facto pela predominância da perspetiva da eficiência na avaliação hospitalar (Figura 1). Isto resulta da preocupação que os diferentes agentes da saúde têm na contenção de recursos, uma vez que a produção hospitalar é, essencialmente, ditada pela procura da população, um fator exógeno ao hospital. Nesta categoria, as variáveis mais utilizadas correspondem ao número de camas, número de médicos, enfermeiros, outros funcionários ou mesmo os custos que estas categorias profissionais representam, assim como os custos com consumíveis (medicamentos e material clínico, principalmente) (Castro, 2011).¹ A maioria dos estudos analisa a eficiência técnica e, tendo em vista uma homogeneização de *outputs*, utilizam-se variáveis que fazem referência à complexidade dos doentes tratados (*case-mix*) ou ao número de serviços dos hospitais (*service-mix*). Devido à inexistência de um modelo de produção globalmente aceite, bem como à heterogeneidade da atividade hospitalar em contextos distintos, não há consenso relativamente ao tipo de rendimentos à escala que deverá ser assumido nas análises, optando-se muitas vezes pelo uso dos dois modelos e posterior comparação.

¹ Em países como Portugal, onde a existência de longas listas de espera é um problema, poder-se-ia optar por uma orientação para os *outputs*, identificando-se, assim, formas para aumentar a produtividade dos hospitais no lado da produção.

Em Portugal, a atividade avaliativa tem vindo a ser objeto de atenção crescente e centrou-se, principalmente, nas recentes transformações que o setor da saúde tem sofrido. Mais recentemente, utilizou-se o DEA em grande medida, ao nível dos hospitais, para avaliar a reforma da transformação dos hospitais públicos em Sociedades Anónimas (SA). Deste universo, podem destacar-se os estudos do Tribunal de Contas (2006), de Moreira (2008) e de Gonçalves (2008). O primeiro tinha por objetivo medir a eficiência económica dos hospitais transformados e medir a qualidade e equidade no acesso aos cuidados de saúde. Só no estudo da eficiência económica se utilizou o DEA. Como conclusões, destaca-se a evidência de diferenças nos níveis de eficiência, quando se comparam os hospitais SA com os do Setor Público Administrativo (SPA) de dimensão semelhante, embora sem resultados totalmente consistentes. Em Moreira (2008), utilizou-se o grupo dos hospitais SPA como um grupo de controlo, criando-se dois tipos de fronteiras. As conclusões, baseadas em dados de 2001 a 2005, apontam para uma maior eficiência média inicial no grupo dos hospitais SPA, o que se reverteu no ano de 2005. Recorrendo à evolução temporal dos valores de eficiência, a análise permitiu identificar uma maior evolução (ligeira, porque as diferenças iniciais eram também ligeiras) dos hospitais SA face aos SPA, em termos médios. O estudo de Gonçalves (2008) obteve um resultado contrário. Utilizando dados referentes ao período entre 2002 e 2004, Gonçalves (2008) verificou, usando os Índices de Malmquist, que a fronteira de eficiência técnica evoluiu positivamente neste período, tendo sido os hospitais SPA os que mais contribuíram para esta evolução, já que apresentaram crescimentos de eficiência maiores que os SA. Em suma, os estudos parecem indicar que os hospitais SA viram a sua eficiência melhorar, globalmente, mas não a um ritmo que os diferencie muito dos hospitais SPA, como esperado inicialmente. Em conjunto, os estudos apresentam evidência da inadequação do método quando se avaliam hospitais, pela impossibilidade da definição de modelos que capturem mais fielmente a sua complexidade. Em Portugal, não existe outra avaliação que, usando DEA, se foque nos serviços dos hospitais.

3. Modelo Avaliativo para os Serviços Médicos/Cirúrgicos dos Hospitais do SNS numa Perspetiva de Eficiência

3.1 Caracterização e Escolha dos Serviços a Analisar

A estrutura dos hospitais não é homogénea. Cada hospital tem a sua estrutura interna definida de acordo com a sua influência geográfica e, principalmente, de acordo com as valências/especialidades implementadas. No entanto, algo é comum a todos: os serviços são as suas unidades funcionais fundamentais. Existem diversas formas de agrupar os diferentes tipos de serviços de um hospital, apresentando-se aqui uma que define quatro grupos de serviços. Os serviços das especialidades médicas ou cirúrgicas² têm a responsabilidade de servir diretamente os pacientes através de internamentos, consultas externas, sessões em hospital de dia ou cirurgias em ambulatório, quatro das cinco principais linhas de produção que um hospital realiza, como é convencionalmente aceite na literatura (falta nesta lista a linha de produção de urgência). A estes serviços podemos chamar de “Serviços Principais”. Os hospitais têm também serviços responsáveis por servir de apoio clínico aos serviços anteriores, tais como o bloco operatório, os serviços médicos apenas responsáveis por meios complementares de diagnóstico e terapêutica ou a farmácia hospitalar. Também têm serviços que prestam cuidados de saúde generalizados, como o serviço de Urgência Geral. Todos estes serviços podem ser denominados “Serviços Clínicos Transversais”. Para um apoio mais geral a todos os serviços hospitalares e unidades, como a limpeza ou manutenção dos edifícios, os hospitais têm ainda os denominados “Serviços de Apoio Geral”. Por fim, os hospitais têm também serviços para gerir toda a sua estrutura, como o conselho de administração e o pessoal administrativo ou outros técnicos. Esses serviços podem ser agrupados nos “Serviços Administrativos”. Neste trabalho, vão analisar-se os serviços principais já que, para além de serem os responsáveis pelo atendimento direto dos utentes, são eles também os responsáveis pelo maior volume de custos que se podem imputar diretamente

² Em Medicina, diferenciam-se as especialidades que são exclusivas dos cirurgiões, especialidades cirúrgicas, das que são da responsabilidade dos não-cirurgiões, especialidades médicas.

a cada serviço, como pode ver-se na Figura 2, com dados dos hospitais do SNS que reportam a 2008.

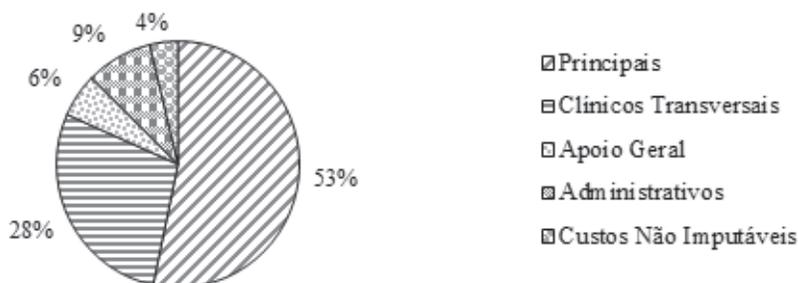


Figura 2 – Distribuição dos Custos Diretos pelos Diferentes Grupos de Serviços.

Neste estudo apresenta-se um modelo “ideal”, que tem como objetivo uma avaliação, numa perspetiva de eficiência, dos serviços principais, considerando apenas a produção das quatro principais linhas de produção. Também é nestes serviços que se realiza a maioria da produção hospitalar e é neles que são maioritariamente imputados todos os custos das restantes secções hospitalares (secções de apoio). É necessário notar que é enquanto dura um internamento que se realizam cirurgias nos pacientes internados em serviços de especialidades cirúrgicas e, por isso, a produção de cirurgias está já contemplada quando se fala em internamentos e se controla para a complexidade das doenças tratadas. Durante o internamento, o serviço pode também colocar os doentes em unidades de cuidados intensivos afetas exclusivamente ao serviço. Aqui, considerar-se-á que estas unidades fazem parte da prestação de cuidados em internamento e, como tal, os recursos utilizados em internamento contemplam os recursos usados nestas unidades especiais dos respetivos serviços, para além dos usados nas unidades mais convencionais. Posteriormente à idealização do modelo, ele será concretizado numa avaliação do serviço de Medicina Interna de hospitais do SNS, para o ano de 2008.

3.2 Proposta de Modelo de DEA para Benchmarking aos Serviços Hospitalares Principais numa Perspetiva de Eficiência

A criação do modelo deste estudo iniciou-se com uma revisão bibliográfica, que teve por finalidade identificar as variáveis mais frequentes nos estudos da área. Esta revisão pode ser consultada em Castro (2011).

Tendo em vista a identificação das variáveis que melhor caracterizam o conjunto dos recursos (*inputs*) utilizados pelos serviços a estudar, identificaram-se, posteriormente, as estruturas de custos médias para os serviços de especialidades médicas e para os de especialidades cirúrgicas, em todos os hospitais do SNS. As variáveis de *input* em causa, num modelo de aferição da eficiência técnica, podem ser escolhidas de tal forma que representem cada elemento da estrutura de custos média. Para concretizar isto mesmo, recorreu-se à informação da contabilidade de gestão de 2008, fornecida pelos hospitais à ACSS. Nesta análise, contabilizaram-se todos os custos incorridos pelas linhas de produção de internamento, consulta externa e hospital de dia, já que a ACSS não dispõe de informação relativa à discriminação dos custos com cirurgia em ambulatório pelos diferentes serviços. No entanto, esta linha de produção apresentou, em 2008, custos residuais face aos restantes.

Como pode ser visto na Figura 3, os tipos de custos mais representativos nos dois grupos de serviços são os mesmos (variando apenas o peso de cada um deles), à exceção do Bloco Operatório, não utilizado pelas especialidades médicas (não realizam cirurgias). As principais categorias de custos são, então, os custos com Pessoal (CP), Matérias Vendidas e Consumidas (CMVMC – que representam principalmente os medicamentos e o material clínico), Meios Complementares de Diagnóstico e Terapêutica (MCDT), Serviços Administrativos (SA), Fornecimentos e Serviços Externos (FSE) e Serviços Hoteleiros (SH), responsáveis por 93% do total de custos nos serviços médicos. Aos serviços cirúrgicos, acrescenta-se o Bloco Operatório (BO) e o total destas categorias representa 94% do total.

Sendo os custos com o pessoal uma categoria bastante heterogénea, por incorporar diferentes categorias profissionais, é necessário avaliá-la de forma mais pormenorizada. Realizando uma análise semelhante à anterior, verificou-

se que 90% e 92% dos custos com pessoal nos serviços médicos e cirúrgicos, respetivamente, são atribuíveis aos médicos, enfermeiros e ao pessoal operário e auxiliar. Estas serão, portanto, as categorias profissionais que interessará considerar como recursos, se se quiser aferir quais as que têm maior impacto nos custos e uma ação direta no funcionamento dos serviços, mantendo ainda o número de variáveis reduzido.

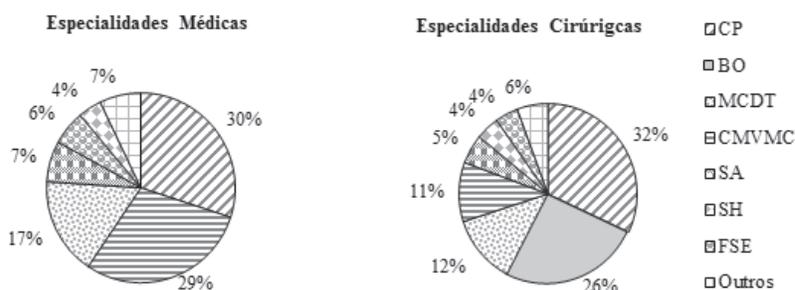


Figura 3 - Estrutura dos Custos Gerais nos Serviços Principais.

Após a análise à estrutura de custos dos serviços principais, é então possível definir uma lista de *inputs* e *outputs* que permita modelar, idealmente, o processo de utilização de recursos para disponibilizar aos clínicos a capacidade de tratar doentes (*outputs* intermédios). Como se pretende calcular a eficiência técnica das unidades em avaliação, definem-se todas as variáveis, de um modelo ideal, em termos de quantidades e não de valores económicos.

Para representar os recursos de pessoal, optou-se por considerar o número de horas semanais disponibilizadas pelos médicos e enfermeiros, como dois dos *inputs* do modelo. Como o pessoal operário e auxiliar tem um peso nos custos bastante inferior ao das classes clínicas, optou-se por representar este recurso por uma variável mais genérica. Pretendeu-se que esta variável representasse outros recursos que se possam assumir proporcionais à dimensão de um serviço. A variável escolhida foi o número de camas, muitas vezes considerada, na literatura, uma *proxy* para a dimensão dos serviços. Além de representar o pessoal operário, a variável “número de camas” representa, neste modelo, os custos com os FSE (muitas vezes considerados, na literatura, proporcionais à dimensão de um serviço), com os serviços administrativos e

com os serviços hoteleiros. O número de camas é, assim, uma variável que funciona como *proxy* para estas quatro categorias de recursos. A alternativa a esta opção seria a consideração de mais variáveis de *input*, o que reduziria a discriminação do modelo, dado o tamanho da amostra de hospitais nacionais. Para atestar esta opção, verificou-se a existência de correlações significativas entre estas variáveis e o número de camas (valores entre 0,703 e 0,937), o que confirma a validade da representação destes recursos pela variável “número de camas”.

Os diferentes serviços de MCDT prestam apoio aos serviços principais, o que leva à necessidade de se contabilizar os procedimentos que os últimos requerem, de forma ponderada, para refletir a diferente complexidade e imputar um maior dispêndio de recursos aos serviços que utilizem procedimentos mais complexos. O ponderador dos procedimentos é definido pelo Ministério na Portaria n.º 839-A/2009. Os MCDT são então representados por esta variável. Para representar os CMVMC, seria ideal contabilizarem-se os medicamentos e material clínico, de forma semelhante ao que é feito com os MCDT. Contudo, não existe uma definição de ponderadores, por parte do Ministério da Saúde, para estes elementos. Assim, na prática e por agora, não resta alternativa à utilização dos custos para a sua ponderação. Na lista de recursos apresentada abaixo permanece a variável que seria ideal utilizar (e que um dia talvez se possa utilizar), enquanto na secção seguinte, em que se demonstra a aplicação prática do modelo teórico aqui proposto, se utilizam os custos incorridos com os CMVMC, como *proxy* para as quantidades de matérias ponderadas usadas pelos serviços.

