



# ADVANCES IN FOREST FIRE RESEARCH 2018

EDITED BY

**DOMINGOS XAVIER VIEGAS**  
ADAI/CEIF, UNIVERSITY OF COIMBRA, PORTUGAL

# Generation and evaluation of ensemble simulations of wildfire spread for probabilistic forecast

Frédéric Allaire<sup>1\*</sup>; Jean-Baptiste Filippi<sup>2</sup>; Vivien Mallet<sup>1</sup>

<sup>1</sup>Inria, 2 rue Simone Iff, Paris, France and CEREMA, Sorbonne Universités, UPMC, Lab. Jacques-Louis Lions UMR CNRS 7598. {frederic.allaire@inria.fr\*, vivien.mallet@inria.fr}

<sup>2</sup>SPE - UMR 6134 CNRS. Université de Corse, France. {filippi@univ-corse.fr};

## Abstract

Typical simulation of a large wildfire requires semi-empirical models for rate of spread (ROS) such as the Rothermel model. ROS models usually take into account several physical parameters such as slope, wind speed, moisture content and other fuel properties (load, height, surface-to-volume ratio, heat of combustion, etc.). However, these parameters can be highly uncertain. In addition, the model itself is far from perfect and the solver is subject to numerical approximations. A promising direction for the prediction of wildfire spread is to rely on an ensemble of simulations instead of a single deterministic simulation.

The generation of an ensemble may follow a Monte Carlo strategy where the input parameters are sampled according to given discrete or continuous probability distributions. Each sample is independent from each other and is used to run one wildfire simulation. For a wildfire starting at a given location and time, we can use the outputs from the ensemble of simulations to estimate a burn probability map. One main challenge of the method lies in the choice of the distributions of the input parameters. Depending on the sensitivity of the model to the input parameters, small changes on the probability distributions could yield highly different outputs. At this stage, a suggestion of distributions is introduced, although calibration should be applied in order to improve simulation outputs.

Similar probabilistic approaches have already been developed in wildland fire and particularly in forest fire behavior models such as the FSPro system. A significant difficulty regarding the evaluation of such systems is to properly take into account their probabilistic aspects. We propose to perform ensemble evaluation considering all the burn probabilities that are predicted at the scale of a specific fire event. It also requires the observed burned surfaces corresponding to the specific event. Then, probabilistic scores that are commonly used in meteorological applications, such as the Brier Score, can be computed. The whole methodology is applied to a wildfire that occurred in Corsica in summer 2017 and could be extended to more fire cases.

Such probabilistic framework could then be completed with a description of ignition probability and vulnerability in order to assess fire risk over a region like Corsica.

**Keywords:** simulation ; wildfire ; uncertainty quantification ; Monte Carlo ; probabilistic score

## 1. Introduction

In an operational context, being able to accurately forecast how a wildfire will spread would provide invaluable help in decision making, not only in a crisis situation where a fire just ignited but also in the assessment of fire risk. In the latter case, one goal is to identify areas that require high-priority land planning actions in the long run or the mobilization of firefighting resources in the short term. In spite of the significant efforts in modeling to provide better and faster-to-compute forecasts of fire spread, systematically making an accurate prediction remains a distant goal (Gollner et al., 2015). One of the reasons lies in the difficulty to predict and describe the weather conditions and the state of the vegetation that will drive the fire spread. In this uncertain context, one may think that a range of likely outcomes might be more useful than a single forecast. This can be achieved by running many

simulations. This strategy is not new to the fire community as an approach relying on several runs of likely fire scenarios was already implemented for assessment of long-term fire risk in systems such as BURN-P3 (Parisien et al., 2005) and FSim (Finney et al., 2011a) where a probability of a zone burning during a year is computed (among others) over a whole region. In the case of a fire that has just ignited, which is the focus of this study, the FSPro (Finney et al., 2011b) system uses an ensemble of simulations based on a time series approach to define several weather inputs and provide a map of burn probability, that is to say, an estimate of the chance that a zone close to ignition point will burn after a given time. Evaluation in FSPro is then performed in two ways. The first method consists in comparing the mean area of the simulated burned surfaces to the observed burned area. The second method, relies on the predicted burn probabilities obtained at the vertices of the polygon describing the observed burned surface.

