HAL
open science

Quantification of wildland fire risk using metamodeling of fire spread

Frederic Allaire

To cite this version:

Frederic Allaire. Quantification of wildland fire risk using metamodeling of fire spread. Mathematics [math]. Sorbonne université, 2021. English. NNT : . tel-03385307v1

HAL Id: tel-03385307

https://inria.hal.science/tel-03385307v1

Submitted on 19 Oct 2021 (v1), last revised 17 Nov 2021 (v2)

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.

Sorbonne Université

École doctorale de Sciences Mathématiques de Paris Centre (ED 386)

Inria

Thèse de Doctorat en Mathématiques appliquées

par

Frédéric Allaire

Quantification of wildland fire risk using metamodeling of fire spread *

Soutenue le 14 juin 2021 devant un jury composé de :

M. Jan Mandel	U. Colorado Denver, USA	rapporteur
M. Anthony Nouy	Centrale Nantes	rapporteur
Mme Rossella Arcucci	Imperial College, UK	examinatrice
M. Mark Finney	USDA, USA	examinateur
M. Patrick Gallinari	Sorbonne Université	président et examinateur
M. Jean-Baptiste Filippi	CNRS	encadrant de thèse
M. Vivien Mallet	INRIA	encadrant de thèse
M. Jacques Sainte-Marie	INRIA	directeur de thèse

*. Quantification du risque incendie par métamodélisation de la propagation de feux de forêt

Remerciements

Je remercie l'Agence Nationale de la Recherche (ANR) d'avoir financé mes travaux de thèse.

Je tiens à remercier dans un premier temps les membres du jury pour avoir accepté d'évaluer mes travaux de thèse, pour leur participation à ma soutenance de thèse, et pour leurs suggestions et commentaires pertinents lors de la soutenance. I thank professor Jan Mandel for accepting to review my thesis. Je remercie également professeur Anthony Nouy d'avoir accepté d'être rapporteur de ma thèse. Merci à tous les deux pour votre avis critique sur mon manuscrit et vos remarques judicieuses. Je tiens à remercier professeur Patrick Gallinari d'avoir accepté d'être examinateur, et plus particulièrement d'avoir présidé ma soutenance. I thank Rossella Arcucci and Mark Finney for accepting to be examiners of my thesis as well. Merci à tous pour le temps consacré à l'évaluation de mes travaux, pour votre disponibilité et votre réactivité, vu les complications liées au contexte sanitaire.

Je remercie vivement Jacques Sainte-Marie pour avoir dirigé ma thèse et pour m'avoir accueilli au sein de l'équipe ANGE à l'INRIA Paris. Je te remercie pour tes conseils avisés et ton intérêt pour les thématiques de l'équipe (anciennement) CLIME, qui ont contribué au bon déroulement de cette thèse.

Je remercie chaleureusement mes encadrants, Vivien Mallet et Jean-Baptiste Filippi d'avoir proposé ce sujet de thèse. J'ai énormément appris au cours de cette thèse, tant sur la métamodélisation, la quantification d'incertitude, que les outils informatiques et les feux de forêt – sujet sur lequel je ne connaissais pas grand-chose avant de commencer la thèse – et plus généralement en tant que chercheur. Vous m'avez accordé votre confiance et de votre soutien, si bien que j'ai pu explorer mon sujet en autonomie tout en étant guidé grâce à vos précieux conseils. Merci pour votre suivi et votre soutien, même en période de télétravail, depuis la Corse ou l'étranger, malgré les complications liées à la situation sanitaire. Vivien, j'ai particulièrement apprécié ta rigueur et tous tes enseignements, notamment en informatique, qui m'ont fait gagner en efficacité et me seront utiles toute au long de ma carrière. Jean-Baptiste, merci pour ton expertise et ton aide pour la prise en main des outils relatifs à la prévision incendie, tu m'as guidé admirablement tant dans le domaine des feux de forêt que lors de mes séjours en Corse.

Je remercie les acteurs impliqués dans le projet ANR FireCaster, coordonné par Jean-Baptiste, dans lequel s'est inscrit ma thèse, pour les discussions et leurs retours fort utiles. Merci donc aux membres du CERFACS, du LA, du CNRM, des UMRs SPE et LISA, de Météo-France, ainsi qu'aux membres de l'Office National des Forêts (ONF) et aux opérationnels de la lutte incendie. Je tiens plus particulièrement à remercier Yolanda Perez pour son expertise en modélisation de la végétation; Antoine Belgodère pour son travail sur l'estimation du coût des incendies, sur lequel j'ai beaucoup apprécié de collaborer; Florence Vaysse et Mathieu Régimbeau pour les échanges et leur contribution sur les indices météo de danger incendie; Mélanie Rochoux pour les discussions sur les méthodes mathématiques liées à la prévision incendie; et Aurélien Costes pour les discussions tantôt scientifiques, tantôt à propos de notre expérience en tant que doctorant, etc.

J'adresse un remerciement particulier à Céline Helbert pour ses cours lors de ma

formation d'ingénieur à l'École Centrale de Lyon, plus spécifiquement ceux sur la métamodélisation et la quantification d'incertitude. Merci pour m'avoir fait découvrir ces sujets passionnants, je n'aurais peut-être pas trouvé l'offre de cette thèse sans tes conseils. Merci également à Christophette Blanchet pour m'avoir proposé mon sujet de fin d'études au cours duquel j'ai pu approfondir mes connaissances en métamodélisation, ce qui m'a incité à poursuivre dans cette voie.

Au cours de ma thèse, j'ai bénéficié d'un environnement de travail idéal pour mener à bien mes travaux. Le contexte de l'ANR FireCaster a favorisé l'accès à des données environnementales, notamment des prévisions météorologiques. J'ai aussi utilisé des ressources de calcul performantes à l'INRIA. Je remercie aussi l'IDRIS, qui m'a accordé l'accès au supercalculateur Jean-Zay. Je suis reconnaissant d'avoir été accueilli convivialement dans les différents instituts de recherche où ma thèse s'est déroulée, en premier lieu à l'INRIA, mais aussi lors de mes nombreux déplacements en Corse. Sur ce dernier point, je tiens à remercier de nouveau Jean-Baptiste Filippi, ainsi que mes collègues de Corte.

Je souhaite aussi remercier mes collègues de l'INRIA pour leur bonne humeur et les nombreuses discussions conviviales lors de déjeuners, pauses-café, afterworks, etc., grâce à qui l'ambiance de travail fut excellente. La liste de personnes que je souhaite remercier étant longue, je tiens par avance à m'excuser pour ceux que j'aurais oublié de citer. Dans un premier temps, un grand merci aux membres de la "branche CLIME", Raphaël, Antoine, Janelle, et Bao. Bao, je te remercie particulièrement pour les échanges fructueux sur les problèmes de quantification d'incertitude et pour tes excellentes recommandations de restaurants de la rue Sainte-Anne. Merci aussi à Nourhène pour avoir été, tout comme Bao, une collègue de bureau extrêmement sympathique. Merci aux doctorants plus expérimentés pour leurs conseils avisés: Raphaël, Noémie, Ruiwei, Hugo, Olivier, Hussam, Mathieu, Léa et Fabien. Léa, merci pour ton aide sans faille, en particulier à propos des enseignements de TD. Fabien, merci de m'avoir accueilli au "B.A.-BA" – clin d'oeil à Philippe et aux autres habitués – et de m'avoir passé le flambeau. Merci aux doctorants et ingénieurs ayant rejoint l'INRIA en cours de route: Virgile, Liudi, Thibault, Mathieu, Chourouk, Matthias et Apolline. Thank you to Jan and Van Thanh for the enjoyable conversations, especially during lunchtime. Merci à Fabien R. pour m'avoir motivé à faire de la course à pied le midi, merci aussi à tous ceux qui ont participé à la course du midi et au fameux "plat du jour". Merci aussi aux chercheurs ANGE et affiliés: Yohan, Martin, Julien, Cindy, Nina, Boris et plus particulièrement Bernard, qui, comme moi, a fait à plusieurs reprises la navette entre Corte et l'INRIA. Merci enfin à Maryse et Julien pour leur aide dans les démarches administratives. Merci à tous pour votre enthousiasme.

Un grand merci à mes amis et ma famille pour avoir suivi, de près ou de loin, mes travaux de thèse, et plus généralement pour les bons moments passés ensemble qui m'ont été d'autant plus précieux dans les périodes les plus difficiles. Je remercie de tout cœur ma mère et ma sœur pour leur écoute, leur patience et leur soutien tout au long de ma thèse.

J'ai une pensée particulière pour Louis qui m'a tant apporté. Merci infiniment pour ta bienveillance et pour la joie que tu as mis dans ma vie.

À tous ceux qui m'ont soutenu jusqu'ici, merci.

Contents

Preamble **9**

I. Introduction to uncertainty quantification, metamodeling, and wildland fires **13**
I.1. Uncertainty quantification . . . 13
I.1.1. Propagation of uncertainty . . . 14
I.1.2. Evaluation and dissimilarity measures of probability distributions 17
I.1.3. Inverse uncertainty quantification . . . 20
I.2. Metamodeling . . . 22
I.2.1. Design of experiments . . . 23
I.2.2. Gaussian processes . . . 26
I.2.3. Neural networks . . . 29
I.3. Wildland fires . . . 32
I.3.1. Description and modeling . . . 32
I.3.2. Wildland fire risk and danger . . . 38

II. Generation and evaluation of an ensemble of wildland fire predictions **47**
II.1. Introduction . . . 49
II.2. Probability distribution of wildfire spread simulations . . . 51
II.2.1. Fire spread simulation . . . 51
II.2.2. Modeling uncertainty in input data . . . 51
II.2.3. Uncertainty propagation in fire spread simulations . . . 53
II.3. Probabilistic evaluation . . . 54
II.3.1. Accuracy . . . 54
II.3.2. Reliability and sharpness . . . 55
II.3.3. Probabilistic resolution . . . 55
II.3.4. Consistency . . . 56
II.3.5. Evaluation domain . . . 57
II.4. Application to seven Corsican fires . . . 57
II.4.1. Data sources . . . 57
II.4.2. Probability distributions . . . 59
II.4.3. Complementary data and error corrections . . . 61
II.4.4. Computational set-up . . . 62
II.5. Results . . . 63
II.5.1. Detailed case: Calenzana fire . . . 63
II.5.2. Performance for all seven fires . . . 65
II.6. Conclusions . . . 68
II.7. Consistency and rank histogram . . . 71
II.7.1. Continuous case . . . 71
II.7.2. Binary case . . . 72

III. A posteriori uncertainty quantification for wildland fire simulation **75**
III.1. Introduction . . . 76
III.2. A posteriori uncertainty quantification . . . 78
III.2.1. Distribution based on Wasserstein distance . . . 79
III.2.2. Emulation . . . 81
III.2.3. Extension to several fire cases . . . 83
III.2.4. Sampling from the calibrated distribution . . . 84
III.3. Application to wildland fire spread . . . 85
III.3.1. Fire spread simulation . . . 85
III.3.2. Prior uncertainty in input data . . . 86
III.3.3. Application to seven Corsican wildland fires . . . 88
III.4. Results . . . 89
III.4.1. Emulation . . . 89
III.4.2. Calibrated distribution . . . 89
III.4.3. Ensemble evaluation . . . 95
III.5. Discussion and conclusions . . . 99
III.6. Appendix: Bootstrap confidence interval for the Brier skill score . . . 100

IV. Emulation of wildland fire simulations with deep learning **103**
IV.1. Introduction . . . 104
IV.2. Simulation of wildland fire spread . . . 107
IV.3. Emulation with deep learning . . . 111
IV.3.1. Design of experiments . . . 111
IV.3.2. Neural network architecture . . . 112
IV.3.3. Accuracy metrics and training strategy . . . 115
IV.3.4. Extraction of the actual emulator . . . 116
IV.3.5. Implementation . . . 117
IV.4. Results and discussion . . . 117
IV.5. Conclusions . . . 127

V. High-resolution fire danger mapping **131**
V.1. Introduction . . . 132
V.2. Method . . . 135
V.2.1. DeepFire: fire size emulation using a deep neural network . . . 135
V.2.2. Deterministic predictions of fire danger maps . . . 137
V.2.3. Probabilistic predictions representing input uncertainty . . . 139
V.2.4. Application to weather forecasts corresponding to 13 Corsican fires 140
V.3. Analysis of predictions for specific fires . . . 143
V.3.1. A detailed fire case: Calenzana 2017 . . . 143
V.3.2. Overview of predictions for all 13 fires . . . 152
V.4. Possible designs for daily fire danger mapping . . . 156
V.4.1. Focus on the day of Calenzana 2017 fire . . . 156
V.4.2. Overview of the predictions for the 12 other fires . . . 161
V.5. Conclusions and perspectives . . . 162

VI. Conclusion **165**
VI.1. Thesis summary . . . 165

VI.2. Research perspectives 168

Appendix A. Hurry! There is a fire **173**
A.1. Introduction 174
A.2. Numerical methods 176
A.2.1. Simulation of wildland fire spread 176
A.2.2. Generation of an ensemble of simulations 178
A.3. Evaluating the cost of fire 180
A.3.1. Land Use in Corsica 180
A.3.2. Cost of a single fire 181
A.3.3. Modeling risk aversion 184
A.4. Evolution of the damage 184
A.5. Discussion 192
A.6. Conclusion 193

Bibliography **195**

Preamble

Context and background

One of the challenges of firefighting is the anticipation of risks related to wildland fires. In the short-term, when environmental conditions are favorable to the occurrence and growth of large wildfires, firefighting resources are preemptively mobilized and deployed over the territory. Optimal allocation should allow to carry out an initial attack as early as possible on fires that are most susceptible to grow fast and cause considerable losses. In the long-term, local land-use planning and fuel management can help in limiting wildfire occurrence and making firefighting easier in case a fire starts in the area. In both situations, identifying areas with highest risk can help taking optimal decisions given the limited resources available.

In this context, wildland fire spread models can be used to simulate the growth of a fire over the landscape given environmental data including weather variables, elevation, land use and characterization of vegetation. Still, accuracy of predictions relying on such simulators, in spite of their wide use, is challenged by the complexity of the phenomenon and the uncertainties involved in the modeling process, notably those stemming from the estimation of the speed at which the flames advance, simplifying assumptions to describe vegetation properties, and use of meteorological forecasts to define weather inputs. A promising direction to address this issue is to rely on an ensemble of simulations instead of a single deterministic simulation, so as to provide an output distribution representative of these uncertainties. Comparison with observations of fire spread, which can be considered deterministic, is not direct in this probabilistic approach, warranting the definition of appropriate properties and scores for evaluation purposes. Seemingly, in a perspective of model improvement, suitable methods need to be developed for calibration of the probabilistic distributions involved in uncertainty quantification.

