

# Generation and evaluation of an ensemble of wildland fire simulations

Frédéric Allaire, Jean-Baptiste Filippi, Vivien Mallet

► **To cite this version:**

Frédéric Allaire, Jean-Baptiste Filippi, Vivien Mallet. Generation and evaluation of an ensemble of wildland fire simulations. *International Journal of Wildland Fire*, CSIRO Publishing, 2020, 29 (2), pp.160-173. 10.1071/WF19073 . hal-02447187

**HAL Id: hal-02447187**

**<https://hal.inria.fr/hal-02447187>**

Submitted on 21 Jan 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Generation and evaluation of an ensemble of wildland fire simulations

Frédéric Allaire<sup>A,\*</sup>, Jean-Baptiste Filippi<sup>B</sup>, and Vivien Mallet<sup>A</sup>

<sup>A</sup> Institut national de recherche en informatique et en automatique (INRIA), 2 rue Simone Iff, Paris, France; Sorbonne Université, Laboratoire Jacques-Louis Lions, France.

<sup>B</sup> Centre national de la recherche scientifique (CNRS), Sciences pour l'Environnement – Unité Mixte de Recherche 6134, Università di Corsica, Campus Grossetti, Corte, France.

\* Email: frederic.allaire@inria.fr

## Contents

<b>1 Introduction</b>	<b>3</b>
<b>2 Probability distribution of wildfire spread simulations</b>	<b>9</b>
2.1 Fire spread simulation . . . . .	9
2.2 Modeling uncertainty in input data . . . . .	10

2.3	Uncertainty propagation in fire spread simulations . . . . .	13
<b>3</b>	<b>Probabilistic evaluation</b>	<b>15</b>
3.1	Accuracy . . . . .	16
3.2	Reliability and sharpness . . . . .	18
3.3	Probabilistic resolution . . . . .	19
3.4	Consistency . . . . .	20
3.5	Evaluation domain . . . . .	22
<b>4</b>	<b>Application to seven Corsican fires</b>	<b>23</b>
4.1	Data sources . . . . .	23
4.2	Probability distributions . . . . .	26
4.3	Complementary data and error corrections . . . . .	31
4.4	Computational set-up . . . . .	33
<b>5</b>	<b>Results</b>	<b>34</b>
5.1	Detailed case: Calenzana fire . . . . .	34
5.2	Performance for all seven fires . . . . .	40
<b>6</b>	<b>Conclusions</b>	<b>44</b>

**Abstract:** Numerical simulations of wildfire spread can provide support in deciding firefighting actions but their predictive performance is challenged by the uncertainty of model inputs stemming from weather forecasts, fuel parameterization and other fire characteristics.

In this study, we assign probability distributions to the inputs and propagate the uncertainty by running hundreds of Monte Carlo simulations. The ensemble of simulations is summarized via a burn probability map whose evaluation based on the corresponding observed burned surface is not obvious. We define several properties and introduce probabilistic scores that are common in meteorological applications. Based on these elements, we evaluate the predictive performance of our ensembles for seven fires that occurred in Corsica from mid-2017 to early 2018.

We obtain fair performance in some of the cases but accuracy and reliability of the forecasts can be improved. The ensemble generation can be accomplished in a reasonable amount of time and could be used in an operational context provided that sufficient computational resources are available. The proposed probabilistic scores are also appropriate in a calibration process to improve the ensembles.

**Keywords:** Monte Carlo, probabilistic score, uncertainty quantification.

## 1 Introduction

Wildfire simulation can be useful in decision support systems in order to estimate how a wildfire will spread right after ignition and anticipate actions [Sullivan(2009)]. Before ignition, it is also used to assess fire likelihood

and help in distributing firefighting resources. Possibly, in the long term, it can indicate what are the areas that require high-priority land planning actions.

However, the ability to make accurate predictions is not trivial. In wild-fire management, there are significant uncertainty sources that may lead to considerable difficulties in determining the most appropriate decisions in an operational context [Thompson *et al.*(2017)]. Generally, "uncertainty" means that there is a lack of knowledge and/or information. Uncertainty assessment may cover the whole decision process in wildfire management. As this study deals with the prediction of surface wildland fire spread, we only focus on the uncertainty in the simulation of wildland fire front dynamics. Modeling (and, consequently, predicting) wildfire spread is a challenging task and variability in the performance of predictions in an operational context can be attributed to an incomplete or over-simplified formulation of the underlying physical and numerical model. This issue is hard to overcome because of two major sources of uncertainty: variability of the environmental conditions and incomplete knowledge the fire spread. Under these conditions, relying on a deterministic prediction of wildland fire spread with no estimation of the error may strongly limit the relevance of the forecasts.

Uncertainty in wildland fire simulations is typically quantified by attributing a probability distribution to the model inputs and propagating it by means of Monte Carlo (MC) methods or more sophisticated approaches. Uncertainty propagation can be performed at different levels ranging from

the rate of spread (ROS) model (e.g., [Salvador *et al.*(2001), Cruz(2010), Liu *et al.*(2015b), Liu *et al.*(2015a), Ervilha *et al.*(2017)]) to fire spread simulations at regional scale [Finney *et al.*(2011a), Parisien *et al.*(2005), Paz *et al.*(2011), Salis *et al.*(2013), Lautenberger(2017)]. Depending on the level of the study, the identification of the uncertain inputs and their probability distributions may differ.

At the level of the ROS model, a typical goal is to identify the inputs that have the most influence on the output of the model. Global sensitivity analysis methods for Rothermel’s rate of spread model [Rothermel(1972)] have been carried out either on one fuel model [Salvador *et al.*(2001), Liu *et al.*(2015b)] or more [Liu *et al.*(2015a)] where wind speed norm and direction, slope, fuel moisture content and most (if not all) fuel model parameters are considered as uncertain inputs. In [Liu *et al.*(2015b), Liu *et al.*(2015a)], all inputs are assumed to be independent and the authors use uniform distributions for which the mean is a reference value and the standard deviation is taken as 5% of the mean. In [Salvador *et al.*(2001)], the authors use data mostly from the scientific literature, field measurements and theory to assign probability distributions to the uncertain inputs, leading to normal, lognormal and “empirical” (i.e., sampled from a large database) distributions that account for the variability in the inputs of Rothermel’s model in shrubland fuels. More recently, Cai *et al.* [Cai *et al.*(2019)] carried out a sensitivity analysis on dynamic fire spread simulations. The authors focused on the uncertainty of fuel model parameters while the other inputs were based on data from an actual

fire used as case study; as for the output, the authors computed the means of some outputs of Rothermel’s model over a simulation of fire spread.

At the regional scale, ensembles of wildland fire simulations are used to generate maps of “burn probability” that represent the potential of a zone to be burned by a wildland fire during a year or a season. Among the systems generating such maps, FSim [Finney *et al.*(2011a)] is commonly used in the United States, and BURN-P3 [Parisien *et al.*(2005)] in Canada. Similar studies were conducted in the Mediterranean region [Paz *et al.*(2011), Salis *et al.*(2013)] and in California [Lautenberger(2017)]. In these studies, the uncertainty in elevation and fuel model parameters other than fuel moisture content is not modeled as a source of uncertainty and/or is considered negligible compared to other sources. Instead, the uncertain inputs are the location of the ignition point of a given fire, the fire weather scenario, and sometimes the duration of a fire. They are typically sampled based on statistical models and historical data of fire and weather records of the region.

The type of “burn probability” maps we are interested in for this study are those associated to a “crisis” situation, that is to say when a wildland fire has just started. In this case, burn probability indicates the probability that a zone will be reached by a specific fire whose time and location of ignition are quite well known. The burn probability map indicates that there is uncertainty in the prediction of wildfire spread, which is not the case when a single deterministic simulation is used. Still, some uncertainty sources (quality of the data, ROS model simplifications, etc.) are harder

to quantify than others. We chose to focus on uncertainty of model inputs (weather forecasts, fuel model parameters, etc.). Our strategy is similar to the approach used in FSPro [Finney *et al.*(2011b)] and other works (e.g., [Miller *et al.*(2015), Pinto *et al.*(2016)]) where the probabilistic prediction stems from an ensemble of fire spread predictions. Our method for the generation of an ensemble makes use of stochastic perturbations of some inputs of our fire spread simulator. The perturbations are sampled from the probability distributions that describe input uncertainty. Another particularity of our method is that the inputs are available in an operational context. The burn probability map obtained after the propagation of this uncertainty in simulations is meant to summarize the potential scenarios of wildfire spread.

A burn probability map must be evaluated, as should any forecast, to establish its credibility. It is fundamental to improve the system as it helps to determine which systems have the best performance. In wildland fire research, several methods have been proposed to compare an observed burned surface with its predicted counterpart. A typical method is to compute one or several metrics whose values will indicate how much the two surfaces match (e.g., [Duff *et al.*(2016), Filippi *et al.*(2013)]). More sophisticated methods exist, for instance some use the distance between the vertices of the fire perimeter [Fujioka(2002), Duff *et al.*(2012)] while others consist in computing scores based on information on the dynamics of the simulated and observed fire surfaces [Filippi *et al.*(2013)]. Evaluation of model per-



formance with such scores can be performed on fire cases by running the fire spread simulations using data known in hindsight, such as knowledge regarding fire suppression actions or observed (sometimes corrected) weather data (e.g., [Duff *et al.*(2018), Salis *et al.*(2016)]). Another possibility is to run the simulation based on data that are available at the time of fire start (e.g., [Filippi *et al.*(2014)]). The former alternative makes use of inputs that are more representative of what occurs during the fire and measures the potential of the fire spread model, whereas the latter evaluates the predictive performance of the model under operational conditions. In this study we focus on the latter situation, where there is more uncertainty.

In any case, with probabilistic prediction systems, evaluation is not as direct as in the deterministic case because probabilities are not directly observable. The aforementioned metrics cannot be computed but it is still possible to compare probabilities with observations by means of probabilistic scores. Such scores are quite common in meteorology but, to our knowledge, their use in the development of a probabilistic forecast is new in wildland fire research.