We propose another method for ensemble simulation in this context. One main difference with FSPro lies in the approach used when generating the inputs for each simulation. Here, we used perturbations on weather forecasts, but also on other input data such as fuel parameters. In addition, we present tools for evaluation that are commonly used in meteorological applications. Indeed, judging the quality of an ensemble implies the confrontation against observed data with appropriate scores. This methodology is applied to the Calenzana fire that occurred in Corsica in August 2017.

The method for generating an ensemble of wildfire simulations is introduced in the first section. In the second section, we define scores for ensemble evaluation. In the third section, we describe the data and perturbations used in the simulations of the fire event. The results are presented and discussed in the last section.

## **2. Generation of an ensemble**

### **2.1. Running one simulation of wildfire spread**

In this study, we used the ForeFire code, described in (Filippi et al., 2009). ForeFire is a front propagation solver based on discrete event simulation together with a Lagrangian front-tracking method where the fire line is described by markers linked to each other. Each marker represents a point in a 2D space that may move according to a given speed vector. The direction is determined by the local geometry of the fire line. On the other hand, the speed norm depends on a formula for rate of spread (ROS) that can be based on local weather, slope, and fuel properties, in addition to fire line geometry.

To start a simulation, the user needs to select not only a ROS formulation, ignition date, time and location, but also elevation, weather and fuel data so that ROS can be computed at any time of the fire spread and at any point in space around ignition point. However, when a fire starts, knowledge regarding these input variables is limited. This uncertainty makes the prediction of wildfire spread especially difficult as a relatively small change on some of the inputs could lead to significantly different results. Here, we decided to focus only on the inputs, but it should be noted that uncertainty also stems from numerical approximation and from the ROS model which is based on simplification assumptions and therefore does not describe perfectly the physics of the phenomenon.

### **2.2. Running an ensemble of simulations**

In a context of high uncertainty, only one simulation may give a unreliable prediction, even if the forecaster chooses the most likely input variables. Another strategy is to make a choice based on several simulations of fire spread that were computed with different inputs, even though these inputs are not all the most likely ones.

Here, we assume that we have a nominal value  $x$  for a given input variable. Depending on our knowledge on the variable, we might want to try a simulation with another value  $x_p$  (the underscript  $p$  stands for “perturbed”) that remains relatively close to  $x$ . Also, given the physical parameter described

by the variable, some values are not admissible for  $x_p$ . One obvious constraint is that many variables cannot be negative. Therefore, choosing a perturbation of wind speed norm in a range of +/- 5 m/s when the nominal value (obtained by a weather forecast, for instance) is 3 m/s is not possible.

In this study we define 3 main classes of perturbations: “*additive*”, “*percentage*” and “*multiplicative*”. Let  $z$  be our perturbation coefficient. The perturbed value  $x_p$  will be computed based on  $x$  and  $z$  according to the perturbation class.

For an additive perturbation, we have  $x_p = x + z$  where  $z$  has the same unit as  $x$  and is real (i.e., it can be either positive or negative).

For a percentage perturbation, we have  $x_p = x * (1 + z)$  where  $z$  has no unit and must be between -1 and 1 ( $z = -0.2$  corresponds to -20%, for instance).

For a multiplicative perturbation, we have  $x_p = x * z$  where  $z$  has no unit and must be positive.

The sampling of perturbations for one input variable is made according to some probability distribution. A given  $z$  is therefore conceived as a realization of a random variable  $Z$ . In most cases,  $Z$  can be described by a probability density function (PDF). In this study, we decided to use mainly “centered uniform” PDFs, which are described by only one parameter. Let  $z_{max}$  be the maximum value that  $z$  can take. For the additive and percentage classes,  $z$  is directly sampled according to a uniform distribution in  $[-z_{max}, z_{max}]$ . In the multiplicative case, we have  $z = \exp(u)$ , where  $u$  is sampled uniformly in the interval  $[-\log(z_{max}), \log(z_{max})]$ . This way,  $z$  is in  $[1/z_{max}, z_{max}]$ . We can talk about a (centered) “log-uniform” distribution by analogy to the log-normal distribution.

In this framework, a perturbation  $Z$  associated to a given input variable  $x$  is entirely described by its class and one parameter  $z_{max}$ .