Para se representar os recursos gastos, pelos serviços, com o bloco operatório (BO), optou-se por escolher a variável que representa o tempo médio semanal que cada especialidade ocupa neste serviço auxiliar. Além do tempo total das intervenções, pode considerar-se o número total de intervenções como produto do BO. O número de horas que um serviço ocupou no bloco será vantajoso relativamente ao total de intervenções, por aferir, simultaneamente, a complexidade e quantidade das intervenções realizadas, considerando-se uma qualidade uniforme dos médicos.

Por fim, é de salientar que as variáveis de recursos escolhidas, e sumariadas na Tabela 1, se referem a 89 e 92% dos custos totais das quatro linhas de produção consideradas nos serviços de especialidades médicas e cirúrgicas, respetivamente.

Tabela 1 - Variáveis de Input dos Modelos de DEA para os Serviços Principais

Nome	Definição Operacional
Nº Horas Médicos	Contabilização do nº de horas semanais de trabalho médico no serviço.
Nº Horas Enfermeiros	Contabilização do nº de horas semanais de trabalho enfermeiro no serviço.
Nº Exames Ponderados	Contabilização do nº de procedimentos de MCDT requeridos pelo serviço, ponderados pelos fatores definidos na Portaria n.º 839-A/2009 do Ministério da Saúde: soma do total de procedimentos multiplicados pelos seus ponderadores.
Nº Matérias Ponderadas	“Contabilização” do nº ponderado de medicamentos e materiais clínicos utilizados pelo serviço.
Nº Camas	Contabilização do nº de camas existentes no serviço.
Nº Horas de Bloco	Contabilização do nº de horas totais das intervenções cirúrgicas realizadas sob a responsabilidade do serviço.

Relativamente às variáveis de *output*, a sua escolha deve refletir o que é fundamental ser medido: o que é produzido pelas diferentes linhas de produção. Nisto incluem-se o número total de doentes atendidos em cada linha de produção, que pode também ser considerado uma medida de acesso aos cuidados de saúde, e a quantidade total de tratamento produzido e disponível para todos os utentes, medida usualmente pelo número de unidades pelas quais se pode dividir cada tipo de tratamento (*proxies* desta quantidade). O número de doentes atendidos, em internamento, é usualmente contabilizado pelo número de altas num ano, em consulta externa, pelo número de primeiras consultas e, em hospital de dia, pela contabilização direta do número de doentes. O número de cirurgias em ambulatório reflete o número de doentes atendidos e as quantidades totais de tratamento produzidas, já que apenas se considera um novo doente sempre que se realiza uma nova cirurgia. Quanto ao total de tratamento produzido nas restantes linhas, ele é usualmente medido, para o internamento, pelo número de dias total que os doentes de um serviço permaneceram internados, para as consultas, pelo total de consultas realizadas (a soma das primeiras com as consultas de acompanhamento) e,

para o hospital de dia, pelo número total de sessões a que todos os utentes tiveram acesso, tudo num determinado período de tempo. Todos estes *outputs* encontram-se sumariados na Tabela 2.

Tabela 2 - Variáveis de Output do Modelo de DEA para os Serviços Principais

Nome	Definição Operacional
Tempo Estadia Total	Soma do nº de dias de estadia dos pacientes que saem do internamento.
Nº Altas Internamento (com correção pelo <i>case-mix</i>)	Contabilização do nº total de altas de internamento ponderadas pelos índices de <i>case-mix</i> do serviço: soma do total de altas multiplicadas pelo índice de <i>case-mix</i> do serviço.
Nº de Consultas Externas	Contabilização do nº de consultas externas do serviço.
Nº 1 ^{as} consultas	Contabilização do nº de primeiras consultas
Nº Sessões Hospital de Dia	Nº total de sessões de Hospital de Dia.
Nº Utentes Hospital de Dia	Contabilização do nº de doentes que tiveram acesso a sessões de hospital de dia do serviço.
Nº Cirurgias Ambulatório (com correção pelo <i>case-mix</i>)	Contabilização do nº de cirurgias em ambulatório do serviço ponderadas pelo respetivo índice de <i>case-mix</i> .

Ao medir ambas as dimensões da produção, comparando duas unidades com o mesmo nível de *inputs* e de produção em quantidade (com os mesmos dias de doentes em internamento), a que apresente o menor acesso de pessoas aos cuidados (menor número de altas), apresenta-se como uma unidade mais ineficiente. No entanto, a um fraco acesso aos cuidados de saúde proporcionado por um hospital (baixo número de altas, por exemplo), pode estar associada uma maior complexidade das doenças que este trata, o que faz prolongar a estadia de um só doente. Para controlar esta heterogeneidade das doenças, utiliza-se aqui a variável mais comum e a única disponível em Portugal: o índice de *case-mix*. Os índices determinam-se calculando o rácio entre o número de doentes equivalentes (definição na Circular Normativa N.º 3/2008 do Ministério da Saúde) ponderados pelos pesos relativos dos respetivos GDH definidos na Portaria n.º 839-A/2009 do Ministério da Saúde, que refletem as diferenças inerentes às doenças. Este índice apenas pode ser calculado para o internamento e para as cirurgias em ambulatório, porque estas são as únicas linhas para as quais existe um registo e determinação de ponderadores, efetuada pelo Ministério da Saúde. O índice não é usado como

um *output per se*, mas como um fator de ajuste, multiplicando-se o seu valor pela variável a ajustar.

É também de sublinhar que, aquando da implementação aos diferentes serviços (concretos), será necessário verificar a existência ou a validade de algumas das variáveis da Tabela 2.

3.3 Descrição do Procedimento de Implementação do Modelo

Concretizou-se, neste estudo, um modelo piloto para avaliar o desempenho do serviço de Medicina Interna. Além de ser o segundo serviço principal mais dispendioso, lida com os casos mais comuns e é o que existe num maior número de hospitais. Para este serviço médico, o modelo “ideal” tem como variáveis de *input*, o “Nº de Horas de Médicos”, o “Nº de Horas de Enfermeiros”, o “Nº de Exames Ponderados”, o “Nº de Matérias Ponderadas” e o “Nº de Camas”. Quanto aos *outputs*, sendo esta uma especialidade onde não existe produção em hospital de dia nem de cirurgias, serão relevantes as variáveis do modelo “ideal” “Tempo Estadia Total”, “Nº Altas Internamento”, “Nº de Consultas Externas” e “Nº 1^{as} consultas”.

Ao tentar implementar o modelo exemplificativo, encontramos dificuldades na obtenção de alguns dados e na fiabilidade de outros. Por esse motivo, entende-se este trabalho como um guião para a implementação prática do modelo proposto e uma exemplificação do cálculo da eficiência dos diferentes serviços dos hospitais, sem se ambicionar alcançar uma imagem absolutamente fidedigna da realidade aqui avaliada (para uma discussão aprofundada do assunto, consultar Castro (2011)). Para além disto, para construir o modelo apresentado na secção anterior, não foi possível obter atempadamente alguns dados, nomeadamente o índice de *case-mix* do serviço de Medicina Interna, a quantidade de MCDT ponderados consumidos e as horas de trabalho de enfermeiros, embora esta informação seja recolhida pela ACSS. Devido a estas restrições, foram utilizados os custos incorridos pelos serviços com os enfermeiros, com os exames ponderados e com as matérias ponderadas, como *proxy* para as quantidades em falta. Para se calcular o custo com os enfermeiros somou-se o seu salário de base aos benefícios imputados ao pessoal do serviço, de forma proporcional. Para as matérias ponderadas

somaram-se os custos com medicamentos e material clínico e para os exames ponderados somaram-se os custos com todos os MCDT incorridos pelos serviços em estudo. O índice de *case-mix* usado foi o índice para as especialidades médicas de cada hospital³.

Contabilizaram-se 54 hospitais que apresentaram dados para o serviço em análise (hospitais que têm, de facto, esta valência), mas apenas 48 foram avaliados. Estes hospitais fazem parte do universo dos 59 hospitais públicos do SNS para os quais a ACSS disponibilizou os dados. Em 2008, havia mais 7 hospitais no SNS mas, por diversos motivos, a ACSS não possuía informação sobre eles. Foram eliminados da amostra os hospitais que apresentavam falhas em qualquer das variáveis, quer por valores em falta, quer terem uma ordem de grandeza incompatível com a realidade medida (para mais informação consultar Castro (2011)). Apresentam-se as estatísticas descritivas dos dados na Tabela 3, que evidenciam a heterogeneidade da amostra analisada.

Tabela 3 - Estatística Descritiva das Variáveis Utilizadas no Modelo de DEA

Variável	Média	Desvio Padrão	Máximo	Mínimo
Inputs				
x_1 - Nº Horas Médicos (h)	1350	1262	6420	46
x_2 - Custos Enfermeiros (€)	1509070	1371092	6099323	27602
x_3 - Custos Exames Ponderados (€)	1476455	1532799	7980923	27687
x_4 - Custos Matérias Ponderadas (€)	1410007	1625900	8805764	297
x_5 - Nº Camas	78	58	260	0
Outputs				
y_1 - Tempo Estadia Total (dias)	28159	21068	102931	0
y_2 - Nº Altas Internamento (com correção pelo <i>case-mix</i>)	2680	2568	15032	0
y_3 - Nº de Consultas Externas	8350	6334	23932	529
y_4 - Nº Primeiras Consultas	1665	1436	8324	108

Utilizou-se o *software* não-comercial EMS (*Efficiency Measurement System*) para calcular os resultados de eficiência das diferentes unidades. Começou-se

³ Em vez de um único valor de *case-mix*, a ACSS discrimina a diferente complexidade de um internamento ou ocorrência em ambulatório devido a doenças do foro cirúrgico e de foro médico.

por calcular o valor da eficiência considerando que se estaria perante uma situação de rendimentos constantes à escala. No entanto, verificou-se que as unidades consideradas eficientes eram todas de pequena dimensão, existindo uma tendência de decréscimo da eficiência com o aumento da dimensão dos serviços, representada pelo número de camas. Assim sendo, adotou-se um modelo de DEA de rendimentos variáveis à escala. Ao calcular os valores de eficiência para este novo modelo, concluiu-se que o número de unidades consideradas eficientes era muito elevado (28 dos 48 serviços de medicina interna), fruto da existência de vários pesos nulos, que eliminaram a influência da respetiva variável. Assim, decidiu-se introduzir restrições aos valores dos pesos, usando o método da região de confiança, de forma a impedir que as unidades pudessem optar por valorizações irreais dos diferentes fatores.

Para se identificar as restrições a aplicar aos pesos das diferentes variáveis, começou por identificar-se a importância relativa de cada *output* e *input*. Começando pelos *outputs*, analisaram-se os custos unitários das duas linhas de produção consideradas. Os resultados apresentam-se na Tabela 4.

Tabela 4 - Custos Unitários por Unidade Produzida nos Serviços de Medicina Interna

Variável de <i>Output</i>	Volume	Custo Unitário Médio
y_2	157.207 altas	2.091,39 €/alta
y_1	1.366.808 dias internamento	240,55 €/dia internamento
y_3	517.972 consultas	155,28 €/consulta

Estes resultados revelam que, na dimensão referente à quantidade total de tratamento produzido, o peso da variável referente ao internamento deverá ser igual a pelo menos 1,5 vezes o peso da variável referente às consultas externas (valor obtido do rácio entre os custos unitários das consultas e dos dias em internamento). Para impedir uma excessiva disparidade na razão entre os pesos destas variáveis, restringiu-se o limite superior desta relação ao valor de 5. A consideração deste limite está relacionada com a manutenção da ordem de

grandeza entre os pesos⁴. Isto permite que os pesos tomem valores de magnitude semelhante, sem, no entanto, se restringir excessivamente a capacidade de cada DMU escolher os seus pesos, o que resulta nas primeiras restrições:

$$1,5 \leq \frac{u_1}{u_3} < 5 \quad (14)$$

De forma a não tornar o modelo demasiado restritivo, apenas se restringiu esta razão entre *outputs* de linhas de produção distintas (consultas e internamento).

Relativamente à dimensão referente ao acesso aos cuidados, apenas foi possível obter informação para a linha de produção de internamento (ver Tabela 4, onde falta y_4). Nesta linha, verifica-se que esta dimensão (y_2) apresenta um custo unitário aproximadamente dez vezes superior à dimensão da quantidade de tratamento (y_1). Assim, optou-se por impor limites à relação entre os pesos que refletissem essa magnitude, embora permitindo também alguma flexibilidade em torno do valor 10 (entre 5 e 50)⁵:

$$5 \leq \frac{u_2}{u_1} < 50 \quad (15)$$

Novamente, esta limitação apenas elimina as possibilidades de produção irreais, deixando ao modelo de DEA a capacidade de dar às unidades uma liberdade de escolha dos pesos plausível. Este é um intervalo muito lato, que reflete apenas uma limitação à magnitude das razões que se pretende obter, para que as unidades não deixem de pesar todos os *outputs*.

⁴ O conceito de ordem de grandeza de um número identifica, na escala decimal, a quantidade de números 10 que um número contém. Convenciona-se internacionalmente que um número x é da ordem de grandeza n se:

$$5 \times 10^{n-1} \leq x < 5 \times 10^n$$

Neste caso, 5 será o valor limite para a relação entre os pesos permanecer na mesma ordem de grandeza: 0, que vem de 10^0 .

⁵ Da mesma forma que antes, para que agora se reflita uma relação de ordem de grandeza 1: 10^1 .

Dada a inexistência de informação quanto à dimensão de acesso na linha de produção de consulta externa (y_4), considerou-se que a relação entre as suas duas dimensões seria semelhante à existente na linha de internamento. Por conseguinte, aplicaram-se nesta linha de produção restrições semelhantes às da linha de produção relativa ao internamento:

$$5 \leq \frac{u_4}{u_3} < 50 \quad (16)$$

Para identificar a importância de cada *input*, analisou-se a distribuição dos custos que cada uma das variáveis representa em todos os serviços de Medicina Interna. Os resultados apresentam-se na Figura 4, que evidencia que cada uma das cinco variáveis de *input* aqui consideradas tem uma importância muito semelhante nos custos totais do serviço em análise.