Risk is traditionally quantified using the expected value $\mathbb{E}[Y]$ of "cost" Y, a random variable whose probability distribution represents the consequences (e.g. damage to structures) of possible scenarios and how likely they are. This raises the question of how to best define or estimate this distribution, whether it be to calculate $\mathbb{E}[Y]$ or to assess risk by means of probabilities, quantiles, or more elaborate quantities derived from the distribution of Y. This ties with uncertainty quantification, whose aim is to provide relevant probability distributions to best represent the uncertainty of the modeling elements, even when the model is deterministic.

Currently, a major focus of short-term risk assessment is the quantification of "fire danger", a term that generally relates to the potential for ignition and spread of a fire at a given location. Traditional fire danger indices, widely used for decision support in an operational context, consist in a unitless value calculated essentially based on weather variables. Making use of output burned surfaces of wildland fire spread simulations

offers an interesting alternative for quantifying fire danger: for instance, the resulting fire size accounts for both weather and terrain (elevation and vegetation) and can be expressed in hectares, a quantifiable result.

Wildfire danger and risk mapping based on ensembles of fire spread simulations for each point in the territory would require too much computational time when aiming for high-resolution maps, even with 1-member ensembles. This calls for the design of a metamodel (aka emulator), i.e. a model that allows to compute an approximation of the simulation output much faster. Still, properly accounting for the variety of influential inputs and data involved in the simulations is a major challenge that must be addressed in order to provide a good approximation.

This thesis is part of the project FireCaster, funded by the ANR (Agence Nationale de la Recherche). Its goal is the development of a prototype forecasting system for wildland fire risk assessment, crisis management and decision-support based on innovative modeling tools and high-performance computing. Although it imposed the constraint of working with real-world data and carrying out faster-than-real-time simulations, this context also facilitated access to the data necessary for the models used in this study, including weather forecasts, and data maps of elevation and land-use. Research axes in the project include: better description of vegetation, data assimilation methods, development of coupled fire-atmosphere models, estimation of socio-economical and ecological costs, and risk/danger assessment.

The latter aspect of the project is the main focus of this thesis, whose final objective is to generate maps of wildfire danger (respectively, risk) that account for potential fire size (resp. cost) for all points in Corsica island (the test area), by means of fire spread simulations. This requires to address several topics: the generation of ensembles of simulations to account for the uncertainties, the evaluation of these ensembles using observations, the calibration of the underlying distributions of the ensembles based on observations, and the design of emulators to make up for the high total computational time of the simulations required in the methods proposed.

Outline of the thesis

The present document is organized as follows:

In Chapter I, we introduce mathematical and physical concepts involved throughout the whole manuscript. Notions involved in uncertainty quantification are presented with a focus on methods involving observations for evaluation and calibration purposes, followed by a background on metamodeling by presenting techniques involved in following chapters, namely Gaussian process modeling and neural networks. The last part of the chapter introduces fundamental aspects of wildland fire science, wildfire spread models and methods for quantification and mapping of fire risk and danger.

In Chapter II, we study direct uncertainty quantification of numerical simulations of wildfire spread of an ongoing fire by attributing probability distributions to the uncertain inputs and propagating them using a Monte Carlo approach. We define several properties and introduce probabilistic scores that are common in meteorological applications to evaluate the probabilistic predictions resulting from the ensemble of simulations. The method is applied to seven fires that occurred in Corsica from mid-2017 to early 2018.

In Chapter III, we build upon the study of the preceding chapter and focus on inverse uncertainty quantification of wildland fire spread predictions. We propose a method to provide a relevant input probability distribution *a posteriori* by integrating the information provided by the observed burned surfaces used for evaluation purposes. We make use of the Wasserstein distance as an evaluation metric between surfaces to define a pseudo-likelihood function involved in the posterior PDF. Due to the high dimension and the computational requirements of the pseudo-likelihood function, a Gaussian process emulator is built to obtain a sample of the calibrated input distribution with a MCMC algorithm. We apply the method using the seven Corsican fires presented in the preceding chapter and investigate the calibrated input distribution as well as the resulting probabilistic predictions of wildland fire spread.

The calibrated input distribution is employed in an interdisciplinary work led by Antoine Belgodère, reported in Appendix A. This study focuses on estimating the economical cost of wildfires and marginal damage as a function of time using simulations of fire spread, and is applied to the aforementionned seven Corsican fires.

In Chapter IV, we focus on the "short-term" situation that requires to assess fire risk and danger for the coming day. The aim is to use the size of the simulated burned surface resulting from a fire starting at a given location at any given time and provide a map with high spatial and temporal resolution. The computational time required to carry out the massive amount of simulations for this task is very high, calling for the use of an emulator that can account for all the possible input conditions, including the input uncertainty that was quantified in the previous chapters. We present the design of a deep neural network that was trained to approximate the simulated fire size, and assess the performance of the resulting emulator.

In Chapter V, we use the emulator of simulated fire size presented in the preceding chapter (and referred to as DeepFire) and study its application to fire danger mapping using actual weather forecasts. Predictions based on DeepFire are generated for 13 relatively big fires that occurred in Corsica are compared with an operational fire danger index used in France. We investigate the insights provided by such predictions characterized by high spatial resolution and frequency that can account for input uncertainty in the form of ensembles. We then address how to best summarize such forecasts and how to derive new daily fire danger ratings.

Finally, conclusions of the thesis are summarized at the end of the present manuscript. Perspectives and future research axes that have emerged from the results of this thesis are also presented.

I. Introduction to uncertainty quantification, metamodeling, and wildland fires

In the present chapter, we introduce mathematical and physical concepts involved in the latter chapters of the manuscript. Notions involved in uncertainty quantification are presented with a focus on methods involving observations for evaluation and calibration purposes, followed by a background on metamodeling, notably by presenting techniques involved in Chapters III and IV. The last part of the chapter introduces fundamental aspects of wildland fire science, wildfire spread models and methods for quantification and mapping of fire risk and danger.

Summary

I.1. Uncertainty quantification 13
- I.1.1. Propagation of uncertainty 14
- I.1.2. Evaluation and dissimilarity measures of probability distributions 17
- I.1.3. Inverse uncertainty quantification 20

I.2. Metamodeling 22
- I.2.1. Design of experiments 23
- I.2.2. Gaussian processes 26
- I.2.3. Neural networks 29

I.3. Wildland fires 32
- I.3.1. Description and modeling 32
- I.3.2. Wildland fire risk and danger 38

I.1. Uncertainty quantification

Some degree of discrepancy is inevitable between model results and observations, and can be significant in some applications. Uncertainty quantification (UQ) focuses on the uncertainty of outputs resulting from numerical models by accounting for the sources of uncertainty that may lead to differences between simulated and observed values. As such, several uncertainty sources may be identified, including:

— input uncertainty, when there is a range of possible values for a given model parameter or data source, possibly due to the use of forecasts or difficulty to estimate a single "best" value;

— model uncertainty, which can be due to modeling assumptions or more generally to the fact that a model offers a simplified representation of the observed phenomenon;
— numerical uncertainty, referring to machine precision of numerical models and the possible errors implied by approximation of mathematical models by numerical ones (e.g. to solve systems of partial differential equations);
— measurement error, due to the accuracy and precision of the system used to provide the observation.

A key aspect of UQ is to represent the uncertainty of the modeling elements, even when the model is deterministic, by means of probability distributions.

I.1.1. Propagation of uncertainty

Let us denote a *deterministic model* as $y = \mathcal{M}(x)$, where $y \in \mathbb{R}$ and $x \in \mathbb{R}^d$. Typically, y is a quantity of interest that is a direct output of a numerical model $\mathcal{M}$ and the value of y is always the same for a given x. A simple UQ framework consists in replacing the deterministic value x with a random vector X whose probability distribution accounts for input uncertainty, yielding a *probabilistic model*:

$$Y = \mathcal{M}(X), \tag{I.1}$$

whose output Y is now a random variable. Other sources of uncertainty can be included in the probabilistic model. For instance, measurement error can be modeled via a noise term ϵ_{meas} (e.g. following a Gaussian distribution) independent from X; whereas model uncertainty can be accounted for via another independent noise term ϵ_{mod} and/or via a multi-model approach by means of z models $\mathcal{M}_1, \ldots, \mathcal{M}_z$ whose choice would be determined according to a random variable Z (e.g. following a uniform distribution in $\{1, \ldots, z\}$ in the case a of a finite number of models). Including these elements in model (I.1) would lead to the following model:

$$Y = \mathcal{M}_Z(X) + \epsilon_{\text{mod}} + \epsilon_{\text{meas}}. \tag{I.2}$$

The rest of the present section focuses on model (I.1) but can be easily extended to model (I.2).

As a first approach, one may be interested in moments of Y. Assuming $\mathbb{E}[|Y|] < \infty$ and that X has probability density function (PDF) f, one may calculate the expected value $\mathbb{E}[Y] = \int_{\mathbb{R}^d} \mathcal{M}(x) f(x) \mathrm{d}x$. Depending on the complexity of the numerical model $\mathcal{M}$ and the high input dimension d, deterministic methods based on quadrature rules for each dimension may not provide a satisfactory approximation of $\mathbb{E}[Y]$, whereas Monte Carlo (MC) methods may be better-suited for this task. MC methods only require an independent identically distributed (i.i.d.) sample; in the case at hand, provided an i.i.d. sample $X_1, \ldots, X_n$ following the probability law $\mathcal{L}(X)$ of X, then the resulting output sample $Y_1 = \mathcal{M}(X_1), \ldots, Y_n = \mathcal{M}(X_n)$ is also i.i.d. and follows $\mathcal{L}(Y)$. Therefore, well-known theorems from probability theory can be applied to the classical estimator $\overline{Y}$, defined as follows:

$$\overline{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i, \tag{I.3}$$

which is an unbiased estimate (i.e. $\mathbb{E}[\overline{Y}] = \mathbb{E}[Y]$). The law of large numbers yields that $\overline{Y}$ converges to $\mathbb{E}[Y]$ as $n \to \infty$ and, assuming that Y has finite variance $\mathrm{Var}[Y]$, the central limit theorem (CLT) states that random variable $\sqrt{n}(\overline{Y} - \mathbb{E}[Y])$ converges in law to a normal distribution $\mathcal{N}(0, \mathrm{Var}[Y])$. More generally, MC methods lead to an empirical distribution represented by $(Y_1, \ldots, Y_n)$, which can be used to approximate the distribution of Y.

Probabilistic convergence of statistical estimators is usually quantified by means of confidence intervals (CI). In the context of numerical models, these intervals can help in determining whether the sample size n is sufficient. A traditional way of providing a CI is to assume that the sample is i.i.d. following a normal law, in which case the following CI at level $1 - \alpha$ (usually 95%) for the mean is obtained, given $\alpha \in]0, 1[$:

$$\mathbb{P}\left[\left|\overline{Y} - \mathbb{E}[Y]\right| \leq z_{\alpha/2}\left(\frac{\mathrm{Var}[Y]}{n}\right)^{1/2}\right] = 1 - \alpha \tag{I.4}$$

where z_q is the quantile of the standard normal distribution $\mathcal{N}(0, 1)$ for probability q. If Y does not follow a normal distribution, an *asymptotic* CI can be obtained based on the CLT, i.e. the left-hand term of Equation (I.4) converges to $1 - \alpha$ as $n \to \infty$. In practice, $\mathrm{Var}[Y]$ is also unknown but can be estimated using S^2 defined as follows:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2, \tag{I.5}$$

and $\sqrt{n}(\overline{Y} - \mathbb{E}[Y])/S$ also converges in distribution to $\mathcal{N}(0, 1)$ as $n \to \infty$, which implies

$$\mathbb{P}\left[\left|\frac{\sqrt{n}(\overline{Y} - \mathbb{E}[Y])}{S}\right| \leq z_{\alpha/2}\right] \underset{n\to\infty}{\longrightarrow} 1 - \alpha, \tag{I.6}$$

and $\overline{Y} \pm z_{\alpha/2} S/\sqrt{n}$ is an asymptotic CI for $\mathbb{E}[Y]$ at level $1 - \alpha$.

Alternatives for CIs may be obtained depending on the properties of Y. For instance, if y is binary and equal to 1 under some condition (e.g. exceeding a given threshold) and equal to 0 otherwise, Y is a Bernoulli variable; therefore, defining parameter $p = \mathbb{P}[Y = 1] = \mathbb{E}[Y]$, Hoeffding's inequality yields:

$$\mathbb{P}\left[\left|\overline{Y} - p\right| \leq \left(\frac{\log(2/\alpha)}{2n}\right)^{1/2}\right] \geq 1 - \alpha. \tag{I.7}$$

In the absence of useful properties of assumptions regarding the distribution of Y, an asymptotic CI can be obtained via *empirical bootstrap* [39], which may apply to other statistical quantities than the mean. Let us denote ψ a statistical quantity of Y and $\widehat{\psi}$ an estimator of ψ based on the i.i.d. sample $(Y_1, \ldots, Y_n)$. Empirical bootstrap consists in drawing a random sample of size n with replacement from $(Y_1, \ldots, Y_n)$, denoted $(Y_1^\star, \ldots, Y_n^\star)$ and called a *bootstrap sample*. From a bootstrap sample, one may compute the estimator of ψ; and by drawing m independent bootstrap samples, an ensemble of m estimators $(\psi_1^\star, \ldots, \psi_m^\star)$ is obtained. Many variants of the bootstrap procedure exist, but the core idea is to draw samples with replacement among the

original sample. Regarding empirical bootstrap, a CI can be derived using the ensemble of m estimators. A first method consists in using the empirical quantiles of ensemble $(\psi_1^\star, \ldots, \psi_m^\star)$. The quantile at probability q being denoted as $\psi_{[q]}^\star$, a CI at level $(1-\alpha)$ for ψ can be approximated using the following interval:

$$\left[\psi_{[\alpha/2]}^\star, \psi_{[1-\alpha/2]}^\star\right], \tag{I.8}$$

which is called the *percentile interval*. A second method consists in computing the estimator of the variance defined in Equation (I.5), but applied to the m-sized sample $(\psi_1^\star, \ldots, \psi_m^\star)$. The standard deviation $\sigma^\star$ of this ensemble can be used to provide the following interval:

$$\widehat{\psi} \pm z_{\alpha/2}\ \sigma^\star, \tag{I.9}$$

which is called the *standard interval*. For both intervals, a large sample size n and a large number of bootstrap samples m is desired. Notably, the standard interval in Equation (I.9) can be proved to be an asymptotic CI at level $1-\alpha$ provided that the functional defining the statistical estimator $\widehat{\psi}$ is smooth enough in regards to distributions (e.g. Fréchet differentiability is a sufficient condition, see [122]).

