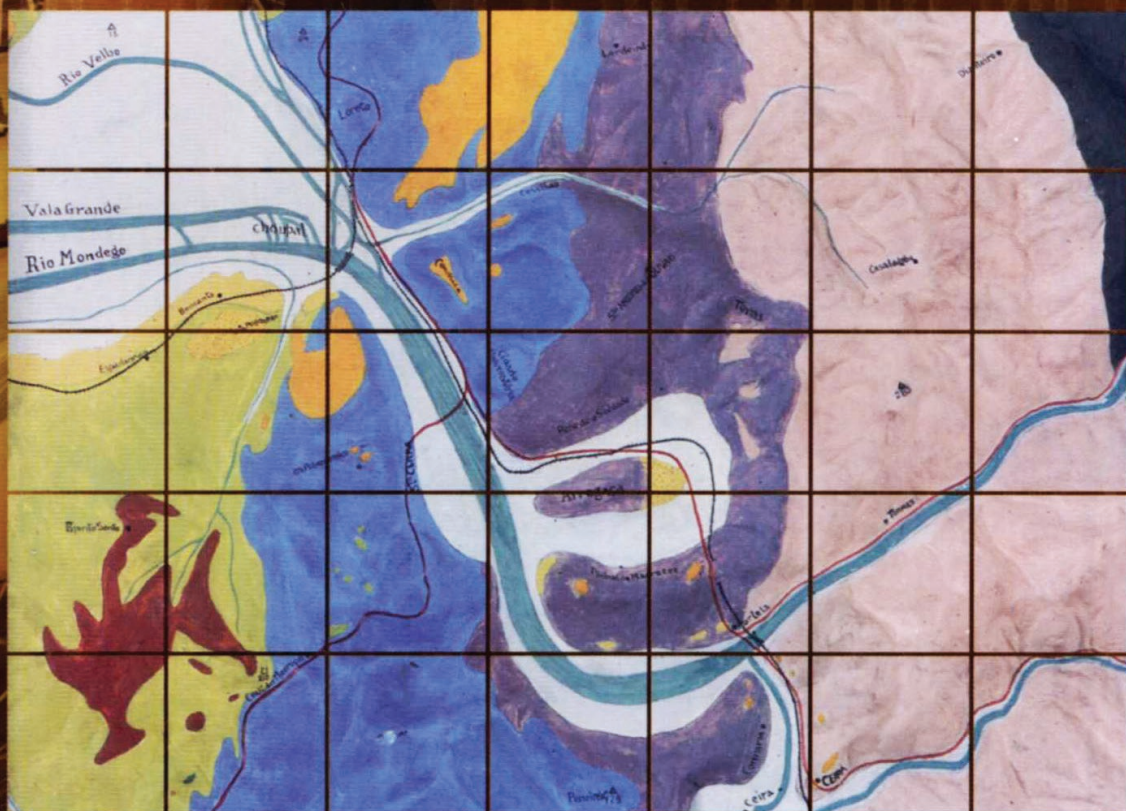


Departamento de Geografia
Centro de Estudos em Geografia e Ordenamento do Território

Cadernos de Geografia



Nº 26/27 - 2007/08

Desempenho de um modelo de máxima entropia na previsão da distribuição do endemismo *Euphorbia piscatoria* na Ilha da Madeira

Albano Figueiredo

Centro de Estudos em Geografia e Ordenamento do Território (CEGOT). geofiguc@gmail.com

Introdução

O conhecimento da corologia das espécies representa um aspecto de grande importância no suporte à definição e aplicação de medidas de gestão direccionadas à preservação da biodiversidade, aspecto que tem beneficiado da crescente aplicação de processos de modelação. A disponibilidade de informação ambiental detalhada, a utilização de novas técnicas estatísticas e a disponibilidade crescente de ferramentas SIG contribuiu, decisivamente, para o desenvolvimento e crescente aplicação de técnicas de modelação no âmbito da distribuição de espécies (GUISAN e ZIMMERMANN, 2000; PHILLIPS *et al.*, 2006). O processo de modelação recorre, em diferentes etapas, a ferramentas disponibilizadas pelos Sistemas de Informação Geográfica (SIG). Uma das contribuições mais importantes destas plataformas reside na possibilidade de apoiar a recolha, produção, armazenamento, gestão e preparação dos dados que serão necessários à calibração dos modelos, como é o caso dos dados relativos à distribuição do organismo em estudo e das variáveis ambientais entendidas como determinantes para a explicação dessa distribuição. Há mesmo a possibilidade de levar a cabo todo o processo de modelação em ambiente SIG, através da acoplagem de módulos desenvolvidos para o efeito.

Ainda que os resultados produzidos pelos modelos devam ser objecto de análise cautelosa (ARAÚJO e NEW, 2007), os resultados que produzem são de grande utilidade e apresentam inúmeras aplicações em vários domínios (biogeografia, evolução, ecologia, conservação). Desde a simples análise das relações entre variáveis ambientais e diversidade (ELITH *et al.*, 2006), até à identificação de territórios sobre ameaça potencial de espécies invasoras (PETERSON, 2003; THUILLER *et al.*, 2005c), delimitação de reservas naturais (POLASKY e SOLOW, 2001; ARAÚJO *et al.*, 2004), avaliação de alterações na distribuição de espécies associadas a mudanças climáticas (THOMAS *et al.*, 2004; ARAÚJO *et al.*, 2005; THUILLER *et al.*, 2005a; THUILLER *et al.*, 2005b; ARAÚJO e RAHBEK, 2006; ARAÚJO *et al.*, 2006), sob

condições climáticas passadas (HUGALL *et al.*, 2002; PETERSON *et al.*, 2004), ou na identificação de áreas para ocorrência potencial de espécie raras, nomeadamente em territórios onde a sua presença não está confirmada.