This methodology is applied to seven big fires that occurred in Corsica between summer 2017 and early 2018. From the perspective of applicability in an operational context, particular attention is paid to the time required to run the simulations. Concerning data sources, all data used for the computations were recorded before or during the fires, and were available during the event, although they might not have been processed by firefighters in the

field.

The strategy for the generation of our probabilistic predictions is detailed in Section 2. Probabilistic evaluation tools adapted to the context of fire spread predictions are presented in Section 3. In Section 4, we mainly describe the features of the fires studied and the simulations. The results are presented and discussed in Section 5, both by focusing on one specific wildfire and by analyzing all studied fires with the evaluation tools.

## 2 Probability distribution of wildfire spread simulations

### 2.1 Fire spread simulation

In this study, we used ForeFire [Filippi *et al.*(2010)] to simulate fire spread. ForeFire is a front propagation solver based on discrete event simulation together with a Lagrangian front-tracking method where the fire front 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 of the vector is determined by the local geometry of the fire front. The norm of the vector is the ROS, which is estimated from the local weather, slope and fuel parameters, in addition to fire front geometry.

To run a simulation, the user needs to select a ROS formulation, ignition date, time, and location, but also elevation, weather, and fuel data so that

ROS can be computed potentially at any time after ignition and at any point in the simulation domain.

The most basic output of a simulation is the shape of the fire front at any instant after fire ignition.

## 2.2 Modeling uncertainty in input data

When a wildfire starts, knowledge regarding the present and future state of the environment is limited. The existing ROS models only give a simplified representation of the physics of fire behavior. The same goes for the corresponding fuel parameterization which only focuses on certain aspects of the vegetation. Moreover, spatial heterogeneity and temporal variability can only be described to a limited extent owing to the spatial and temporal resolutions of the data. Even if interpolations are performed, some approximations are unavoidable. Additionally, in a crisis context, the time and location of fire ignition are not known perfectly either. Most importantly, the *future* state of the environment, especially wind speed, is highly uncertain even if weather forecasts are available. The input data may be inaccurate themselves, but even accurate data could lead to error in the prediction of the ROS at a given time and location. An accumulation of errors in a single simulation could result in an unsatisfactory prediction of wildfire spread, even if the forecaster chose the most likely values for the model inputs. Another strategy available for the forecast consists in running several simulations of fire spread that are based on different inputs, even though they are not all the most likely.

Here, we assume that we have a nominal value  $x$  for a given input variable. Given the uncertainty in the variable, one can be tempted to try a simulation with another value  $x_p$  (where subscript  $p$  stands for "perturbed"). Depending on the physical quantity described by the variable, some values are not admissible for  $x_p$ . One obvious constraint is that many variables cannot be negative. Therefore, if the reference wind speed norm is  $15 \text{ m s}^{-1}$ , a perturbation of  $-5 \text{ m s}^{-1}$  is admissible, but not if the reference value is  $3 \text{ m s}^{-1}$ . An absolute perturbation that does not depend on  $x$  is not appropriate for all the inputs, so we need to consider relative perturbations. There are also qualitative inputs that cannot be modified by means of arithmetic operations. Therefore, in this study, we define three classes of perturbations : *additive*, *multiplicative* and *transition*. The first two classes can be applied to scalar inputs. If we denote  $z$  as the perturbation coefficient, the perturbed value  $x_p$  will be computed based on  $x$  and  $z$  as follows:

- for an additive perturbation,  $x_p = x + z$ , where  $z$  has the same unit as  $x$  and can be either positive or negative,
- for a multiplicative perturbation,  $x_p = x \times z$ , where  $z$  has no unit and must be positive.

For a transition perturbation, as the name implies, we have a transition from  $x$  to  $x_p$ . The choice of  $x_p$  is made among the (finite set of) possible values for the input.

Then, we use probability distributions to quantify the uncertainty in the

inputs. First, we assume that a perturbation in one input is independent from the other perturbations. For a given simulation, the perturbation on a given input is sampled according to its marginal probability distributions. With the first two perturbation classes, a given  $z$  is therefore conceived as a realization of a random variable  $Z$ , with a given probability density function (PDF). In this study, we use truncated normal distributions for the additive and most multiplicative perturbations. More precisely, the underlying Gaussian PDF having  $\mu$  and standard deviation  $\sigma$ , we choose  $\mu = 0$  (respectively,  $\mu = 1$ ) for additive (resp, multiplicative) perturbations and the PDF is restricted to  $[\mu - 2\sigma, \mu + 2\sigma]$ . For the multiplicative perturbation on wind speed, we use a truncated log-normal distribution. The truncation is made so that the perturbation is sampled in  $[1/z_{max}, z_{max}]$ , with  $z_{max} > 1$ . Without the truncation,  $Y$ , the logarithm of  $Z$ , follows a normal distribution with 0 mean and standard deviation  $\sigma = 0.5 \log z_{max}$ . To account for the truncation, the Gaussian PDF is restricted to  $[-\log z_{max}, \log z_{max}]$ . In these cases, the distributions are symmetric, either directly or after taking the logarithm, and are described by just one uncertainty parameter  $\sigma$ . Finally, the probability distribution for a transition from a given value  $x$  is described by transition probabilities to each possible value (including  $x$  itself). Each transition probability from a given value must be between 0 and 1 and their sum must be equal to 1.

The use of normal and log-normal distributions is quite common in uncertainty quantification studies. Here, the distributions are symmetric and

their medians correspond to a case with no perturbation, which means we do not favor overestimation nor underestimation of the inputs. It also means we are fairly confident of the unperturbed inputs. Also, the use of truncation helps to avoid sampling extreme values that may be unrealistic. Other distribution shapes such as uniform, triangular or even empirical distributions are possible alternatives to represent input uncertainty.

For simplicity, we assumed there is no correlation between the perturbations. Although some parameters may be correlated, properly quantifying their correlation and that of their perturbations is not easy and requires a lot of data. Also, weak correlations may have little influence in the generation of the ensemble.

### **2.3 Uncertainty propagation in fire spread simulations**

To run a perturbed simulation, we sample a set of perturbations according to the probability distributions assigned to the inputs, we apply it to the available input data, and the simulation is run based on the perturbed inputs. An ensemble of  $n$  simulations is simply obtained by repeating this process  $n$  times with independently sampled sets of perturbations. Each simulation returns the geometry of the fire front at several instants in time. These  $n$  scenarios of fire spread may be looked at individually, but analyzing each scenario in detail can be time-consuming and may prove counter-productive in an operational context.

To summarize the information provided by the  $n$  ensemble members, we

consider a 2D spatial domain that covers a sufficiently wide area around the presumed ignition point. For any point  $x_i$  in this map, we can assign a number  $b_i$  equal to 1 if we predict that this area will burn, 0 otherwise for any instant after ignition. The value of  $b_i$  will depend on the member chosen among our ensemble. In our probabilistic framework, a set  $(b_1, \dots, b_N)$  can be understood as the realization of a random vector  $(B_1, \dots, B_N)$  where each component  $B_i$  follows a Bernoulli law of parameter  $q_i = \mathbb{P}[B_i = 1]$  that represents the predicted burn probability at the point  $x_i$ .

For a given  $i$ , the exact value of  $q_i$  is unknown but it can be approximated based on the different values of  $b_i$  in our ensemble, which are independent. We have the following estimate of  $q_i$ :

$$p_i = \frac{n_i}{n}, \tag{1}$$

where  $n_i$  is the number of simulations for which  $b_i = 1$ .  $p_i$  is the MC estimate of  $q_i$  and converges to it as  $n$  increases. It is possible to quantify this convergence. For all  $\alpha \in ]0, 1[$ , we have the following inequalities:

$$\mathbb{P} \left[ |p_i - q_i| \leq \left( \frac{q_i(1 - q_i)}{\alpha n} \right)^{1/2} \right] \geq 1 - \alpha, \tag{2}$$

and

$$\mathbb{P} \left[ |p_i - q_i| \leq \left( \frac{\log(2/\alpha)}{2n} \right)^{1/2} \right] \geq 1 - \alpha \tag{3}$$

that yield confidence intervals with a confidence level of  $100(1 - \alpha)\%$  (at

least). Equation (2) relies on Bienaymé-Tchebychev inequality and requires the knowledge of  $q_i$  to compute the intervals, of which we only have an approximation. In place of the term  $q_i(1 - q_i)$ , which is unknown, we could use its upper bound  $1/4$ . By doing so, the inequality still holds, but the length of the interval may be overestimated. However, the interval resulting from Equation (3), which relies on Hoeffding’s inequality, does not depend on  $q_i$  and can be computed easily.

In the following, we call  $p_i$  the *burn probability*. We can produce the 2D field of burn probabilities, thereafter called the *burn probability map*, based on  $x_i$  and  $p_i$ . This map is the main output of the ensemble of simulations. A location  $x_i$  with higher burn probability  $p_i$  indicates that, based on our fire spread model and the associated input uncertainties, we estimate it is more likely that this location will burn. It is possible to compute a burn probability map at different times. Nonetheless, because the observed data available when studying a fire consists mostly of a 2D burned surface observed after the fire has completely stopped, we decided to focus on the burn probability map based on the end result of each simulation.

### 3 Probabilistic evaluation

Now that we can generate an ensemble, we need a method to evaluate its performance. It is possible to evaluate each simulation output individually based on deterministic scores (e.g., as in [Filippi *et al.*(2014)]) . However,



it is highly unlikely that each prediction will be analyzed in detail in an operational context, especially if  $n$ , the ensemble size, is high. For decision support, we will look, rather, at the burn probability map. Its comparison with the corresponding observed burned surface is not trivial because we are dealing with two items of different natures, namely the realization an event and its predicted probability.