To run an ensemble of simulations, for a given input variable we simply draw  $M$  independent samples  $z_1, \dots, z_M$  following the distribution of  $Z$ . Then, for  $m$  in  $\{1, \dots, M\}$ , we run the fire spread simulation with  $x$  perturbed by  $z_m$ .

In the case of wildfires, we define a fourth perturbation class specifically for fuel types: “*transition*”. Indeed, physical parameters of the fuel are usually determined by its type. This perturbation allows a given fuel type to be changed to another type (and therefore, receive new values of physical fuel parameters) with some probability  $p$  in  $[0, 1]$ . We have the constraint that all transition probabilities for a given fuel type (including transition towards itself, i.e., “no change”) are non-negative and sum to 1.

Choosing the probability distributions of the perturbations is a way of quantifying the uncertainty regarding the input variables.

#### Outputs of the ensemble of simulations

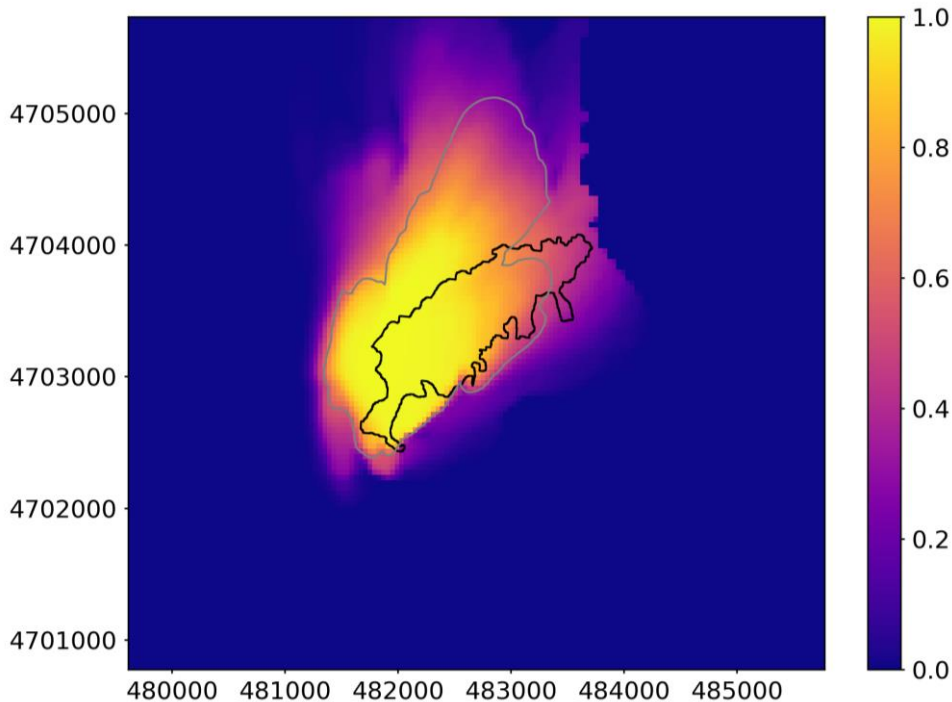
At the end of a simulation, ForeFire may return the geometry of the fire line, potentially at any time during the fire event, as well as an arrival time of the fire at any point in space. Since the observed data available when studying a fire event consists mostly of a 2D burned surface observed after the fire has completely stopped, we decided to focus on the fire line obtained at the end of the simulation. In the following, a simulation domain is described by a set of  $N$  points  $x_1, \dots, x_N$  where each  $x_i$  is associated to a number  $b_i$  that takes the value 1 if  $x_i$  is burned, 0 otherwise.

In a probabilistic framework, a set  $(b_1, \dots, b_N)$  associated to a given domain can be understood as a realization of a random vector  $(B_1, \dots, B_N)$  where each component  $B_i$  follows a Bernoulli law of parameter  $P[B_i = 1]$ , which represents the burn probability (conditionally to fire ignition).

With our ensemble of  $M$  simulations, since the inputs are generated independently, we can give a Monte Carlo estimate of  $P[B_i = 1]$ :

$$p_i = (\text{Number of simulations where } b_i = 1) / M.$$

We define the *burn probability map* for a given fire as the set  $\{(x_i, p_i) \mid i = 1, \dots, N\}$ . It is one of the main products of the ensemble of simulations. An example is given in Figure 1.