Uma parte significativa dos métodos de modelação baseia-se no conceito de *nicho ecológico*. Com base num conjunto de pontos, que representam a ocorrência de uma determinada espécie num território (presenças), e um conjunto de variáveis ambientais, que servirão como variáveis predictoras (factores bióticos e abióticos), procura-se determinar a correlação entre os dois conjuntos. Na verdade, essa correlação permite criar um perfil ecológico da espécie com base nas variáveis utilizadas, de forma a poder avaliar o grau de adequação do território às suas exigências ecológicas, o que remete para a probabilidade de ocorrência do organismo no território em análise ou noutros territórios. Desta forma, a selecção de variáveis e a representatividade da amostra relativa à ocorrência da espécie determinam a qualidade dos resultados obtidos, os quais são ainda determinados pelo método de modelação em uso e pelo tratamento a que podem estar sujeitos os *outputs* produzidos pelos modelos. A interpretação dos resultados de modelos preditivos deve, portanto, ser cautelosa, uma vez que os mesmos estão dependentes de diversos factores.

Material e Métodos

Método de modelação

O principal objectivo deste trabalho é avaliar o desempenho do método de modelação *Maxent* na previsão da distribuição da espécie endémica *Euphorbia piscatoria* na Ilha da Madeira. Esta avaliação passa por determinar em que medida a dimensão da amostra condiciona os resultados dos modelos, visto que a taxa de ocorrência pode afectar o desempenho do método de modelação (MANEL *et al.*, 2001; JIMENEZ-VALVERDE e LOBO, 2007). Segundo MANEL *et al.* (2001), as presenças tendem a ser previstas de forma mais efectiva quando a taxa de ocorrência aumenta, verificando-se a situação inversa ao nível das ausências, o que tem influên-

cia em termos de erros (estima em excesso ou por defeito as áreas adequadas à ocorrência).

A aplicação deste método, de base estatística e baseado no princípio da máxima entropia, baseou-se na utilização do software *Maxent*, versão 3.2.1 (PHILLIPS *et al.*, 2006; PHILLIPS e DUDIK, 2008). Enquadrado no âmbito dos *habitat suitability models*, apresenta uma previsão com base na probabilidade de distribuição por máxima entropia, partindo de uma correlação entre os dados de presença da espécie e as variáveis preditivas seleccionadas. Classificado como do tipo "*machine learning*", revela bom desempenho na modelação preditiva de distribuição de espécies quando comparado com outros métodos de modelação (ELITH *et al.*, 2006; HERNANDEZ *et al.*, 2006; PHILLIPS *et al.*, 2006).

Dados de calibração dos modelos

Para a modelação da distribuição da espécie *Euphorbia piscatoria* utilizam-se apenas dados de presença. A opção por um modelo que inclui apenas informação relativa a *presenças* não coloca em causa a validade dos resultados, sendo este tipo de modelos considerado como válido para a avaliação da distribuição potencial de espécies (GRAHAM *et al.*, 2004; ELITH *et al.*, 2006). Além disso, a utilização de dados de ausência em territórios fortemente perturbados pela actividade antrópica reveste-se de algum risco para a fiabilidade dos resultados de modelos que prevêem a distribuição potencial. No caso do território de ocorrência potencial da euforbiácea em estudo, a ausência em determinados sectores está frequentemente relacionada com o factor perturbação antrópica, não traduzindo, portanto, condições ecológicas inadequadas para a ocorrência da espécie, o que representa uma situação em que a utilização de *ausências* comporta um valor questionável para o rigor dos resultados dos modelos (ANDERSON *et al.*, 2003).

Para testar a capacidade de previsão deste método de modelação utilizam-se diferentes subconjuntos de dados, os quais derivam de uma mesma base de dados original, cujas ocorrências para a espécie *Euphorbia piscatoria* foram obtidas com base em fotointerpretação de fotografia aérea (georreferenciada e ortorrectificada, cor verdadeira, escala 1/2000, ano de 2007¹) em ambiente SIG. Para validação da metodologia utilizada no levantamento das ocorrências da espécie procedeu-se à confirmação em campo de 30% das ocorrências detectadas.

A criação de amostras de diferente efectivo baseou-se na selecção aleatória de presenças, correspondendo a fracções da base de dados que contém o total de ocorrências (5%, 10%, 25%, 50%, 75%). Assim, foi possível comparar o desempenho de modelos que apenas dispõem de uma fracção da informação (distribuição da espécie e condições associadas) com um modelo que dispõe de todos os pontos para o processo de calibração.

Para o processo de calibração de cada modelo utilizou-se uma fracção das ocorrências (70%), tendo a restantes (30%) sido utilizadas no processo de avaliação dos resultados. Apenas um modelo foi calibrado com o total de ocorrências, sem que tenha sido utilizada uma fracção para validação dos resultados. Este modelo serviu apenas para comparação visual em termos de área prevista, com o objectivo de avaliar de que forma o processo de modelação pode estar comprometido quando a calibração não dispõe de toda a informação disponível sobre a ocorrência da espécie em causa. Isto porque, um modelo óptimo, capaz de produzir a melhor previsão, deverá poder tirar partido de toda a informação disponível sobre a distribuição da espécie aquando da calibração (PHILLIPS *et al.*, 2006).

De forma a avaliar a influência da auto-correlação espacial dos dados no enviesamento dos resultados, a validação dos modelos foi feita com base em duas aproximações distintas, ainda que contendo a mesma fracção de dados (30%). Assim, além de se avaliar o modelo com base numa fracção (30%) da amostra seleccionada aleatoriamente pelo software *Maxent*, utilizou-se ainda um ficheiro independente, cujas presenças foram seleccionadas aleatoriamente numa plataforma SIG.