In this section, we define some desirable properties of an ensemble as well as tools that are suited to the evaluation of such properties. Most of them are suited to any probabilistic prediction system, except for consistency, which is only relevant when the system relies on an ensemble of predictions. We refer the reader to [Wilks(2011)] for a general overview in the field of atmospheric sciences.

For any event  $A_i$  (i.e., in most cases, the event ‘location  $x_i$  was burned’), we define the occurrence variable  $o_i$  as follows:  $o_i = 1$  if the event occurs,  $o_i = 0$  otherwise. The forecast probability associated with  $A_i$  is denoted as  $p_i$ .

### 3.1 Accuracy

For any prediction system, accuracy corresponds to the overall agreement between predictions and observations. If we evaluate  $N$  distinct events  $A_1, \dots, A_N$ , then accuracy can be measured by the Brier score, denoted as

$BS$  and defined as follows:

$$BS = \frac{1}{N} \sum_{i=1}^N (o_i - p_i)^2. \quad (4)$$

This score ranges between 0 and 1 and is negatively oriented: the lower it is, the better the prediction. If we have a prediction system whose Brier score is  $BS_{ref}$ , it is common to compare the accuracy of the new prediction system with the Brier skill score ( $BSS$ )

$$BSS = 1 - \frac{BS}{BS_{ref}}. \quad (5)$$

$BSS$  is lower than 1. The more accurate the new prediction system, the higher  $BSS$ . If  $BSS$  is positive, it means that  $BS \leq BS_{ref}$ , i.e., the new prediction is more accurate than the reference.

Among the systems that assign the same forecast probability for all events, it is easy to prove that the one that minimizes (4) is obtained *a posteriori* with the probability

$$p_c = \frac{1}{N} \sum_{i=1}^N o_i. \quad (6)$$

A typical reference used for comparison is the one based on the optimal constant probability  $p_c$  from Equation (6). In this case, the Brier score is equal to  $p_c(1 - p_c)$ .

## 3.2 Reliability and sharpness

Probabilities can be interpreted in different ways. With a frequentist interpretation, assuming that we test the occurrence of an event with probability  $p$  on several occasions, the proportion of trials where the event occurs is approximately equal to  $p$ . Based on this interpretation, a desirable property of a probabilistic prediction system is *reliability*. For  $p$  in  $[0, 1]$ , we define the relative frequency  $f(p)$  as the proportion of the events that occurred among the events for which the system assigned a forecast probability  $p$ . A prediction system is reliable if,  $\forall p \in [0, 1], f(p) = p$ . For instance, with a reliable prediction system, among the events that obtained a probability of  $p = 0.2$  according to the prediction system, exactly  $f(p) = 20\%$  of them actually occurred (i.e., were observed).

Reliability is assessed with a *reliability diagram*, which is simply the plot of  $f(p)$  against  $p$ . In an ideally reliable case, we obtain a curve that is overlaid on the first bisector. For a complete analysis, the reliability diagram should be displayed together with a *sharpness graph*. We define  $g(p)$  as the proportion of events that are assigned a probability  $p$  among all evaluated events. The sharpness graph is simply the plot of  $g(p)$  against  $p$ . When  $g(p)$  is too low,  $f(p)$  cannot be considered a good estimation of the frequency of occurrence in the observations. The reliability diagram and the sharpness diagram taken together represent the joint distribution of the pair observation-predicted probabilities  $(o, p)$ . Whereas  $g(p)$  indicates the distribution of  $p$ ,  $f(p)$  indicates the distribution of the conditional variable  $o|p$ . Therefore, to

get a full representation of the joint distribution between observation and forecast, the two figures are displayed together. Also  $g(p)$  quantifies the frequency at which the probability  $p$  is produced by the system, which can be helpful information in itself. For instance, the system could very rarely yield probabilities outside the range  $[0.3, 0.7]$ , which would be indicated by high bars at the center of the sharpness graph.

### 3.3 Probabilistic resolution

*Probabilistic resolution* is the capacity of the system to yield relative frequencies that are different from the reference probability  $p_c$ . Usually, the term “resolution” is used, but we specify “probabilistic resolution” to avoid confusions with spatial and temporal resolutions. If the system has high probabilistic resolution, it can therefore distinguish between different events (i.e., producing different probabilities in different conditions), instead of always producing the same probability  $p_c$ . Probabilistic resolution can be assessed with the reliability diagram by looking at the deviation between  $f(p)$  and  $p_c$ . As for reliability, the sharpness graph should also be taken into account when considering the influence of these deviations on accuracy as they will be more significant if  $g(p)$  is high. Murphy [Murphy(1973)] showed that the Brier score could be partitioned into a sum of three terms, pointing out the contribution of resolution and reliability to the accuracy of the forecast. In the case of our  $n$ -member ensemble, there are  $n + 1$  possible values for the predicted probabilities that we denote for all  $i \in \{0, 1, \dots, n\}$  as  $p^i = \frac{i}{n}$ ,

so with the notation introduced in the previous section the partition of the Brier score can be written as follows:

$$BS = \underbrace{\sum_{j=0}^n (p^j - f(p^j))^2 g(p^j)}_{\text{reliability}} - \underbrace{\sum_{j=0}^n (f(p^j) - p_c)^2 g(p^j)}_{\text{probabilistic resolution}} + \underbrace{p_c(1 - p_c)}_{\text{uncertainty}}. \quad (7)$$

The last term, *uncertainty*, is independent of the forecast and is representative of the variability of the observed events. If the system is reliable, Equation (7) becomes

$$BS = - \sum_{j=0}^n (p^j - p_c)^2 g(p^j) + p_c(1 - p_c), \quad (8)$$

and probabilistic resolution corresponds to the ability of the system to predict probabilities  $p$  that are far from  $p_c$ .

### 3.4 Consistency

The idea of consistency for an ensemble was presented by Anderson [[Anderson\(1997\)](#)] as follows: ‘if the verifying truth is indistinguishable from a randomly selected member of an ensemble over a large set of forecast cases, the ensemble forecasts are said to be consistent with the truth’. In particular, an ensemble whose members always forecast the observation perfectly should be consistent. However, care should be taken as consistency does not guarantee accuracy.

Consistency can be studied by the means of a rank histogram. However,

the rule that is commonly used to establish this histogram is not appropriate when the studied variables can only take two values. We propose a definition that is suited to a binary context.

For the event  $A_i$ , based on the observation  $o_i$  and the ensemble probability  $p_i = \frac{n_i}{n}$ , we define  $w_{i,j}$  the weight of rank  $j$  as follows:

$$\text{if } o_i = 0 : \begin{cases} \forall j \in \{0, \dots, n - n_i\}, & w_{i,j} = \frac{1}{n - n_i + 1}, \\ \forall j \in \{n - n_i + 1, \dots, n\}, & w_{i,j} = 0, \end{cases}$$

$$\text{if } o_i = 1 : \begin{cases} \forall j \in \{0, \dots, n - n_i - 1\}, & w_{i,j} = 0, \\ \forall j \in \{n - n_i, \dots, n\}, & w_{i,j} = \frac{1}{n_i + 1}. \end{cases}$$

If we define the vector  $w_i = (w_{i,0}, \dots, w_{i,n})$ , we obtain the following representation:

$$w_i = \begin{cases} \left( \begin{array}{ccccccc} \frac{1}{n - n_i + 1} & \dots & \dots & \frac{1}{n - n_i + 1} & 0 & \dots & 0 \end{array} \right) & \text{if } o_i = 0, \\ \left( \begin{array}{ccccccc} 0 & \dots & 0 & \frac{1}{n_i + 1} & \dots & \dots & \frac{1}{n_i + 1} \end{array} \right) & \text{if } o_i = 1. \end{cases} \quad (9)$$

$$\begin{array}{ccccc} \longleftarrow & \longleftrightarrow & \longleftrightarrow & & \\ & & & & \\ & n - n_i & 1 & n_i & \end{array}$$

Finally, to make an overall evaluation based on all the events  $A_1, \dots, A_N$ , we sum the weights of each event by computing  $y_j = \sum_{i=1}^N w_{i,j}$ . The *rank histogram* of the ensemble forecasts is obtained by plotting  $y_j$  against  $j$  for

$j \in \{0, \dots, n\}$ .

Although the definition is different than in the usual scalar case, its main property remains: a consistent ensemble yields a flat rank histogram. However, the reverse is not necessarily true. This means that even if we obtain a quite flat histogram, it does not guarantee that the ensemble is consistent. However, deviations from an ideal flat histogram may reveal biases in the prediction system. The typical cases are different from the ones encountered when we deal with continuous variables. With binary events, when the ensemble has a tendency to assign probabilities that are too high, we will observe higher bars for low values of the rank  $j$ . Conversely, a tendency to forecast probabilities that are too low will result in higher bars for high values of  $j$ .

### 3.5 Evaluation domain

For each fire case in the present study, we defined a rectangular spatial domain covering the observed burned surface. This domain was chosen rather large: its area was at least 10 times bigger than that of the observed burned surface. Still, it was chosen independently of the simulation results, so some simulations could go beyond the boundaries of the domain. Then, it was divided into cells with a side length of approximately 20 m. The centers of these cells formed a regular grid of  $N$  points:  $x_1, \dots, x_N$ . The elementary evaluation event was  $A_i$ : ‘ $x_i$  belongs to the burned surface’.

## 4 Application to seven Corsican fires

### 4.1 Data sources

The fuel layer data are derived from Corine Land Cover (CLC) data [Feranec *et al.*(2016)] coupled with data from the IGN (Institut Géographique National) product BD TOPO<sup>®</sup> for road and drainage networks. The Digital Elevation Model (DEM) is extracted from another IGN product: BD ALTI<sup>®</sup> which has a 25-m resolution. The area of study is the island of Corsica (France) for which the fields were combined and formatted for fire simulation in raster format at 80-m resolution.