Remark: In the context of statistical estimation, a CI provides a measure of probabilistic convergence, which is not to be mistaken for other estimated quantities such as the interval between empirical quantiles of $(Y_1, \ldots, Y_n)$. Uncertainty relates to the probability distribution of Y and a possible "measure" of uncertainty can be expressed by its variance $\mathrm{Var}[Y]$ or the interval between its quantiles for instance. The use of a finite sample to estimate these quantities can be considered as a source of uncertainty but is usually neglected in UQ compared to other sources, provided that n is large enough.

An application related to propagation of uncertainty is sensitivity analysis (SA), which consists in quantifying the influence of the uncertain inputs on model output. There is a distinction between local SA which focuses on local variation, usually by means of partial derivatives at an input value $x = (x_1, \ldots, x_d)$, and global SA where the whole definition domain of the input is studied. Sensitivity can be investigated qualitatively via pair plots and quantitatively by means of traditional statistical techniques such as linear regression, fractional plans, variation of input components one-at-a-time, etc. Depending on the complexity of the model (whether it is linear or monotonous, how regular it is), more elaborate SA techniques may be required. For global SA, which can be carried out to complement uncertainty quantification, Sobol' indices [126] provide measures of sensitivity, based on the decomposition of variance. Assuming that the components of the random input vector X are i.i.d. following a uniform distribution in $[0, 1]$, the variance can be decomposed as follows:

$$\mathrm{Var}[Y] = \sum_{i=1}^{d} V_i + \sum_{i<j}^{d} V_{ij} + \ldots + V_{1\ldots d} \tag{I.10}$$

where $V_i = \mathrm{Var}\big[\mathbb{E}[Y|X_i]\big]$ is the variance of the conditional mean value of Y on X_i (i.e. the expected value is taken relatively to all input components except x_i, and the

variance is therefore taken relatively to x_i only), $V_{ij} = \mathrm{Var}\big[\mathbb{E}[Y|X_i, X_j]\big] - V_i - V_j$ expresses the influence of the interaction of input components x_i and x_j on the output, and so on. Dividing by $\mathrm{Var}[Y]$ in Equation (I.10) yields the Sobol' indices:

$$\sum_{i=1}^{d} S_i + \sum_{i<j}^{d} S_{ij} + \ldots + S_{1\ldots d} = 1 \tag{I.11}$$

where $S_i = V_i/\mathrm{Var}[Y]$, and so on. For a given i, $S_i \in [0,1]$ is called the *first-order sensitivity index* and measures the overall influence of varying input variable x_i while ignoring the interaction effects with other input components measured by the higher-order indices S_{ij}, S_{ijk}, and so on. To account for these interaction effects, one may consider the *total-effect index*, which is defined for a given i as follows:

$$S_{T_i} = 1 - \frac{\mathrm{Var}\big[\mathbb{E}[Y|X_{\sim i}]\big]}{\mathrm{Var}[Y]}, \tag{I.12}$$

where $X_{\sim i}$ stands for the subset of all components of X except X_i. S_{T_i} measures the fraction of variance that is due to not conditioning input component x_i and is equal to the sum of the first-order index S_i and all higher-order components where i belongs to the subset of conditional inputs; hence, if $S_{T_i} = 0$, x_i can be considered to be non-influential, and if $S_{T_i} = S_i$, one can consider that the interactions between x_i and the other variables are non-existent. Estimation all 2^d Sobol' indices is too costly when d is high, so the analysis if often restricted to the computation of the d first-order and d total-order indices. Estimation can be carried out by MC methods, but more efficient computational methods (i.e. requiring lower sample size) have been proposed.

Alternatives to MC can be considered, notably when the computational time of the numerical model is high implying that only a low sample size n will be obtained at reasonable expense of time. In this situation, quasi-Monte Carlo methods (e.g. using low discrepancy sequences) can be used for statistical estimation. Another alternative when a high number of model calls are required (e.g. for the estimation of Sobol' indices) is to use a metamodel. More details on these methods can be found in Section I.2.

I.1.2. Evaluation and dissimilarity measures of probability distributions

A predictive model can be evaluated by comparing its outputs with observations. Considering a deterministic model with N outputs $y^{(1)}, \ldots, y^{(N)}$ and corresponding observations $y_{\mathrm{obs}}^{(1)}, \ldots, y_{\mathrm{obs}}^{(N)}$, one may use common error metrics such as the mean squared error (MSE), defined as follows:

$$\mathrm{MSE} = \frac{1}{N} \sum_{i=1}^{N} \left(y^{(i)} - y_{\mathrm{obs}}^{(i)} \right)^2. \tag{I.13}$$

In the case of a probabilistic model, evaluation is less direct. A possible approach is to use an sample of n sets of outputs (e.g. obtained via a MC method) and compute the deterministic score for each of the n sets then study its distribution as if it were the quantity of interest mentioned in Section I.1.1. Another approach is to use so-called probabilistic scores. In a binary setting where the i-th observation corresponds to the

occurrence of an event A_i, let $y_{\text{obs}}^{(i)}$ be equal to 1 if A_i occurs, 0 otherwise. For simplicity, using notation $y_{\text{obs}}^{(i)} = o_i$ and defining $p_i = \mathbb{P}[A_i]$ the probability of A_i according to the model, the Brier score [18], denoted as BS, is defined as follows:

$$\text{BS} = \frac{1}{N}\sum_{i=1}^{N}(o_i - p_i)^2. \tag{I.14}$$

This score measures the *accuracy* of the probabilistic models, that is to say the overall agreement between output and observation. BS ranges between 0 and 1 and is negatively oriented: the lower BS, the more accurate the model. The definition in Equation (I.14) can be seen as an *empirical* estimation of an expected value, i.e. for large N:

$$\text{BS} \approx \mathbb{E}\big[(O-P)^2\big], \tag{I.15}$$

where O and P are both random variables defined on the same probability space with values in $\{0,1\}$ and $[0,1]$, respectively. A classical decomposition of the Brier score [97] can be derived from its probabilistic expression as follows:

$$\begin{aligned}
\mathbb{E}\big[(O-P)^2\big] &= \mathbb{E}\big[\,\mathbb{E}[(O-P)^2|P]\big] \\
&= \mathbb{E}\big[\,\mathbb{E}[(O^2-2OP+P^2)|P]\big] \\
&= \mathbb{E}\big[\,\mathbb{E}[O^2|P] - 2P\mathbb{E}[O|P] + P^2\big] \\
&= \mathbb{E}\big[\,\mathbb{E}[O|P] - 2P\mathbb{E}[O|P] + \mathbb{E}[O|P]^2 - \mathbb{E}[O|P]^2 + P^2\big] \\
&= \mathbb{E}\big[\,(P-\mathbb{E}[O|P])^2 + \mathbb{E}[O]^2 - \mathbb{E}[O]^2 - 2\mathbb{E}[O]\mathbb{E}[O|P] + 2\mathbb{E}[O]\mathbb{E}[O|P] \\
&\qquad + \mathbb{E}[O|P] - \mathbb{E}[O|P]^2\big] \\
&= \mathbb{E}\big[\,(P-\mathbb{E}[O|P])^2 - (\mathbb{E}[O|P]-\mathbb{E}[O])^2 + \mathbb{E}[O]^2 - 2\mathbb{E}[O]\mathbb{E}[O|P] + \mathbb{E}[O|P]\big] \\
&= \mathbb{E}\big[\,(P-\mathbb{E}[O|P])^2\big] + \mathbb{E}\big[(\mathbb{E}[O|P]-\mathbb{E}[O])^2\big] + \mathbb{E}[O]^2 + (1-2\mathbb{E}[O])\mathbb{E}\big[\mathbb{E}[O|P]\big] \\
&= \mathbb{E}\big[\,(P-\mathbb{E}[O|P])^2\big] + \mathbb{E}\big[(\mathbb{E}[O|P]-\mathbb{E}[O])^2\big] + \mathbb{E}[O]\mathbb{E}[O] + (1-2\mathbb{E}[O])\mathbb{E}[O] \\
&= \mathbb{E}\big[\,(P-\mathbb{E}[O|P])^2\big] + \mathbb{E}\big[(\mathbb{E}[O|P]-\mathbb{E}[O])^2\big] + \mathbb{E}[O](1-\mathbb{E}[O]),
\end{aligned}$$

which leads to three additive components: *reliability*, (probabilistic) *resolution*, and *uncertainty*:

$$\mathbb{E}\big[(O-P)^2\big] = \underbrace{\mathbb{E}\big[(P-\mathbb{E}[O|P])^2\big]}_{\text{reliability}} - \underbrace{\mathbb{E}\big[(\mathbb{E}[O|P]-\mathbb{E}[O])^2\big]}_{\text{probabilistic resolution}} + \underbrace{\mathbb{E}[O](1-\mathbb{E}[O])}_{\text{uncertainty}}. \tag{I.16}$$

The conditional mean $\mathbb{E}[O|P]$ can be understood as as the proportion of events associated to a given value of modeled probability P that actually occur; when they are both equal regardless of the value of P, the probabilistic model is called *reliable* and the first component of Equation (I.16) is null. Here, the uncertainty component is equal to the variance of O, and can be understood as the inherent variability of the observed events. The first step of the decomposition relies on conditioning by P, which, in terms of empirical estimation, consists in sorting the events by the possible values of P. Assuming a discrete setting where there are $n+1$ possible values of P denoted as $p^0, \ldots, p^n$, such that $g(p^j)$ is the proportion of instances with modeled probability p^j, $f(p^j)$ is the proportion of events that occur among those attributed with modeled probability p^j, and p_c is the mean of O estimated from the events, the following discrete

decomposition of the Brier score is obtained:

$$BS = \underbrace{\sum_{j=0}^{n}(p^j - f(p^j))^2 g(p^j)}_{\text{reliability}} - \underbrace{\sum_{j=0}^{n}(f(p^j) - p_c)^2 g(p^j)}_{\text{probabilistic resolution}} + \underbrace{p_c(1-p_c)}_{\text{uncertainty}}, \tag{I.17}$$

which can be obtained from Equation (I.14) by re-arranging the sum of over index j instead of i. More details on desirable properties of a probabilistic prediction systems such as accuracy and reliability are provided in Section II.3.

The Brier score can be seen as an indicator of the accuracy between two probabilistic models with the particularity that, while P may take any value in $[0, 1]$, the "observed" probabilistic model may only return 0 or 1. Dissimilarity between probability distributions can be indicated by a distance between them, such as the Hellinger distance. In the case of two discrete probability distributions with respective probabilities $(p_1, \ldots, p_K)$ and $(q_1, \ldots, q_K)$, the square of Hellinger distance is expressed as follows:

$$H(p, q)^2 = \frac{1}{2}\sum_{k=1}^{K}\left(\sqrt{p_k} - \sqrt{q_k}\right)^2, \tag{I.18}$$

but how to use it to measure the similarity between a probabilistic model and observation is not obvious, while probabilistic scores such as the Brier score are suited to the evaluation of probabilistic models.

When it makes sense to represent the observation by a probability distribution, however, a distance between probability can provide a relevant dissimilarity measure for evaluation. As such, in Chapter III, the Wasserstein distance is proposed as a deterministic score to compare an observed surface with its modeled counterpart, and the choice of this distance is motivated by its use in the field of optimal transport (see for instance [120] for more details). Given two separable metric spaces $\mathcal{X}$ and $\mathcal{Y}$ on which are defined the measures μ and ν respectively, the optimal transport problem as formulated by Kantorovitch consists in finding the infimum

$$\inf\left\{\int_{\mathcal{X}\times\mathcal{Y}} c(x, y)\ d\gamma(x, y)\ \middle|\ \gamma \in \Gamma(\mu, \nu)\right\}, \tag{I.19}$$

where $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_+ \cup \{\infty\}$ is a measurable function and $\Gamma(\mu, \nu)$ is the ensemble of the measures defined on $\mathcal{X} \times \mathcal{Y}$ such that the marginal measure on $\mathcal{X}$ is μ and the marginal measure on $\mathcal{Y}$ is ν. The optimal transport problem can be defined in different equivalent ways such as the dual formulation and the Benamou-Brenier formulation, possibly up to some factor. The original formulation by Monge, on the other hand, is a bit different and not always equivalent to (I.19). The function c can be interpreted as a cost and γ as a mapping, so that $c(x, y)$ quantifies what is required to move x to y and $\gamma(x, y)$ is the amount of mass that is moved. The Wasserstein distance is obtained in a specific case where $\mathcal{X} = \mathcal{Y} = \mathbb{R}^s$ and c is the squared Euclidean distance on $\mathbb{R}^s$. We denote it as $\mathcal{W}_2(\mu, \nu)$ and it is the square root of the infimum of the optimal transport

problem, i.e.

$$\mathcal{W}_2^2(\mu,\nu) = \inf\left\{ \int_{\mathbb{R}^s\times\mathbb{R}^s} ||x-y||_2^2 \, d\gamma(x,y) \;\middle|\; \gamma \in \Gamma(\mu,\nu) \right\}, \tag{I.20}$$

where $||.||_2$ is the Euclidean distance in $\mathbb{R}^s$. Typically $s \leq 3$ and for comparison of surfaces, we have $s = 2$. $\mathcal{W}_2^2(\mu,\nu)$ can be interpreted as the minimum amount of energy that is required to move the mass that is distributed according to μ so that, after transport, it is distributed according to ν.

I.1.3. Inverse uncertainty quantification

Propagating uncertainty in the probabilistic model (cf. Equation (I.1)) can be referred to as solving a problem of *direct* uncertainty quantification. When used for prediction purposes, performance of such a model can be evaluated by comparing the output probability distribution with observations using probabilistic scores. One may also consider using the knowledge of the observations to determine a probability distribution of the random input vector X that is more relevant. This approach that focuses on the input can be referred to as solving a problem of *inverse* UQ, by contrast to the direct problem that focused on the probability distribution of output Y. Still, as the strategy for updating the input distribution consists in providing better similarity between model output and observation, a consequence of inverse UQ is updating the output distribution so as to increase predictive performance.