*Figure 1 - Burn probability map obtained for Calenzana (2017.08.05) wildfire.  
The colorbar gives the value of the burn probability  $p_i$ . Axes give the point coordinates  $x_i$  in EPSG:32632 (m).  
Black line: contour of observed burned area (obtained from Office de l'Environnement de la Corse).  
Grey line: contour of the burned area returned by a simulation without perturbation.*

### 3. Evaluation

Now that we have proposed a method for ensemble generation, we need to define desirable properties of an ensemble.

The ultimate goal of a forecaster is to give a perfect prediction of what is to come. For an ensemble, this would mean that all simulation results are the same as the observations, in any case. However, not only is it hard (if not impossible) to come close to this goal, but making an ensemble of simulations would be irrelevant in that case as only one forecast would be enough.

A significant difficulty when dealing with an ensemble is that one needs to compare a given observation to a set of simulations that yields a probability. The comparison between two items of different nature, namely the realization of an event and a corresponding probability, is not trivial. That is why in this section we define scores commonly used in meteorology that enable to evaluate the quality of an ensemble.

#### 3.1. Brier score

In order to compare an observation with one (or more) forecast(s), numerical scores can be computed. If we had just one forecast, some surface comparison scores such as those defined in Filippi et al. (2013) could be employed. However, we cannot compare a burn probability map with a burned surface by the means of such scores as they are meant for comparison with one simulation. For a numerical comparison with an ensemble forecast, we need to use probabilistic scores.



A score that seems appropriate in the case of a wildfire which is essentially binary (i.e., “burn or no burn”) is the *Brier score* (BS). Let us consider the *burn event* : “A given elementary surface burns”, we note  $o = 1$  if the event occurred, 0 otherwise. Let  $p$  be the probability of the *burn event* that was predicted by our ensemble. In this case, the Brier score is defined as  $BS = (o - p)^2$ .

In the general case, if we have  $N$  observations  $o_1, \dots, o_N$  for the burn event and  $N$  associated predicted probabilities  $p_1, \dots, p_N$ , then BS is defined as follows:

$$BS = \frac{1}{N} \sum_{i=1}^N (o_i - p_i)^2.$$

This score is positive and negatively oriented: the closer it is to 0, the better the ensemble forecast. For a detailed presentation of probabilistic scores, we refer the reader to Brocker (2009). In particular, the Brier score has the advantage of being strictly proper.

Using the notations of section 2.3,  $o_i$  will be the value of the observed  $b_i$ .  $N$  is the number of points in the study area and  $p_i$  is the Monte Carlo estimate of the burn probability  $P[B_i = 1]$  obtained with the ensemble.

The attentive reader might notice that proceeding this way, we have one observation for each one of the  $N$  burn events and not  $N$  observations of one burn event. The Brier score was introduced with the latter view, which corresponds to a probabilistic expectation. The way we propose to compute it, the Brier score corresponds to a spatial average.

### 3.2. Rank histogram

Another tool used in ensemble forecasting is the rank histogram. The rank is a variable used to test whether the observation is distinguishable from the predictions or not. When the distinction cannot be made, we say that the ensemble prediction system is *consistent*.

Let us define a variable  $y$  that corresponds to the observation and  $y_1, \dots, y_M$  that correspond to the predictions obtained from the simulations. The rank  $R(y)$  of  $y$  among  $y_1, \dots, y_M$  is usually defined as follows: if there are exactly  $m$  of the  $y_j$  that are strictly lower than  $y$  (and therefore the remaining  $M - m$  of the  $y_j$  are equal to or higher than  $y$ ), then  $R(y) = m$ . The rank can therefore take values in  $\{0, 1, \dots, M\}$ . Let us give a few examples. If all the  $y_j$  are strictly lower than  $y$ , then  $R(y) = M$ . Conversely, if  $y$  is lower than all the  $y_j$ , then  $R(y) = 0$ . If only  $y_3$  and  $y_5$  are strictly lower than  $y$ , then  $R(y) = 2$ . One interesting property of the rank is that if we evaluate it on many occasions and if the ensemble prediction system is consistent, each value of the rank (i.e., each integer in the set  $\{0, 1, \dots, M\}$ ) should be obtained approximately  $100 / (M + 1) \%$  of the time.