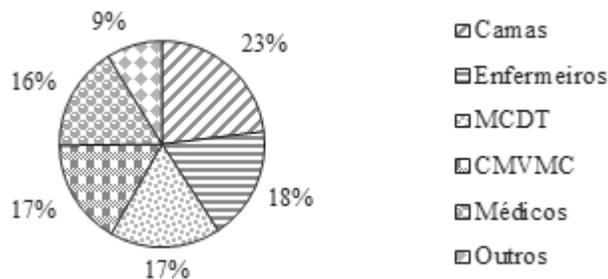


Figura 4 - Distribuição dos Custos Gerais nos Serviços de Medicina Interna Representados pelas Variáveis de Interesse.

Para que se reflita nos resultados a importância relativa semelhante de cada *input*, sem se restringir demasiado a liberdade de escolha de cada DMU (como se fez no lado dos *outputs*), optou-se por forçar os hospitais a pesarem todos os *inputs*, já que nenhum parece ter uma importância superior aos restantes e não seria “admissível” que um hospital se tornasse eficiente por pesar apenas alguns *inputs*, negligenciando os restantes. Nesse sentido, limitaram-se as relações entre todos os *inputs* e o *input* “número de horas médicos” (x_1), permitindo mesmo assim margens de variação largas entre os pesos virtuais de cada *input* e de x_1 . De notar que neste caso as restrições impostas são entre os

pesos virtuais de um hospital fictício, usando *inputs* médios. A opção por restrições aos pesos virtuais deveu-se às diferentes unidades de medida das variáveis. Como se pode ver em (17) estas restrições virtuais convertem-se em *Assurance Regions*.

$$0,5 \leq \frac{v_i \bar{x}_i}{v_1 \bar{x}_1} < 5 \Leftrightarrow$$

$$0,5 \times \frac{\bar{x}_1}{\bar{x}_i} \leq \frac{v_i}{v_1} < 5 \times \frac{\bar{x}_1}{\bar{x}_i} \quad (17)$$

Dado que se pretendia apenas limitar a magnitude das razões entre os pesos de forma lata, arredondou-se o rácio dos valores médios dos *inputs* para a potência de base 10 mais próxima⁶. Assim, resultaram as seguintes restrições, aplicadas aos pesos dos *inputs*:

$$0,5 \times 10^{-3} \leq \frac{v_2}{v_1} < 5 \times 10^{-3} \quad (18)$$

$$0,5 \times 10^{-3} \leq \frac{v_3}{v_1} < 5 \times 10^{-3} \quad (19)$$

$$0,5 \times 10^{-3} \leq \frac{v_4}{v_1} < 5 \times 10^{-3} \quad (20)$$

$$0,5 \times 10^1 \leq \frac{v_5}{v_1} < 5 \times 10^1 \quad (21)$$

3.4 Discussão dos Resultados da Avaliação de Eficiência

Na Figura 5 apresenta-se a distribuição das eficiências obtidas para os 49 hospitais analisados. A média dos resultados obtidos foi de 74,2%, com um desvio padrão de 20,1%. O valor mínimo de eficiência obtido foi de 38,4%. Dos resultados, destaca-se a grande amplitude dos valores de eficiência, o que evidencia uma grande disparidade de desempenhos. Aproximadamente um quinto dos hospitais foi considerado eficiente (10 em 48) e é semelhante o número de hospitais com um nível de eficiência baixo (7, abaixo dos 50%) e alto (9, entre os 80% e 100%). A grande maioria dos hospitais, 22, encontra-se no entanto num nível médio de eficiência (entre 50% e 80%). Esta análise permite concluir que há uma grande margem de melhoria da eficiência dos serviços de Medicina Interna, o que justifica a relevância deste assunto para a gestão hospitalar (para mais detalhes de resultados consultar Castro (2011)).

⁶ Novamente, entra aqui o conceito de ordem de grandeza de um número.

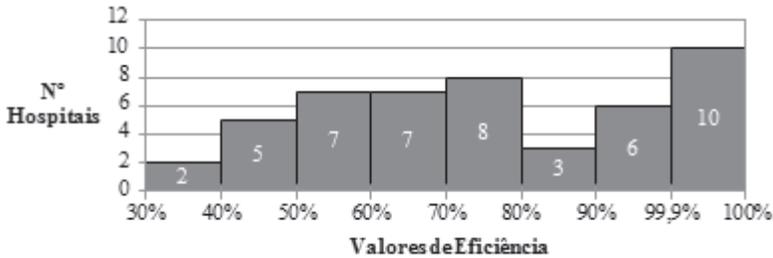


Figura 5 – Distribuição dos Valores de Eficiência.

Calculando a razão entre a média dos valores das variáveis de *input* e de *output* das unidades mais ineficientes (eficiência inferior a 80%) e as eficientes, como se apresenta na Figura 6, normalizando para os valores destas últimas, percebemos que os serviços de medicina interna mais ineficientes apresentam tempos de internamento e número de consultas externas semelhantes aos serviços eficientes, apresentando menos altas de internamento e menor número de primeiras consultas. Ainda assim, estas unidades apresentam custos bastante superiores aos das unidades eficientes, particularmente em matérias ponderadas, cujos custos são cerca de 1,5 vezes superiores. Os custos com exames ponderados e com enfermeiros também são superiores, mas não se verificam muitas diferenças no nº de camas e nas horas de médicos entre os serviços eficientes e ineficientes.

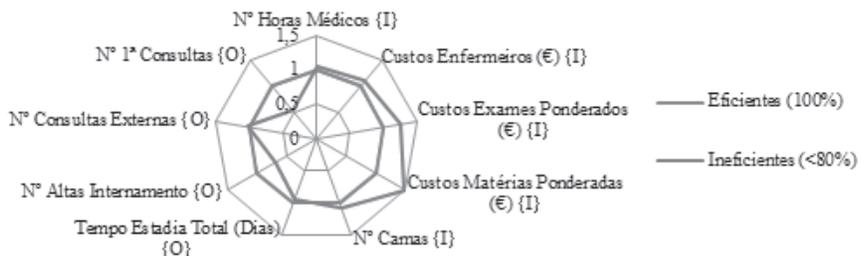


Figura 6 - Comparação entre Valores Médios de Unidades Eficientes e Ineficientes.

Nos hospitais avaliados, o total das poupanças potenciais obtidas através da eliminação das ineficiências é apresentado na Figura 7. Pode concluir-se que,

tal como verificado anteriormente, as matérias ponderadas, os enfermeiros e os MCDT são as áreas onde será possível obter maiores poupanças

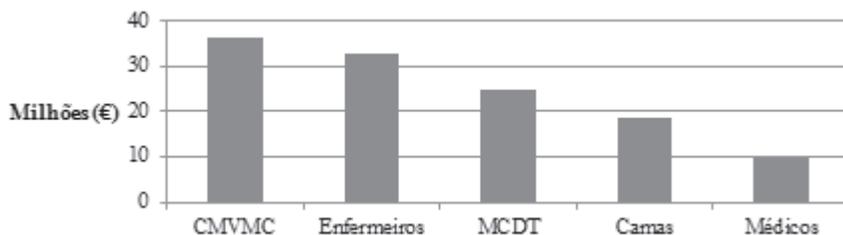


Figura 7 – Potencial de Poupanças a Obter Identificadas pelo Modelo.

A análise com DEA permite identificar as oportunidades de melhoria de hospitais, individualmente. Apresenta-se na Figura 8, a título de exemplo, a razão entre os valores das variáveis para um hospital com resultado de eficiência de 73,1% e a média dos valores para os seus *benchmarks* (três hospitais). Neste exemplo, é possível verificar como este hospital se compara com uma unidade ideal (uma ponderação dos seus três *benchmarks*) que tem praticamente o seu nível de *outputs*. Assim, é possível identificar quais os excessos de *inputs* que este hospital apresenta e que o afastam da fronteira das melhores práticas. No caso dos CMVMC, a categoria em que este hospital deve apostar mais para melhorar o seu desempenho neste serviço, verifica-se que a razão entre o valor do input observado e o ideal é cerca de 4,5.

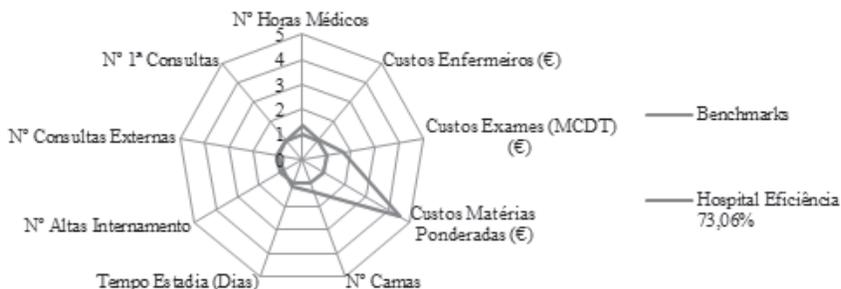


Figura 8- Comparação entre Valores Médios de uma Unidade e seus *Benchmarks*.

4. Conclusões e Trabalho Futuro

O trabalho aqui desenvolvido foi realizado no âmbito do projeto “*Benchmarking* de Hospitais Portugueses”, que tem como objetivo desenvolver uma plataforma *web* de *benchmarking* para os hospitais portugueses. Pretende-se que esta possa ser utilizada numa efetiva avaliação dos hospitais, permitindo observar-se e melhorar continuamente o seu desempenho, assim como permitir ao público em geral escrutinar e até escolher os prestadores dos serviços de saúde. A plataforma ambiciona apresentar uma panóplia alargada de indicadores que incidam sobre as várias atividades hospitalares e recorrer a técnicas de fronteira, para agregar alguma dessa informação em indicadores globais que permitam uma mais fácil comparação de hospitais.

Neste trabalho idealizou-se um modelo para avaliar os diferentes serviços principais, na ótica da eficiência hospitalar. Este modelo teve por base uma revisão bibliográfica que permitiu identificar as variáveis mais comuns e passíveis de ser usadas num modelo de DEA, e um estudo exaustivo da contabilidade analítica dos hospitais do SNS, que permitiu identificar as variáveis responsáveis pela maior proporção de custos. Como exemplificado na implementação prática, quando aplicado a cada um dos serviços de um hospital, o modelo tem de ser adaptado mas não deixa de ter uma identidade genérica.

A aplicação do modelo de DEA ao serviço de Medicina Interna, permitiu comparar os níveis de eficiência deste serviço em hospitais distintos. O modelo permitiu identificar, para os serviços ineficientes, serviços eficientes com características semelhantes, que podem ser usados para compreender melhor as melhorias que cada unidade pode fazer, no sentido de caminhar para um melhor desempenho. No caso da Medicina Interna, verificou-se que os maiores ganhos de eficiência podem ser realizados através de reduções nos custos dos medicamentos e material clínico.

Há ainda muito trabalho pela frente para se operacionalizar uma avaliação aos hospitais. Em primeiro lugar, faltará validar o modelo proposto junto dos principais utilizadores, os gestores hospitalares. Faltará também averiguar as razões para a falta de coerência apresentada pelas diferentes fontes de informação fornecidas pelo mesmo organismo do Estado e tentar identificar

uma forma de se solucionar este problema. Posteriormente será necessário idealizar uma forma de agregar os resultados de eficiência de cada serviço numa medida única a atribuir a cada hospital, de forma a permitir uma comparação global entre os hospitais (e perceber o desempenho dos gestores hospitalares). Tendo os serviços transversais e auxiliares um peso significativo nos custos diretos dos hospitais, será também importante fazer-se a sua avaliação.

Será fundamental aferir-se a eficácia no tratamento dos pacientes, através de modelos de avaliação do tratamento das doenças. Sem isto, o modelo aqui apresentado perde algum do fundamento de existir. Para isto será fundamental o envolvimento de clínicos, dada a sua capacidade única de aferir a capacidade das variáveis do modelo representarem adequadamente o que se pretende medir. Por outro lado, a sua colaboração é essencial para validar e apoiar a aceitação por parte da classe médica dos modelos a propor. Será também interessante verificar se existe ou não antagonismo nos valores dos indicadores agregadores das medidas de eficiência e de eficácia dos hospitais.

Apesar de todos os constrangimentos, os desenvolvimentos alcançados com este trabalho foram significativos, nomeadamente com a concretização de um modelo de avaliação ao nível dos serviços hospitalares, unidades verdadeiramente comparáveis mas até aqui não abordadas na literatura nacional. A aplicação de DEA permite comparar os hospitais recorrendo a uma medida única, mas tendo por base um conjunto alargado de indicadores, cuja importância individual é definida no interesse de cada avaliado. Espera-se, por isso, que este trabalho contribua para uma mais efetiva avaliação dos hospitais e se consiga, assim, promover a melhoria da sua eficiência, sem prejuízo da sua eficácia.

Agradecimentos

Agradecemos à Fundação para a Ciência e a Tecnologia (FCT) que, através da bolsa de doutoramento SFRH/BD/76634/2011 e do projeto HOBE - Benchmarking Hospitais Portugueses (PTDC/EGE-GES/112232/2009), financiou este trabalho.