Although it is possible to implement any ROS formula with ForeFire, in this study we only used the semi-empirical model of Rothermel [Rothermel(1972)]. Several points on the implementation of the model in the present study must be noted. First, we used a wind adjustment factor of 0.4 and implemented the wind limit function recommended by Andrews [Andrews *et al.*(2013)] instead of the one originally proposed by Rothermel. Second, the mineral damping coefficient is always 1. Finally, the moisture of extinction is set to 0.3 for all fuel types.

Some of the fuel types in the CLC classification were aggregated. More precisely, types 211, 221, 222, 223, 241, 242, 243, and, 244, corresponding to various crop types, were regrouped as one type that we called ‘Agricultural’. We made a correspondence between the CLC classification and fuel models derived from those proposed by Scott and Burgan [H Scott and E Burgan(2005)]



to determine the values of the fuel parameters, as summarized in Table 1.

Spatially high resolution wind fields for the cardinal and intermediate directions were precomputed with the mass conserving preconditioner from the atmospheric forecasting system Meso-NH [Lac *et al.*(2018)] to take into account orographic effects and save computational time. For a fire case, the forcing values to get high resolution wind speed, wind direction, and dead fuel moisture were derived from Meso-NH Local Area Model forecasts initialized from the French national AROME model [Termonia *et al.*(2018)]. Although these fields have a spatial resolution of 600 m, only the values at the presumed ignition point were used. The meteorological forecasts are run everyday with an origin time of 00:00 UTC over a range of 42 hours, and only the outputs every three forecast hours are stored. Here, only atmospheric model outputs corresponding to the time frames of the ForeFire simulations were used. The forecasts require approximately 12 hours to be computed (from the availability of large-scale model output to the delivery of the output data). To be representative of the conditions in an operational context, only data that were (or could have been) available at the time of the fire were used in these simulations. The main consequence is that for a fire that started in the morning of day  $D$ , the weather forecast used to run the simulations was the one whose origin date was  $D - 1$ . This does not apply for fires that occurred after midday.

Some information regarding the fires was derived from the Prométhée database (<http://www.promethee.com/>, [verified 26 November 2019]), a French

CLC	Model	$S_v$ ( $\text{m}^{-1}$ )	$e$ (m)	$\sigma_d$ ( $\text{kg m}^{-2}$ )	$\rho_d$ ( $\text{kg m}^{-3}$ )	$\Delta H$ ( $\text{MJ kg}^{-1}$ )
141	GS2 <sup>A</sup>	6562	0.46	0.22	512	18.6
Agricultural	GR2 <sup>A</sup>	6560	0.30	0.13		
231	GR6 <sup>A</sup>	7218	0.45	0.40		
311	TU5	4922	0.30	0.90		
312	TU5	4922	0.30	0.90		
313	TU5	4922	0.30	0.90		
321	GR2 <sup>A</sup>	5905	0.46	0.54		
322	SH5 <sup>A</sup>	2460	1.20	0.80		
323	SH5	2460	1.80	0.80		
324	SH5	2460	1.80	0.80		
333	GR2	6560	0.30	0.20		
334	GR2	6560	0.30	0.20		
412	GR3	4922	0.60	0.02		

Table 1: Fuel parameterization.

Corresponding Corine Land Cover (CLC) classification and fuel models from [H Scott and E Burgan(2005)] (<sup>A</sup>: with slight alterations).  $S_v$ : surface-volume ratio;  $e$ : fuel height;  $\sigma_d$ : fuel load;  $\rho_d$ : particle density;  $\Delta H$ : heat of combustion.

repository of wildfire observations. In all cases but one, the time of first alert was used as the presumed ignition time. We also retrieved the ignition point, when available. The use of this information in the simulations and their associated uncertainty will be discussed in the next section.

The observed burned surfaces were obtained from private communication with the ‘Office de l’Environnement de la Corse’ (<http://www.oec.corsica/>, [verified 26 November 2019]). These surfaces were observed after the fires ended and estimated based on either spatially high resolution (20 m) satellite data imagery [Drusch *et al.*(2010)] or local global positioning system contouring in the field. With either method, the resulting shape had a spatial resolution similar to or higher than that used in the simulations ( $\sim 20$  m). Only this final observation was evaluated as we did not have enough information to estimate the location of the fire front at intermediate times. Although the time of the end of intervention was available in Prométhée for all seven fires, the fires usually stopped spreading several days before. This time was not representative of the main propagation phase of the fire and only provided an upper bound for the time of fire end in our simulations.

## 4.2 Probability distributions

The probability distributions of our perturbations follow one of the three classes described in section 2.2. Table 2 describes how the scalar inputs are perturbed, whereas Table 3 describes how the transitions between fuel types are sampled.

In Table 2, we qualified some perturbations as global or individual. The meaning of these perturbations depends on the input. For wind direction, wind speed norm, and dead fuel moisture, it means that at each time step the same perturbation coefficient  $z$  is applied to the forecast obtained from Meso-NH. A similar perturbation strategy for wind direction and wind speed norm was used by Hanna et al. [Hanna *et al.*(1998)] to account for the uncertainty in air quality predictions. For the fuel parameters other than moisture content, ‘global’ means that the same perturbation coefficient is applied for all fuel types, while ‘individual’ means that a perturbation coefficient is sampled independently for each fuel type. The value of the uncertainty parameter  $\sigma$  was chosen based on expert knowledge and the range of values found in the scientific literature.

The remaining three scalar inputs require further explanation. First, there is uncertainty in the coordinates of the ignition point. We point out that the initial geometry of the fire front in the ForeFire simulation must be a closed polygon. In our simulations, we used a regular octagon with a surface area of 0.45 ha centered around the perturbed ignition point. This perturbed ignition point was sampled inside a circle with the presumed ignition point as its center and with radius  $2\sigma$ . The chance of sampling a new point close to the center of the circle is higher than at the disc boundaries and no direction was favored. Nonetheless, if at least one of the vertices of the octagon happened to fall in a non-burnable zone, another point was sampled.

Fire start corresponds to the time assigned to the initial fire front in the

Input	Unit	Perturbation	$\sigma$	Range
Wind direction	$^{\circ}$	Additive, global	30	$[-60, 60]$
Wind speed norm	$\text{m s}^{-1}$	Multiplicative, global (log)	$0.5 \log 1.5$	$[2/3, 1.5]$
Dead fuel moisture		Multiplicative, global	0.15	$[0.7, 1.3]$
Heat of combustion	$\text{MJ kg}^{-1}$	Additive, global	1.0	$[-2.0, -2.0]$
Particle density	$\text{kg m}^{-3}$	Additive, global	100	$[-200, 200]$
Fuel height	m	Multiplicative, individual	0.15	$[0.7, 1.3]$
Fuel load	$\text{kg m}^{-2}$	Multiplicative, individual	0.15	$[0.7, 1.3]$
Surface-volume ratio	$\text{m}^{-1}$	Multiplicative, individual	0.15	$[0.7, 1.3]$
Ignition point	m	Additive	$\{50, 150, 250\}$	In a radius of $2\sigma$
Time of fire start	min	Additive	$\{5, 7.5, 15, 30\}$	$[-2\sigma, 2\sigma]$
Time of fire end	min	Additive	$\{5, 30, 60, 90, 120\}$	$[-2\sigma, 2\sigma]$

Table 2: Properties of the perturbations on scalar inputs.

For the last three inputs, the value of the uncertainty parameter  $\sigma$  is among the set but depends on the fire case. All multiplicative perturbations are sampled from a truncated normal, except for wind speed, which follows a truncated log-normal.

simulation. At first glance, our symmetric probability distributions may seem inappropriate, as a time of fire start that is posterior to the first alert (which happens after ignition) could be issued. Nonetheless, in our simulations, the initial burned surface covers 0.45 ha, which could be lower or higher than the actual burned surface at the time of the alert. Moreover, variation of the ignition time is also representative of the temporal uncertainty in the meteorological data.

Fire end corresponds to the time that marks the end of a simulation. In this study, the simulations represent a fire with free spread that could continue to grow indefinitely unless the ROS reaches 0 everywhere on the fire front. Firefighting actions were not modeled, even though they occurred in reality and had an effect on the final observed burned surface. In the absence of observed burned surfaces at specific times during the fire, we decided to choose a time that is representative of the end of the main propagation phase and accounts for most of the final observed burned area to end our simulations. It is a rough approximation and this is arguably the main source of uncertainty in the evaluation method, which explains the fairly high values of  $\sigma$  for almost all fire cases. The shape and the area of the observed fire surfaces have some uncertainty, but based on the spatial resolution of the data (at most 20 m) we assumed that this uncertainty was negligible compared to the sources of uncertainty in the simulation and decided not to take it into account in this study.

We tried to account for most of the uncertainty in the fire simulator in-

Origin type	Target type	Transition probability
Agricultural	323	0.5
333	323	0.5
334	323	0.5

Table 3: Properties of the transition perturbations on Corine Land Cover (CLC) fuel types

puts, but other approaches that focus on different inputs and use different probability distributions are possible. In [Miller *et al.*(2015)], the authors assign normal distributions to fuel and wind speed and the value of each cell in the simulation domain is sampled independently for a given simulation in the ensemble. In FSPro, long-lasting fires (more than 1 day) are considered; uncertainty is assumed to stem mostly from weather inputs. Historical data are used to obtain a probability distribution from which a sequence of daily wind values is sampled. For humidity, both predicted values and a time series analysis are used to obtain a scenario of daily values for a given simulation. In [Pinto *et al.*(2016)], the weather inputs (relative humidity, wind speed and direction) are perturbed according to an additive coefficient sampled from independent normal distributions. Uncertainty regarding ignition location is also considered and a probability distribution ranging between 0.33 and 3 for the ROS adjustment factor is defined for each fuel model. Our method for the generation of an ensemble is similar in the sense that it makes use of stochastic perturbations of some inputs of our fire spread simu-

lator. A sophisticated approach regarding input uncertainty quantification is presented in [Benali *et al.*(2016)] where empirical distributions are obtained based on comparison between different sources of data and should be quite representative of the potential error stemming from the inputs.