Variáveis

Para o processo de modelação foram seleccionadas as variáveis ambientais que apresentavam maior contribuição para a explicação da distribuição actual da espécie em estudo. Esta selecção baseou-se na aplicação da técnica estatística de *jackknifing*, tendo sido avaliada a contribuição de todas as variáveis em conjunto, de cada uma das variáveis isoladamente e ainda de todas as variáveis à excepção de uma. Para tal calibrou-se um modelo com todas as variáveis e todas as ocorrências para a espécie em estudo.

De um conjunto de seis variáveis foram seleccionadas para calibração dos modelos 3 variáveis quantitativas contínuas, previamente normalizadas (altitude - 41,9%; precipitação média anual - 31,8% e temperatura média anual - 5% (Atlas do Ambiente Digital - Instituto do Ambiente, 2008), e uma qualitativa (vegetação natural potencial - 15.7% (CAPELO *et al.*,

¹ - Fotografia aérea cedida ao abrigo de protocolo estabelecido com a Direcção Regional de Geografia e Cadastro, Secretaria Regional do Equipamento Social da Região Autónoma da Madeira (Licença de utilização n.º 24/2008).

2004), transformada em variável quantitativa discreta de escala ordinal. O modelo classificou como pouco importante a variável exposição (0.3%), e apesar de atribuir ao declive (5.4%) uma contribuição superior à temperatura, esta variável permite explicar uma percentagem da distribuição muito reduzida quando utilizada individualmente. A exclusão desta variável deve-se ainda ao facto de a correlação elevada entre declives acentuados e a ocorrência da espécie estar associada ao facto de as vertentes abruptas, onde a ocupação agrícola não foi possível, terem funcionado como refúgio para a flora indígena, pelo que o papel desta variável na explicação da ocorrência da espécie é questionável em termos ecológicos.

A resolução espacial utilizada para as variáveis foi de 100X100 m, uma resolução concordante com aquela que orientou a recolha de informação sobre a distribuição actual da espécie em estudo. A alteração da resolução espacial das variáveis baseou-se no método *linear interpolation*, dado ser considerado mais realista que a opção *nearest-neighbour interpolation* (PHILLIPS *et al.*, 2006).

Ecologia da espécie

A espécie *Euphorbia piscatoria* (figueira-do-inferno) é um endemismo do Arquipélago da Madeira (excepto Selvagens). Trata-se de um arbusto perene, que pode atingir 2 m de altura, suculento, caules muito ramificados e desprovidos de folhas na parte inferior, as quais são linear-lanceoladas ou linear-oblongas, glaucas, alternas, sésseis e caducas no Verão. As flores, pequenas e de cor esverdeada a amarelo-alaranjado, dispõem-se numa inflorescência umbeliforme que pode variar de simples a composta. Os frutos (cápsulas) são avermelhados quando maduros.

Na Ilha da Madeira tem preferência por territórios de carácter inframediterrâneo seco a termomediterrâneo sub-húmido, o que lhe garante maior representatividade nos sectores de menor altitude (< 300m) da encosta meridional da ilha. Dado o elevado grau de perturbação da sua área de ocorrência potencial, em função da ocupação agrícola, apresenta maior representatividade em sectores de declive mais acentuado.

Do ponto de vista sinfitossociológico, as comunidades dominadas por *Euphorbia piscatoria* constituem uma etapa subserial (mato baixo) de diferentes séries de vegetação (CAPELO *et al.*, 2004): zambujal inframediterrâneo seco (*Mayteno umbellatae-Oleo maderensis sigmetum*), matagal inframediterrâneo sub-húmido de marmulano (*Helichryso melaleucisideroxylo marmulanae sigmetum*), e em ambas as fácies da série da laurissilva infra-termomediterrânea,

sub-húmida a húmida, do barbusano (*Semele androgynae-Apollonio barbujae sigmetum*).

Resultados

Avaliação dos modelos

A avaliação do desempenho do método de modelação assenta na apreciação dos resultados produzidos pelos diferentes modelos. Para a avaliação da performance dos modelos utilizaram-se medidas dependentes e não-dependentes da definição de limiar. A definição de limiar resulta do facto de os resultados dos modelos se apresentarem numa escala contínua (0 a 1), correspondente ao grau de adequação do território à ocorrência da espécie, em função das variáveis seleccionadas. Para a aplicação de algumas medidas, que permitem a comparação dos resultados dos diferentes modelos, é necessária a transformação numa escala binária (0,1), processo que implica a definição de um limiar. Perante a adopção de um determinado limiar, cuja definição pode estar baseada em diferentes aproximações (LIU *et al.*, 2005; FREEMAN e MOISEN, 2008), todos os píxeis com valor superior são classificados como representando condições adequadas à ocorrência da espécie, enquanto todos os que apresentarem valores inferiores traduzem condições não adequadas à sua presença. Em termos de medidas dependentes da definição de limiar optou-se pelas medidas *taxa de omissão* e *área prevista proporcional* (PHILLIPS *et al.*, 2006). A *taxa de omissão* corresponde à fracção de presenças utilizadas para avaliação dos resultados que coincidem com áreas não previstas como adequadas pelo modelo, ou seja, são pontos em que a presença da espécie foi confirmada mas que o modelo prevê como áreas não adequadas à sua ocorrência, pelo que representa uma taxa de erro. A *área prevista proporcional* corresponde à percentagem da área prevista pelo modelo como adequada à ocorrência da espécie em relação à área total, que, tal como a medida anterior, é afectada pela definição do limiar, mas em ordem inversa.