REFERÊNCIAS

- Banco Central Europeu, Comissão Europeia, Fundo Monetário Internacional, República Portuguesa e Ministério das Finanças e da Administração Pública. 2011. Portugal: Memorando de Entendimento sobre as Condicionalidades de Política Económica (tradução de 17 de Maio de 2011). Disponível em: infoeuropa.eu/rocid.pt/registo/000046743/
- Banker, R.D., A. Charnes e W.W. Cooper. 1984. Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Management Science* 30(9): 1078-1092.
- Barros, P.P. e J.d.A. Simões. 2007. Portugal: Health System Review. *Health Systems in Transition*. Vol. 9(5): 1-140.
- Camp, R.C. 1989. The Search for Industry Best Practices that Lead to Superior Performance. Milwaukee, Wisconsin: ASQC Quality Press.
- Castro, R.A.S. 2011. *Benchmarking de Hospitais Portugueses - Modelação com Data Envelopment Analysis* Dissertação de Mestrado, Faculdade de Engenharia, Universidade do Porto, Porto.
- Charnes, A., W.W. Cooper e E. Rhodes. 1978. Measuring Efficiency of Decision-Making Units. *European Journal of Operational Research* 2(6): 429-444.
- Chilingerian, J.A. e H.D. Sherman. 2004. Health Care Applications: From Hospitals to Physicians, From Productive Efficiency to Quality Frontiers. em *Handbook on Data Envelopment Analysis*, editado por Cooper, W. W., L. M. Seiford e J. Zhu. Boston / Dordrecht / London: Kluwer Academic Publisher.
- Cooper, W.W., L.M. Seiford e K. Tone. 2007. *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*. (2ª ed.). Nova Iorque: Springer Science and Business Media, LLC.
- Costa, C. e S. Lopes. 2007. *Avaliação do Desempenho dos dos Hospitais Públicos em Portugal Continental - 2005*. Escola Nacional de Saúde Pública - Universidade Nova de Lisboa.
- Dismuke, C. e V. Sena. 2001. Is there a Trade-Off between Quality and Productivity? The Case of Diagnostic Technologies in Portugal. *Annals of Operations Research* 107(1): 101-116.
- Donabedian, A. 1980. *The definition of quality and approaches to its assessment*. Vol. 1. Ann Arbor, Michigan: Health Administration Press.
- Fried, H.O., C.A.K. Lovell e S.S. Schmidt. 2008. Efficiency and Productivity. em *The Measurement of Productive Efficiency and Productivity Growth*, editado por Fried, H. O., C. A. K. Lovell e S. S. Schmidt. New York: Oxford University Press.
- Gonçalves, L.F.M. 2008. Análise da eficiência dos hospitais SA e SPA segundo uma abordagem de fronteira de eficiência. Tese de Doutoramento, ISCTE, Lisboa.
- Hollingsworth, B. 2008. The measurement of efficiency and productivity of health care delivery. *Health Economics* 17(10): 1107-1128.
- Hollingsworth, B., P.J. Dawson e N. Maniadakakis. 1999. Efficiency measurement of health care: a review of non-parametric methods and applications. *Health Care Manag Sci* 2(3): 161-172.
- Kontodimopoulos, N. e D. Niakas. 2005. Efficiency measurement of hemodialysis units in Greece with data envelopment analysis. *Health Policy* 71(2): 195-204.
- Koopmans, T.C. 1951. *Activity Analysis of Production and Allocation*. Editado por Sons, J. W. a. New York.

Marshall, M., S. Leatherman e S. Mattke. 2004. Selecting Indicators for the Quality of Health Promotion, Prevention and Primary Care at the Health Systems Level in OECD Countries. *OECD Health Technical Papers*. OECD Publishing. No. 16.

Moreira, S. 2008. Análise da Eficiência dos Hospitais-Empresa: Uma Aplicação da "Data Envelopment Analysis". *Boletim Econômico*, editado por Banco de Portugal. Vol. Primavera.

Nayar, P. e Y.A. Ozcan. 2008. Data Envelopment Analysis comparison of hospital efficiency and quality. *Journal of Medical Systems* 32(3): 193-199.

O'Neill, L., M. Rauner, K. Heidenberger e M. Kraus. 2008. A cross-national comparison and taxonomy of DEA-based hospital efficiency studies. *Socio-Economic Planning Sciences* 42(3): 158-189.

Oliveira, M., S.N.d. Silva, J. Pereira, J. Simões, P.P. Barros, A.S. Ferreira, P.S. Ferreira, A.P. Hespagnol e M. Teixeira. 2007. *A Sustentabilidade Financeira do Serviço Nacional de Saúde*. editado por Ministério da Saúde - Secretaria Geral

Pordata. 2011. *Pordata - Base de Dados Portugal Contemporâneo*. Fundação Francisco Manuel dos Santos [último acesso a 25 de Fevereiro de 2011]. Disponível em: www.pordata.pt.

Puig-Junoy, J. 1998. Technical efficiency in the clinical management of critically ill patients. *Health Economics* 7(3): 263-277.

Simões, J. 2004. Retrato Político da Saúde - Dependência do Percurso e Inovação em Saúde: da Ideologia ao Desempenho. Livraria Almedina.

Thanassoulis, E., M.C. Portela e R. Allen. 2004. Incorporating Value Judgements in DEA. em *Handbook on Data Envelopment Analysis*, editado por Cooper, W. W., L. M. Seiford e J. Zhu: Kluwer Academic Publishers.

Thompson, R.G., F.D. Singleton, R.M. Thrall e B.A. Smith. 1986. Comparative site evaluations for locating a high-energy Physics lab in Texas. *Interfaces* 16(6): 35-49.

Triantafillou, P. 2007. Benchmarking in the public sector: A critical conceptual framework. *Public Administration* 85(3): 829-846.

Tribunal de Contas. 2006. Relatório Global de Avaliação do Modelo de Gestão dos Hospitais do SEE – Período de 2001-2004. *Relatório n.º 20/06 – Audit*. Vol. I-II.

Voss, C.A., P. Åhlström e K. Blackmon. 1997. Benchmarking and operational performance: some empirical results. *Benchmarking for Quality Management & Technology* 4(4): 273-285.

CASO 21

AVALIAÇÃO E ACOMPANHAMENTO DE PROJETOS DE I&D NUMA EMPRESA DE TELECOMUNICAÇÕES

Joana Fialho

Área Científica de Matemática, Escola Superior de Tecnologia e Gestão de Viseu
Instituto de Engenharia de Sistemas e Computadores de Coimbra, INESC-C
jfialho@estv.ipv.pt

Pedro Godinho

Faculdade de Economia da Universidade de Coimbra
Grupo de Estudos Monetários e Financeiros, GEMF
pgodinho@fe.uc.pt

João Paulo Costa

Faculdade de Economia da Universidade de Coimbra
Instituto de Engenharia de Sistemas e Computadores de Coimbra, INESC-C,
jpaulo@fe.uc.pt

Ricardo Afonso

PT Inovação
ricardo@ptinovacao.pt

RESUMO

Neste capítulo, apresenta-se uma solução para a avaliação e seleção de projetos de Investigação e Desenvolvimento (I&D) de uma empresa de telecomunicações (a PT Inovação), bem como o processo que a ela conduziu. A empresa dedica-se, por excelência, à inovação, investigação e desenvolvimento, pelo que uma correta alocação de recursos, baseada num modelo de avaliação das diferentes propostas com que a empresa se depara, é fundamental para que possa prosperar. No entanto, a avaliação deste tipo de projetos apresenta dificuldades específicas de índole diversa, e deve também ser devidamente enquadrada na política da empresa, pelo que se entendeu ser desejável definir uma solução de raiz, em vez de se optar por modelos predefinidos. A solução resulta de um processo iterativo que envolveu, por um lado, um conjunto de investigadores das áreas das finanças e investigação operacional e, por outro, vários elementos da empresa. A solução tem por base a avaliação comparativa das propostas com que a empresa se depara, e parte daí para a alocação de recursos entre os projetos selecionados. Os recursos mais importantes desta empresa são as pessoas, e a solução definida permite orientar os decisores relativamente à alocação das pessoas aos projetos.

PALAVRAS-CHAVE

Análise de Decisão Multicritério, Avaliação de projetos de I&D, AHP

1. Introdução

Os projetos de Inovação, Investigação e Desenvolvimento promovem o crescimento e o sucesso de uma empresa. É, pois, fundamental avaliar corretamente estes projetos para que a empresa possa estabelecer quais os que deve realizar e qual a quantidade de recursos a utilizar. Para avaliar projetos de investigação e desenvolvimento (I&D), são necessárias ferramentas que tenham em conta o valor do projeto resultante da flexibilidade operacional que lhe é intrínseca, e a totalidade das fontes de incerteza a que pode estar sujeito. A avaliação depende da informação disponibilizada e, quando não existe informação quantitativa rigorosa, pode recorrer-se a fatores qualitativos, muitas vezes fornecidos por especialistas e gestores (Carlsson e Outros, 2007). A avaliação de projetos de I&D depende, também, da informação que se vai obtendo ao longo do tempo de vida do projeto. Quando a informação é quantitativa, de natureza económica e rigorosa, podem aplicar-se métodos económicos, tais como a teoria das opções reais ou outros mecanismos que fornecem resultados mais rigorosos, já que existe informação financeira suficiente para tal. No caso da informação disponível ser qualitativa ou inferida a partir de fatores qualitativos, é necessário aplicar outro tipo de métodos que consiga explorar e avaliar os projetos, partindo desse tipo de informação. Pode também acontecer que a avaliação dos projetos seja feita recorrendo, simultaneamente, a fatores qualitativos e quantitativos.

Neste capítulo, pretende-se apresentar uma abordagem de avaliação de projetos de I&D para uma empresa de telecomunicações - a PT Inovação. Os projetos de I&D desta empresa apresentam características específicas, de índole diversa, que devem ser devidamente enquadradas numa abordagem de avaliação, pelo que se entendeu ser desejável definir, de raiz, esta abordagem em vez de se optar por outras já desenvolvidas. A abordagem resultou, assim, de um processo iterativo, que envolveu investigadores das áreas financeira e de investigação operacional, bem como elementos da própria empresa. A solução encontrada propõe a comparação dos diversos projetos disponíveis e, posteriormente, a partir dos resultados obtidos, apoia a alocação dos recursos aos projetos a empreender. Note-se que os recursos mais importantes desta empresa são pessoas com diferentes tipos e níveis de especialidade. A solução

preconizada ajuda, portanto, os decisores a definirem a alocação das pessoas aos projetos.

A abordagem desenvolvida foi inspirada no processo desenvolvido para a British Aerospace Military Aircraft and Aerostructures, apresentado em Farrukh e Outros (2000). Tal como nesse processo, as reuniões com elementos da PT Inovação foram frequentes e fundamentais. Estas reuniões serviram, entre outros, para identificar o tipo de informação existente na empresa, bem como o tipo de informação que é razoável solicitar, a curto prazo, aos responsáveis pelas ações e pelos *clusters* (unidades orgânicas da empresa definidas na secção 2). Esta informação é essencial para a construção do modelo de avaliação. As reuniões serviram ainda para definir que se deveriam usar critérios gerais e subcritérios mais específicos, e para definir estes, bem como os parâmetros que devem integrar a sua avaliação.

No processo de avaliação, foi utilizada uma estrutura hierárquica baseada no processo analítico hierárquico (AHP) (Saaty, 1980). O método AHP é muito utilizado, inclusivamente na avaliação de projetos de I&D. Por exemplo, Shin e Outros (2007) aplicam este método para avaliar projetos nucleares de I&D na Coreia. Poh e Outros (2001) também utilizam o método AHP para comparar métodos de avaliação de projetos de I&D.

No entanto, na abordagem apresentada, alguns aspetos do AHP foram modificados, de forma a poder ultrapassar algumas dificuldades, tais como o elevado número de comparações e a integração de critérios quantitativos e qualitativos. Existem outras dificuldades relacionadas com o AHP, como a alteração da ordem das alternativas, a escolha e utilização das escalas ou a escolha da forma como se deriva o vetor das prioridades. A alteração da ordem das alternativas, quando se introduz ou remove uma alternativa, é um problema do AHP que está muito debatido na literatura. Existem autores que criticam este fenómeno (Dyer, 1990a, 1990b; Holder, 1990, 1991; Stam e Silva, 2003) e outros que o legitimam (Harker e Vargas, 1987, 1990; Saaty, 1986, 1990, 1991, 1994a, 2006). Saaty (1997) defende que, usualmente, a ordem das alternativas é mantida se as alternativas são comparadas entre si, mas relativamente a um padrão ideal. Caso as alternativas sejam comparadas entre si, diretamente, sem nenhum padrão ideal definido, a alteração da ordem pode

acontecer. Quando não há padrão ideal, preservar a ordem das alternativas, quando se modificam os critérios ou alternativas, mas não os julgamentos, é muito difícil. É da responsabilidade do agente de decisão saber se se deve ou não manter a ordem, quando se alteram critérios e/ou alternativas. Saaty (1994a, 1997) afirma que a alteração da ordem não é um problema, desde que o método AHP seja corretamente aplicado. No caso específico da abordagem aqui apresentada, os autores consideraram que a alteração da ordem das alternativas não seria, em geral, um problema grave. Apesar desta opinião dos autores, foi tido em conta que a possibilidade de ocorrer uma alteração da ordem podia causar desconforto nalguns utilizadores. Assim, foi definida uma forma de utilização da solução em que esta alteração não ocorre. Esta forma de utilização consiste na avaliação individual de cada alternativa em relação a um conjunto de alternativas padrão ou, usando a terminologia associada à abordagem, alternativas “de referência”. Se os avaliadores optarem por esta forma de utilização da solução, então o aparecimento de uma nova alternativa nunca poderá causar a alteração da ordem das restantes, uma vez que o processo de avaliação de cada alternativa é independente das restantes (com exceção da normalização final das avaliações).

Apesar da existência de alguns problemas associados à utilização do método AHP, este consegue atingir o compromisso entre a usabilidade e uma boa modelação (Ishizaka e Labib, 2011). Na prática, este método é bastante utilizado na tomada de decisões, já que a sua estrutura hierárquica é simples de utilizar e bastante intuitiva. A utilização deste método permite a avaliação verbal e a análise de coerência dessa avaliação. Este é, também, um método que, facilmente, se integra num *software* ou numa folha de cálculo. Em relação à escolha da escala, manteve-se a escala proposta por Saaty, mas com a possibilidade de utilizar outros valores que não estejam contemplados nessa escala. A escala de Saaty serve de suporte ao decisor, de forma a existir um significado verbal para os valores, mas qualquer valor real positivo pode ser utilizado para comparar duas alternativas. Para a derivação do vetor das prioridades, utilizou-se o vetor próprio normalizado (com o modo distributivo) associado ao valor próprio máximo.