### 4.3 Complementary data and error corrections

For most fires, we tried to rely on the raw data as much as possible in order to perform simulations in a realistic context, which led us to make some choices that were quite ‘arbitrary’. As discussed in the previous section, the most uncertain input is the time of fire end, so we usually attributed a fairly high value to the uncertainty parameter  $\sigma$  for its perturbation distribution. The uncertainty parameter for the ignition point and for the time of fire start are also quite arbitrary as we do not know how much information was at the disposal of the fire managers when the fires occurred. Our choices are summarized in Table 4. The Calenzana fire was the smallest of the seven and supplementary information was documented by the firefighting services. This led us to reduce the uncertainty in temporal inputs in this case.

Other sources of information about the cases included news articles and Fire Information for Resource Management System (FIRMS) active fire data products from MODIS (Moderate-Resolution Imaging Spectroradiometer) [Giglio *et al.*(2016)] (MODIS Collection 6 NRT Hotspot / Active Fire Detections MCD14DL) and VIIRS (Visible Infrared Imaging Radiometer Suite) [Schroeder *et al.*(2014)] (NRT VIIRS 375 m Active Fire product VNP14IMG). These mainly helped



in determining an approximate time of fire end. It also helped us to either increase our confidence in some of the raw data on the fires or make some corrections. The latter only occurred with the Ville di Paraso fire, which started the first time on 20 October 2017, but stayed dormant and only burned a few hectares that day, while another start occurred 2 days later. Most of the final burned area was due to the fire spread on this subsequent day. For this reason, in this case, we did not use the raw data from the Prométhée database. Instead, in our simulations we considered a reference time of fire start of 1100hours on 22 October 2017 (local time, instead of 0927 hours on 20 October 2017) and the reference ignition point was slightly shifted eastwards and northwards. The coordinates of the new point were  $(42.567^\circ, 9.005^\circ)$  instead of  $(42.569^\circ, 8.998^\circ)$ . These modifications came with an increase in the uncertainty parameter  $\sigma$  on the uncertainty distribution for these inputs. We also relied on VIIRS data for the Ghisoni fire to approximate the ignition point as the data was missing in the Prométhée database. The coordinates  $(42.035^\circ, 9.164^\circ)$  were assigned together with a relatively high value of  $\sigma$ .

One last modification was applied to Ghisoni fire for which several starts occurred and three main separate burned areas were observed in the end. We chose not to include the smallest area, which resulted from a preemptive fire initiated by the firefighters. The other two were kept, but we only implemented one fire start in our simulations. We expect the predictions from the simulations to be quite different from the observations in this case. Indeed, not only did several fire starts occur, but the fire was not wind-driven and

Fire name	Start time	End time	$\sigma_{loc}$	$\sigma_{start}$	$\sigma_{end}$
Olmata di Tuda	24-07-2017 1249 hours	24-07-2017 2200 hours	50 m	7 min 30 s	60 min
Calenzana	05-08-2017 1742 hours	05-08-2017 1842 hours	50 m	5 min	5 min
Nonza	11-08-2017 0045 hours	11-08-2017 1700h ours	50 m	15 min	120 min
Ville di Paraso	22-10-2017 1100 hours	22-10-2017 2100 hours	250 m	30 min	30 min
Ghisoni	26-10-2017 1537 hours	27-10-2017 2100 hours	250 m	15 min	60 min
Sant'Andrea di Cotone	02-01-2018 0513 hours	02-01-2018 1500 hours	150 m	15 min	90 min
Chiatra	03-01-2018 1944 hours	04-01-2018 0500 hours	50 m	15 min	60 min

Table 4: Start time and ‘arbitrary’ information about the fire cases.

The end time is the one used in the ForeFire simulations when no perturbation is applied. The uncertainty parameters are applied to the following inputs : ignition point ( $\sigma_{loc}$ ), time of fire start ( $\sigma_{start}$ ), and time of fire end ( $\sigma_{end}$ ).

lasted for a long time, making it hard to predict.

#### 4.4 Computational set-up

For each of the seven fire cases, an ensemble of 500 simulations was run on a distributed Linux system with a total of 30 cores. We decided to implement a time limit of 30 min for each simulation so that a simulation that was not completed fast enough was aborted and its outputs were not saved. Although this may lead to a smaller ensemble, it ensures that a burn probability map for one fire case will be obtained (unless all simulations exceed the time limit) within 30 min, if 500 cores are available in an operational context. It could also limit overprediction as the longest simulations are usually those that

return the largest burned area. In this study, the time limit only had effect in the case of the Ghisoni fire where 324 simulations out of 500 were successful. For each case, reading the input data and sampling the perturbation coefficients were done sequentially as this process is rather fast ( $\sim 1$  min). Running 500 ForeFire simulations usually takes more time, so this step was run in parallel. When possible, some parts of the evaluation procedure were run in parallel as well.

## 5 Results

In this section, we investigate the performance of the probabilistic predictions. First, the case of the Calenzana fire, which is the smallest of the seven cases (119 ha), is analyzed in details through the perspective of the evaluation tools. Then, the results of all seven cases are discussed.

### 5.1 Detailed case: Calenzana fire

The burn probability map in this first case is shown in Figure 1 together with the observed burned surface. Here, the figure covers the whole evaluation domain. Where the background colors are visible, the predicted burn probability is 0. Most of the points in the evaluation domain fall in this category and it appears in the sharpness diagram in Figure 2b: a probability of 0.05 or less was attributed to about 86% of the evaluation domain. Based on the map, it seems that of the points with a probability lower than 0.1,

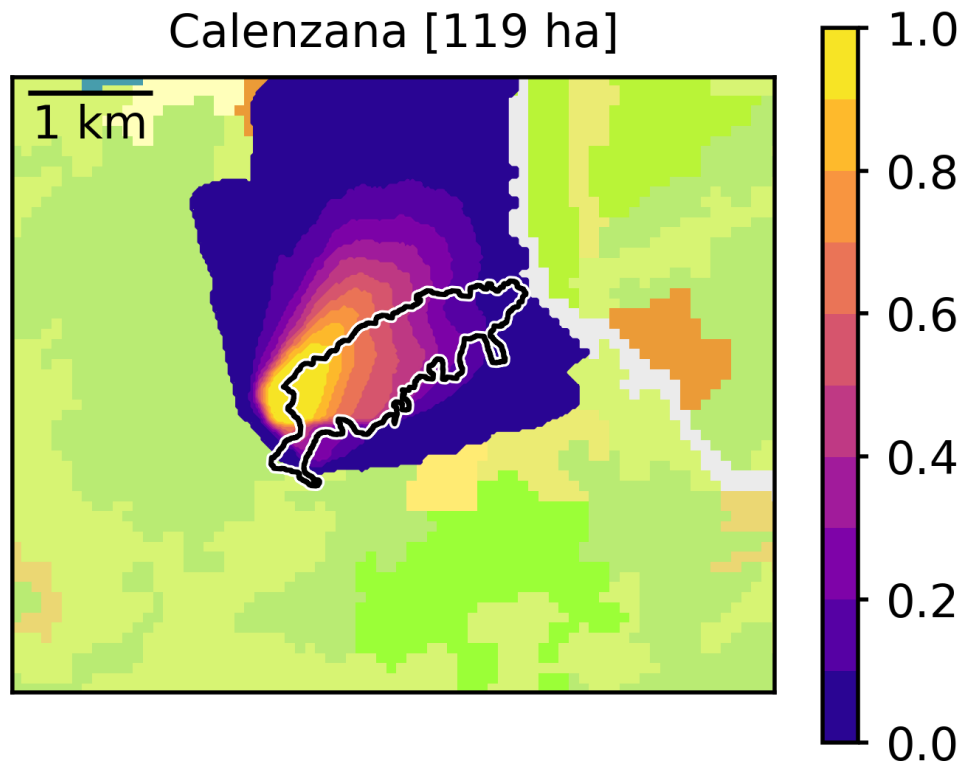


Figure 1: Burn probability map obtained in the case of Calenzana fire (5 August 2017).

The colorbar indicates the predicted burn probability ; black and white line is the contour of the observed burned surface ; background colors represent the Corine Land Cover data

most are located outside the observed burned surface, which indicates a good reliability for low probabilities. This is supported by the reliability diagram in Figure 2a: for  $p \leq 0.1$ , deviation from the ideally reliable case is low. Actually, reliability is very good for  $p$  up to 0.6. For higher predicted probabilities, the deviation is larger, so the prediction is not completely reliable in this case.

The burn probability map appears qualitatively to have a good coverage of the observed burned surface. Although the predicted probability may be a bit too high in some of the areas outside of the observed burned surface, there is no clear sign that the ensemble has a tendency to give probabilities that are either too high or too low. This is supported by the rank histogram in Figure 3, which is slightly U-shaped.

The Brier score obtained by the ensemble is about 0.027. As it is much closer to 0 than 1, one could hastily conclude that the accuracy of the system is excellent. However, as the fire duration was relatively short, predicting a probability close to 0 for points far enough from the presumed ignition point does not give much information to the forecaster. Here, the evaluation domain is quite large as the observed burned surface represents only 4% of the domain. As defined in Section 3.1, by taking the system with the best constant probability  $p_c$  as reference, we obtain a Brier skill score of 0.269: the system has better accuracy than the reference.

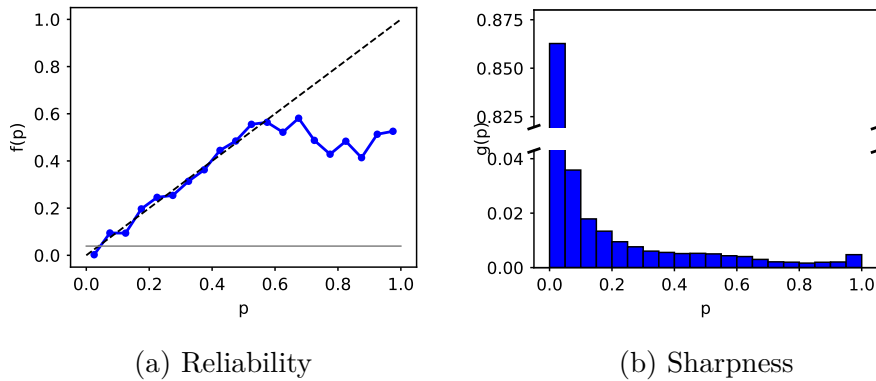


Figure 2: Reliability and sharpness diagrams obtained in the case of Calenzana fire (5 August 2017).