A avaliação da performance dos modelos com base em medidas dependentes de um limiar tem alguns problemas associados, uma vez que o processo de transformação tem implicações significativas em termos de área prevista como adequada à ocorrência da espécie (FREEMAN e MOISEN, 2008). Isto porque, a utilização de um valor demasiado alto como limiar vai favorecer a redução da área prevista como adequada pelo modelo, uma vez que vai seleccionar apenas as áreas em que a probabilidade de ocorrência da espécie é mais alta, dado tratarem-se das áreas mais adequadas. No entan-

to, esta situação promove um aumento do erro de omissão (*taxa de omissão*), ou seja, não inclusão de áreas que, apesar de apresentarem condições menos adequadas à presença da espécie, têm presenças associadas, o que favorece menor sucesso em termos de correcta previsão pelo modelo do subconjunto de dados utilizado para avaliação dos resultados. Já a utilização de um limiar muito baixo reduz a *taxa de omissão*, pois inclui áreas mais extensas do território, mesmo as que apresentam condições menos adequadas à ocorrência da espécie, o que aumenta o erro associado à previsão de área em excesso, favorecendo uma situação em que o modelo estima por excesso. Seria de considerar a utilização de um valor intermédio de 0.5, numa escala que varia entre 0 e 1 (*logistic output*), no entanto, a utilização deste limiar carece de fundamentação ecológica e representatividade estatística, pelo que não assegura necessariamente maior rigor nos resultados de previsão da distribuição, nomeadamente para situações em que as espécies apresentam uma frequência muito baixa ou muito elevada (FREEMAN e MOISEN, 2008). Como a definição do limiar está dependente de vários factores, como a dimensão da amostra utilizada no processo de calibração e a contribuição de cada ponto para o processo de calibração do modelo, bem como o contexto no qual os resultados são utilizados (PHILLIPS *et al.*, 2006), tem sido demonstrado que este limiar não é uma boa opção na transformação de uma escala contínua numa escala binária (JIMENEZ-VALVERDE e LOBO, 2007). A sua aplicação será adequada no caso de uma distribuição simétrica, ou seja, em que a base de dados detém o mesmo número de *presenças* e *ausências*, e no caso de se atribuir igual peso aos erros de omissão e previsão por excesso (LOBO *et al.*, 2008). Com base neste argumento, este limiar foi imediatamente rejeitado para o caso em estudo. Isto porque, uma vez que não se disponibilizaram dados sobre as ausências da espécie, o *software* utiliza como *ausências*, na verdade *pseudo-ausências*, locais em que a presença da espécie não foi detectada. Durante o processo de calibração o logaritmo seleccionou sempre um número significativamente superior de *ausências* (*pseudo-ausências*) comparativamente às presenças disponibilizadas pela amostra. Como, em termos de probabilidade de ocorrência, os resultados dos modelos são enviesados no sentido do tipo de dados mais comum (CRAMER, 1999), neste caso, como as *pseudo-ausências* apresentam um efectivo superior em todos os modelos, os resultados acabam por avançar probabilidades que estão enviesadas no sentido das ausências, ou seja, tendem a favorecer as ausências em relação às presenças.

Como o modelo não dispõe de ausências confirmadas, esta situação poderia constituir um problema, pois as *pseudo-ausências* poderiam ser ausências reais, associadas a limitações de carácter ambiental, e

ausências fictícias, associadas a perturbação de carácter humano e incorrecta classificação devido a sub-amostragem. Neste caso, reduziu-se o problema das falsas ausências associadas a sub-amostragem utilizando um sistema de amostragem que implica um esforço homogéneo em toda a área em estudo. Assim, só as ausências fictícias associadas a perturbação antrópica podem contribuir para aumentar a taxa de erro associada à distribuição potencial da espécie prevista pelos modelos. Erro que, aliás, deve ser considerado na modelação da distribuição potencial da espécie *Euphorbia piscatoria*, dada a coincidência espacial entre os sectores mais perturbados pela actividade agrícola e o território de ocorrência potencial da espécie.

Desta forma, como em teoria nem todas as *pseudo-ausências* correspondem a ausências por determinantes de carácter ambiental, foi necessário diminuir o seu peso na determinação do limiar utilizado para conversão dos resultados numa escala binária, valorizando-se a capacidade do modelo para classificar correctamente as presenças, o que permite reduzir a *taxa de omissão*. Neste sentido, utilizou-se um limiar que valoriza a capacidade do modelo para prever correctamente as presenças utilizadas para testar o modelo, ou seja, maximizou-se a taxa de sucesso associada à correcta classificação de presenças confirmadas (*Maximum test sensitivity and specificity*). Além de reduzir a taxa de omissão ao mínimo, não coloca problemas em termos de excesso de área prevista, aspecto verificado através de comparação com a distribuição de todas as presenças disponíveis para a distribuição da espécie. Poder-se-ia utilizar como limiar o valor mínimo de probabilidade de ocorrência em que foi detectada uma presença com base no conjunto de calibração (*minimum training presence*), o que reduziria a taxa de omissão a um valor muito baixo ou mesmo nulo. No entanto, esta opção implicaria, para a maior parte dos modelos, um aumento significativo da área prevista, considerando mesmo sectores em que a ocorrência é pouco provável, favorecendo um reforço do erro de estimativa por excesso.

Como todo o processo implica algum grau de subjectividade, e de forma a evitar a interferência da definição de limiar nos resultados dos modelos, utilizou-se uma medida de avaliação de desempenho que não está dependente deste processo, a medida AUC de ROC (*Area under the curve of Receiver-Operating Characteristics analysis*). De uso adequado (FIELDING e BELL, 1997) e muito frequente na avaliação da performance dos modelos, esta medida avalia a capacidade do modelo para discriminar os diferentes graus de adequação do território à ocorrência da espécie. Para tal avalia, para todos os limiares possíveis, a capacidade do modelo para prever correctamente a presença

da espécie e o erro em termos de excesso de área prevista como adequada. Além de ser uma medida que não está dependente da definição de limiar (JIMENEZ-VALVERDE e LOBO, 2007), apresenta ainda a vantagem de não ser afectada pela frequência com que a espécie ocorre (MANEL *et al.*, 2001; MCPHERSON *et al.*, 2004).