A estrutura deste capítulo é a seguinte: na secção 2, é feito um enquadramento da abordagem de avaliação no seio da empresa; na secção 3, é descrita a estrutura e a parte técnica da solução proposta para a avaliação; na secção 4, o processo de utilização da abordagem de avaliação é apresentado; na secção 5, é dado um exemplo simples de aplicação; na secção 6, o impacto desta abordagem de avaliação na empresa é analisado e a secção 7 conclui.

2. Enquadramento da abordagem de avaliação na empresa

Numa fase preliminar, foi necessário compreender a estrutura da empresa na área da inovação, investigação e desenvolvimento. A apreensão desta estrutura, que em seguida se relata de forma sintética, ocorreu com o auxílio de elementos da empresa.

A empresa comporta dois grandes setores: serviços e projetos. Existem serviços de engenharia, de testes e consultoria, de formação e de divulgação da inovação. Os projetos podem incluir estudos e investigação aplicada, para aquisição de conhecimento, e desenvolvimento de sistemas e serviços integrados de soluções de telecomunicações.

Os serviços e projetos dividem-se em grandes grupos com características próprias, aproximadamente definidos por áreas de conhecimento e denominados *clusters*. Cada *cluster* é, por sua vez, composto por várias ações. Uma ação é um conjunto de tarefas com elevada inter-relação e com características e objetivos específicos. Os problemas de decisão que se colocam ao nível da ação são fundamentalmente diferentes dos que se colocam ao nível do *cluster*: enquanto os *clusters* requerem um planeamento estratégico a médio e longo prazo, as ações requerem um planeamento tático a curto prazo.

Dentro de um *cluster*, podem, por vezes, identificar-se linhas de investigação diferentes, que se focam em obter serviços ou produtos específicos e independentes entre si. Justifica-se, portanto, um tratamento analítico individual de cada uma destas linhas de investigação, até porque uma mesma linha pode ser partilhada por diferentes *clusters*, por haver necessidade de incorporar no mesmo produto ou serviço conhecimentos de áreas diferentes, e para duas linhas que pertençam totalmente ao mesmo cluster, pode ser

favorável seguir uma delas e não outra. Este tipo de considerações fragiliza uma análise feita exclusivamente ao nível do *cluster*.

De forma a tratar estas linhas de investigação, foi definida uma nova entidade para as representar: o agregado. Sem existência formal na empresa, o agregado representa o que habitualmente se designa por projeto. A abordagem de avaliação desenvolvida baseia-se neste conceito que é, posteriormente, integrado nas estruturas formais existentes na empresa (as ações e os *clusters*).

Dividiram-se os agregados em dois tipos:

- “estratégico”, em que se assumem objetivos de médio ou longo prazo,
- “de negócio”, em que se considera importante a obtenção de lucros a curto prazo.

Esta diferenciação permite que cada tipo de agregado seja tratado de forma diferente, de acordo com as suas características específicas. No entanto, ambos os tipos de agregado são planeados, em geral, a médio prazo (4 ou 5 anos).

As ações têm como objetivo adquirir conhecimento, desenvolver e/ou integrar serviços ou soluções, prestar serviços de engenharia, entre outros. Foram identificados vários tipos de ações:

- Ações de investigação exploratória (tipo 1), que se destinam a explorar e dominar tecnologias, mesmo que não venham a ser imediatamente aplicadas. Este tipo de ação é, pois, destinada ao desenvolvimento e/ou obtenção de novas competências.

- Ações de desenvolvimento experimental (tipo 2), que têm uma aplicação bem definida, estão integradas numa linha de investigação definida, e que são relevantes para o avanço dessa mesma linha. A par de desenvolvimento, pode haver também a aquisição de competências.

- Ações de desenvolvimento de produto para venda imediata e serviços de engenharia (tipo 3), que visam desenvolver um produto ou serviço (para venda imediata) ou que visam fornecer apoio à utilização de um produto já desenvolvido. Estas ações têm uma forte componente comercial, ainda que continuem a incluir o desenvolvimento ou a manutenção de um produto ou serviço. Neste desenvolvimento, pode, por vezes, não haver um comprador definido, mas o objetivo será gerar interesse imediato e clientes.

- Ações de gestão (tipo 4), que servem a estrutura organizacional da empresa, tais como ações ligadas à contabilidade ou ao planejamento. Devido à sua natureza, estas ações não foram integradas no processo de avaliação.

A diferenciação dos tipos de ações tornou-se essencial, pois cada tipo corresponde à prossecução de objetivos diferentes e a exigências diferentes de informação: por exemplo, uma ação do tipo 3 requer mais informação financeira do que uma ação dos tipos 1 ou 2. Na sequência de contactos com elementos da empresa, foi ainda estabelecido que se podem criar novos tipos de ações se vierem a surgir, no futuro, algumas ações que não se enquadrem em nenhum dos tipos definidos atrás.

No que se refere ao processo de decisão organizacional, a empresa possui diversos pontos de tomada de decisão: decisões tomadas a nível de direção/administração da empresa; decisões tomadas pelos responsáveis de projeto (a nível do agregado); e decisões tomadas pelos responsáveis pelas ações. No entanto, foram apenas identificados dois níveis de decisão ou escolha: o do agregado e o da ação. O ponto de decisão associado à direção/administração da empresa foi considerado através da parametrização de alguns critérios e ainda através da definição de agregados e ações típicas vistas como 'de referência'.

A abordagem de avaliação permite separar os dois níveis de decisão, sem esquecer a sua inter-relação. As decisões sobre a atribuição de recursos são separadas entre decisões estratégicas, quando da alocação dos recursos ao nível dos projetos (agregados), e decisões táticas, quando da alocação de recursos pelas ações que compõem cada agregado. Em ambos os níveis, foi definida uma estrutura hierárquica que é usada, em primeiro lugar, para definir e estruturar os critérios relevantes, e determinar as suas importâncias relativas. Estas importâncias ou pesos devem refletir as preferências institucionais, sendo assim, definidos por representantes da administração da empresa. A avaliação dos agregados e das ações nos diferentes critérios é feita pelos respetivos responsáveis. A partir destes dados e da importância dos critérios e subcritérios, obtém-se uma avaliação global do agregado ou da ação. A avaliação dos agregados serve de base a decisões gerais sobre alocação de recursos, enquanto a avaliação das ações serve de base a decisões sobre alocação de

recursos dentro do respetivo agregado. A estrutura hierárquica considerada para as ações tem em conta a importância de cada ação no agregado correspondente, permitindo, assim, considerar a interdependência entre os dois níveis de avaliação.

Nas diferentes reuniões realizadas, constatou-se que a flexibilidade operacional se reveste de diferentes formas nos dois níveis de avaliação considerados. A flexibilidade operacional pode ser vista como a capacidade que a empresa tem para proceder a correções ou adaptações no decurso de um projeto, quando ocorrem acontecimentos que não eram expectáveis no cenário inicialmente considerado. A flexibilidade operacional adiciona valor ao projeto, maximizando os ganhos, se os desenvolvimentos ocorridos forem positivos, e minimizando as perdas, se esses desenvolvimentos forem negativos. É, pois, importante integrar a flexibilidade operacional na avaliação, de forma a incorporar corretamente o acréscimo de valor que a empresa terá, se responder de forma correta às situações que poderão ocorrer. As perspetivas relevantes sobre flexibilidade operacional diferem de acordo com os níveis de avaliação considerados.

A incorporação da flexibilidade operacional na avaliação dos agregados foi realizada através de subcritérios dos critérios operacionais e financeiros: flexibilidade da solução, dependência de terceiros, criticidade/escassez de recursos necessários, perda esperada por abandono e possibilidade de adiamento. A flexibilidade da solução refere-se à possibilidade de a solução adotada poder ser adaptada para utilizações diferentes das que estavam inicialmente previstas. A dependência de terceiros pretende penalizar agregados, cuja resposta a acontecimentos imprevistos, por parte da empresa, seja reduzida, por esta não controlar na totalidade o processo de desenvolvimento. Apesar de se valorizar a relação com terceiros, é necessário ter em conta que um maior controlo do processo de desenvolvimento por parte da empresa leva a uma maior capacidade de resposta a acontecimentos imprevistos. A criticidade/escassez de recursos necessários considera a contribuição do agregado para a diminuição da flexibilidade da empresa como um todo. A utilização de recursos escassos, por parte de um agregado, pode inviabilizar o desenvolvimento de outros projetos. Estes recursos podem referir-

se, quer a pessoal muito especializado, quer a equipamentos muito específicos. A perda esperada por abandono considera a probabilidade de abandono e a perda que a empresa incorre no caso de abandono do projeto. Quanto maior a perda por abandono, mais irreversível será o projeto, e, portanto, menor será a flexibilidade operacional existente no mesmo. A probabilidade de abandono reflete a possibilidade de o projeto correr tão mal que valha a pena ser abandonado. A possibilidade de adiamento favorece o desenvolvimento rápido dos agregados que possam ser afetados por atrasos, podendo levar ao adiamento na alocação de recursos a agregados, cujo desenvolvimento possa sofrer menos com tal atraso. O adiamento pode, contudo, ter consequências positivas, como sendo a obtenção de mais informação, evitando o gasto desnecessário de recursos e permitindo o avanço com mais certeza de sucesso, ou o não avanço, em caso de ambiente desfavorável.

Estes subcritérios, espelhando as fontes de flexibilidade, foram os identificados nas reuniões com elementos da empresa. A abordagem permite, no entanto, se necessário, acrescentar outros subcritérios relacionados com outras fontes de flexibilidade. A importância da flexibilidade operacional difere consoante o tipo de agregado em consideração. Por exemplo, num agregado com características de negócio, a possibilidade de ajustamento rápido a uma situação imprevista é, em geral, mais importante do que num agregado com características mais estratégicas (as relações com terceiros, questões tecnológicas ou competências adquiridas serão mais importantes). Esta diferença de importância é tida em conta durante o processo de avaliação.

Para incorporar a flexibilidade operacional na avaliação das ações, os subcritérios identificados foram a dependência (operacional) de terceiros, criticidade/escassez de recursos necessários, perda esperada por abandono e a possibilidade de adiamento. A dependência (operacional) de terceiros é muito semelhante ao subcritério “dependência de terceiros” nos agregados. No entanto, para as ações, esta dependência é puramente operacional, já que no caso dos agregados, a dependência poderá ser “política”, na medida em que o processo de decisão possa não ser exclusivo da PT Inovação. A criticidade/escassez de recursos necessários também é semelhante ao subcritério com o mesmo nome nos agregados, sendo que nas ações os

recursos referem-se aos existentes dentro do agregado a que ação pertence. A perda esperada por abandono e a possibilidade de adiamento têm também uma definição semelhante à que foi feita para os agregados.

Repare-se que o subcritério flexibilidade de solução não se encontra nas ações, pois este subcritério refere-se a projetos globais e não tanto a tarefas mais específicas como aquelas que definem, em geral, as ações. Se forem identificadas outras fontes de flexibilidade, estas podem ser incorporadas através de novos subcritérios, como foi sugerido no caso dos agregados.

Também para as ações, a importância das fontes de flexibilidade é diferente mediante o tipo de ação considerada. Por exemplo, numa ação de desenvolvimento experimental é mais importante saber reagir a acontecimentos imprevistos do que numa ação de investigação exploratória. O modelo de avaliação incorpora estas diferenças de importância nos diversos subcritérios.

Durante o tempo de vida de um projeto, novas oportunidades de ações podem ocorrer ou pedidos inesperados de clientes podem acontecer. Estes acontecimentos não previstos e não contemplados na avaliação inicial requerem uma reafetação dos recursos dentro do agregado, e conseqüentemente, nas ações que o integram, para dar seguimento às oportunidades ou solicitações que podem aparecer. O modelo de avaliação apresentado permite a integração destas situações, através da sua aplicação intercalar dentro dos agregados. Assim, deve proceder-se à avaliação das novas solicitações ou oportunidades como novas ações; reavaliam-se as ações que mostrem diferenças significativas relativamente ao plano inicial; utilizam-se estas avaliações e reavaliações para decidir se se deve ou não empreender as novas ações, e caso esta decisão seja positiva, deve decidir-se sobre os recursos a alocar a estas novas ações e onde devem ser retirados recursos.

Assim, é possível reagir a novas situações a partir do modelo, de forma transparente, e usando a mesma estrutura de decisão utilizada inicialmente.

Após a avaliação dos agregados e a avaliação das ações que pertencem a cada agregado, é necessária a utilização de ferramentas que ajudem na alocação dos recursos de cada agregado pelas diferentes ações que o compõem. Pretende-se, pois, encontrar a melhor alocação de recursos que permita responder às prioridades que vão sendo criadas. Os recursos mais

importantes na empresa são as pessoas, pelo que a avaliação apresentada pretende estabelecer a melhor alocação possível das pessoas às ações e agregados.

A avaliação das ações permite obter um valor de desempenho ou prioridade para cada ação. Este desempenho é então utilizado para determinar os recursos a atribuir a cada ação de entre os que foram atribuídos ao agregado a que as ações pertencem. Esta atribuição pode ser feita usando um de dois procedimentos:

1. Definição de prioridades (classificações) globais para cada ação, e seleção de ações por ordem de prioridade. Neste processo de seleção, podem existir restrições como níveis máximos e/ou mínimos de alocação.

2. Utilização de programação matemática por forma a maximizar a soma das prioridades das ações selecionadas, tendo em conta restrições de recursos e outras consideradas relevantes, como limites de utilização de recursos em cada tipo de ação, mínimos para a pontuação agregada em alguns critérios ou subcritérios, etc.

Estes procedimentos não substituem o papel e o julgamento dos decisores, mas antes completam e ajudam na tomada de decisões, já que tais procedimentos dão indicações valiosas, não descurando da sensibilidade dos decisores para eventuais ajustes na construção de um portfólio equilibrado de ações.