A traditional way to fit the distribution of a random variable based on observations is to use the likelihood. Let us assume that the law $\mathcal{L}(Y)$ is in a given family of distributions parameterized by $\theta \in \Theta \subset \mathbb{R}^r$ (e.g. for a normal distribution, θ would be the mean and variance of Y). Based on the realization $y_1, \ldots, y_N$ of a i.i.d. sample following $\mathcal{L}(Y)$, the likelihood denoted as $\mathcal{L}(y_1, \ldots, y_N|\theta)$ is simply the value of the joint PDF of the random vector $(Y_1, \ldots, Y_N)$ corresponding to this realization. In the case of a distribution admitting PDF $p_\theta : y \in \mathbb{R} \mapsto p_\theta(y)$, the likelihood is expressed as the following product due to the independence of the Y_i:

$$\mathcal{L}(y_1, \ldots, y_N|\theta) = \prod_{i=1}^{N} p_\theta(y_i). \tag{I.21}$$

Maximum likelihood estimation consists in finding $\theta \in \Theta$ that maximizes (I.21) and the resulting estimate, denoted as $\widetilde{\theta}$, is then chosen to characterize the law of Y.

In the context of UQ, it can be difficult to determine a given family of distributions for Y depending on the complexity of $\mathcal{M}$. Also, even with a suitable distribution of Y obtained using MLE, providing a distribution for X that satisfies $Y = \mathcal{M}(X)$ is not trivial. This difficulty of inverse UQ can be addressed by adopting a Bayesian framework. Denoting $p(x,y)$ the joint PDF of (X,Y) at point (x,y) and $p(y)$ the marginal PDF of Y at point y, the conditional PDF of X given Y is defined for y such that $p(y) \neq 0$ as:

$$p(x|y) = \frac{p(x,y)}{p(y)}, \tag{I.22}$$

which yields the following expression of Bayes' rule for random variables X and Y:

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}, \tag{I.23}$$

which is also valid for random vectors. Under this framework, $x \mapsto p(x)$ is referred to as the *prior* PDF of X and the conditional PDF $x \mapsto p(x|y)$ as its *posterior* PDF. In UQ, the prior distribution of X is generally determined by expert knowledge, and inverse UQ will consist in using the posterior distribution of X instead of the prior that would have been used in direct UQ. There are links between inverse UQ and Data Assimilation (DA) in the sense that the "analysis" of Y provided by DA is analogous to a posterior distribution of Y and is generally determined via the assumption of a Gaussian distribution and by focusing on the first two moments of Y.

A crucial step of inverse UQ is to define $p(y|x)$; which is typically derived from a relevant assumption regarding the relationship between X and the observation, which will also be represented by random variable Y for the purpose of notation. It is common to assume that given the input uncertainty of X the difference between Y and $\mathcal{M}(X)$ follows a normal distribution, therefore one may write:

$$p(y|x) = p(y|\mathcal{M}(x)) = \mathcal{L}(y - \mathcal{M}(x)|\theta) = p_\theta(y - \mathcal{M}(x)), \tag{I.24}$$

where θ stands for the mean and variance of the normal distribution, and $p(y|x)$ is then referred to as the likelihood of the observation knowing x and θ. In the case of multiple observations, we may extend Equation (I.24) to a multivariate case where $\mathcal{M}(x), y \in \mathbb{R}^N$ by taking p_θ as the PDF of a Gaussian vector of size N, and θ would refer to the vector of mean and the correlation matrix.

Regardless of the choice for the prior and the likelihood in Equation (I.23), dealing the posterior PDF is not trivial, especially when the input dimension d is high. Depending on model complexity, derivatives of the likelihood – which depends on $\mathcal{M}(x)$ – may not be obtained easily. Also, even though the denominator $p(y)$ may be expressed as the integral of the numerator over the definition domain of x, this quantity may not be tractable. Monte Carlo Markov chain (MCMC) methods can be used to address problems such as finding x so that the maximum of $p(x|y)$ is reached, or drawing a sample following the posterior distribution. Using the fact that the denominator does not depend on x, i.e. $p(x|y) \propto p(y|x)p(x)$, the former problem can be solved by means of a simulated annealing technique, whereas a Metropolis-Hastings (MH) algorithm can be used to solve the latter. Considering a random vector with PDF $x \mapsto f(x)$, one may construct a Markov chain with homogeneous transition kernel admitting an invariant probability measure corresponding to f. MH algorithms consist in generating a realization of the said Markov chain based on an initial state x^0 and, iterate by generating a candidate x_c based the previous state x^k via an instrumental distribution that may be accepted or not according to the transition kernel. When applied to the posterior PDF $x \mapsto p(x|y)$, this yields Algorithm 1. As can be seen in Equation (I.25), the target PDF of the MH algorithm only needs to be known up to a constant to calculate ratio τ, which is the case of the posterior PDF.

Although MCMC techniques are suited to inverse UQ, classical expressions of the likelihood may not be relevant for any type of observations. For instance, a surface can

Algorithm 1 Metropolis-Hastings algorithm applied to $p(x|y)$ from Equation (I.23)

Define K, and an instrumental distribution of PDF $q : \boldsymbol{x} \mapsto q(x|x')$
Choose a starting point x^0
for $k = 1, \ldots, K$ **do**
  Sample a candidate $x_c \sim q(.|\boldsymbol{x}^{k-1})$
  Compute the ratio

$$\tau = \frac{p(x_c|y)q(x^{k-1}|x_c)}{p(x^{k-1}|y)q(x_c|x^{k-1})} = \frac{p(y|x_c)p(x_c)q(x^{k-1}|x_c)}{p(y|x^{k-1})p(x^{k-1})q(x_c|x^{k-1})} \tag{I.25}$$

  if $\tau \geq 1$ **then**
    $x^k \leftarrow x_c$ (accept the candidate)
  else
    (accept the candidate with probability τ)
    Sample $u \sim \mathcal{U}(0, 1)$
    if $u \leq \tau$ **then**
      $x^k \leftarrow x_c$
    else
      $x^k \leftarrow x^{k-1}$
    end if
  end if
end for
return $(x^0, \ldots, x^K)$

be defined in a binary fashion where any point in space either belongs to the surface or not, and the closer the points, the higher the correlation; these aspects are not well represented by assuming a normal distribution and independence. Also, satisfactory results from MCMC techniques generally require a high number of iterations that cannot be easily carried out in parallel, involving many subsequent numerical model calls, and possibly high computational time.

I.2. Metamodeling

A *metamodel* is, as the name implies, a "model of a model". In the context of numerical models, metamodeling consists in building an approximation function $\widehat{\mathcal{M}}$ of the underlying function $\mathcal{M}$ that links the input x to the quantity of interest y. The resulting metamodel, which may also be referred to as *surrogate model* or *emulator* is intended to have much lower computational time than the original model $\mathcal{M}$, for the purpose of UQ for instance. Many methods can be used for metamodeling, including linear regression, nonlinear regression, partial least squares regression, support vector machine, interpolation techniques (by nearest neighbor, splines, etc.), random forests, radial basis functions, etc. The metamodeling procedure can summarized in three main steps:

1. Generate a set of inputs $(x_1, \ldots, x_n)$ and compute their corresponding outputs $(y_1 = \mathcal{M}(x_1), \ldots, y_n = \mathcal{M}(x_n))$, therefore constituting a *training dataset*.

2. Build the metamodel on the basis of the training dataset, typically by fitting the parameters of the metamodel so that $\widehat{\mathcal{M}}(x_i) \approx \mathcal{M}(x_i)$ for all i (there is equality in the case of interpolation methods).
3. Evaluate the metamodel, which can be done both in terms of gain in computational time (compared to the use of $\mathcal{M}$), and in terms of approximation error. The latter aspect should be quantified not only for the training dataset but ideally for any possible value of input x.

Depending on the complexity of model $\mathcal{M}$, classical methods such as linear regression, may lead to poor approximation. The present section is divided into three parts. First, some techniques used to generate relevant datasets for metamodeling are presented. The following two parts each focus on a possible approach that can be used to build a metamodel, namely using Gaussian processes and using neural networks.

I.2.1. Design of experiments

The notion of *design of experiments* (DOE) refers to a task that aims to study one or several quantities of interest given the variation of conditions that are assumed to have an influence on these quantities. In practice, a DOE determines the conditions for carrying out given number of physical or numerical experiments. A traditional procedure is to carry out randomized experiments; but for applications of the DOE such as sensitivity analysis or studying response surfaces, one may seek a DOE that is better-suited to the application. In linear regression for instance, factorial designs can be used to study the individual effects of the input variables (first-order) and their interactions at the second-order.

In the context of metamodeling, one of the two goals is to build $\widehat{\mathcal{M}}$ that ensures low approximation error. In the following, it is assumed that the hypersquare $[0,1]^d$ is the definition domain of the input [1]. Given that metamodels are characterized by low (perhaps null) approximation error at points in the training sample, the error should be low as well in their neighborhood if both $\mathcal{M}$ and $\widehat{\mathcal{M}}$ are continuous. Therefore, an intuitive strategy is to choose all the points at the nodes of an orthogonal grid over $[0,1]^d$ having the same resolution for all d components. However, for high dimension d, this implies a number of points that may be unreasonably high even at relatively low grid resolution; for instance, 10 possible values for each input component will lead to a training sample of size 10^d. It follows that a desirable DOE should provide a sample that is space filling, i.e. that provides good cover of $[0,1]^d$, even with limited sample size n and high dimension d.

Several criteria can indicate how well the points of the design are distributed. Discrepancy measures provide a quantitative assessment of the deviation of the design from a uniform distribution. A commonly used discrepancy measure is the L^∞-star discrepancy, denoted as $D^\star_\infty$ for a design $(x_1, \ldots, x_n)$ of points in $[0,1]^d$ and expressed

1. More generally, the model input x may be defined over a hyperrectangle of dimension d: $\prod_{l=1}^d [L^l, U^l]$, with $L^l < U^l \in \mathbb{R}$ for all $l \in \{1, \ldots, d\}$. The DOE should be carried out for $x' \in [0,1]^d$ obtained from x by means of affine transformations applied to each input component independently. The corresponding values of x can then be easily obtained from x' via the inverse affine transformations.

as follows:

$$D^{\star}_{\infty} = \sup_{z\in[0,1]^d} \left| \frac{1}{n} \#\left\{ i \middle| x_i \in \prod_{l=1}^{d} [0, z^l] \right\} - \prod_{l=1}^{d} z^l \right|, \tag{I.26}$$

where the z^l are the d components of a given vector $z \in [0,1]^d$, and $\#\left\{ i \middle| x_i \in \prod_{l=1}^{d} [0, z^l] \right\}$ is the cardinal of the ensemble of points in the design that belong to the hyperrectangle whose edges are delimited by the null vector in $\mathbb{R}^d$ and z. For a given z, the difference involved in Equation (I.26) is referred to as *local discrepancy* given that the aforementioned cardinal divided by n corresponds to the empirical estimate of the volume of the associated hyperrectangle provided an i.i.d. uniform sample. Therefore, the L^{∞}-star discrepancy identifies the highest possible absolute value of this local discrepancy over $[0,1]^d$ and a desirable property of the DOE is to have low discrepancy. Other measures of discrepancy may rely on other norms than the L^{∞}-norm, and the family of subsets of $[0,1]^d$ for the definition of local discrepancy may also differ [67] (in Equation (I.26), the said family is constituted of the hyperrectangles delimited by the null vector and any point in $[0,1]^d$). For computational purposes, the L^2-norm is preferred to the L^{∞}-norm, and a remarkable measure is the centered L^2-discrepancy, denoted as C^2, whose square has the following closed-form expression:

$$\begin{aligned} C_2^2 =& \left(\frac{13}{12}\right)^d - \frac{2}{n} \sum_{i=1}^{n} \prod_{l=1}^{d} \left(1 + \frac{1}{2}|x_i^l - 0.5| - \frac{1}{2}|x_i^l - 0.5|^2\right) \\ &+ \frac{1}{n^2} \sum_{i,j=1}^{n} \prod_{l=1}^{d} \left(1 + \frac{1}{2}|x_i^l - 0.5| + \frac{1}{2}|x_j^l - 0.5| - \frac{1}{2}|x_i^l - x_j^l|\right). \end{aligned} \tag{I.27}$$

Other criteria may be defined on the distance between points of the sample, such as the minimal distance between any two points of the DOE, defined as follows:

$$\min_{1\leq i<j\leq n} ||x_i - x_j||_2, \tag{I.28}$$

where $||.||_2$ is the Euclidean distance in $\mathbb{R}^d$. A space filling DOE should have a high value of this distance as this implies that the closest neighbors of any point in the sample are located relatively far from it, which should help in avoiding redundant of information (close value the output for close input points). As this criterion is intended to be maximized, it is often referred to as the *maximin* criterion.

A simple DOE consists in drawing an i.i.d. sample following a uniform distribution in $[0,1]^d$, but such samples are known to be poorly space filling. A possibly better approach is to use a Latin Hypersquare Sample (LHS). A LHS of size n relies on a subdivision in n intervals for each input component, denoted as $a_0^l < a_1^l < \ldots < a_n^l$ for $l \in \{1, \ldots, d\}$. In general, the subdivision is chosen so that the intervals have equal probability based on the probability distribution associated to the input component. In our case, we have uniform distributions in $[0,1]$, hence $\forall l \in \{1, \ldots, d\}, \forall i \in \{0, \ldots, n\}$, $a_i^l = i/n$. A sample $(x_1, \ldots, x_n)$ is a LHS if $\forall l \in \{1, \ldots, d\}$, exactly one sample point has it j-th component in each of the n intervals of the subdivision $[a_0^l, a_1^l[, \ldots, [a_{n-1}^l, a_n^l[$. Inside its designated interval, the l-th input component of point sample i may be sampled according to a uniform distribution in our case. As can be

seen in the three examples in Figure I.1, such a DOE may be poorly space filling, so it is common practice to optimize a discrepancy or distance criterion of a LHS; this can be done either by simply draw several of them and selecting the best sample or by using stochastic optimization techniques such as simulated annealing, as proposed in [30]. Numerical implementation of stochastic LHS optimization algorithms is proposed in the R-package DiceDesign [38].