Numa escala de referência entre 0 e 1, o valor 1 indica capacidade perfeita de discriminação, AUC=0.5 indica fraca capacidade de discriminação, e valores de AUC > 0.9 indicam muito boa capacidade do modelo para discriminar entre áreas adequadas e inadequadas para a ocorrência da espécie (ELITH *et al.*, 2006; PEARSON *et al.*, 2006), ainda que não permita avaliar o grau de confiança associado aos resultados dos modelos em relação à distribuição real e potencial da espécie. Por esta razão, esta medida não é aplicada de forma isolada, dado o facto de modelos classificados como apresentando bom desempenho com base na medida AUC, na verdade poderem apresentar resultados poucos satisfatórios em termos de previsão da distribuição (LOBO *et al.*, 2008).

Por fim, com o objectivo de confirmar a performance do método de modelação, de uma forma geral, disponibilizaram-se os resultados dos modelos a um avaliador. Detentor de conhecimento rigoroso da distribuição da espécie *Euphorbia piscatoria* e dos condicionantes ambientais que limitam a sua distribuição, não lhe foi cedida informação sobre os dados utilizados na calibração dos modelos nem sobre os resultados da avaliação dos modelos com base nas medidas de precisão seleccionadas. Foi-lhe pedido que, com base numa escala qualitativa ordinal, classificasse os modelos em função da sua adequação à distribuição potencial da espécie em estudo, devendo justificar a seriação apresentada. A sua classificação teve por base os cartogramas correspondentes a cada um dos modelos, onde constava apenas a área prevista pelo modelo como adequada à ocorrência da espécie.

Discussão dos resultados

Vários são os factores que interferem na precisão dos resultados obtidos com base em modelos de distribuição de espécies, condicionando o grau de fiabilidade dos resultados (FIGUEIREDO, 2008); desde as condições de amostragem, que interferem na representatividade da amostra, o método de modelação, o tratamento dos resultados (conversão de resultados de escala contínua numa escala binária) ou o número/tipo de variáveis predictoras seleccionadas.

Em relação aos resultados obtidos pelos diferentes modelos, e com base nas medidas *taxa de omissão* e *área prevista proporcional*, medidas utilizadas para avaliar a capacidade dos modelos para prever a distribuição da espécie *Euphorbia piscatoria*, verifica-se uma assinalável

proximidade em termos de resultados (Tabela 1), aspecto reforçado pela elevada coincidência espacial.

Todos os modelos apontam para uma presença potencial muito mais significativa na encosta Sul da ilha, o que duplica o comportamento observado em termos de distribuição actual (Figura 1A).

Verifica-se que os modelos 4 (Figura 1B) e 11 apresentam o pior desempenho em termos de medidas de avaliação, sendo identificados pelo avaliador como os que apresentam uma estimativa menos adequada, uma vez que não prevêem a presença da espécie em áreas onde foram identificadas comunidades, como é o caso da foz da Ribeira da Janela, na encosta Norte da ilha. Este aspecto é confirmado pelos valores de AUC que, ainda que permitam classificar o modelo como óptimo, indicam menor capacidade do que os restantes modelos para diferenciar condições adequadas de inadequadas à ocorrência da espécie (Tabela I).

Já os modelos 5 e 6 (Figura 2A) são seleccionados pelo avaliador como os mais realistas em termos de previsão da distribuição da espécie em estudo, situação concordante com o bom desempenho que demonstram ao nível das medidas de avaliação.

Da comparação entre a classificação do avaliador e o desempenho obtido com base na utilização das medidas de avaliação destaca-se apenas uma situação de discordância, relativamente aos resultados dos modelos 7 e 9. Ainda que as medidas de precisão os classifiquem como modelos que apresentam bom desempenho, nomeadamente devido às baixas taxas de omissão, foram classificados como pouco adequados pelo avaliador, dado preverem a presença da espécie em praticamente toda a linha de costa do litoral Norte da ilha (Fig. 2B). Além de a sua presença actual não ter sido detectada em sectores significativos da encosta Norte (Fig. 1A), a maior parte dos modelos também não previu este padrão de ocorrência potencial, o que reforça a posição do avaliador.

Destaca-se ainda o facto de não se verificar uma diferença significativa entre o modelo calibrado com todas as ocorrências conhecidas para a espécie (modelo 13 - figura 3A) e o que foi calibrado com apenas 5% das presenças conhecidas (modelo 12 - figura 3B), embora no pormenor sejam detectadas algumas diferenças.

Com base nos valores de AUC, todos os modelos apresentam resultados muito satisfatórios em termos de capacidade para discriminar os diferentes graus de adequação do território à ocorrência da espécie *Euphorbia piscatoria* (Tabela I), o que confirma o bom desempenho do método de modelação *Maxent* na previsão da distribuição de espécies. Na análise dos valores de AUC deve-se, no entanto, ter em conta o facto de a espécie em estudo apresentar uma distribuição restrita em termos de condições ecológicas,

Tabela I
Resultados das medidas de precisão para os diferentes modelos.

Modelos	Calibração (n.º pontos)	Avaliação (n.º pontos)	AUC	Omissão (%)	Área (%)
1	1093	468	0.920	4	23
2*	1090	471	0.920	5	21
3	108	46	0.911	4	21
4*	108	46	0.909	13	18
5	277	118	0.922	4	20
6*	276	119	0.922	3	19
7	546	234	0.920	3	23
8*	546	234	0.924	2	23
9	816	349	0.927	1	24
10*	812	353	0.923	3	22
11	54	23	0.915	9	16
12*	53	24	0.933	0	19
13	1561	-	0.923	-	21

* Modelos cujo processo de avaliação de resultados se baseou na utilização de ficheiro independente. Os restantes modelos foram avaliados com base num sub-conjunto seleccionado automaticamente pelo *software*, à excepção do modelo 13 que serviu apenas para comparação visual da área prevista.