3. Estrutura e desenvolvimento da solução proposta

Nesta secção, descreve-se a estrutura da solução proposta, bem como as técnicas utilizadas para proceder à avaliação dos agregados e das ações. A avaliação dos agregados e a avaliação das ações baseiam-se em estruturas hierárquicas, e, a partir destas estruturas, são calculados os pesos dos critérios e subcritérios. O cálculo dos pesos dos critérios e subcritérios é feito com base no método AHP. A avaliação dos agregados e das ações em cada critério e subcritério também é feita com base no método AHP. No entanto, quando se consideram valores financeiros, a avaliação dos agregados e das ações pressupõe um tratamento especial, de forma a encontrar uma avaliação comparável à que se obtém quando se aplica o método AHP em critérios e

subcritérios não financeiros, e também para garantir um tratamento uniforme dos valores financeiros em todas as ações e agregados. Para evitar um grande número de comparações, quando se infere o peso dos critérios e dos subcritérios ou as prioridades dos agregados ou ações, são utilizadas matrizes incompletas de comparações (Harker, 1987).

Após a determinação das prioridades dos agregados e das ações de cada agregado, é necessário encontrar ferramentas que ajudem na alocação dos recursos a cada ação de entre os recursos atribuídos ao respectivo agregado.

3.1 A estrutura da solução proposta

A avaliação dos agregados resulta na atribuição de uma prioridade a cada um, sendo que uma maior prioridade equivale a uma maior atratividade do agregado, e portanto, maior será o grau de satisfação com a alocação de recursos. A determinação destas prioridades baseia-se numa estrutura hierárquica.

Tabela 1 - Critérios e subcritérios na hierarquia dos agregados

Critérios	Estratégico	Operacional	Financeiro
Subcritérios	<ul style="list-style-type: none"> - Contribuição para a imagem da empresa - Liderança do mercado - Competências adquiridas - Parcerias estratégicas - Importância da credibilidade da empresa junto do cliente - Importância da tecnologia desenvolvida para o grupo PT 	<ul style="list-style-type: none"> - Incerteza técnica - Criticidade/escassez dos recursos necessários - Flexibilidade da solução - Dependência de terceiros - Satisfação imediata do cliente 	<ul style="list-style-type: none"> - Valor financeiro (valor atual líquido de médio/longo prazo, ou valor atual líquido a um ano + perspectivas de crescimento + tendência de mercado) - Perda esperada por abandono - Possibilidade de adiamento

A estrutura hierárquica para a avaliação dos agregados é a seguinte: no primeiro nível estão os tipos de agregado. A classificação utilizada, como descrito anteriormente, divide os agregados em “tipo estratégico”, em que se assumem objetivos de médio ou longo prazo, e “tipo negócio”, em que se considera importante a obtenção de lucros a curto prazo. Esta diferenciação permite que cada tipo de agregado seja tratado de forma diferente, de acordo com as suas características específicas.

No segundo nível, encontram-se os critérios. Nas reuniões realizadas, obtiveram-se três critérios: estratégico, operacional e financeiro. No terceiro nível da hierarquia, encontram-se os subcritérios, que estão definidos na Tabela 1. Alguns destes subcritérios correspondem aos critérios que refletem a flexibilidade operacional. Finalmente, no último nível da hierarquia, encontram-se os agregados.

A avaliação das ações baseia-se num processo semelhante ao dos agregados, resultando na atribuição de uma prioridade a cada uma. Os valores destas prioridades serão utilizados para determinar os recursos atribuídos a cada ação, de entre os recursos que foram atribuídos ao agregado a que a ação pertence.

Tabela 2 - Critérios e subcritérios na hierarquia das ações

Critérios	Estratégico	Operacional	Financeiro
Subcritérios	<ul style="list-style-type: none"> - Contribuição para a imagem da empresa - Liderança do mercado - Competências adquiridas - Parcerias estratégicas - Importância da satisfação do cliente - Importância da tecnologia desenvolvida para o grupo PT 	<ul style="list-style-type: none"> - Incerteza técnica - Criticidade/escassez dos recursos necessários - Dependência (operacional) de terceiros - Importância da ação para o agregado 	<ul style="list-style-type: none"> - Valor financeiro (valor atual líquido a um ano) - Perspetivas de crescimento - Tendência de mercado - Perda esperada por abandono - Possibilidade de adiamento

A determinação destas prioridades baseia-se também numa estrutura hierárquica, onde no primeiro nível estão os tipos de ação. Relembre-se que, na avaliação, não constam as ações do tipo 4. Esta diferenciação permite que cada tipo de ação seja tratado de forma diferente, de acordo com as suas características específicas. No segundo nível da hierarquia, encontram-se os critérios, cuja identificação é a mesma do que no caso dos agregados: estratégico, operacional e financeiro. No terceiro nível da hierarquia, encontram-se os subcritérios, que estão identificados na Tabela 2. Como no caso dos agregados, alguns destes subcritérios correspondem aos critérios identificados que refletem a flexibilidade operacional. Finalmente, no último nível da hierarquia, encontram-se as ações.

Após a determinação dos pesos dos critérios e dos subcritérios, cada agregado (respetivamente ação) é avaliado em cada subcritério. Para os subcritérios qualitativos ou intangíveis, a determinação das prioridades dos agregados/ações, em cada subcritério, é feita de forma semelhante à utilizada para inferir os pesos dos critérios e os pesos dos subcritérios, que é explicada, em detalhe, na secção 3.2. A avaliação dos agregados/ações, nos subcritérios tangíveis, que correspondem aos subcritérios financeiros, é explicada na secção 3.3, sendo que o resultado dessa avaliação corresponde, como nos restantes subcritérios, a prioridades. A prioridade global, G , dos agregados/ações corresponde à agregação das prioridades dos elementos, em cada subcritério, com os pesos dos critérios e com os pesos dos subcritérios. Isto é, a prioridade global de uma ação/agregado de índice k é:

$$G_k = \sum_{j=1}^{ncrit} W_j \left(\sum_{i=1}^{ns_j} (W_{ij} W_{ijk}) \right)$$

onde:

$ncrit$ é o número de critérios presentes no nível 2 da hierarquia;

W_j é o peso do critério de índice j do nível 2 da hierarquia,
 $j = 1, \dots, ncrit$;

ns_j é o número de subcritérios do nível 3 da hierarquia referente ao critério de índice j , $j = 1, \dots, ncrit$;

W_{ij} é o peso do subcritério de índice i do nível 3 da hierarquia pertencente ao critério de índice j , $i = 1, \dots, ns_j$, $j = 1, \dots, n_{crit}$;

W_{ijk} é a prioridade do agregado/ação, de índice k , em relação ao subcritério de índice i do nível 3 da hierarquia que pertence ao critério de índice j , $i = 1, \dots, ns_j$, $j = 1, \dots, n_{crit}$.

Note-se que esta agregação é aditiva e a sua utilização pressupõe a independência preferencial dos critérios e subcritérios, ou seja, é necessário que as preferências num critério ou subcritério, ou num grupo de critérios ou subcritérios, não sejam afetadas pelas preferências noutros critérios ou subcritérios. A independência preferencial dos critérios é garantida pela forma como foi definida a estrutura hierárquica e os seus elementos: ao incluir, num primeiro nível, o tipo de agregado e ação, foi possível garantir esta independência. Note-se que, sem incluir este primeiro nível era possível que ocorressem violações da independência preferencial: em agregados com uma prioridade elevada no critério estratégico, uma pontuação alta no critério operacional e baixa financeiro podia ser preferível a uma pontuação baixa no primeiro critério e alta no segundo; no entanto, se a prioridade do critério estratégico for baixa, estas preferências poderiam inverter-se. Ao definir o tipo de agregado, que indica qual a vocação prioritária de cada agregado, esta possível dependência desaparece. Um raciocínio semelhante podia também ser feito para as ações. Por outro lado, para os subcritérios, houve o cuidado de os definir por forma a que dois subcritérios meçam características diferentes que não apresentem dependência entre si.

3.2 Determinação dos pesos dos critérios em cada nível das hierarquias

Para proceder à avaliação, é necessário, numa primeira fase, definir os pesos de cada critério e subcritério. A determinação dos pesos é feita com base no método AHP. Para tal, o responsável de decisão em cada nível deve fazer comparações entre os critérios. No procedimento apresentado, esse responsável precisa apenas de fazer algumas comparações, evitando, assim, requerer-lhe demasiada informação. Este procedimento é baseado na inferência dos pesos dos critérios, a partir de uma matriz de comparações incompleta

(Harker, 1987): o responsável preenche, então, parte de uma matriz de comparações, $A = [a_{ij}]_{i,j=1,\dots,n}$, onde n é o número de critérios (subcritérios) e a_{ij} representa a comparação entre o critério i e o critério j .

Para uma matriz completa de comparações, é usual utilizar o modelo AHP original de Saaty, na determinação do peso dos critérios. Este método também analisa a coerência das comparações feitas, isto é, uma matriz completada por um agente de decisão pode conter alguma incoerência devida à possível existência de intransitividade entre as comparações. Saaty (1994b) propôs a medição da incoerência, através do índice de coerência (CI), definido por

$$CI = \frac{\lambda_{\max} - n}{n - 1}$$

onde λ_{\max} representa o valor próprio de maior norma da matriz de comparações. A interpretação do índice de coerência depende do tamanho da matriz (n). Saaty (1994b) propôs dividir o índice de coerência pela média de índices de coerência de matrizes de comparações geradas aleatoriamente. Este valor médio é designado por índice de coerência aleatório (RI) e o quociente entre CI e RI é designado como rácio de coerência (CR) e é interpretado como a percentagem de incoerência na matriz de comparações (Godinho e Outros, 2011). Em geral, considera-se que a incoerência é aceitável para um rácio de coerência de valor máximo de 10%.

Ora, no caso de uma matriz incompleta, existem duas possibilidades no seu preenchimento: preencher $n-1$ comparações ou mais do que $n-1$ comparações, mantendo a matriz incompleta.

Para inferir, de forma coerente, todos os restantes valores da matriz de comparações, a partir de $n-1$ comparações, é suficiente que essas $n-1$ comparações estejam em colunas diferentes da matriz, e apenas num dos triângulos. Note-se que, por exemplo, numa matriz de comparações em que $a_{jk} = r$ e $a_{jh} = s$, então, temos que $a_{kh} = s / r$ e $a_{hk} = r / s$. Neste sentido, é possível inferir todos os valores da matriz de comparações de forma coerente. Numa matriz de comparações totalmente coerente, o valor próprio máximo é n , e o seu vetor próprio associado normalizado corresponde às prioridades das alternativas em comparação. Seja ele $W = [W_1 \ W_2 \ \dots \ W_n]$. Numa matriz de

comparações totalmente coerente, $a_{ij} = W_i / W_j$, $i=1, \dots, n$ e $j=1, \dots, n$, e o índice de coerência é, naturalmente, 0 (Kablan, 2004).

Quando existem mais do que $n-1$ comparações, pode acontecer que a matriz de comparações não seja totalmente coerente, mesmo se for incompleta. Para inferir prioridades a partir de matrizes incompletas, foi utilizado o método apresentado por Harker (1987). O método começa por criar uma matriz nova F , baseada na matriz incompleta de comparações A . A matriz F é definida da seguinte forma:

$$f_{ij} = \begin{cases} a_{ij}, & \text{se } a_{ij} > 0 \text{ e } i \neq j \\ \text{número de células vazias da linha } i, & \text{se } i = j \\ 0, & \text{caso contrário} \end{cases}$$

O vetor próprio normalizado associado ao valor próprio máximo da matriz $F+I$, onde I representa a matriz identidade, é uma aproximação do vetor das prioridades das alternativas. Com este vetor de prioridades, pode calcular-se o rácio de coerência e perceber se se devem ou não rever julgamentos. Se a matriz for suficientemente coerente, pode utilizar-se o vetor de prioridades assim calculado.

3.3 Tratamento de subcritérios financeiros

Para os subcritérios financeiros, como sendo o VAL ou outra perspectiva financeira, considerou-se que a forma como os valores financeiros são convertidos para prioridades devia refletir as preferências organizacionais, e não ser feito de forma casuística consoante as ações ou agregados a serem avaliados. A solução encontrada passou por definir valores financeiros de referência, com o apoio da gestão de topo da empresa. A comparação entre estes valores financeiros de referência é, também, feita pela gestão de topo da empresa.

O intervalo que contém os valores financeiros de referência deve ter uma amplitude, tal que, qualquer previsão para valores financeiros de um agregado ou ação deva estar no interior desse intervalo. Assume-se também que as preferências são lineares para todos os valores pertencentes ao intervalo entre dois valores de referência consecutivos. Esta hipótese permite construir uma

matriz ampliada de comparações com os valores financeiros de referência e com os valores financeiros para ações ou agregados. A partir dessa matriz de comparações, que pode ser incompleta, e aplicando o mesmo procedimento da secção anterior, é possível inferir as prioridades de todos os valores. Retirando apenas as prioridades dos valores financeiros das ações ou agregados, normalizam-se essas prioridades e obtém-se um vetor de prioridades para os valores financeiros das ações ou agregados em avaliação. Estas prioridades normalizadas podem ser utilizadas de forma idêntica às prioridades que se obtém nos outros subcritérios não tangíveis.