(a) The dashed line indicates an ideal reliable case ; the gray line has  $p_c$  for ordinate. The ensemble has good reliability, especially for  $p \leq 0.6$ , and predicts a burn probability close to 0 for most of the points in the evaluation domain.

(b) For most points in the evaluation domain, the ensemble predicted low probabilities (86% points with a probability between 0 and 0.05).

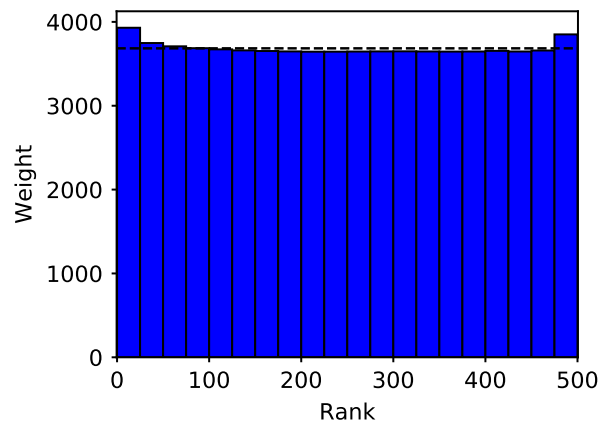


Figure 3: Rank diagram obtained in the case of the Calenzana fire (5 August 2017).

The 501 ranks are regrouped in 20 bins of 25 consecutive ranks (26 for the first bin). The dashed line indicates the flat histogram that would be obtained with a consistent ensemble. Overall, the ensemble appears fairly consistent.

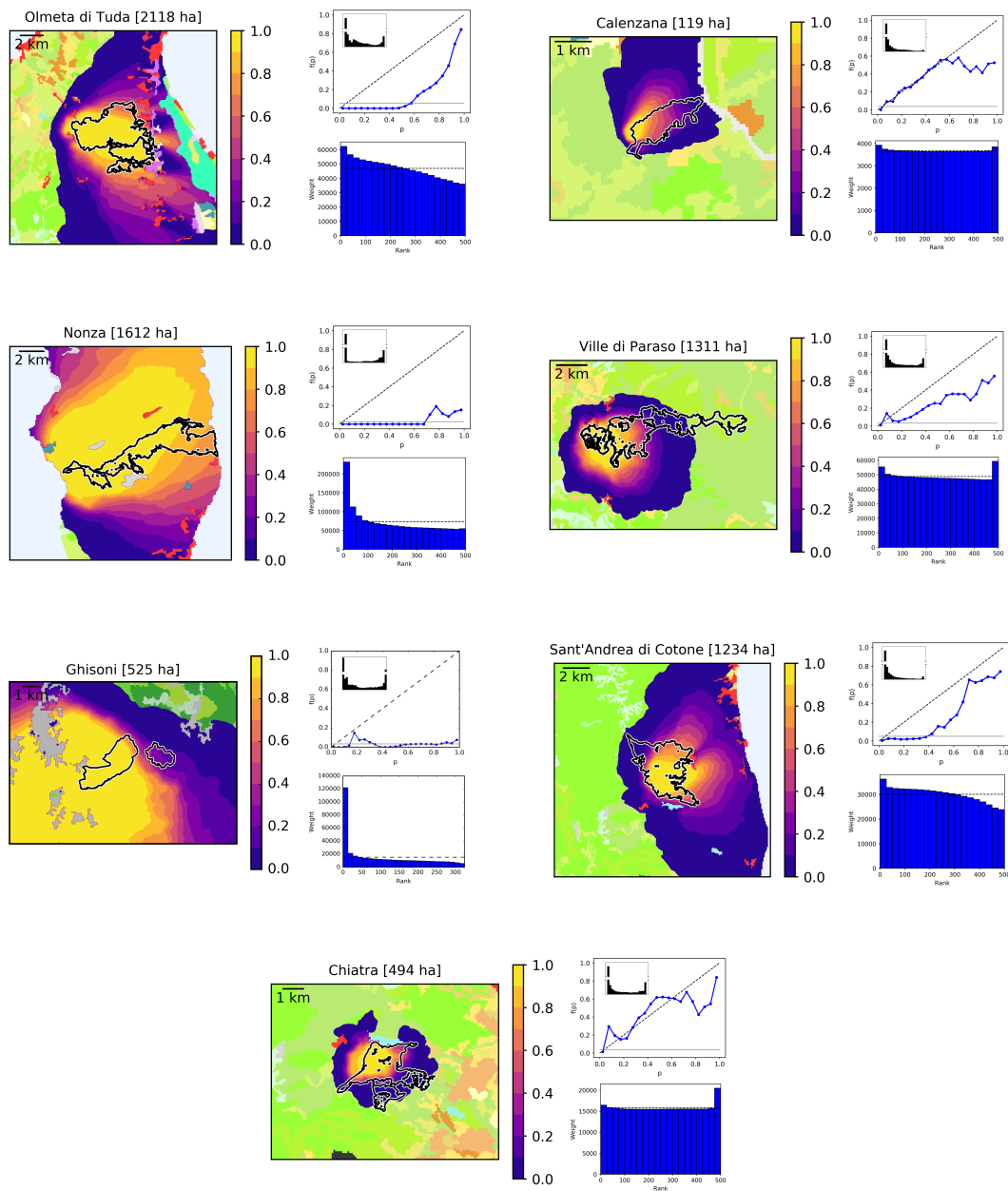


Figure 4: Simulations and evaluation for all fires.

For a given fire case, the burn probability map is on the left; the reliability diagram is in the upper right, with the sharpness histogram overlaid in black; the rank histogram is in the lower right. For almost all cases, only a part of the evaluation domain is covered by the map.



## 5.2 Performance for all seven fires

The system showed good performance in the case of the Calenzana fire, but one case is not sufficient for a robust evaluation. The figures presented in the previous section were plotted for all fires and are regrouped in Figure 4. Most maps in Figure 4 do not cover the entire evaluation domain. Characteristics and performance indicators of the ensembles are summarized in Table 5. Overall, the ensembles are not reliable and we can distinguish three situations based on the Brier skill score computed against the system with constant probability  $p_c$ . For the Calenzana, Sant’Andrea di Cotone and Chiatra fires, we obtain better accuracy than the reference. For Ville di Paraso and Olmeta di Tuda, accuracy is almost the same as that of the reference. And for the remaining cases of Nonza and Ghisoni, poor accuracy is obtained. In the latter two cases, fire spread is clearly overpredicted and the ensemble shows poor reliability. The Ghisoni fire had several starts and burned relatively slowly during several days. This behavior is difficult to predict with ForeFire, so poor performance of the prediction was expected.

There is an obvious overprediction for the Nonza and Ghisoni fires based on the burn probability maps as a large part of the evaluation domain outside the observed burned areas received high burn probabilities. This also appears in the rank histograms, where the weights are highest for ranks close to 0 and decrease when the rank increases. In the case of Nonza (respectively, Ghisoni) fire, only the first 4 out of 20 (resp. first 3 out of 25) bins show weights above that of the ideal uniform rank distribution. In the case of the Olmeta

di tuda and Sant'Andrea fires, both rank histograms show a weight that decreases with rank, which clearly indicates a tendency to overprediction that is not so obvious based on the inspection of the burn probability map alone. Overprediction is lower than in the two previous cases as the histograms are closer to the ideal uniform case. We note that ten (resp. 12) bins out of 20 show weights above that of the ideal uniform rank distribution. In the three remaining cases, the rank histogram is close to uniform. In the Ville di Paraso case, we can see that the weights decrease with rank except for the highest ranks and the bin of the highest ranks has the highest weight among all bins. This is partly due to the non-negligible eastern part of the observed burned area that received a burn probability of 0 or barely higher ; except for this part, the ensemble tends to slightly overpredict burn probability. A high weight for the highest ranks is also obtained in the Chiatra case for which very low probabilities are also obtained in a non-negligible part of the observed burned area, mainly in its southern part.

The most reliable ensembles are obtained for the Calenzana, Chiatra and Ville di Paraso fires. We also notice that their rank histograms are the closest to a uniform diagram among the seven cases. This is explained by the close link between the concepts of reliability and consistency. A reliable ensemble will lead to a rank histogram that is (almost) uniform. Here, we notice that when the reliability diagram is close to the bisector of the ideally reliable ensemble, the rank histogram is also close to a uniform histogram. Lower reliability is obtained for the Olmeta di Tuda and Sant'Andrea di Cotone

fires where the rank histogram deviates more from the uniform case, and the lowest reliability is obtained for the Nonza and Sant'Andrea di Cotone fires where the rank histogram deviates even more from the uniform case.

In all seven cases, the sharpness diagrams indicate that probabilities lower than 0.05 are the most represented in the evaluation domain: they represent between 65% and 95% of the domain depending on the fire case, except for Ghisoni fire where they represent 35% of the domain while 27% of the domain received a probability higher than 0.95. Each of the remaining probability bins represents a portion of the evaluation domain between 0.1% and 10%. This is not surprising as the evaluation domain is much larger than the observed burned surface to ensure that it will include most of the simulated burned areas, regardless of the main spread direction.