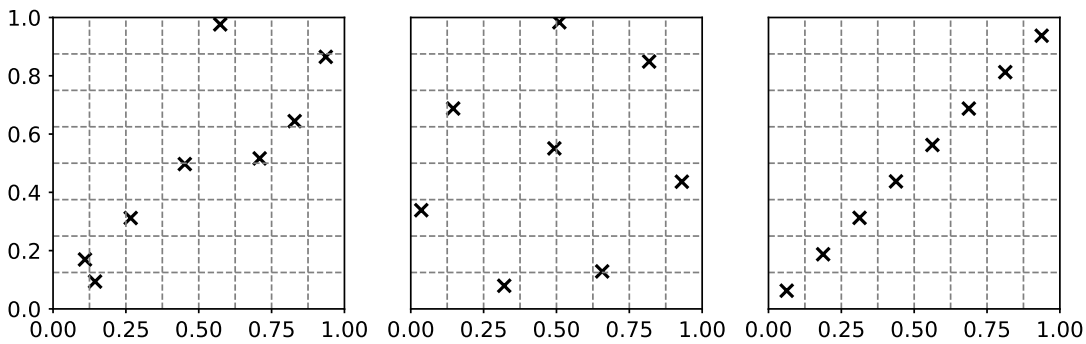


Figure I.1. – Examples of LHS of size $n = 8$ in dimension $d = 2$.
Left: randomly drawn LHS; middle: LHS with highest maximin out of 10,000 randomly drawn samples; right: LHS with points aligned on the first bisector.
The design in the middle has much higher maximin, therefore being more space filling than the other two.

Low-discrepancy sequences such as Halton [61], Hammersley [63], Sobol [125], Faure [41] or Neiderreiter [98] sequences are another type of space filling design. Similarly to traditional pseudo-random number generators, a sequence of points is generated, but in a fashion to ensure that the resulting sample has low discrepancy, as illustrated in Figure I.2. Due to the choice of the parameters for some of the low-discrepancy sequences, some numerical implementations may have limited dimension d. One may use the C++ implementation of Joe and Kuo [2] for generating Sobol sequences in high dimension [72, 73].

Ideally, the metamodel approximation should be evaluated by metrics such as the maximum error in absolute value, but such quantities may be difficult to estimate when $\mathcal{M}$ is a complex model. At the very least, the approximation error can be computed individually for each member of the training dataset and one may consider metrics such as the MSE (cf. Equation (I.13) but applied to the differences between the $\widehat{\mathcal{M}}(x_i)$ and the $\mathcal{M}(x_i)$ for $i \in \{1, \ldots, n\}$). Still, these empirical metrics may not be representative of the overall metamodel error in the entire input space, given that a metamodel is built so that there is little approximation error at the points of the training dataset (the error is null when interpolation techniques are used). Because some *overfitting* of the metamodel may be unavoidable, one should evaluate approximation error using a dataset constituted of points that were not used for training, referred to as a *test dataset*. A simple way to obtain the test dataset is by *splitting* the whole available dataset in two

2. available at https://web.maths.unsw.edu.au/~fkuo/sobol/, last checked 2021.01.18.

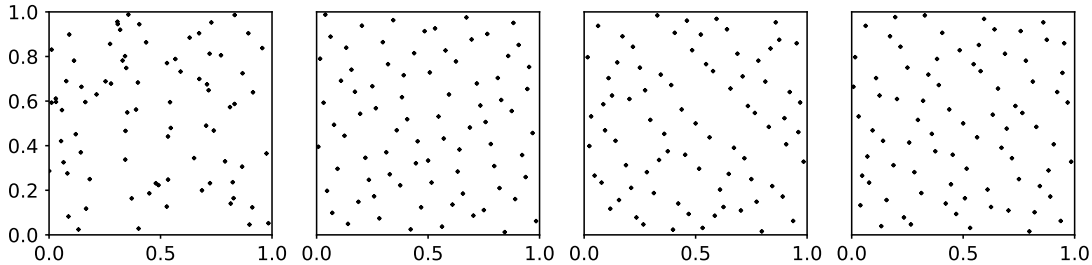


Figure I.2. – Examples of uniform samples of size $n = 80$ in $[0, 1]^2$ generated by (quasi-)random number sequences.
From left to right: i.i.d. sample using a traditional RNG; sample using a Halton sequence; sample using a Sobol sequence; sample using a Faure sequence.
Using the low-discrepancy sequences lead to a sample that appears to be more "representative" of the uniform distribution.

subsets respectively constituting the training and test datasets (typically in proportions 80% and 20%, respectively), which is common practice when the full dataset is imposed. In metamodeling, since it is possible to construct one's own dataset, the test dataset may be generated so that is complements the training dataset using points where the approximation error is expected to be high. The complementary dataset could therefore include points that are located far from those of the training dataset, while also having good space filling properties; a procedure to generate such a dataset was proposed by Iooss et al. [70]. A common metric to evaluate approximation error based on a training dataset of size n_{test} based on the input points $x_1^{\text{test}}, \ldots, x_m^{\text{test}}$ is the Q_2 metric defined as follows:

$$Q_2 = 1 - \frac{\sum_{i=1}^{n_{\text{test}}} \left(\widehat{\mathcal{M}}(x_i^{\text{test}}) - \mathcal{M}(x_i^{\text{test}}) \right)^2}{\sum_{i=1}^{n_{\text{test}}} \left(\widehat{\mathcal{M}}(x_i^{\text{test}}) - \overline{\mathcal{M}} \right)^2}, \tag{I.29}$$

where $\overline{\mathcal{M}}$ is the empirical mean of model $\mathcal{M}$ computed using the training dataset. In the case of uniformly distributed samples, the Q_2 can be understood as the fraction of variance in the model that is explained by the metamodel. It is positively oriented and equal to 1 when the approximation error is null. For a "naive" constant metamodel that always returns the model mean ($\forall x,\ \widehat{\mathcal{M}}(x) = \overline{\mathcal{M}}$), then $Q_2 = 0$ [3]. For a metamodel with lower MSE on the test dataset than that of the naive metamodel, then $Q_2 \in [0, 1]$, otherwise $Q_2 < 0$.

I.2.2. Gaussian processes

Gaussian process (GP) modeling, which may also be referred to as kriging, can be used for metamodeling, as proposed in [116]. This section describes the main aspects of metamodeling using GP, more details on GP and kriging can be found in [108, 115].

3. the empirical mean should have approximately the same value when using the test dataset, implying $Q_2 \approx 0$.

A stochastic prossess Y_x with values in $\mathbb{R}$ indexed by $x \in \mathbb{R}$ is a Gaussian process if, for any $n < \infty$, $\forall x_1, \ldots, x_n \in \mathbb{R}^d$, the finite random vector $[Y_{x_1}, \ldots, Y_{x_n}]^T$ follows a Gaussian multivariate distribution [4]. A GP Y_x is fully characterized by its mean (aka trend) function $a : x \in \mathbb{R}^d \mapsto a(x) = \mathbb{E}[Y_x] \in \mathbb{R}$ and its covariance kernel $C : (x, x') \in \mathbb{R}^d \times \mathbb{R}^d \mapsto C(x, x') = \mathrm{Cov}[Y_x, Y_{x'}] \in \mathbb{R}$.

Simple kriging (SK) applies to a stochastic process (not necessarily Gaussian) with known mean and covariance functions based on the values of the process at a finite number of points. Denoting the realization vector $\boldsymbol{y} = [y_1, \ldots, y_n]^T$ associated to points $x_1, \ldots, x_n$, SK consists in providing the best linear unbiased estimate of Y_x for any x based on the said realization. This yields the so-called SK mean and variance, respectively defined as follows for a given $x \in \mathbb{R}^d$:

$$m_{SK}(x) = a(x) + \boldsymbol{c}(x)^T \boldsymbol{C}^{-1}(\boldsymbol{y} - \boldsymbol{a}), \text{ and} \tag{I.30}$$

$$s^2_{SK}(x) = C(x, x) - \boldsymbol{c}(x)^T \boldsymbol{C}^{-1} \boldsymbol{c}(x), \tag{I.31}$$

where $\boldsymbol{a} = [a(x_1), \ldots, a(x_n)]^T$, $\boldsymbol{c}(x) = [C(x, x_1), \ldots, C(x, x_n)]^T$, and $\boldsymbol{C} = (C(x_i, x_j))_{1 \leq i,j \leq n}$, provided that this matrix is invertible. The SK variance is positive and does not depend on the observation $\boldsymbol{y}$. Metamodeling using the SK mean, i.e. $\widehat{\mathcal{M}} = m_{SK}$, ensures interpolation because the components of $\boldsymbol{c}(x_j)^T \boldsymbol{C}^{-1}$ are all null except for the j-th one, which yields $\forall j \in \{1, \ldots, n\}, m_{SK}(x_j) = a(x_j) + (y_j - a_j) = y_j$. Similarly, the SK variance is null for all x_j; it can be understood as an indicator of the error that is likely to be made using the SK mean as a predictor, which can be used to provide confidence intervals associated to the prediction if one further assumes that process Y_x is Gaussian. Actually, when the process is assumed Gaussian, conditioning random variable Y_x at fixed x given realization $\boldsymbol{y}$ of random vector $\boldsymbol{Y} = [Y_{x_1}, \ldots, Y_{x_n}]$ yields a random variable that is also Gaussian and whose mean and variance correspond to those of simple kriging, i.e. $Y_x | \boldsymbol{Y} = \boldsymbol{y} \sim \mathcal{N}(m_{SK}(x), s^2_{SK}(x))$.

Gaussian	$\rho(x - x') = \exp\Big(-\frac{(x-x')^2}{2\theta^2}\Big)$
Matérn 5/2	$\rho(x - x') = \Big(1 + \frac{\sqrt{5}\lvert x-x'\rvert}{\theta} + \frac{5(x-x')^2}{3\theta^2}\Big) \exp\Big(-\frac{\sqrt{5}\lvert x-x'\rvert}{\theta}\Big)$
Matérn 3/2	$\rho(x - x') = \Big(1 + \frac{\sqrt{3}\lvert x-x'\rvert}{\theta}\Big) \exp\Big(-\frac{\sqrt{3}\lvert x-x'\rvert}{\theta}\Big)$
Exponential	$\rho(x - x') = \exp\Big(-\frac{\lvert x-x'\rvert}{\theta}\Big)$

Table I.1. – Examples of 1D stationary kernels ρ used as correlation functions in GP modeling, based on scale parameter $\theta > 0$ ($x, x' \in \mathbb{R}$).

A common choice for GP modeling is to choose a stationary covariance kernel, i.e. $\forall (x, x') \in \mathbb{R}^d \times \mathbb{R}^d C(x, x') = \sigma^2 \rho(x - x')$, where $\sigma^2 > 0$ and ρ is referred to as the correlation function of the process. Such a function is invariant to translation and denoting $\boldsymbol{R} = (\rho(x_i - x_j))_{1 \leq i,j \leq n}$ and $\boldsymbol{r}(x) = [r(x - x_1), \ldots, r(x - x_n)]^T$, Equa-

4. here, and in the following, superscript T indicates transposition.

tions (I.30) and (I.31) can be respectively written as follows:

$$m_{SK}(x) = a(x) + \boldsymbol{r}(x)^T \boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{a}), \tag{I.32}$$

$$s^2_{SK}(x) = \sigma^2(1 - \boldsymbol{r}(x)^T \boldsymbol{R}^{-1}\boldsymbol{r}(x)). \tag{I.33}$$

The covariance kernel needs to positive definite to be admissible, which, in the case of a stationary covariance function, requires ρ to be of positive type. Popular choices of the correlation function in 1D (i.e. $d = 1$) rely on a scale parameter $\theta > 0$ such that the lower θ, the lower the correlation between Y_x and $Y_{x'}$ at two different points $x, x' \in \mathbb{R}$. A few examples of such functions include exponential, Matérn, and exponential kernels, defined in Table I.1. A notable property of such kernels is the resulting regularity of process Y_x which is indefinitely differentiable with a Gaussian kernel, differentiable respectively twice and once with Matérn 5/2 and 3/2 kernels, and only continuous with an exponential kernel. Knowledge or assumptions regarding the regularity of $\mathcal{M}$ may motivate the choice of a given kernel. A key aspect of GP modeling is the choice of scale parameter θ as shown in Figure I.3 showing the influence of θ in the simple case of approximating a cosine function.

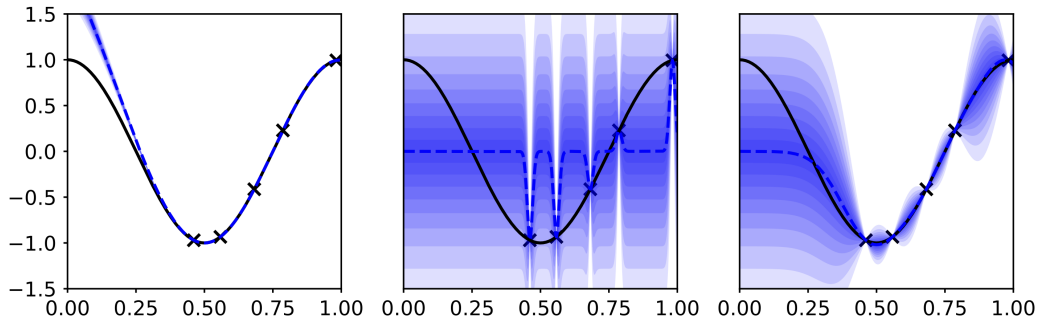


Figure I.3. – Examples of GP metamodeling of function $x \in [0, 1] \mapsto \cos(2\pi x)$ (black solid line) using a Gaussian kernel based on 5 uniformly sampled points in [0,1] (black crosses).
The KS mean (dotted blue) and variance are calculated assuming constant null trend $a(x) = 0$, variance parameter $\sigma^2 = 1$ and variable scale parameter $\theta > 0$. The shaded areas indicate confidence at levels $10\%, 20\%, \ldots, 90\%$.
Left: $\theta = 0.5$, the high value of θ leads to a very low kriging variance, hence the narrow confidence intervals.
Middle: $\theta = 0.01$, the low value of θ leads to a KS mean close to 0 and high KS variance, except in a close neighborhood of the training points.
Right: $\theta = 0.1$, the KS mean and variance are better suited to the approximation of $\mathcal{M}$, although the absence of training points for low x leads to higher error and KS variance.