Apresenta-se agora um exemplo simples. Quer este exemplo quer a sua apresentação seguem de perto Godinho e Outros (2011). Suponha-se que os valores de referência são -100, 0 e +100. A matriz de comparações é a seguinte:

$$\begin{array}{c}
 \begin{array}{ccc}
 -100 & 0 & 100 \\
 -100 & \left[\begin{array}{ccc}
 1 & 1/3 & 1/6 \\
 3 & 1 & 1/2 \\
 6 & 2 & 1
 \end{array} \right] \\
 0 \\
 100
 \end{array}
 \end{array}$$

Esta matriz de comparações leva às seguintes prioridades para os valores de referência: $W = [0.10 \ 0.30 \ 0.60]$, respetivamente para os valores -100, 0 e 100. Introduzam-se, agora, dois valores financeiros, -25 e 50. Assim, tendo em conta as prioridades dos valores de referência e assumindo a linearidade entre as preferências dos valores de referência -100 e 0, tem-se que, como $-25 = -100 + 0.75(0 - (-100))$, então a prioridade do valor de -25 é $0.1 + 0.75(0.30 - 0.10) = 0.25$. Analogamente, pode determinar-se a prioridade do valor 50. Ora, $50 = 0 + 0.50(100 - 0)$. Assumindo a linearidade nas preferências dos valores de referência, a prioridade do valor 50 é $0.30 + 0.50(0.60 - 0.30) = 0.45$.

Com esta informação, pode construir-se uma matriz ampliada de comparações que inclui os valores de referência e os valores financeiros introduzidos. A partir das prioridades encontradas, podem determinar-se as comparações par a par entre todos os valores de referência e cada um dos valores financeiros introduzidos. Por exemplo, a comparação entre o valor -100 e o valor -25 será o quociente entre a prioridade de -100 e a prioridade encontrada de -25. Isto é, a comparação entre -100 e -25 será $0.1/0.25$. O

mesmo raciocínio se aplica nas comparações entre os valores de referência e cada um dos novos valores introduzidos (-25 e 50). Note-se que muito dos valores resultantes das comparações não pertencem à escala definida pelo método AHP (inteiros positivos de 1 a 9 e seus inversos). No entanto, de acordo com Saaty (1997), a escala por ele definida é apenas indicativa e serve para ajudar os agentes de decisão a estabelecerem comparações. Mas se outros valores entrarem na matriz de comparações, o procedimento para encontrar as prioridades das alternativas mantém-se. A matriz ampliada de comparações será, então

	-100	-25	0	50	100
-100	1	$0.1/0.25$	$1/3$	$0.1/0.45$	$1/6$
-25	$0.25/0.1$	1	$0.25/0.3$	-	$0.25/0.6$
0	3	$0.3/0.25$	1	$0.3/0.45$	$1/2$
50	$0.45/0.1$	-	$0.45/0.3$	1	$0.45/0.6$
100	6	$0.6/0.25$	2	$0.6/0.45$	1

Esta matriz de comparações não está completa, pois não foram introduzidas as comparações entre os valores financeiros introduzidos -25 e 50. Aplicando o procedimento apresentado por Harker (1987), para matrizes incompletas de comparações, é possível inferir as prioridades de cada um dos valores.

A partir da metodologia apresentada por Harker (1987), obtém-se o seguinte vetor de prioridades ou importâncias relativas: [0.059 0.147 0.176 0.265 0.353]. Como só existe interesse nas prioridades dos valores financeiros de -25 e 50, as prioridades correspondentes podem ser normalizadas, de forma a que a sua soma seja 1, levando a que a prioridade do valor de -25 seja 0.357 e a que a prioridade do valor de 50 seja de 0.643.

Neste exemplo, o vetor de prioridades apresentado acima podia ainda ser obtido por simples normalização do vetor que inclui as prioridades inicialmente determinadas para os valores de referência e para os valores

financeiros a avaliar, isto é, por simples normalização do vetor [0.10 0.25 0.30 0.45 0.60], não sendo necessário recorrer à segunda matriz de comparações. No entanto, isto ocorre apenas quando a matriz de comparações dos valores de referência apresenta perfeita coerência, isto é, quando o seu índice de coerência é nulo. No caso geral, quando não se pode assumir que a matriz de comparações dos valores de referência é perfeitamente coerente, é necessário efetuar todo o procedimento descrito.

Note-se que, por forma a definir de forma mais detalhada as preferências relativas a valores financeiros é necessário incluir um maior número de valores de referência. Na implementação efetuada na empresa, o processo foi simplificado pelo facto de os representantes deste terem dado a indicação de que se podiam assumir preferências lineares entre valores financeiros positivos. Com este pressuposto, basta incluir no conjunto de valores de referência o valor zero e um (ou nenhum) valor positivo. Se existir um valor positivo nos valores de referência, as prioridades correspondentes a este valor e ao valor zero definem uma linha de tendência que permite realizar todas as comparações que envolvem valores positivos na matriz ampliada. Se não for incluído nenhum valor positivo, então a linha de tendência usada para definir as comparações com valores positivos recorre às preferências relativas ao valor zero e ao maior valor de referência negativo.

3.4 Alocação dos recursos às ações de cada agregado

Após a avaliação dos agregados e das ações, torna-se necessário encontrar ferramentas que ajudem na alocação dos recursos às ações que compõem cada agregado. Note-se que, no caso dos agregados, a solução implementada prevê que a distribuição de recursos tenha em consideração a sua avaliação, mas não foi definida uma regra explícita de distribuição, uma vez que pode ser razoável atribuir preferencialmente recursos de tipo diferente a agregados diferentes.

As prioridades globais encontradas para cada ação, G_k , $k = 1, \dots, n$, sendo n o número de ações avaliadas, no processo de avaliação, podem servir de base à alocação de recursos. A atribuição dos recursos a cada ação pode ser feita através da seleção das ações por ordem das prioridades globais encontradas. Num tal procedimento, pode ainda impor-se que as ações de cada

tipo tenham uma quantidade mínima e uma quantidade máxima de recursos alocados. Neste procedimento, mais simples, as ações com prioridades mais altas são selecionadas e veem, em primeiro lugar, as suas necessidades satisfeitas, em termos de recursos.

A programação matemática também é uma alternativa para encontrar a melhor alocação de recursos às ações de cada agregado. Na formulação de um tal problema, a função objetivo deve conduzir à maximização da prioridade global das ações selecionadas para o agregado, e as restrições devem definir que a quantidade máxima a alocar é aquela que foi estabelecida para o agregado. Em geral, existem vários tipos de recurso que podem ser atribuídos às ações. Cada ação pode precisar de quantidades específicas de cada tipo de recursos. Neste caso, suponha-se que existem N tipos de recurso com quantidades disponíveis de $Q_j, j = 1, \dots, N$. Para cada nível de recursos j , assume-se que cada ação k necessita de uma quantidade q_{kj} para ser executada, $k = 1, \dots, n; j = 1, \dots, N$.

Neste caso, o problema a resolver é

$$\begin{aligned} & \max \sum_{k=1}^n G_k y_k \\ & \text{s.a.} \quad \sum_{k=1}^n q_{kj} y_k \leq Q_j, \quad j = 1, \dots, N \\ & \quad y_k \in \{0, 1\}, \quad k = 1, \dots, n \end{aligned}$$

sendo y_k uma variável dicotômica, que assume o valor 1 se a ação de índice k for selecionada, e 0 se não for. Num agregado, podem existir outras relações de dependência entre as ações a empreender, que podem ser representadas através de outras restrições. Por exemplo, se a ação 3 apenas puder ser empreendida se a ação 1 também o for, a restrição a incluir será $y_1 \geq y_3$; se, de entre as ações 1 e 3, só uma puder ser selecionada, a restrição será $y_1 + y_3 \leq 1$, etc.

Podem, ainda, definir-se outros tipos de restrições relevantes para cada agregado, tais como estabelecer mínimos para as prioridades em diferentes subcritérios, incluir restrições relativas a tipos de recursos especialmente

escassos (por exemplo, relativas a equipamentos, que existam em quantidades muito limitadas na empresa), incluir limites relativos às ações de cada tipo a selecionar, entre outras que possam surgir. Este tipo de restrições é, também, linear e pode, facilmente, incluir-se no(s) problema(s) acima descrito(s).

4. Utilização da solução proposta

Antes da avaliação dos agregados e das ações, é necessário definir o peso dos critérios e dos subcritérios de cada uma das hierarquias definidas, respetivamente, para os agregados e para as ações. É importante que a avaliação passe por diferentes níveis de decisão em toda a organização para que possa corresponder às necessidades da empresa. Assim, a definição dos pesos dos critérios deve ser feita pela gestão de topo, já que, para diferentes tipos de agregados e diferentes tipos de ações, a importância desses critérios é diferente. Esta diferenciação deve-se aos diferentes objetivos que os agregados e as ações possam ter (mais relacionados com a área de negócio ou mais relacionados com a área de investigação e desenvolvimento, onde é mais preponderante a aquisição de novas competências).

Em cada nível de decisão das hierarquias, os respetivos responsáveis, através do seu conhecimento sobre os interesses e motivação de cada tipo de agregado/ação, podem fazer comparações entre os critérios (subcritérios), para que, através dos métodos descritos na secção anterior, se possam obter os pesos de cada um desses critérios (subcritérios).

Após a definição dos pesos de todos os critérios e subcritérios em ambas as hierarquias, pode proceder-se à avaliação dos agregados e das ações que os compõem. Esta avaliação baseia-se na obtenção das prioridades de cada agregado/ação em cada subcritério, definida a partir de um conjunto de comparações. Com estas prioridades e com os pesos dos critérios e subcritérios, é possível determinar a prioridade global de cada um dos agregados/ações do conjunto em avaliação.

Embora a solução implementada permita o tipo de comparações comum no AHP, isto é, a comparação de cada agregado/ação com os restantes em cada subcritério, foi definida uma outra alternativa de avaliação, cuja utilização se considera mais recomendável no contexto da empresa considerada. Esta

avaliação tem por base um conjunto de agregados ou ações de referência, que são agregados ou ações fictícios, definidos pela administração, que representam os tipos de agregados e ações que usualmente são considerados na empresa. As comparações entre agregados de referência, e entre ações de referência, são definidos pela administração, permitindo a esta definir padrões que reflitam as políticas da empresa.

Quando se utilizam estes agregados (ações) de referência, é possível proceder à avaliação de cada agregado (ação) individualmente. O agregado/ação em avaliação individual é comparado com as referências nos diferentes subcritérios, a partir das metodologias apresentadas anteriormente, evitando, assim, um número elevado de comparações. O valor a reter do resultado da avaliação será o desempenho global do agregado ou ação em avaliação. Se vários agregados/ações forem avaliados individualmente em relação às referências, os valores resultantes dessas avaliações podem indicar o nível de atratividade de cada agregado/ação. Note-se que, quando a solução é utilizada desta forma, o problema de alteração da ordem das alternativas quando da introdução de uma nova alternativa, referido na Introdução, não ocorre.

Os agregados (ações) de referência podem também ser usados para avaliar simultaneamente vários agregados (ações). Neste caso, mesmo se o gestor tiver dificuldade em comparar entre si os agregados (ações) em avaliação, não necessita de o fazer – apenas necessita de comparar cada agregado/ação em avaliação com pelo menos uma referência e construir matrizes amplas e incompletas de comparações, que incluem as referências e os agregados em avaliação. Para cada subcritério, o vetor das prioridades inclui, quer as referências, quer os agregados/ações em avaliação, mas devem reter-se e normalizar-se as prioridades dos agregados/ações em avaliação. Note-se que, nesta forma de utilização da solução poderá, novamente, ocorrer o problema de alteração da ordem das alternativas quando da introdução de uma nova alternativa.

A avaliação de um conjunto de ações/agregados, usando ou não as referências, conduz a uma avaliação relativa dos seus elementos. Os valores assim obtidos serão a base para a alocação dos recursos humanos entre os

diferentes agregados e para a alocação de recursos entre as diferentes ações que compõem um agregado. Além disso, é possível, com a avaliação de um conjunto de agregados, ter uma noção sobre o nível de atratividade de cada um, e uma orientação para a atribuição de recursos a estes.

5. Exemplo de aplicação

Nesta secção, apresenta-se um exemplo de aplicação, para a avaliação de dois agregados, com base num conjunto de referências. Suponhamos que se pretendem avaliar dois agregados do tipo estratégico, Ag1 e Ag2, e que existem 3 agregados de referência, Ref1, Ref2 e Ref3. As referências foram definidas e comparadas, em todos os subcritérios, pela administração de acordo com as políticas da empresa. Como foi referido atrás, existem duas possibilidades para avaliar os dois agregados em relação às referências: em cada subcritério, compara-se, individualmente, cada agregado com as referências, utilizando duas matrizes de comparações, ou constrói-se uma matriz ampla de comparações. Esta matriz ampla de comparações deve conter já as comparações entre as referências (fornecidas aquando da sua definição) e, portanto, o responsável pela avaliação apenas tem que comparar cada agregado com, pelo menos, uma das referências, sem ter que comparar os agregados entre si. Suponha-se que a matriz de comparações das referências, no subcritério competências adquiridas relativo ao critério estratégico, é a seguinte:

	Ref1	Ref2	Ref3
Ref1	1	1/3	1/6
Ref2	3	1	1/2
Ref3	6	2	1

O agente de decisão não necessita, como foi referido, de fazer todas as comparações entre as referências e os agregados em avaliação. O responsável apenas tem que inserir, pelo menos, uma comparação, entre cada agregado em avaliação e uma das referências. Suponha-se que a matriz ampla de

comparações contém as comparações entre cada agregado em avaliação e as referências, mas não a comparação direta entre ambos os agregados:

	Ref1	Ref2	Ref3	Ag1	Ag2
Ref1	1	1/3	1/6	1/3	1/3
Ref2	3	1	1/2	1/4	3
Ref3	6	2	1	1	2
Ag1	3	4	1	1	-
Ag2	3	1/3	1/2	-	1

Aplicando a metodologia de Harker (1987), obtém-se o seguinte vetor de prioridades para este subcritério: [Ref1 Ref2 Ref3 Ag1 Ag2]=[0.06 0.175 0.297 0.351 0.116]. Como apenas se pretendem as prioridades dos agregados em avaliação, fazendo a normalização das prioridades dos agregados, vem que [Ag1 Ag2]=[0.75 0.25]. Em todos os subcritérios não tangíveis (qualitativos), adota-se esta metodologia. Em relação ao subcritério tangível (financeiro) adota-se a metodologia apresentada na subsecção 3.3. Para agregar todas as prioridades e poder calcular a prioridade global para cada agregado, suponha-se que os três critérios, estratégico, operacional e financeiros, têm os seguintes pesos, respetivamente, [0.5 0.3 0.2]. Suponha-se ainda os seguintes pesos para os subcritérios de cada critério:

Estratégico: [0.20 0.15 0.20 0.15 0.10 0.20], para os subcritérios respetivos, apresentados na Tabela 1.