In most applications,  $y, y_1, \dots, y_M$  are defined in a real interval and the case where  $y$  is equal to one or more of the  $y_j$  is not problematic. However, in the context of forest fires we may consider the burn event and the variables can only take two values : 0 or 1. For instance, in the case where  $M = 10$ , if 7 of the  $y_j$  are equal to 0 (the point is not burned in 7 out of 10 simulations) and  $y = 0$  (we observe that the fire did not reach this point), one may wonder which rank to pick in  $\{0, \dots, 7\}$ . In fact, it is not satisfactory to pick one rather than the others. This is why, in the binary case, we follow a different strategy. Instead of considering a single value for the rank, we define a *weight* for all the possible ranks in  $\{0, 1, \dots, M\}$ .

The weights of the rank of  $y$  among  $y_1, \dots, y_M$  are given by the vector  $(w_0, w_1, \dots, w_M)$  of size  $M+1$ . Assuming that the variables  $y, y_1, \dots, y_M$  can only take the value 1 or 0 (“burned” or “not burned”), this vector is defined as follows:

Let us assume that  $\sum_{j=1}^M y_j = m$  (i.e., exactly  $m$  of the  $y_j$  are equal to 1).

Then if  $y = 0$ ,  $\forall k \in \{0, \dots, M - m\}$ ,  $w_k = \frac{1}{M+1-m}$ .

Else (if  $y = 1$ ),  $\forall k \in \{M - m, \dots, M\}$ ,  $w_k = \frac{1}{m+1}$ .

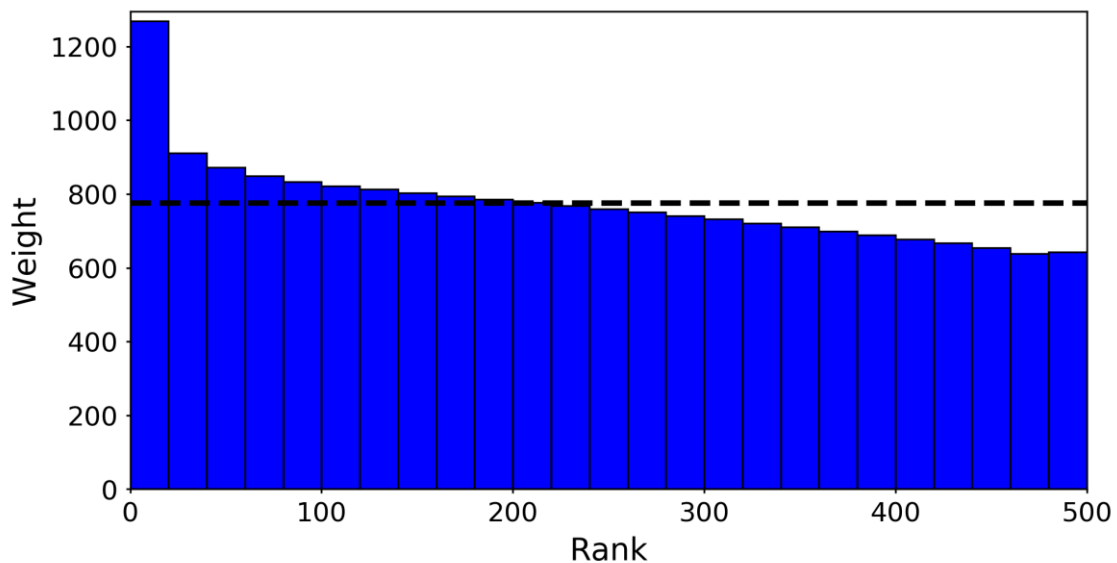
In both cases, for the remaining values of  $k$ ,  $w_k = 0$ .