In the multivariate case $d > 1$, the correlation function may be taken as a product of 1D stationary correlation functions (usually of the same type) with scale parameters $\theta_1, \ldots, \theta_d > 0$ for each input component. Determining an adequate value for these parameters, the variance parameter σ^2 and the trend function a is all the more difficult

in high dimension d, which motivates the use of universal kriging (UK). Assuming that the trend a is a linear combinations of p known functions $f_1, \ldots, f_p$, i.e. $a(x) = \sum_{k=1}^{p} \beta_k f_k(x)$ where vector $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_p]^T \in \mathbb{R}^p$ is unknown, the UK equations for mean and variance are respectively expressed as follows given $x \in \mathbb{R}^d$ and covariance function C:

$$m_{UK}(x) = \boldsymbol{f}(x)^T \widehat{\boldsymbol{\beta}} + \boldsymbol{c}(x)^T \boldsymbol{C}^{-1}(\boldsymbol{y} - \boldsymbol{F}\widehat{\boldsymbol{\beta}}), \text{ and} \tag{I.34}$$

$$s^2_{UK}(x) = s^2_{SK}(x) + (\boldsymbol{f}(x)^T - \boldsymbol{c}(x)^T \boldsymbol{C}^{-1}\boldsymbol{F})^T (\boldsymbol{F}^T \boldsymbol{C}^{-1}\boldsymbol{F})^{-1} (\boldsymbol{f}(x)^T - \boldsymbol{c}(x)^T \boldsymbol{C}^{-1}\boldsymbol{F}), \tag{I.35}$$

where $\boldsymbol{f}(x) = [f_1(x), \ldots, f_p(x)]^T$, $\boldsymbol{F} = (f_k(x_i))_{\substack{1 \le i \le n \\ 1 \le k \le p}}$ and $\widehat{\beta} = (\boldsymbol{F}^T \boldsymbol{C}^{-1}\boldsymbol{F})^{-1} \boldsymbol{F}^T \boldsymbol{C}^{-1} \boldsymbol{y}$ are respectively the trend vector of the p functions at point x, the $n \times p$ matrix of the trend vector for n training points, and the best linear estimator of vector $\boldsymbol{\beta}$. Using the UK mean also ensures interpolation and the UK variance is null at points $x_1, \ldots, x_n$. Using stationary kernels for each input component, one may estimate parameters $\boldsymbol{\theta} \in \mathbb{R}^d$, $\boldsymbol{\beta} \in \mathbb{R}^p$, and $\sigma > 0$ using maximum likelihood estimation [101]. Although this estimation method yields the estimator $\widehat{\boldsymbol{\beta}}$, it depends on $\boldsymbol{\theta}$ and the system of equations for this parameter and σ^2 is not easy to solve. A numerical solver is proposed in the R-package DiceKriging [115] to address this complex optimization problem.

GP metamodeling may provide good interpolation even with a low number of points, and it is common practice to use a space filling DOE to constitute the training dataset (cf. Section I.2.1). However the higher dimension d, the larger the training sample may be required, which may not be reasonable using numerical models with high computational time. Increasing the size of covariance matrix $\boldsymbol{C}$ may also result in numerical difficulties due to RAM limitations or high computational time of the kriging mean, defeating the purpose of metamodeling. These issues can be addressed using dimension reduction techniques such as principal components analysis (PCA) in conjunction with GP metamodeling as in [90], or using other methods that are better suited to high-dimensional inputs.

I.2.3. Neural networks

Artificial neural networks, commonly called neural networks (NN) for simplicity, refer to a type of models encountered in computer science whose structure was originally inspired from the functioning of biological neurons. This structure can be considered as a set of "layers" composed of several "neurons" where each neuron (usually called a *node*) is influenced by the neurons in other layers.

A notable common component of NNs is the *fully connected* (aka dense) layer. Given a layer characterized by nodes with values $x^1, \ldots, x^d \in \mathbb{R}$ and $i \in \{1, \ldots, p\}$, the i-th node of the fully connected layer calculates its value z_i as follows:

$$z_i = f\left(\sum_{j=1}^{d} w_{i,j} x^j + \beta\right), \tag{I.36}$$

where the $w_{i,j} \in \mathbb{R}$ are called weights of the layer and both function $f : \mathbb{R} \to \mathbb{R}$ and weight $\beta \in \mathbb{R}$ do not depend on i and are respectively referred to as the *activation*

function and the *bias*. A basic type of NN is the multilayer perceptron (MLP), which consists in a succession of fully connected layers, each layer having its own weights and activation function. For such a system, the very first (aka input) layer will be used to determine the values of the nodes in the first intermediate (aka hidden) layer, and so on until the very last (aka output) layer. In the particular case of a single node $y \in \mathbb{R}$ in the output layer and identity activation functions, a MLP amounts to a linear regression model, whose direct coefficients between input and output could be fitted based on a training dataset using maximum likelihood estimation, for instance, which amounts to a simpler problem than fitting the weights of all intermediate layers. Actually, the activation functions of a MLP are non-linear, potentially providing a better fit to the data, although fitting the layer weights is a more complicated problem than in the linear case. Traditional activation functions are the sigmoid $f : u \in \mathbb{R} \mapsto \frac{1}{1+\exp(-u)} \in]0, 1[$, hyperbolic tangent $f : u \in \mathbb{R} \mapsto \tanh(u) = \frac{\exp(2u)-1}{\exp(2u)+1} \in]-1, 1[$, and more recently the Rectified Linear Unit (ReLU), $f : u \in \mathbb{R} \mapsto \max(u, 0) \in \mathbb{R}_+$.

The weights of a NN are fitted to data through the minimization of a loss function. Let us denote the assumed relationship between input $x \in \mathbb{R}^d$ and output $y \in \mathbb{R}$ via model $\mathcal{M}$, which may be more general than a mathematical or numerical deterministic model; for instance $\mathcal{M}$ it may indicate how many persons are in a picture. Denoting the input-output relationship of the NN via model $\widehat{\mathcal{M}}_w$ where $w \in \mathbb{R}^p$ stands for the model weights, a natural criteria for fitting w is to minimize a mean error, for instance using square differences:

$$\mathbb{E}\big[(\widehat{\mathcal{M}}_w(X) - \mathcal{M}(X))^2\big], \tag{I.37}$$

where random variable X stands for the input distribution. Usually, $\mathcal{M}$ is too complex or even unknown, so an empirical approximation of the mean error is used instead. Given a training dataset of size n, $(x_1, y_1 = \mathcal{M}(x_1)), \ldots, (x_n, y_n = \mathcal{M}(x_n))$, and denoting $\widehat{y}_i = \widehat{\mathcal{M}}_w(x_i)$ for all $\forall i \in \{1, \ldots, n\}$, the MSE based on n these members is used instead:

$$\frac{1}{n}\sum_{i=1}^{n}(\widehat{y}_i - y_i)^2 = \frac{1}{n}\sum_{i=1}^{n}(\widehat{\mathcal{M}}_w(x_i) - y_i)^2. \tag{I.38}$$

A characteristic of a NN is that the function determining a node value based on the previous layer nodes is relatively simple. For a fully-connected layer (cf. Equation (I.36)), it is a simple composition of an affine function (relatively to the weights) by the activation function f. The derivative of the latter is generally simple; for instance $f'(u) = f(u)(1 - f(u))$ for the sigmoid, $f'(u) = 1 - f(u)^2$ for hyperbolic tangent and it is simply the sign of $u \in \mathbb{R}^*$ for the ReLU. In other words, partial derivatives of a node relatively to its layer's weights can be easily computed using the formula for the derivative of a composition of two functions given the value of the nodes in the preceding layer. By extension, for a given $x \in \mathbb{R}^d$, $\widehat{\mathcal{M}}_w(x)$ can be expressed as a succession of compositions, so its gradient relatively to the weights w and more importantly that of the mean error (I.38) can be computed fairly easily. As the computation of the gradient first requires to compute $\widehat{\mathcal{M}}_w(x)$, the computation of the gradient is referred to as *backpropagation* because it is carried out from output layer to input layer. Traditional "deterministic" gradient descent techniques, however, are not suited to minimizing (I.38) due to the complexity of $\widehat{\mathcal{M}}_w$ and high number of parameters p, as they may result in being stuck in a local minimum. Optimization is instead carried out

using a *stochastic gradient descent* algorithm, an iteration of which can be summarized by the following steps:

1. Randomly partition the training dataset into B subsets (usually of equal size) called *batches*.
2. For $b = 1, \ldots, B$:
   a) Compute the sample loss (I.38) but using based on batch b instead of the whole dataset.
   b) Compute the gradient of the previous quantity relatively to w.
   c) Update the value of weights w based on the value of the gradient.

The idea behind step 2 is that a batch should provide a fair approximation of (I.37). Also, the use of a stochastic algorithm usually works well even when the NN (and/or the loss function) presents singularities such as the local non-derivability of ReLU in 0. Actually, a reason behind the popularity of the ReLU is that it suffers less than other activation functions from vanishing or exploding gradient problems, which are characterized by very low (resp. high) absolute values of the partial derivatives in some instances (usually in the first layers of the network) resulting in poor optimization.

Several iterations of the algorithm may be required until the loss is low enough. Because NNs are known to induce overfitting, it is common practice to compute the loss function based on a test dataset at the end of each iteration, which is not used to update the weights. Usually, the loss will tend to decrease for both datasets until a given number of iterations starting from which it may oscillate around a given value or even increase for the test dataset whereas it continues to decrease for the training dataset; which is when the algorithm is generally stopped. Ideally, a NN with low error in the training dataset should be *generalizing*, meaning that for any other dataset the error should also be low. For this purpose, several *regularization* techniques that can limit overfitting may be used, such as loss penalization (e.g. adding a term proportional to the L^2-norm of weight w to the loss function), adding layers carrying out dropout or batch normalization (e.g. during training, dropout consists in randomly setting a fixed proportion of the previous layer's nodes at null value, with the intention for the NN output to depend less on the value of specific weights). Data augmentation techniques can also be used to artificially increase the size of the training dataset; for instance, when the output is invariant to a transformation ϕ of the input then the $(\phi(x_i), y_i)$ can be added to the training dataset.

Possible "architectures" of NNs extend much further than MLP and regularization layers, but the idea of fitting the weights through the optimization of a loss function for a training dataset using a stochastic gradient descent algorithm remains the same. For instance, recurrent neural networks are commonly used to process text data, whereas convolutional neural networks (CNN) are commonly used for image classification. In the latter application, a CNN usually processes an input image by means of layers performing convolutions to identify patterns in the image or the layers together with "pooling" operations, which reduces the data size by computing average or maximum values of nodes, and its output can consist of several nodes summing to 1, where each node provides a probability of the image being labeled in a given category. NNs can be used to address a wide variety of problems and proved to be successful in many applications (e.g. [78, 142, 124]).

The number of weights in a NN is generally very high, but the operations performed by the individual nodes are relatively simple, therefore its output can be computed relatively fast. Also, although NN implementation and training requires to properly configure each node, how to compute the gradients, etc., this can be done quite easily using existing software, such the notable Python libraries TensorFlow and PyTorch. In TensorFlow, it is faster to carry out the computations for several members of the dataset at the same type (e.g. over a batch) thanks to the use of tensor representation of the data. Regarding hardware, it is common practice to use one or several GPUs to speed-up the computations in the case of CNNs, as they can carry out convolutions faster than CPUs. Another noteworthy aspect regarding hardware is the RAM, which can impose an upper limit of the batch size, as the computation of layers with a high number of nodes for many dataset members at the same time may cause a memory error. In TensorFlow, more flexibility regarding batch management can be obtained using Python's generators (i.e. functions that involve statement *yield*), in conjunction with object "tf.data.Dataset.from_generator", or more generally using so-called "input pipelines" [5].

NNs are suitable candidates for metamodeling given the accessible programming libraries as well as their capacity to address complex modeling problems involving various types of input, high-dimensional inputs included. One should expect approximation to be better the larger the training dataset, but constituting a very large dataset may be unreasonable if the original model has high computational time. Even though these models generally involve a high number of parameters and operations, they should also have lower computational time than the original model, especially if the output is computed for several input values regrouped in a batch instead of one after another.

I.3. Wildland fires

Wildland fires are distinguished among fires by their occurrence in vegetation or natural fuels, even though such fires may reach artificial structures and urban areas. "Wildland fire" is a term that includes both prescribed fires, which are planned, authorized, and aim to meet certain objectives, and wildfires, which are not prescribed. In the present section, we first provide a brief description of the physics of the phenomenon, the influential factors, what can be observed during an event, and how it is accounted in mathematical and numerical models. The second part deals with the notions of danger and risk and presents notable methods to associate quantities to these concepts and generate maps in an operational context.

I.3.1. Description and modeling

During fire spread, the combustion of fuel can be divided in several phases corresponding to different degradation phenomenons. The fuel first heats until it reaches the temperature of evaporation and consumes energy until being dried out. For heating and evaporation, the amount of heat required increases with the quantity of water, typically quantified by the fuel moisture content (FMC), which is defined as the ratio between mass of water and mass of dried fuel; other quantities involved in the drying

5. cf. https://www.tensorflow.org/guide/data, last checked on 2021.01.14

process are C_p and C_w, respectively the heat capacity of fuel and water, as well as the heat of evaporation Δh. Distillation may also occur, leading to the emission of volatile organic components. Then, pyrolysis, which consists in the chemical decomposition of fuel, mainly leads to the production of combustible, volatile gases and reactive char. High enough temperature and gas flow resulting from the decomposition of fuel lead to flaming combustion, arising from the oxidation of emitted gases at a high enough temperature so that hot gases and soot residues generate radiation in a wide spectrum (thermal infrared to visible); this exothermic chemical reaction can be quantified by its heat of combustion ΔH. Meanwhile, smoldering or glowing combustion which consists in the oxidation of reactive char also occurs and this reaction continues after flaming combustion stops, until extinction. Propagation is the result of the amount of energy transferred by radiation (flame) and convection (hot gases) that contributes to the evaporation and pyrolysis of fuel in front of this flame.

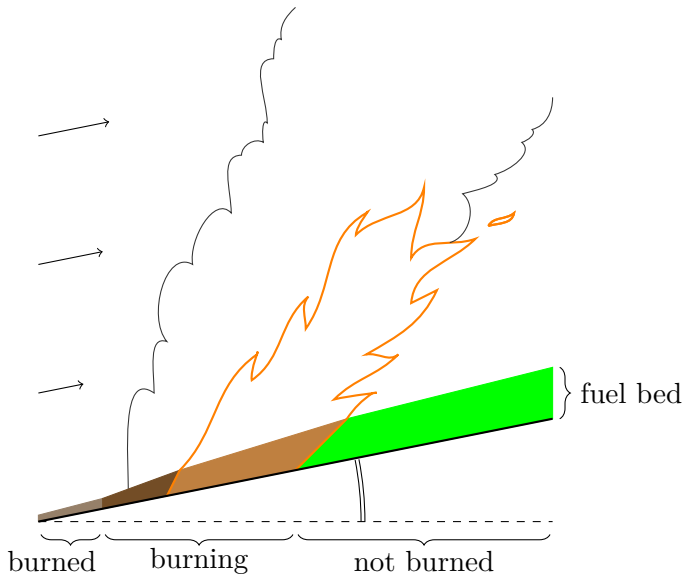


Figure I.4. – Simplified representation of wildland fire spread from profile perspective in the 2D plan spanned by the direction of spread and the vertical. Here, there is a positive slope in the direction of spread. Also, wind speed (black arrows) "drives" both the flames and the smoke plume forward.