Operacional: [0.20 0.15 0.25 0.20 0.20], para os subcritérios respetivos, apresentados na Tabela 1.

Financeiro: [0.40 0.30 0.30], para os subcritérios respetivos, apresentados na Tabela 1.

A metodologia apresentada para calcular as prioridades dos agregados no subcritério competências adquiridas é repetida para os restantes subcritérios não tangíveis. Para o subcritério tangível, é seguida a metodologia já apresentada (subsecção 3.3). Suponha-se, então, que as prioridades para cada um dos subcritérios são as apresentadas na Tabela 3. Agregando as prioridades

com os pesos dos subcritérios e dos critérios, temos que as prioridades globais dos agregados são, respetivamente, $G_{Ag1} = 0.513$ e $G_{Ag2} = 0.487$.

Agregados Subcritérios	Ag1	Ag2
Contribuição para a imagem da empresa	0.6	0.4
Liderança do mercado	0.5	0.5
Competências adquiridas	0.75	0.25
Parcerias estratégicas	0.4	0.6
Importância da credibilidade da empresa junto do cliente	0.7	0.3
Importância da tecnologia desenvolvida para o grupo PT	0.45	0.55
Incerteza técnica	0.4	0.6
Criticidade/escassez de recursos	0.5	0.5
Flexibilidade da solução	0.3	0.7
Dependência de terceiros	0.5	0.5
Satisfação imediata do cliente	0.6	0.4
Valor financeiro	0.3	0.7
Perda esperada por abandono	0.5	0.5
Possibilidade de adiamento	0.7	0.3

Tabela 3 – Prioridades dos agregados nos diferentes subcritérios

É necessário distribuir os recursos existentes pelos dois agregados. Suponha-se que existem três tipos de recurso, T_1 , T_2 e T_3 , representando três perfis de mão-de-obra, com competências diferenciadas. A disponibilidade dos três tipos de recursos é, respetivamente, $q_1 = 100$; $q_2 = 200$; $q_3 = 250$ (valores em homens.mês). Como foi indicado na subsecção 3.4, não foi definida uma regra explícita de distribuição de recursos pelos agregados. Assim, vamos supor que, tendo em conta avaliação global do agregado Ag1, as características deste e a

quantidade global de recursos existente, foi decidido atribuir a este 58 homens.mês do tipo T_1 , 117 homens.mês de T_2 e 146 homens.mês de T_3 , e que este agregado tem 5 ações que podem ser empreendidas, $Ac_1, Ac_2, Ac_3, Ac_4, Ac_5$. Na Tabela 4, apresentam-se as necessidades de cada ação em relação a cada tipo de recurso, bem como as prioridades que se obtiveram, depois de aplicado o processo de avaliação (em tudo semelhante ao que foi explicado para os agregados).

Ação	Quant. necessária de T_1	Quant. necessária de T_2	Quant. necessária de T_3	Prioridade da ação
Ac_1	10	20	25	0.125
Ac_2	15	30	36	0.201
Ac_3	10	12	20	0.089
Ac_4	18	36	45	0.325
Ac_5	11	20	30	0.260

Tabela 4 – Quantidades necessárias de recursos e prioridades das ações

Com estes dados, e assumindo que não se pretendem definir limites para as ações de cada tipo a implementar nem outras restrições, o programa linear a resolver para seleccionar as ações a empreender do agregado Ag1 seria o seguinte:

$$\begin{aligned} \max & 0.125y_1 + 0.201y_2 + 0.089y_3 + 0.325y_4 + 0.260y_5 \\ \text{s.a.} & 10y_1 + 15y_2 + 10y_3 + 18y_4 + 11y_5 \leq 58 \\ & 20y_1 + 30y_2 + 12y_3 + 36y_4 + 20y_5 \leq 117 \\ & 25y_1 + 36y_2 + 20y_3 + 45y_4 + 30y_5 \leq 146 \\ & y_i \in \{0,1\} \quad i = 1, \dots, 5 \end{aligned}$$

A solução ótima deste problema é $[y_1 \ y_2 \ y_3 \ y_4 \ y_5] = [1 \ 1 \ 0 \ 1 \ 1]$. Para o responsável pelo agregado Ag1, as ações a executar seriam Ac_1, Ac_2, Ac_4, Ac_5 .

Note-se que, no final do procedimento, é normal que as ações selecionadas não utilizem a totalidade dos recursos disponíveis, pois haverá quantidades residuais dos recursos que sobram em cada agregado – no caso do exemplo, é fácil ver que ficam por utilizar 4 homens.mês do tipo T_1 , 11 homens.mês de T_2 e 10 homens.mês de T_3 . Estas quantidades seriam redistribuídas a um número pequeno de agregados, de acordo com as pontuações e características destes, e o problema de distribuição de recursos a ações seria novamente resolvido para estes agregados. O procedimento de redistribuição de recursos sobrantes seria repetido até que as sobras fossem suficientemente pequenas para não permitir a seleção de mais ações.

6. Impacto para a empresa

A abordagem apresentada permite apoiar a obtenção de decisões solidamente fundamentadas. A existência de uma estrutura de preferências explícita permite compreender (e justificar) melhor as decisões tomadas, além de permitir também detetar com mais facilidade incoerências no processo de decisão. Esta abordagem permite, pois, uma maior responsabilização dos decisores, ajudando-os a perceber melhor onde poderão ter ocorrido erros de avaliação quando se constate que as decisões tomadas não foram as mais adequadas. O facto desta abordagem de avaliação necessitar da intervenção de vários agentes de decisão, correspondentes a vários níveis de gestão da empresa, leva a que esses agentes de decisão (que vão desde a gestão de topo aos responsáveis pelos agregados/projetos) pensem, discutam e explicitem os seus pressupostos. Além disso, perante os resultados e processo de avaliação, os decisores podem alterar a opinião ou a estrutura de prioridades. As estruturas hierárquicas apresentadas levam a que agentes de decisão de um determinado nível da hierarquia não possam intervir diretamente nas decisões de outros níveis. Esta diferenciação conduz, assim, a resultados que consideram variados pontos de vista e objetivos por parte da empresa.

Através do conceito de agregado, a abordagem permite também a harmonização de decisões sobre ações inter-relacionadas que estão situadas em *clusters* diferentes.

O modelo de avaliação proposto permite ainda apoiar decisões intercalares (entre dois momentos de planejamento e decisão) quando surgem situações imprevistas, como o aparecimento de novas oportunidades não previstas no momento do planejamento. Numa situação como esta, tais oportunidades podem ser integradas facilmente no modelo, e no caso de se decidir aproveitar tais oportunidades, a realocação dos recursos também pode ser feita com base no modelo.

Esta abordagem também ajuda a incentivar a identificação de oportunidades estratégicas e de flexibilidade operacional nas ações e agregados da empresa, através da definição dos critérios (subcritérios). Estas oportunidades estratégicas e de flexibilidade operacional passam a ser um fator relevante nas decisões a tomar, levando os responsáveis pelas propostas a procurá-las ativamente.

A solução foi aplicada na empresa a um conjunto de agregados e ações de teste, a partir de um protótipo desenvolvido em Microsoft Excel. Para tal, foi determinada a estrutura de preferências, com auxílio de elementos ligados à gestão da empresa, e esta foi utilizada para fazer um conjunto de avaliações. Os resultados assim obtidos foram analisados por elementos da empresa, por forma a verificar se estavam de acordo com as expectativas de atratividade das ações e agregados considerados. Esta análise permitiu fazer uma primeira calibração do modelo, após a qual se considerou que os resultados estavam em consonância com as expectativas.

Após esta primeira validação, passou-se à fase de construção da aplicação a ser usada pela empresa, e da sua integração nos sistemas da empresa. Esta integração permitiu que a solução desenvolvida pudesse passar a ser usada requerendo um esforço muito menor de introdução de dados. Mais uma vez, os testes efetuados a um conjunto de ações e agregados, com base na estrutura de preferências eliciada a elementos ligados à gestão da empresa, mostraram que a solução contribui para tornar claro e objetivo o processo de alocação de recursos, conduzindo a uma utilização mais eficientes dos recursos.

7. Conclusões

A abordagem de avaliação apresentada permite avaliar os projetos de Inovação, Investigação e Desenvolvimento de uma empresa de

telecomunicações: a PT Inovação. Esta avaliação é feita em dois níveis distintos de decisão: a ação e o agregado de ações. Enquanto a ação é um conceito já utilizado na empresa, o agregado é um novo conceito introduzido e que corresponde a um conjunto de ações inter-relacionadas vocacionadas para um mesmo serviço ou produto. Em cada um dos níveis de decisão, são utilizados diferentes critérios, estruturados de forma hierárquica. São considerados diferentes tipos de ações e de agregados e o modelo prevê uma diferenciação dos pesos dos critérios consoante o tipo de ação ou agregado em avaliação. O modelo de avaliação apresentado permite considerar critérios, quer qualitativos, quer quantitativos (como sendo o caso de alguns critérios financeiros). Para além deste facto, o modelo permite também lidar com um grande número de alternativas. Esta abordagem não necessita que o decisor ou responsável de projeto ou ação proceda a todas as comparações entre as alternativas apresentadas, mas antes permite que sejam feitas apenas um número mínimo de comparações, evitando demasiado esforço e informação requerida ao decisor.

O modelo de avaliação destina-se, principalmente, a apoiar os decisores na tomada de decisões sobre alocação de pessoas a projetos. O modelo tem um papel importante na deteção de incoerências no processo de avaliação, integração de decisões intercalares neste processo, e identificação de fontes de erros de avaliação. A hierarquização do modelo permite também a identificação da estrutura de preferências dos diferentes órgãos de decisão que participam no processo de avaliação.

Este modelo de avaliação pode também ajudar a incentivar na identificação de oportunidades estratégicas e de flexibilidade operacional nos projetos da empresa.

BIBIOGRAFIA

CARLSSON, C. e Outros, *A fuzzy approach to R&D project portfolio selection*. International Journal of Approximate Reasoning, 44: p. 93-105, 2007.

DYER, J.S., *A clarification of "remarks on the analytic hierarchy process*. Management Science, 36(3): p. 274-275, 1990a.

DYER, J.S., *Remarks on the analytic hierarchy process*. Management Science, 36(3): p. 249-258, 1990b.

- FARRUKH, C. e Outros, Developing a process for the relative valuation of R&D programmes. *R&D Management*, 30(1): p. 43-53, 2000.
- GODINHO, P. e Outros, *Some issues about the application of the analytic hierarchy process to R&D project selection*. *Global Business and Economics Review*, 13(1): p. 26-41, 2011.
- HARKER, P.T., *Incomplete pairwise comparisons in the analytic hierarchy process*. *Mathematical Modelling*, 9(11): p. 837-848, 1987.
- HARKER, P. e L. Vargas, *The theory of ratio scale estimation: Saaty's analytic hierarchy process*. *Management Science*, 33: p. 1383-1403, 1987.
- HARKER, P. e L. Vargas, *Reply to 'Remarks on analytic hierarchy process'*. *Management Science*, 36: p. 269-273, 1990.
- HOLDER, R., *Some comment on the analytic hierarchy process*. *Journal of the Operational Research Society*, 41: p. 1073-1076, 1990.
- HOLDER, R., *Response to holder's comments on the analytic hierarchy process: Response to the response*. *Journal of the Operational Research Society*, 42: p. 914-918, 1991.
- ISHIZAKA, A. e A. Labib, Review of the main developments in the analytic hierarchy process. *Expert Systems with Applications*, 38: p. 14336-14345, 2011.
- KABLAN, M.M., *Decision support for energy conservation promotion: an analytic hierarchy process approach*. *Energy Policy*, 32(10): p. 1151-1158, 2004.
- POH, K., B.W. Ang, e F. Bai, *A comparative analysis of R&D project evaluation methods*. *R&D Management*, 31(1): p. 63-75, 2001.
- SAATY, T.L., *The Analytic Hierarchy Process: Planning, Priority, Setting Resource Allocation*, New York: McGraw-Hill, 1980.
- SAATY, T.L., *Axiomatic foundation of the analytic hierarchy process*. *Management Science*, 32: p. 841-855, 1986.
- SAATY, T.L., *An exposition of the AHP in reply to the paper: 'Remarks on the analytic hierarchy process'*. *Management Science*, 36: p. 259-268, 1990.
- SAATY, T.L., *Response to holder's comments on the analytic hierarchy process*. *Journal of the Operational Research Society*, 42: p. 909-929, 1991.
- SAATY, T.L., *Highlights and critical points in the theory and applications of the analytic hierarchy process*. *European Journal of Operational Research*, 74: p. 426-477, 1994a.
- SAATY, T.L., *Fundamentals of Decision Making and Priority Theory with the Analytic Hierarchy Process*, RWS Publications, 1994b.SAATY, T.L., *That is not the analytic hierarchy process: what the AHP is and what it is not*. *Journal of Multi-Criteria Decision Analysis*, 6(6): p. 324-335, 1997.
- SAATY, T.L., *Rank from comparisons and from ratings in the analytic hierarchy/network processes*. *European Journal of Operational Research*, 168: p. 557-570, 2006.
- SHIN, C.-O., S.-H. Yoo, e S.-J. Kwak, *Applying the analytic hierarchy process to evaluation of the national nuclear R&D projects: The case of Korea*. *Progress in Nuclear Energy*, 49(5): p. 375-384, 2007.
- STAM, A. e P. Duarte Silva, *On multiplicative priority rating methods for AHP*. *European Journal of Operational Research*, 145: p. 92-108, 2003.

(Página deixada propositadamente em branco)

IMPrensa DA UNIVERSIDADE DE COIMBRA
COIMBRA UNIVERSITY PRESS
2014