AUC = Area under the curve of ROC plot; Omissão = Taxa de omissão; Área = Área prevista proporcional.

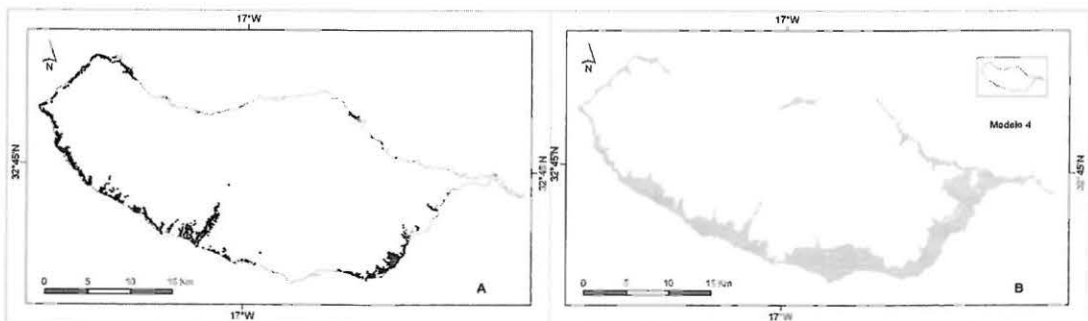


Figura 1
A) Distribuição actual do endemismo *Euphorbia piscatoria* na Ilha da Madeira; B) Previsão de distribuição potencial para a espécie pelo modelo que apresenta o pior desempenho.

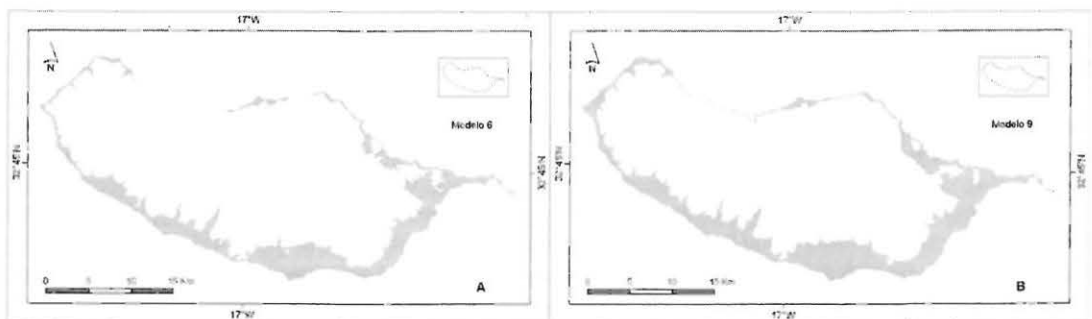


Figura 2
Previsão de distribuição potencial para a espécie *Euphorbia piscatoria*: A) modelo com bom desempenho; B) modelo classificado pelo avaliador como pouco adequado, apesar do bom desempenho

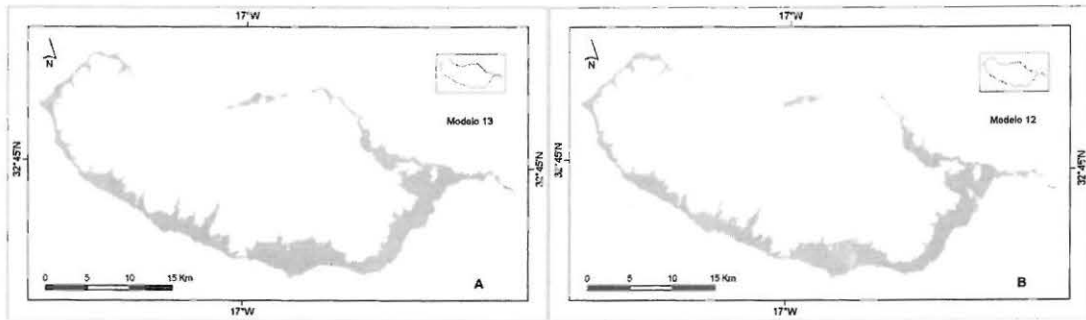


Figura 3

Previsão de distribuição potencial para a espécie *Euphorbia piscatoria*: A) resultados para modelo calibrado com todas as presenças conhecidas; B) resultados para modelo calibrado com apenas 5% das presenças conhecidas.

representando uma pequena fracção da área em estudo, o que promove valores mais elevados em termos de avaliação por esta medida. Segundo LOBO *et al.* (2008), o valor de AUC deve mesmo ser interpretado como uma indicação sobre o tipo de distribuição da espécie em relação às condições ecológicas, restrita ou generalista, e não como uma medida que avalia o desempenho do modelo, ideia que é reforçada pelo elevado valor de AUC obtido pelo modelo calibrado com apenas 5% da amostra original (Modelo 12 - Figura 3B).

A comprovar o bom desempenho dos modelos está ainda a aplicação de um teste não paramétrico, o Qui-quadrado (χ^2), cuja aplicação permitiu verificar que todos os modelos apresentam melhor desempenho que um modelo aleatório na previsão do sub-conjunto de ocorrências utilizadas para avaliação, aspecto reforçado pelos resultados da aplicação do teste binomial ($p < 0.001$, one tailed¹).

Em termos de variáveis ambientais, a altitude apresenta-se como um factor determinante à distribuição em todos os modelos, de tal forma que as previsões de distribuição potencial para a encosta Sul, na maior parte dos modelos, seguem com grande proximidade a cota dos 450 metros. A identificação de um comportamento muito distinto na encosta Norte e extremo ocidental da Ilha vem reforçar a ideia de bom desempenho pela maior parte dos modelos.