The aforementioned steps can be understood as occurring from right to left in the simplified representation of fire spread provided in Figure I.4. Two major heat transfer mechanisms, radiation and convection, are involved in heating the fuel that is not burning until combustion, but may also result from contact with flames or fire brands. Provided that heat is produced and transferred at a sufficient rate, compensating the heat required for preheating to pyrolysis, the aforementioned thermal degradation processes will be sustained, and the fire can continue to spread.

Several factors are known to influence fire spread. Positive terrain slope in the direction of spread leads to higher radiation transfer and higher rate of spread (ROS, i.e. the speed at which the flaming front advances). Incoming wind is also a highly influential factor as it leads to increasingly high ROS with increasingly high wind

speed norm if oriented towards the fuel that is not burned, whereas a decrease in ROS is observed if wind direction opposes that of spread. Given the occupation of space by vegetation, fuel can be considered a porous medium; its characteristics are usually quantified via fuel load σ_f (i.e. the mass of the fuel bed over a surface) and its depth or height h_f. At the smaller scale, particles of fuel are characterized by their particle density ρ_p and surface-volume ratio S_v. The density of the fuel bed may be characterized by the packing ratio β_f which is the ratio between particle density and density of the fuel bed at a larger scale, which can be approximated as $\beta_f = \frac{\sigma_f}{h_f \rho_p}$.

These factors may interact with each other, and more generally, wildland fires can be understood as the interaction of terrain, weather, and fuel from local to global scale. For instance, the lower the packing ratio β_f, the more heat may be transferred by radiation to the bottom of the fuel bed; fuel moisture content will depend on the size of fuel particles and weather through relative humidity, rainfall and air temperature, which vary during the day; the direction of the slope (i.e. the aspect of the terrain) influences the amount of solar heat absorbed by fuel during the day, which can be reduced for lower fuel layers due to the shade provided by the top of the trees; valleys will create boundary conditions, which has an influence on wind currents; regions with different climates may present different types of vegetation, etc.

At the origin of the fire is an (initial) ignition. In wildland fire spread models, fuel ignition is often characterized by an ignition temperature T_i, that fuel must reach for ignition to occur. This parameter provides a simplified representation of ignition; for instance, ignition and T_i itself may depend on the rate at which the fuel is heated, its thickness, environmental conditions, etc. [27, 48]. A fundamental assumption proposed by Fons [51] for modeling wildland fire spread is to consider that fire spread is a succession of ignitions from burning fuel to neighboring fuel that is not burning by means of heat transfer, hence the use of T_i among other parameters in models. In some ROS models, an example of parameter characterizing fuel ignition properties is the so-called moisture of extinction, denoted as m_χ, which is a threshold value for FMC over which ignition is assumed to be impossible.

Categories of the causes for a wildfire include lighting strikes, accidental, intended (malevolence), unintended, or simply unknown, and complementary categories and subcategories may be considered [6].

During a wildfire event, what is observed during spread may be much more complex than the simplified representation in Figure I.4 (see for instance [84]). The intensity of combustion and the flame length are generally positively correlated with the ROS, but behind the flaming front, patches of fuel may also undergo flaming combustion. Over the event, additional ignitions may occur, which can be due to spotting fires, i.e. ignitions due to the transport of fire brands. In contrast with the simplified representation of vegetation as a homogeneous fuel bed, heterogeneities include the variety of species, particles, discontinuities in fuel parameters and/or presence, as well as several vertical layers across which fire may spread.

Regarding the latter, three main types of spread are commonly identified: ground fires, surface fires (including debris on the surface, which includes dead branches, leaves, and low vegetation) and crown fires (occurring at the top of tree and bushes). Never-

6. see for instance the French wildfire database Prométhée https://www.promethee.com/, last checked 2021.01.21)

theless it is commonly considered [114] that small fuel particles contribute most to fire spread, so that surface fire is generally the main driver of a flaming front; but crown fires can also contribute to an increase in ROS in which case they are qualified as active (e.g. fire spreading from a tree top to adjacent ones), whereas they are qualified as passive when they have little contribution to the advance of the fire front.

Firefighting actions also have obvious impacts on fire spread, and the absence of fuel or any other environmental condition across the landscape that is not favorable to fire spread (e.g. fuel with high FMC) can also halt fire. Still, fire spotting or simply high flames may allow a fire to cross obstacles such as roads or reach the other side of a valley. Other "extreme" phenomenon may also occur, such as fire whirls, eruptive fires, and formation of pyrocumulonimbus clouds with extreme atmospheric convection and fire induced winds. Designing models, prediction systems, and decision support systems is challenging given that wildfires are such complex phenomena characterized by variable environmental conditions over time and across the land, while there is limited knowledge about when and where a fire will ignite.

At the local scale, ROS as well as other quantities such as the geometry of the flame or reaction intensity may be expressed as a function of environmental parameters. Most ROS models can be understood as a function r whose inputs typically include terrain as local slope α, meteorological conditions (wind speed W, fuel moisture content m_c) and quantities characterizing the vegetation denoted here as v (which may include $\Delta H, \sigma_f, S_v, \ldots$), so that one may write the following:

$$\mathrm{ROS} = r(W, \alpha, m_c, v). \tag{I.39}$$

A notable and widely use model of ROS for surface fire spread is that of Rothermel [114], classified as an empirical model of wildland fire spread [130] whose coefficients have been fitted on the basis of a high number of controlled laboratory and field experiments. Inputs of such ROS models may be measured easily in the context of experiments, but are more difficult to assess right during a wildfire. Regarding vegetation characteristics, determining some of these inputs is commonly addressed using fuel models which provide values v for several types or classes of vegetation; a notable example being the fuel models of Scott and Burgan [60] for use in conjunction with the model of Rothermel.

ROS models, which provide a local speed, can be included in models of wildland fire spread representing the advance of the flames across a landscape over time. As such, several approaches can be considered, including markers methods, level-set methods, and cellular automata. The first two approaches consist in considering that fuel is either burned or not, so that fire spread amounts to the evolution of the fire front, i.e. the interface between burned surface and the rest of the spatial domain; whereas cellular automata (e.g. [54]) consist in a raster-based representation of the spatial domain with cells that may have three possible states during fire spread: burned, burning, or not burned, so that the fire front can be represented by the interface between the burning and not burned areas.

The equations at the origin of markers methods describe the evolution of the fire front and may be derived from Huygens' wavelet principle where each point of the fire front is a source of fire spread emitting an infinitesimal burned surface, usually assuming that the said surface is elliptical with dimensions expressed based on the ROS. In any case, associated numerical models imply a discretization of the fire front

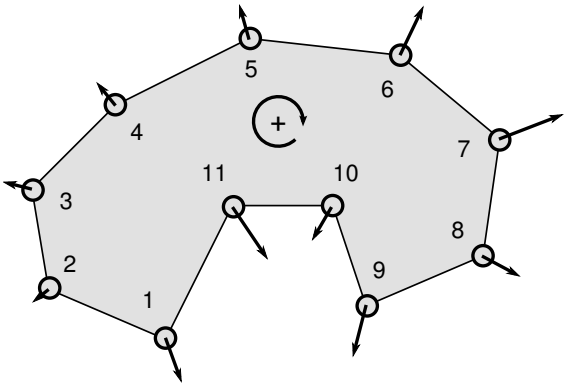


Figure I.5. – Example of a small fire front discretization with ordered markers. The shaded gray area corresponds to the burned surface and the markers may advance iteratively toward the rest of the domain.

by updating the position and time of the vertices (i.e. the markers) from their initial state accordingly to the ROS function and the geometry of the fire front, as illustrated in Figure I.5. Notable solvers of fire spread using markers method include FARSITE [47], Prometheus [137] and Phoenix [134], commonly used in the US, Canada, and Australia, respectively; as well as ForeFire [45].

Level-set methods rely on a function of location and time $\Phi : \mathbb{R}^s \times [0, T] \to \mathbb{R}$ used for front-tracking by defining the interface as its level zero, so that at a given time $t \in [0, T]$ the front is defined as the ensemble $\{x \in \mathbb{R}^s | \Phi(x, t) = 0\}$. Function Φ is evolved according to a velocity field $V : \mathbb{R}^s \times [0, T] \to \mathbb{R}^s$ and defined as the solution of the following partial differential equation:

$$\frac{\partial \Phi}{\partial t} + V \cdot \nabla \Phi = 0. \tag{I.40}$$

The dot product can be simplified using the normal component of V, i.e. its projection along gradient Φ. In wildfire modeling, by denoting the normal component of the ROS as R_N, Equation (I.40) can be rewritten as follows:

$$\frac{\partial \Phi}{\partial t} + R_N ||\nabla \Phi||_2 = 0, \tag{I.41}$$

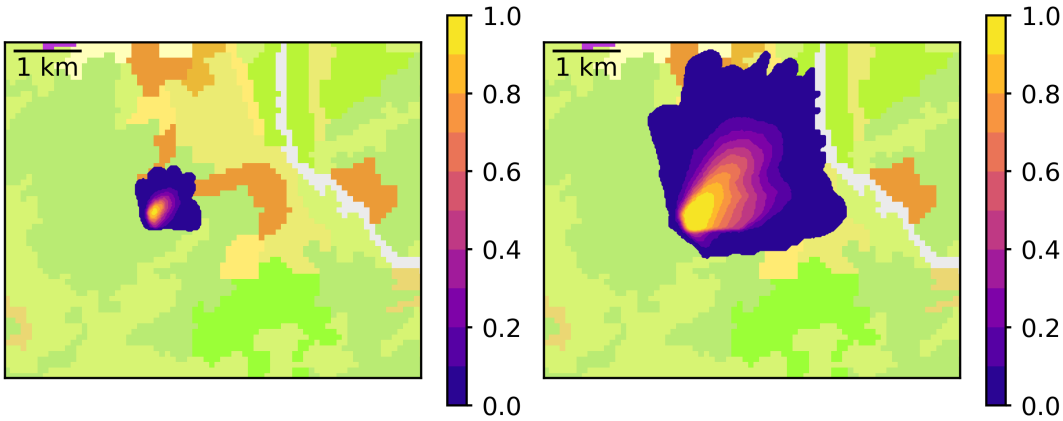
where $||.||_2$ denotes the Euclidean norm in $\mathbb{R}^s$ and $s = 2$ here. Solving Equation (I.41) yields the evolution of the fire front by identifying the zero level of Φ; a notable solver of wildland fire spread using the level-set method is SFIRE [91].

In this thesis, we use ForeFire as a solver of wildland fire spread because of the development carried out regarding system evaluation with this solver [43], which has been applied to a database of 80 Corsican fires [44].

Another category of wildland fire spread models propose a more comprehensive description of the underlying physics. As such, they may be referred to as physical models and consist in a closed system of differential equations with initial and boundary conditions (e.g. WFDS [92]), including equation of continuity, equations of conservation (moment, energy, and chemical species), and may account for turbulence, combustion

chemistry, etc. In such models, ROS is not directly computed, but is instead diagnosed from the evolution of a temperature or fuel consumption field computed at a high resolution. While being more detailed, the high computational requirements of such methods is nevertheless an obstacle to simulating wildfire spread in an operational context, especially for large fires spreading over several thousands of hectares.

Wildland fire predictions may be carried out at different time scales; for simplicity, we identify three categories in the present work: "long-term", "short-term" and "crisis" situations [7]. In the latter, the typical scenario is that a fire has just started, its time and location of ignition are known fairly accurately and the goal is to predict what areas this specific fire is likely to reach as time passes if it spreads "freely", i.e. if no firefighting actions occur.



(a) At T0+10 minutes. (b) At T0+50 minutes.

Figure I.6. – Example of burn probability maps obtained from an ensemble of 500 fire spread simulations for a "crisis" situation where the estimated time of fire start is T0.
The colorbar indicates the probability of the location to be burned at the specified time.
Colored cells in the background corresponds to land-use according to the classification of the Corine Land Cover

Numerical simulators using an underlying ROS model can be used to make such predictions in a reasonably short amount of time, although there is significant uncertainty notably due to the estimation of environmental inputs via fuel models and weather forecasts. Accounting for input uncertainty can be addressed using ensemble methods, which can be seen as an example of propagation of uncertainty based on a Monte Carlo approach because an i.i.d. ensemble of simulations is generated (cf. Section I.1.1) instead of generating a single deterministic succession of burned surfaces; the ensemble may then be summarized via estimated "burn probabilities" [8] as a function of time, an example of which is represented in Figure I.6. A notable implementation of this ap-

7. the first two situations are discussed further below in Section I.3.2.
8. here, referring to the probability that a location may belong to the simulated burned surface at a given time for a specific fire that has already started

proach is the FSPro system [49], which relies on an ensemble of FARSITE simulations, and other similar methods have been proposed [94, 96].

I.3.2. Wildland fire risk and danger

As underlined by Miller et al. [93], *risk* and related terms, namely hazard, exposure, threat, vulnerability, and fire danger, have been frequently used by the wildfire community, although with variable or unspecified definitions. The following definitions of risk and wildland fire risk proposed by Bachmann and Allgöwer [12], which is consistent with common risk assessment frameworks:

— risk: "The probability of an undesired event and its outcome. (...)" [9] and
— wildland fire risk: "The probability of a wildland fire occurring at a specified location and under specific circumstances, together with its expected outcome as defined by its impacts on the objects it affects".

In the latter definition, the "occurrence of fire" may refer not only to the location of initial ignition but also to those reached by subsequent fire spread; whereas the "outcome" can be simply understood as fuel being burned, but any other consequence on socioeconomical or ecological values which may be attributed to the fire can be considered, including lives, structures, property, economical activity, tree mortality, smoke emissions, soil heating and erosion, etc. Despite the usually negative connotation of risk, positive consequences may be considered such as removal of accumulated fuels, creation of plant establishment sites, enhancement of habitat diversity, etc. [95].

From a mathematical perspective, common risk assessment frameworks can be understood as modeling the possible consequences of the undesired event via a random variable Y. Let us consider a simple binary setting where either no fire occurs or a fire occurs with negative consequences summarized by a cost value $y \in \mathbb{R}_+$ and probability $p \in [0, 1]$, such that $\mathbb{P}[Y = y] = p$ and $\mathbb{P}[Y = 0] = 1 - p$. In this case, risk is typically quantified using (and sometimes defined as) the expected value of Y: $\mathbb{E}[Y] = p \times y$, which is consistent with the idea that there is no risk if the event has no chance of happening (i.e. if $p = 0$), or if there is no negative consequence on the values at stake (i.e. if $y = 0$) [10].