In practice, we proceed with this calculation for each point  $x_i$ , which yields the corresponding weight vectors  $(w_0^i, w_1^i, \dots, w_M^i)$ . For each rank  $k$  in  $\{0, \dots, M\}$ , we define the total weight  $W_k$  as

$$W_k = \sum_{i=1}^N w_k^i,$$

and we obtain the total weight vector  $(W_0, W_1, \dots, W_M)$ . We define the *rank histogram* as the histogram where the total weight vector  $(W_0, \dots, W_M)$  is plotted against the rank vector  $(0, \dots, M)$ . An example is given in Figure 2.

If ideally we have a consistent ensemble prediction system and  $N$  is high enough, then each of the  $W_j$  should be approximately equal to  $N / (M + 1)$ . It implies that the rank histogram will be flat. When the ensemble over-predicts burn probability, there will be higher bars on the left and lower bars on the right (compared to the ideal flat histogram). Conversely, under-predicting the burn probability will lead to higher bars on the left and lower bars on the right.



*Figure 2 Rank histogram obtained for Calenzana (2017.08.05) wildfire. The 501 ranks are regrouped in 25 bins of 20 consecutive ranks (21 for the first bin). The dashed line indicates the flat histogram that would be obtained with a consistent ensemble. The shape of the diagram indicates that the ensemble overestimates burn probability.*

## 4. Application

### 4.1. ROS model and data sources

The area of study is Corsica island (France). Spatial data consist of an elevation field combined with a field of fuel type. The latter was obtained from Corine Land Cover (CLC) land use maps. Both fields are at raster format with a resolution of 80 m.

In this study, we decided to use the semi-empirical model of Rothermel (Rothermel, 1972), although any ROS model can be implemented (we could use the one described in (Balbi et al., 2009), for instance). A few points on the implementation of the model in this study must be noted. First, the

mineral damping coefficient is always 1. Second, all fuel types are assumed to have a moisture of extinction of 0.3.

For a given CLC fuel type, all ROS input parameters, except for wind speed, wind direction, and slope, are the same everywhere. Although the discrete event simulation method allows for change of any ROS input parameter in time, only wind speed, wind direction and dead fuel moisture were allowed to change during a fire simulation.

Spatial wind fields for the cardinal and intermediate directions were precomputed to take into account the elevation field and to save computational time. Wind speed, wind direction and dead fuel moisture data for a given wildfire are obtained from weather forecasts of the atmospheric model MesoNH. Forecasts are run everyday with an origin time of 00:00 UTC over a range of 42 hours with a time step of 3 hours. Spatial resolution of the fields is around 600 m.

## **4.2. Fire case**

The simulations were run based on a fire that occurred in Corsica on August 8<sup>th</sup> 2017. Fire started near the town of Calenzana, at approximately 17:42 (local time). Consequently, the MesoNH forecasts with origin date and time 2017.08.05, 00:00 UTC were used. The observed burned surface after the event was around 120 ha. The burned surface together with the corresponding (assumed) ignition time and location were provided by Office de l'Environnement de la Corse. We decided to quantify uncertainty regarding ignition time with an additive uniform perturbation of +/- 10 min for ignition time. For ignition location, we first defined a square of 200 m side length centered around the reference point. The ignition locations used in the simulations were randomly chosen among the centers of the CLC squares that were contained in the 200 m x 200 m square.

In general, one of the major sources of uncertainty regarding a fire event is its duration. Since we do not take into account the firefighting actions, ForeFire simulates a fire with free spread. Therefore, the only way for a simulation to end was either to reach the end of the input fire duration, or to have a ROS of almost 0 at every point of the fire line (which did not happen in our simulations). With this method, it is important to note that the simulation returns the burned surface of a fire that could continue to spread while the observation represents the burned surface of a fire that is not burning anymore. An important implication is that evaluation will highly depend on the prescribed fire duration. One way to address this difficulty is to choose a fire duration that accounts for most of the burned area. In the Calenzana case, although the fire spread was almost stopped 3 hours after ignition time, it seems that most of the surface was burned during the first hour. Therefore, we chose a reference duration of 1 hour with a uniform perturbation of 10 minutes.

## **4.3. Perturbation parameters**

The perturbations on ignition time, ignition location and fire duration were addressed in the previous section because they depend on our knowledge of the event, which may differ in another fire case.