For Ville di Paraso, the fire started at approximately 1100hours (local time), so we had to rely on the weather forecast of the previous day. Here, if the Meso-NH prediction of wind speed vector had a stronger eastwards component, better accuracy of the probabilistic prediction would be achieved. Our perturbation distributions on weather forecasts do not depend on the time difference from the origin of the forecast, so uncertainty may be underestimated in this situation. Another issue with the weather forecasts is seen in the case of the Chiatra fire, for which strong winds were measured by a nearby weather station while they were not predicted by Meso-NH. The highest value of the Brier skill score is obtained for this fire. Still, because firefighting actions are not modeled in the simulations, overprediction of burn

probabilities is perhaps more desirable than their underprediction. From this perspective, the prediction appears poorer for this fire. Conversely, based on the rank histograms, a tendency to overpredict burn probabilities is indicated in the case of the Sant’Andrea di Cotone and Olmeta di Tuda fires, which may be more representative of how the fire would have spread in the absence of firefighting actions. In the latter case, although accuracy is almost the same as the reference, the prediction seems rather appropriate.

The convergence of the MC method of our 500-member ensembles can be investigated for each individual estimated burn probability  $p_i$ . For a 95% confidence level (at least), Equation (3) yields the interval  $\left[ \max(0, p_i - 0.061), \min(1, p_i + 0.061) \right]$ . For the Ghisoni wildfire, considering only the 324 successful simulations, we obtain a slightly larger interval :  $\left[ \max(0, p_i - 0.075), \min(1, p_i + 0.075) \right]$ .

The runtime necessary to carry out the generation and evaluation procedure highly depends on the fire case and the availability of computing cores. As expected, it increases with the mean size of the simulated fires and in some cases, would be prohibitive in an operational context with limited resources. Still, runtime could be lower in practice with more computing cores. Except for the Ghisoni fire, if 500 cores were available, the computations could be carried out in less than 10 minutes for all the studied fire cases. Also, the input duration of the fires is higher than 9 h in all our cases but the Calenzana fire. At the start of the fire, a time horizon of 3 h would be more relevant for fire managers. Running simulations of shorter fires would decrease fire size

Fire name	Duration	$p_c$	Observed size	Mean size	BS	BSS	Runtime
Olmata di Tuda	9h11	0.054	2118 ha	5744 ha	0.048	0.063	2420 s
Calenzana	1h00	0.039	119 ha	135 ha	0.027	0.269	71 s
Nonza	16h15	0.027	1612 ha	17 972 ha	0.164	-5.323	8202 s
Ville di Paraso	10h00	0.032	1311 ha	1600 ha	0.031	0.021	498 s
Ghisoni	29h27	0.033	525 ha	26 801 ha	0.352	-9.986	25 305 s
Sant'Andrea di Cotone	9h47	0.050	1234 ha	2653 ha	0.038	0.190	725 s
Chiatra	9h16	0.038	494 ha	366 ha	0.025	0.324	148 s

Table 5: Performance of the ensemble for all fires on 30 cores.

The duration is that used in a wildfire simulation when no perturbation is applied. The observed size is that of the final observed burned area, while the mean size is that of the simulated surfaces of the ensemble. BS is the Brier score and BSS is the associated Brier skill score, where the reference probabilistic forecast has constant probability  $p_c$ . Runtime is the computational time taken to carry out the generation and evaluation procedure with 30 cores, in seconds.

and, consequently, runtime. Under these conditions, a 500-member ensemble could be computed fast enough.

## 6 Conclusions

In this paper, we proposed and evaluated a new method to generate a probabilistic prediction of wildfire spread in a realistic operational context. We considered a database of seven fires whose final burned surfaces were known.

Uncertainty in the weather forecasts, vegetation cover, and fuel model parameters used as input in our simulations was described by means of perturbations and probability distributions that were common for all fires. For the start and the end of the fire, uncertainty varied depending on the quality of the information available for the fire case. Evaluation required the use of tools that were adapted to probabilistic simulations.

Overall, the probabilistic prediction method meets computational requirements of an operational context. All computations were handled with a realistic set of inputs that are consistent with the information available at the time when a fire is detected. We set a time limit of 30 min for each individual simulation which was only exceeded in one complex case. Given a 500-core supercomputer or a set of computers with a total of 500 cores, the computations could typically be completed in less than 10 min. Still, low predictive performance was obtained in some cases. Better accuracy is desirable before actual use in operational conditions.

The robustness of the evaluation could be improved with better information on the observed dynamics of the fire spread. For almost all cases, a major uncertainty source was the duration of the fire to use in the simulations. For long and large fires, it is likely that an evaluation with probabilistic scores will be of more interest between observed burned surfaces at known intermediate times and the corresponding ensemble of simulated surfaces. The uncertainty on fire duration would be less significant and in the early

stages of the fire, suppression activities could be absent or have had low effectiveness. A possible downside is that the observation may only cover a small part of the actual burned surface or not be available at all. Also, depending on the means of acquiring the data, uncertainty could increase at other levels (e.g., data with low spatial resolution would make the location of the observed burned zones rather uncertain). More fire cases would also be beneficial to the evaluation procedure as they would help the user to identify the performance of the prediction system in a wide range of situations. Here, the system showed good performance in the Calenzana case, which is the shortest of the seven. Arguably, uncertainty increases with the duration of the fire, so we believe the system is likely to deliver better predictions of fire spread over a few hours after ignition rather than in the later stages of the fire. Nevertheless, an advantage of the probabilistic prediction is that it represents this growing uncertainty, which is not the case with a single deterministic simulation.

In this work, we have clearly identified the uncertain inputs in our wild-fire simulations. The probability distributions could be refined by comparing different sources of data [Benali *et al.*(2016)] to better represent the error that is likely to pertain to input data. Another refinement could be to take into account the correlations between inputs. For instance, we can imagine that a positive correlation between fuel depth and fuel load in a given fuel model would be realistic. Also, more sophisticated methods for the propagation of the uncertainty exist. The MC approach has the advantage of being

easy to implement on a wide range of probability distributions, the only requirement being the ability to sample independent sets. Alternatively, some quasi-MC methods [Caffisch(1998)] (e.g., low-discrepancy sequences) may converge faster but are hardly suited to distributions with correlated inputs.

Current research perspectives now aim at calibrating the PDFs of these probability distributions by directly taking the observations into account. Even with the limited information constraint, accuracy and reliability of the probabilistic forecasts can still be improved, and evaluation scores will help to select the most relevant ensembles. Calibration methods adapted to probabilistic models such as Bayesian inference are the focus of a separate work to enhance forecast quality. By the means of a likelihood function, Bayesian inference may account for model uncertainty, which was not taken into account in this study where we focused on input uncertainty. A promising application of calibration following the Generalized Likelihood Uncertainty Estimation (GLUE) methodology was presented in [Benali *et al.*(2017)].

Our probabilistic prediction system is designed to serve in a crisis context and be part of a risk assessment framework. Within this framework, another means of using these forecasts is to generate fire danger maps based on probabilistic potential fire simulations to support firefighting decision before a fire starts. An advantage of these daily maps is to estimate potential fire size and they could help in the distribution of firefighting resources, while long-term maps could identify the zones that require high-priority land-planning actions.



## **Acknowledgments**

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

## **Conflicts of interest**

The authors declare no conflicts of interest.

## References

- [Anderson(1997)] Anderson JL (1997) The Impact of Dynamical Constraints on the Selection of Initial Conditions for Ensemble Predictions: Low-Order Perfect Model Results. *Monthly Weather Review* **125**, 2969–2983, doi:10.1175/1520-0493(1997)125<2969:TIODCO>2.0.CO;2
- [Andrews *et al.*(2013)] Andrews PL, Cruz MG, Rothermel RC (2013) Examination of the wind speed limit function in the Rothermel surface fire spread model. *International Journal of Wildland Fire* **22**
- [Benali *et al.*(2016)] Benali A, Ervilha AR, Sá AC, Fernandes PM, Pinto RM, Trigo RM, Pereira JM (2016) Deciphering the impact of uncertainty on the accuracy of large wildfire spread simulations. *Science of The Total Environment* **569-570**, 73 – 85, doi:<https://doi.org/10.1016/j.scitotenv.2016.06.112>
- [Benali *et al.*(2017)] Benali A, Sá AC, Ervilha AR, Trigo RM, Fernandes PM, Pereira JM (2017) Fire spread predictions: Sweeping uncertainty under the rug. *Science of The Total Environment* **592**, 187 – 196, doi: <https://doi.org/10.1016/j.scitotenv.2017.03.106>
- [Caffisch(1998)] Caffisch RE (1998) Monte Carlo and quasi-Monte Carlo methods. *Acta Numerica* **7**, 1–49, doi:10.1017/S0962492900002804
- [Cai *et al.*(2019)] Cai L, He H, Liang Y, Wu Z, Huang C (2019) Analysis of the uncertainty of fuel model parameters in wildland fire modelling of

- a boreal forest in north-east China. *International Journal of Wildland Fire* **28**, 205–215, doi:10.1071/WF18083
- [Cruz(2010)] Cruz MG (2010) Monte Carlo-based ensemble method for prediction of grassland fire spread. *International Journal of Wildland Fire* **19**, 521–530, doi:10.1071/WF08195
- [Drusch *et al.*(2010)] Drusch M, Gascon F, Berger M (2010) GMES Sentinel-2 Mission Requirements Document. EOP-SM/1163/MR-dr. Issue 2.1
- [Duff *et al.*(2018)] Duff TJ, Cawson JG, Cirulis B, Nyman P, Sheridan GJ, Tolhurst KG (2018) Conditional Performance Evaluation: Using Wildfire Observations for Systematic Fire Simulator Development. *Forests* **9**, doi:10.3390/f9040189
- [Duff *et al.*(2012)] Duff TJ, Chong DM, Taylor P, Tolhurst KG (2012) Procrustes based metrics for spatial validation and calibration of two-dimensional perimeter spread models: A case study considering fire. *Agricultural and Forest Meteorology* **160**, 110 – 117, doi:<https://doi.org/10.1016/j.agrformet.2012.03.002>
- [Duff *et al.*(2016)] Duff TJ, Chong DM, Tolhurst KG (2016) Indices for the evaluation of wildfire spread simulations using contemporaneous predictions and observations of burnt area. *Environmental Modelling & Software* **83**, 276 – 285, doi:<https://doi.org/10.1016/j.envsoft.2016.05.005>