Conclusões

A informação de carácter territorial adquire especial significado aquando da modelação da distribuição de espécies à escala local, nomeadamente as condições de perturbação. No caso do endemismo *Euphorbia piscatoria*,

a sua distribuição actual está muito condicionada pela extensão dos processos de perturbação, nomeadamente pela actividade agrícola e área social. Esta razão explica a importância do factor declive na distribuição da espécie, uma vez que as áreas sem utilidade agrícola registaram menor perturbação, garantindo a presença de vegetação natural, entre a qual comunidades dominadas por *Euphorbia piscatoria*.

Para o conjunto dos modelos, e de uma forma geral, a técnica de modelação *Maxent* apresenta bons resultados em termos de previsão da distribuição da espécie *Euphorbia piscatoria*, evidenciando baixa sensibilidade à variação da dimensão da amostra, pois o modelo calibrado com a amostra de menor dimensão apresenta o valor de AUC mais elevado, uma das taxas de omissão mais baixas e uma área prevista dentro da média (Modelo 12).

Apesar dos bons resultados dos modelos e bom desempenho do método de modelação, não é possível assegurar com total confiança que a distribuição potencial prevista seja a mais próxima da potencial real, tendo em conta a intensa perturbação a que o território de ocorrência potencial da espécie *Euphorbia piscatoria* esteve sujeito durante séculos. Esta situação pode colocar em dúvida se a amostra utilizada é representativa de toda a amplitude ecológica da espécie, apesar da utilização de um sistema de identificação de presenças que reduz o erro associado a um insuficiente esforço de amostragem. Esta desconfiança acaba, no entanto, por perder algum peso quando se faz uma análise da capacidade de regeneração da espécie após abandono agrícola. Dada a assinalável capacidade de regeneração da espécie em várias parcelas afectadas pelo abandono, a sua ausência nas mesmas condições de perturbação, e decorrido tempo de abandono suficiente, pode ser interpretado como ausência de condições ecológicas adequadas, o que permite

¹ - Teste aplicado automaticamente pelo software MAXENT.

reforçar a validade dos modelos. É claro que haveria que avaliar a interferência da perturbação agrícola e a disponibilidade de propágulos em termos de taxas de regeneração ao longo de um gradiente ecológico que incluísse toda a amplitude da espécie, de forma a determinar a capacidade de regeneração nas áreas onde a espécie se encontra no limite da sua distribuição ecológica.

Em termos de resultados, e com base nas medidas de precisão e comparação visual, apesar de não se identificar uma diferença significativa entre os modelos, parece ser mais prudente a avaliação ou validação dos modelos com base na utilização de um ficheiro de dados independente. Isto porque, da análise dos resultados obtidos para cada modelo, e produzidos pelo próprio *software*, verifica-se que os dados apresentam problemas em termos de auto-correlação espacial, pelo que seria pertinente testar o resultado dos modelos com base num ficheiro independente cuja selecção se baseie na divisão espacial da amostra (*spatial split*).

Agradecimentos

Investigação apoiada pela Fundação para a Ciência e a Tecnologia (SFRH/BD/27810/2006).

Um agradecimento ao Doutor Miguel Pinto da Silva Menezes de Sequeira do Departamento de Biologia da Universidade da Madeira que, além de todo o apoio prestado, desempenhou ainda o papel de avaliador dos modelos. Um agradecimento à Direcção Regional de Geografia e Cadastro - Secretaria Regional do Equipamento Social da Região Autónoma da Madeira - pela cedência de fotografia aérea.