For the other inputs, we used constant perturbation coefficients. It means for example that in a given simulation, even though wind speed may change in time, the same wind speed perturbation will be applied at the different steps of the simulation. Similarly, a perturbation on fuel parameters will be the same for all fuel types. Finally, a fuel type transition affects all the CLC cells with fuel type. For instance, it means that if a fuel transition from vineyards to sclerophyllous plants is applied in a simulation (which happens with a probability  $p$ ), ForeFire will behave as if all vineyard cells (and not just a proportion  $p$  of these cells) had become sclerophyllous plant cells.

For a given simulation, the different perturbation coefficients were sampled independently. Still, the same (percentage) perturbation was applied on the following fuel parameters: height, load, heat of combustion (positively) and particle density (negatively). For example, if we obtain a value of +15% for this global perturbation, then height, load and heat of combustion undergo a 15% increase while



the particle density decreases by 15%. From a physical perspective, there is no reason for the uncertainties regarding these four fuel parameters to be fully correlated. The motivation behind this choice is more pragmatic. Indeed, increasing either fuel height, fuel load or heat of combustion should increase the ROS, while increasing the particle density should decrease the ROS. Consequently, with that global fuel coefficient, we expect to obtain an upper bound on the uncertainty regarding the area of the simulated burned surfaces.

The range of the perturbations on wind was chosen after the assumed uncertainties listed in (Hanna et al., 1998) with the difference that we chose uniform PDFs instead of normal ones. As for perturbation on fuel parameters, it was based on common values encountered in the literature and expert judgment. The implementation of a fuel transition from vineyards (CLC type 221) to sclerophyllous plants (CLC type 323) was motivated by the presence of vineyards in our fuel data near the ignition point, where the ROS is much lower than for other fuel types. In this case, data were outdated, since there was no vineyard in this zone at the time of the fire. Transition allows to account for the classification error.

The distributions of the main perturbations that were used in this study are given in Table 1. We stress that at the moment the distributions are just an *a priori* suggestion. To obtain more reliable distribution parameters, some calibration method based on the Brier score and/or a Bayesian approach could be implemented. Calibration based on just one fire is not likely to give good evaluation results for other events, therefore it is out of the scope of this paper.

*Table 1 - Main perturbations used in the ForeFire simulations with the classification described in section 2.2.  
 \*For these parameters, the perturbation may differ depending on the fire event.*

<b>Perturbation</b>	<b>Type</b>	<b><math>z_{\max} / p</math></b>
Wind direction	Additive	40°
Wind speed norm	Multiplicative	1.5
Dead fuel	Percentage	30%
Surface-volume	Multiplicative	2
Fuel global	Percentage	20%
Ignition time*	Additive	10 min
Fire duration*	Additive	10 min
221 → 323	Fuel transition	0.8

Following these distributions, an ensemble of 500 simulations was run.

## 5. Results and discussion

The simulations were run on a Linux kernel with 8 cores, allowing for parallel computations. The whole computational process took about 5 minutes. It should be noted that given the relatively short fire duration, one simulation took a few seconds. But if we wanted to obtain a 3-hour probability map, simulation time would have increased. Also, in other fire events, higher fire duration and higher average ROS could lead to larger burned area, and therefore runtime could reach one minute or more for one simulation. In any case, this indicates that with appropriate computational resources to account for the simulation of potentially large fires (e.g., more cores to run the simulations in parallel), this

method has potential for use in an operational context. For instance, hourly burn probability maps over a few hours could be computed in a few minutes.

The burn probability map with coordinates in EPSG:32632 Mercator conformal projection system is represented in Figure 1 together with two lines, one encompassing the observed burned surface while the other encompasses the simulated burned surface obtained when there is no perturbation on the inputs. Qualitatively, the burn probability seem to have a better coverage of the observed burned area than the unperturbed contour. This analysis is supported by the Brier score, that we computed for the ensemble of 500 simulations and for an ensemble where the only member is the unperturbed simulation. It is much better in the ensemble case (0.055) than in the latter (0.072). It also appears that the predicted probability might be too high in some areas outside of the burned surface.

The rank histogram is represented in Figure 2. An aggregation was carried out so that each one of the 25 rectangles represents 20 ranks, except for the first one which is associated to 21 ranks. Without aggregation, the overall shape of the histogram remains similar. It appears that the histogram has higher bars on the left, which means that the burn probability is overestimated in several locations.