- [Ervilha *et al.*(2017)] Ervilha A, Pereira J, Pereira J (2017) On the parametric uncertainty quantification of the Rothermel's rate of spread model. *Applied Mathematical Modelling* **41**, 37 – 53, doi:<https://doi.org/10.1016/j.apm.2016.06.026>
- [Feranec *et al.*(2016)] Feranec J, Soukup T, Hazeu G, Jaffrain G (2016) European landscape dynamics: CORINE Land Cover Data. (CRC Press, Boca Raton, USA)
- [Filippi *et al.*(2013)] Filippi JB, Mallet V, Nader B (2013) Representation and evaluation of wildfire propagation simulations. *International Journal of Wildland Fire* **23**, 46–57, doi:[10.1071/WF12202](https://doi.org/10.1071/WF12202)
- [Filippi *et al.*(2014)] Filippi JB, Mallet V, Nader B (2014) Evaluation of forest fire models on a large observation database. *Natural Hazards and Earth System Sciences Discussions* **2**, 3077–3091, doi:[10.5194/nhessd-2-3219-2014](https://doi.org/10.5194/nhessd-2-3219-2014)
- [Filippi *et al.*(2010)] Filippi JB, Morandini F, Balbi JH, Hill DR (2010) Discrete Event Front-tracking Simulation of a Physical Fire-spread Model. *SIMULATION* **86**, 629–646, doi:[10.1177/0037549709343117](https://doi.org/10.1177/0037549709343117)
- [Finney *et al.*(2011a)] Finney M, McHugh C, Grenfell I, Riley K, Short K (2011a) A simulation of probabilistic wildfire risk components for the continental United States. *Stochastic Environmental Research and Risk Assessment* **25**, 973–1000, doi:[10.1007/s00477-011-0462-z](https://doi.org/10.1007/s00477-011-0462-z)

- [Finney *et al.*(2011b)] Finney MA, Grenfell IC, McHugh CW, Seli RC, Trethewey D, Stratton RD, Brittain S (2011b) A Method for Ensemble Wildland Fire Simulation. *Environmental Modeling & Assessment* **16**, 153–167, doi:10.1007/s10666-010-9241-3
- [Fujioka(2002)] Fujioka F (2002) A new method for the analysis of fire spread modeling errors. *International Journal of Wildland Fire - INT J WILDLAND FIRE* **11**, 193–203, doi:10.1071/WF02004
- [Giglio *et al.*(2016)] Giglio L, Schroeder W, Justice C (2016) The collection 6 MODIS active fire detection algorithm and fire products. *Remote Sensing of Environment* **178**, 31–41, doi:10.1016/j.rse.2016.02.054
- [H Scott and E Burgan(2005)] H Scott J, E Burgan R (2005) Standard Fire Behavior Fuel Models : A Comprehensive Set for Use with Rothermel’s Surface Fire Spread Model. *The Bark Beetles, Fuels, and Fire Bibliography*
- [Hanna *et al.*(1998)] 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**, 3619 – 3628, doi:https://doi.org/10.1016/S1352-2310(97)00419-6
- [Lac *et al.*(2018)] Lac C, Chaboureau JP, Masson V, Pinty JP, Tulet P, Escobar J, Leriche M, Barthe C, Aouizerats B, Augros C, Aumond P,

Auguste F, Bechtold P, Berthet S, Bieilli S, Bosseur F, Caumont O, Cohard JM, Colin J, Couvreur F, Cuxart J, Delautier G, Dauhut T, Ducrocq V, Filippi JB, Gazen D, Geoffroy O, Gheusi F, Honnert R, Lafore JP, Lebeau-pin Brossier C, Libois Q, Lunet T, Mari C, Maric T, Mascart P, Mogé M, Molinié G, Nuissier O, Pantillon F, Peyrillé P, Pergaud J, Perraud E, Pianezze J, Redelsperger JL, Ricard D, Richard E, Riette S, Rodier Q, Schoetter R, Seyfried L, Stein J, Suhre K, Taufour M, Thouron O, Turner S, Verrelle A, Vié B, Visentin F, Vionnet V, Wautelet P (2018) Overview of the Meso-NH model version 5.4 and its applications. *Geoscientific Model Development Discussions* **11**, 1929–1969, doi:10.5194/gmd-11-1929-2018

[Lautenberger(2017)] Lautenberger C (2017) Mapping areas at elevated risk of large-scale structure loss using Monte Carlo simulation and wildland fire modeling. *Fire Safety Journal* **91**, 768 – 775, doi:<https://doi.org/10.1016/j.firesaf.2017.04.014>. Fire Safety Science: Proceedings of the 12th International Symposium

[Liu *et al.*(2015a)] Liu Y, Hussaini MY, Ökten G (2015a) Global sensitivity analysis for the Rothermel model based on high-dimensional model representation. *Canadian Journal of Forest Research* **45**, 1474–1479, doi:10.1139/cjfr-2015-0148

[Liu *et al.*(2015b)] Liu Y, Jimenez E, Hussaini M, Ökten G, Goodrick S (2015b) Parametric uncertainty quantification in the Rothermel model

with randomised quasi-Monte Carlo methods. *International Journal of Wildland Fire* **24**, 307–316, doi:10.1071/WF13097

[Miller *et al.*(2015)] Miller C, Hilton J, Sullivan A, Prakash M (2015) SPARK – A Bushfire Spread Prediction Tool. In R Denzer, RM Argent, G Schimak, J Hřebíček (Eds.) *Environmental Software Systems. Infrastructures, Services and Applications*, pp. 262–271 (Springer International Publishing, Cham)

[Murphy(1973)] Murphy AH (1973) A New Vector Partition of the Probability Score. *Journal of Applied Meteorology (1962-1982)* **12**, 595–600

[Parisien *et al.*(2005)] Parisien M, Kafka V, Hirsch K, Todd J, Lavoie S, Maczek P (2005) Mapping Wildfire Susceptibility with the BURN-P3 Simulation Model. Canadian Forest Service, Northern Forestry Centre

[Paz *et al.*(2011)] Paz S, Carmel Y, Jahshan F, Shoshany M (2011) Post-fire analysis of pre-fire mapping of fire-risk: A recent case study from Mt. Carmel (Israel). *Forest Ecology and Management* **262**, 1184 – 1188, doi:<https://doi.org/10.1016/j.foreco.2011.06.011>

[Pinto *et al.*(2016)] Pinto R, Benali A, Sá A, Fernandes P, Soares P, Cardoso R, Trigo R, Pereira J (2016) Probabilistic fire spread forecast as a management tool in an operational setting. *SpringerPlus* **5**, 1205, doi:10.1186/s40064-016-2842-9

- [Rothermel(1972)] Rothermel RC (1972) A mathematical model for predicting fire spread in wildland fuels. Res. Pap. INT-115. Ogden, UT: U.S. Department of Agriculture, Intermountain Forest and Range Experiment Station. 40 p
- [Salis *et al.*(2013)] Salis M, Ager A, Arca B, Finney M, Bacciu V, Duce P, Spano D (2013) Assessing exposure of human and ecological values to wildfire in Sardinia, Italy. *International Journal of Wildland Fire* **22**, 549–565, doi:10.1071/WF11060
- [Salis *et al.*(2016)] Salis M, Arca B, Alcasena F, Arianoutsou M, Bacciu V, Duce P, Duguay B, Koutsias N, Mallinis G, Mitsopoulos I, Moreno J, Perez J, Urbietta I, Xystrakis F, Zavala G, Spano D (2016) Predicting wildfire spread and behavior in Mediterranean landscapes. *International Journal of Wildland Fire* **25**, 1015–1032, doi:10.1071/WF15081
- [Salvador *et al.*(2001)] Salvador R, Piñol J, Tarantola S, Pla E (2001) Global sensitivity analysis and scale effects of a fire propagation model used over Mediterranean shrublands. *Ecological Modelling* **136**, 175 – 189, doi:https://doi.org/10.1016/S0304-3800(00)00419-1
- [Schroeder *et al.*(2014)] Schroeder W, Oliva P, Giglio L, Csiszar I (2014) The New VIIRS 375 m active fire detection data product: Algorithm description and initial assessment. *Remote Sensing of Environment* **143**, 85–96, doi:10.1016/j.rse.2013.12.008



- [Sullivan(2009)] Sullivan AL (2009) Wildland surface fire spread modelling, 1990–2007. 1: Physical and quasi-physical models. *International Journal of Wildland Fire* **18**, 349–368, doi:10.1071/WF06143
- [Termonia *et al.*(2018)] Termonia P, Fischer C, Bazile E, Bouyssel F, Brožková R, Bénard P, Bochenek B, Degrauwe D, Derková M, El Khatib R, Hamdi R, Mašek J, Pottier P, Pristov N, Seity Y, Smolíková P, Španiel O, Tudor M, Wang Y, Wittmann C, Joly A (2018) The ALADIN System and its canonical model configurations AROME CY41T1 and ALARO CY40T1. *Geoscientific Model Development* **11**, 257–281, doi:10.5194/gmd-11-257-2018
- [Thompson *et al.*(2017)] Thompson M, Calkin D, Scott JH, Hand M (2017) Uncertainty and probability in wildfire management decision support: An example from the United States. In K Riley, P Webley, M Thompson (Eds.) *Natural Hazard Uncertainty Assessment: Modelling and Decision Support*, volume 223 of *Geophysical Monograph*, chapter 4, pp. 31–41 (American Geophysical Union), first edition
- [Wilks(2011)] Wilks D (2011) Chapter 7 - Statistical Forecasting. In DS Wilks (Ed.) *Statistical Methods in the Atmospheric Sciences*, volume 100 of *International Geophysics*, pp. 215 – 300 (Academic Press), third edition, doi:<https://doi.org/10.1016/B978-0-12-385022-5.00007-5>