Referências Bibliográficas

- ANDERSON, R. P., LEW, D. e PETERSON, A. T. (2003) - "Evaluating predictive models of species' distributions: criteria for selecting optimal models". *Ecological Modelling*, 162(3), pp. 211-232.
- ARAÚJO, M. B.; CABEZA, M.; THULLER, W.; HANNAH, L. e WILLIAMS, P. H. (2004) - "Would climate change drive species out of reserves? An assessment of existing reserve-selection methods". *Global Change Biology*, 10(9): 1618-1626.
- ARAÚJO, M. B. e NEW, M. (2007) - "Ensemble forecasting of species distributions". *Trends in Ecology & Evolution*, 22(1), pp. 42-47.
- ARAÚJO, M. B.; PEARSON, R. G.; THULLER, W. e ERHARD, M. (2005) - "Validation of species - climate impact models under climate change". *Global Change Biology*, 11(9), pp. 1504-1513.
- ARAÚJO, M. B. e RAHBEK, C. (2006) - "How does climate change affect biodiversity?". *Science*, 313(5792), pp. 1396-1397.
- ARAÚJO, M. B.; THULLER, W. e PEARSON, R. G. (2006) - "Climate warming and the decline of amphibians and reptiles in Europe". *Journal of Biogeography*, 33(10), pp. 1712-1728.
- CAPELO, J.; SEQUEIRA, M.; JARDIM, R. e COSTA, J. C. (2004) - "Guia da excursão geobotânica dos V Encontros ALFA 2004 à Ilha da Madeira". In CAPELO, J. - *A paisagem vegetal da Ilha da Madeira*. Quercetea. 6, pp.5-46.
- CRAMER, J. S. (1999) - "Predictive Performance of the Binary Logit Model in Unbalanced Samples". *The Statistician, Journal of the Royal Statistical Society (Series D): The Statistician*, 48(1), pp. 85-94.
- EUTH, J., GRAHAM, C. H., ANDERSON, R. P., DUDIK, M. & FERRIER, S., GUIGAN, A., HUIJAN, R. J., HUETTSMANN, F., LEATHWICK, J. R., LEHMANN, A., LI, J., LOHMANN, L. G., LOISELLE, B. A., MAMON, G., MORITZ, C., NAKAMURA, M., NAKAZAWA, Y., OVERTON, J. McC., PETERSON, A. T., PHILLIPS, S. J., RICHARDSON, K. S., SCACHETTI-PEREIRA, R., SCHAPIRE, R. E., SOBERON, J., WILLIAMS, S., WISZ, M. S. e ZIMMERMANN, N. E. (2006) - "Novel methods improve prediction of species' distributions from occurrence data". *Ecography* 29(2): 129-151.
- FIELDING, A. H. e BELL, J. F. (1997) - "A review of methods for the assessment of prediction errors in conservation presence/absence models". *Environmental Conservation* 24(01): 38-49.
- FIGUEIREDO, A. (2008) - "Avaliação de impactes das mudanças climáticas na vegetação natural da Ilha da Madeira. Questões e opções metodológicas". *Actas do 1º Congresso Nacional sobre Alterações Climáticas*. Disponível: http://www.apea.pt/xFiles/scContentDeployer_pt/docs/Doc140.pdf [22 Novembro, 2008].
- FREEMAN, E. A. e MOISEN, G. G. (2008) - "A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa". *Ecological Modelling* 217(1-2): 48-58.
- GRAHAM, C. H., FERRIER, S., HUETTSMANN, F., MORITZ, C. e PETERSON, A. T. (2004) - "New developments in museum-based informatics and applications in biodiversity analysis". 19(9): 497-503.
- GUIGAN, A. e ZIMMERMANN, N. E. (2000) - "Predictive habitat distribution models in ecology". *Ecological Modelling* 135(2-3): 147-186.
- HERNANDEZ, P. A., GRAHAM, C. H., MASTER, L. L. e ALBERT, D. L. (2006) - "The effect of sample size and species characteristics on performance of different species distribution modeling methods." *Ecography* 29(5): 773-785.
- HUGALL, A., MORITZ, C., MOUSSALLI, A. e STANISIC, J. (2002) - "Reconciling paleodistribution models and comparative phylogeography in the Wet Tropics rainforest land snail *Gnarosophia bellendenkerensis* (Brazil 1875)". *Proceedings of the National Academy of Sciences of the United States of America* 99(9): 6112-6117.
- JIMENEZ-VALVERDE, A. e LOBO, J. M. (2007) - "Threshold criteria for conversion of probability of species presence to either-or presence-absence." *Acta Oecologica-International Journal of Ecology* 31(3): 361-369.
- LIU, C. R., BERRY, P. M., DAWSON, T. P. e PEARSON, R. G. (2005) - "Selecting thresholds of occurrence in the prediction of species distributions." *Ecography* 28(3): 385-393.

- LOBO, J. M., JIMÉNEZ-VALVERDE, A. e REAL, R. (2008) - "AUC: a misleading measure of the performance of predictive distribution models." *Global Ecology and Biogeography* 17(2): 145-151.
- MANEL, S., WILLIAMS, H. C. e ORMEROD, S. J. (2001). "Evaluating presence-absence models in ecology: the need to account for prevalence." *Journal of Applied Ecology* 38(5): 921-931.
- MCPHERSON, J. M., JETZ, W. e ROGERS, D. J. (2004) - "The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact?" *Journal of Applied Ecology* 41: 811-823.
- PEARSON, R. G., THULLER, W., ARAÚJO, M. B., MARTINEZ-MEYER, E., BROTONS, L., McCLEAN, C., MILES, L., SEGURADO, P., DAWSON, T. P. e LEES, D. C. (2006) - "Model-based uncertainty in species range prediction". *Journal of Biogeography* 33(10): 1704-1711.
- PETERSON, A. T. (2003) - "Predicting the Geography of Species Invasions via Ecological Niche Modeling." *The Quarterly Review of Biology* 78(4): 419-433.
- PETERSON, A. T., MARTINEZ-MEYER, E. e GONZALEZ-SALAZAR, C. (2004) - "Reconstructing the Pleistocene geography of the *Aphelocoma* jays (Corvidae)." *Diversity and Distributions* 10(4): 237-246.
- PHILLIPS, S. J., ANDERSON, R. P. e SCHAPIRE, R. E. (2006) - "Maximum entropy modeling of species geographic distributions." *Ecological Modelling* 190(3-4): 231-259.
- PHILLIPS, S. J. e DUDIK, M. (2008) - "Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation." *Ecography* 31(2): 161-175.
- POLASKY, S. e SOLOW, A. R. (2001) - "The value of information in reserve site selection." *Biodiversity and Conservation* 10(7): 1051-1058.
- THOMAS, C. D., CAMERON, A., GREEN, R. E., BAKKENES, M., BEAUMONT, L. J., COLLINGHAM, Y. C., ERASMUS, B. F. N., DE SIQUEIRA, M. F., GRAINGER, A., HANNAH, L., HUGHES, L., HUNTLEY, B., VAN JAARSVELD, A. S., MIDGLEY, G. F., MILES, L., ORTEGA-HUERTA, M. A., TOWNSEND PETERSON, A., PHILLIPS, O. L. e WILLIAMS, S. E. (2004) - "Extinction risk from climate change." *Nature* 427(6970): 145-148.
- THULLER, W., LAVOREL, S. e ARAÚJO, M. B. (2005a) - "Niche properties and geographical extent as predictors of species sensitivity to climate change." *Global Ecology and Biogeography* 14(4): 347-357.
- THULLER, W., LAVOREL, S., ARAÚJO, M. B., SYKES, M. T. e PRENTICE, I. C. (2005b) - "Climate change threats to plant diversity in Europe." *Proceedings of the National Academy of Sciences of the United States of America* 102(23): 8245-8250.
- THULLER, W., RICHARDSON, D. M., PYSEK, P., MIDGLEY, G. F., HUGHES, G. O. e ROUGET, M. (2005c) - "Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale." *Global Change Biology* 11(12): 2234-2250.