In spite of those promising results, a few difficulties regarding the methodology need to be discussed.

First, the choice of the number  $M$  of members in the ensemble is not obvious. Since we used a Monte Carlo method, we know that a given burn probability estimator  $p_i$  converges as  $M$  grows and confidence intervals can be computed for this estimator. However, to our knowledge, there is no reference for confidence levels in this context. Also, as  $M$  increases, computational time increases as well, which will limit  $M$  and the narrowness of the confidence intervals.

Second, many points in the study area were not burned and obtained a predicted burn probability of 0. If we were to extend the study area, the number of such points would increase. One implication is that the rank histogram would be flattened. Similarly, the Brier score would get lower because  $N$  would increase while the sum of the errors would be unchanged. In some sense, it is possible to cheat with a configuration where the burn probability is easy to predict over most of the domain (since most points are far from the ignition point), because in this case, evaluation will lead to better scores.

Finally, we noticed that some points outside of the observed burned surface obtained a burn probability of 1. A very plausible explanation is that the simulations did not model the firefighting actions. Still, this case can be viewed as a “failure” of the prediction system. The same can be said of the case where a point of the observed burned area receives a burn probability of 0, although it did not happen in this study. This raises the question of the relevance of extreme probabilities (i.e., close to 0 or close to 1) obtained via the ensemble. A possible way of circumventing this issue could be the use of probability thresholds (e.g., 0.1 and 0.9) on the burn probability maps. These thresholds could be interpreted as a consequence of a part of uncertainty that is not explained by the perturbations.

An application of this method on several cases of bigger fires is currently in process to make the evaluation more robust.

## 6. Acknowledgement

This research was developed within the ANR-16-CE04-0006 FIRECASTER.

## 7. References

- Balbi J-H, Morandini F, Silvani X, Filippi J-B, Rinieri F (2009) A physical model for wildland fires. *Combustion Flame* 156 (12), 2217-2230.
- Brocker J, (2009) Reliability, sufficiency, and the decomposition of proper scores. *Quarterly Journal of the Royal Meteorological Society* 135 (643), 1512-1519.

- Filippi J-B, Morandini F, Balbi J-H, Hill D (2009) Discrete event front-tracking simulation of a physical fire spread model. *Simulation* 86 (10), 629-644.
- Filippi J-B, Mallet V, Nader B (2013) Representation and evaluation of wildfire propagation of wildfire simulations. *International Journal of Wildland Fire* 2 (34), 2633-2640.
- Finney MA, McHugh CW, Grenfell IC, Karin LR, Short KC (2011a) A simulation of probabilistic wildfire risk components for the continental United States. *Stochastic Environmental Research and Risk Assessment* 25 (7), 973-1000.
- Finney MA, Grenfell IC, McHugh CW, Seli RC, Trethewey D, Stratton RD, Brittain S (2011b) A method for ensemble wildland fire simulation. *Environmental Modeling and Assessment* 16 (2), 153-167.
- Gollner M, Trouve A, Altintas I, Block J, De Callafon R, Clements C, Cortés A, Ellicott E, Filippi J-B, Finney M, Ide K, Jenkins MA, Jimenez D, Lautenberger C, Mandel J, Rochoux M, Simeoni A (2015) Towards Data-Driven Operational Wildfire Spread Modeling : A Report of the NSF-Funded WIFIRE Workshop. DOI: 10.13016/M2Z70R.
- Hanna SR, Chang JC, Fernau ME (1998) Monte Carlo estimates of uncertainties in predictions by a photochemical grid model (UAM-IV) due to uncertainties in input variables. *Atmospheric Environment* 32 (21), 3619-3628.
- Parisien MA, Kafka VG, Hirsch KG, Todd JB, Lavoie SG, Maczek PD (2005) Mapping wildfire susceptibility with the BURN-P3 simulation model. Natural Resources Canada, Canadian Forest Service, Northern Forestry Center, Information Report NOR-X-405 (Edmonton, Alberta).
- Rothermel R (1972) A mathematical model for predicting fire spread in wildland fuels. USDA Forest Service, Intermountain Forest and Range Experiment station, Research Paper RP-INT-115 (Ogden, UT).