Extending the binary setting to more than two outcomes can be easily carried out by modeling Y as a discrete or a continuous random variable. A single variable may not be satisfying to summarize different types of values (e.g. economical cost may not easily include ecological consequences) in which case Y may be defined as a random vector whose components each account for a given category of consequences. In any of the cases previously mentioned, the expected value can still be used as a quantity representing risk. Using a single statistic such as $\mathbb{E}[Y]$ has the advantage of providing a summarized quantification of risk, but even the simple binary setting involves the definition of probability p and cost y which raises the question of how they should be determined. More generally, considering a "constructive" approach where $\mathbb{E}[Y]$ is calculated or estimated as a result of the knowledge of the probability distribution of

9. the rest of the definition is: "An undesired event is a realization of a hazard." which relates to their proposed definition for hazard: "A process with undesirable outcomes.", which is a term that is not used in the present work.

10. one may also consider that there is no risk in both "deterministic" cases $p = 1$ and $p = 0$, which can be understood as absolute certainty as to whether the event will occur or not.

Y, the main problem of risk quantification is to determine a "relevant" probability distribution. One may also question the use of the expected value to quantify risk and choose to consider other moments of the probability distribution of Yn quantiles, or more elaborate quantities. For instance, in the binary setting, a distribution with high probability p and low cost y may be considered preferable to a distribution with low p and high y even if they have equal expected value $\mathbb{E}[Y]$; this echoes to the well-known notion of "risk aversion" in economics.

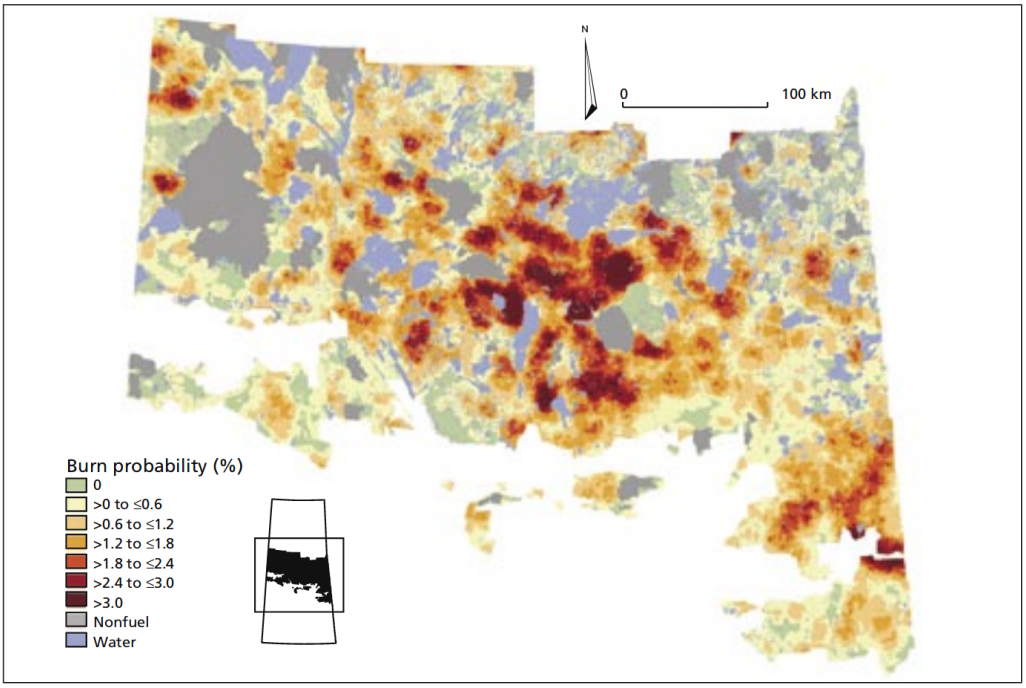


Figure I.7. – Example of burn probability map, reproduction of Figure 12 in [99]. The map was obtained with BURN-P3 based on 500 iterations for the study area of central Saskatchewan, Canada.

Wildland fire risk assessment can be decomposed into three major components: ignition, spread, and exposure. Here, "exposure" refers to the values at stake that can be affected as a consequence of the fire, which were discussed previously. "Ignition" refers to the initial ignition of a fire, which has been the object of a large number of studies in the recent years (see for instance [104, 53]) and is typically quantified by means of an ignition probability model which, for simplicity, can be considered uniform over areas of the landscape where fuel is present, but may be more complex and aim to involve influential factors and provide a better representation of observed ignition patterns.

In the latter case, the model is constructed on the basis of data on past observed fires and may depend on the cause of ignition as well as environmental and/or socioeconomical data using approaches including logistic regression [31, 59] and point processes [32, 13]. While modeling ignition and identifying exposure can be carried out independently, spread can be seen as the link between these two components of risk in the sense that it results from ignition and that the consequences on exposure depend on where and how the fire spreads. These three components of wildland fire

risk are intertwined in a complex fashion: for instance, environmental conditions may be favorable to both ignition and spread (e.g. low FMC), structures reached by the fire may not be completely damaged but only partially, therefore depending not only on the ROS but also on fire intensity, which are correlated.

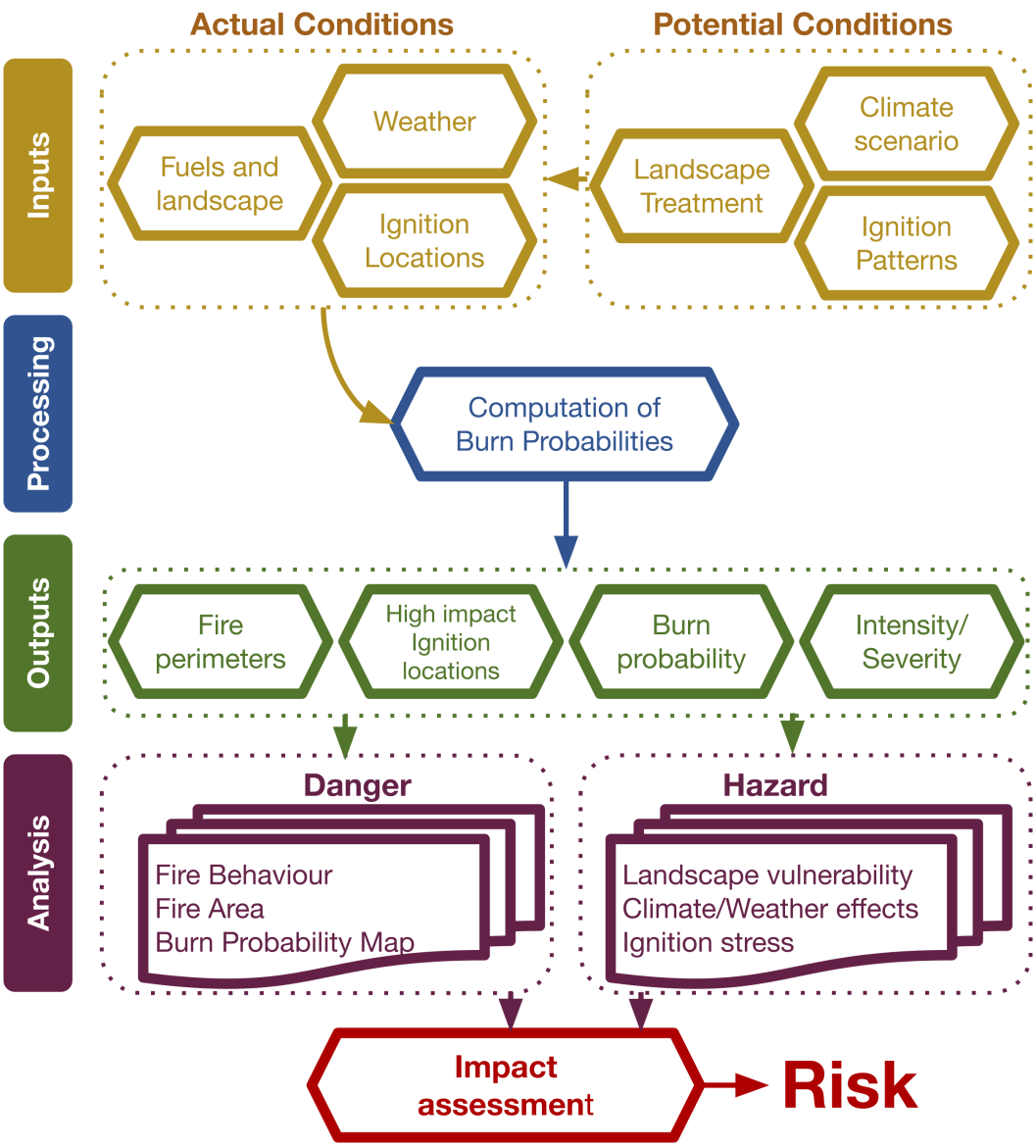


Figure I.8. – Process of using burn probability models, largely inspired from [100].

A popular approach that accounts for both ignition and spread and may be used for wildland fire risk assessment is burn probability (BP) modeling, where the burn probability of a specific area represents its potential to be burned by a wildland fire during a year or a season [11].

11. although the same term was used in the previous section, they refer to different events: here

BP modeling consists in propagating input distributions of environmental characteristics (weather and fuels) and fire characteristics (ignition, fire duration, number of fires in a year) via several simulations of wildland fire ignition and spread, this process being repeated independently several times via a MC method so as to represent potential scenarios of wildland fire activity in a region and/or a country.

Notable implementations include Burn-P3 [99] which relies on the Prometheus solver and FSim [46] which relies on FARSITE. A large number of studies have proposed and/or applied BP modeling (see for instance [100] and references therein). A major output of BP modeling is the burn probability map, an example of which being provided in Figure I.7 although simulated ignition and fire spread patterns as well as maps of fire behavior can also be analyzed. Applications include risk quantification, use of BP modeling outputs in conjunction with other models, and direct analysis of the output distribution, which may be carried out with varying inputs (e.g. fuel treatments, different ignition patterns) so as to study the effect of several fuel management decisions; a summary of the overall process is represented in Figure I.8.

It follows that a local risk can be defined as the expected value of the cost Y_i associated to a given location $x_i \in \mathbb{R}^2$ (including positive and negative consequences) in a binary fashion given burn probability p_i at the said location via $\mathbb{E}[Y_i] = p_i \times y_i$, where y_i represents the exposure specific to location x_i and p_i (and so $\mathbb{E}[Y_i]$) can be estimated by BP modeling. A limit of this approach is limited by the difficulty to quantify via y_i non-local and non-static aspects even though they are identified such as presence of people (evacuations), reduced air quality, impacts on the economical activity due a park closing, or even regeneration of ecosystems and mitigation of negative consequences; but it still provides a spatialized quantification of risk, which can be used to determine priority among areas that require fuel management, for instance.

From an operational perspective, BP modeling can be seen as a decision support tool for "long-term" planning. For "crisis" situations, FSPro is a notable component of the Wildland Fire Decision Support System (WFDSS) [25] to assess burn probabilities of a specific fire. WFDSS also includes assessment of the values at risk and may be combined with FSPro.

One may identify an intermediate "short-term" situation where one or several wildfires are likely to occur in the current day and/or the following one(s), warranting scheduling of patrols, aerial monitoring, public warnings, and planning how many and where firefighting resources should be allocated in order to provide an optimal response in case a wildfire starts. For these purposes, fire danger indices and ratings were developed and are currently widely used in an operational context. Here, a distinction is made between a fire danger index, which we consider a scalar quantity, and a rating, which we consider a category (generally determined among five or so possible categories, generally including "low", "moderate", "high", and "extreme").

More generally, fire danger rating systems include assessment of one or several fire danger indices, which are used among other information to provide a rating, so the notion of "rating" may also refer to the scalar indices composing the system. Regardless of the type of quantity used to assess it, fire danger relates to the proneness for ignition,

we are in a "long-term" situation where there is considerable uncertainty regarding where a fire may ignite, while the previously mentioned burn probability referred to the "crisis" situation for a specific fire that has already started.

spread and/or intensity of a wildfire according to the state of the vegetation and its environment at a given time, therefore reflecting how difficult it may be to control fire. Assessment of fire danger can help in determining and prioritizing which areas need protection, early detection, early initial attack, etc. [12] Development of fire danger rating methods has led to the implementation of systems at the national scale about 50 years ago in Canada and in the US, with the Canadian Forest Fire Danger Rating System (CFFDRS [83], whose development started in 1968) and the National Fire Danger Rating System (NFDRS [17], first published in 1972 but whose development as well as fire danger rating methods dates further back, see for instance [65]).

Maps of fire danger ratings are usually generated every day to assess the situation of the current day and forecast the situation for the day(s) to come. Such maps may be available among other data via internet-based information systems; for instance, covering the US as part of the Wildland Fire Assessment System (WFAS [20]) [13], covering Canada as part of the Canadian Wildland Fire Information System (CFWIS) [14], covering Europe and the Mediterranean area as part of the European Forest Fire Information System (EFFIS) [15], or even covering the globe as part of the Global Wildfire Information System (GWIS) [16], which builds on activities of EFFIS. A widely used fire danger index is the Canadian Fire Weather Index (FWI [138]), which is the main output of a system computing several indices pertaining to fuel moisture and potential fire behavior. The FWI is used in itself and to derive fire danger ratings, not only in CFFDRS and CFWIS, but also in EFFIS and GWIS. Another notable fire danger rating system is based on the Australian Fire Danger Index (FDI), which can be distinguished according to the type of fuel as Forest Fire Danger Index (FFDI) and Grassland Fire Danger Index (GFDI); forecasted maps of these indices are routinely issued by the Australian Bureau of Meteorology [33] and are available online [17].

Calculation of the aforementioned indices mostly depends on weather inputs and, to some extent, fuel moisture, which is generally derived from weather information. For instance, the FFDI is computed as follows:

$$\mathrm{FFDI} = 2\exp\{-0.45 + 0.987\log(DF) - 0.0345RH + 0.0338T_a + 0.0234W\}, \quad \text{(I.42)}$$

where wind speed W, relative humidity RH, and air temperature T_a are respectively expressed in km/h, %, and °C in Equation (I.42), whereas the drought factor DF (unitless) ranges from 0 to 10 and is partly based on soil moisture deficit accounts for the influence of recent temperatures, and rainfall. The indices of the FWI system depend on similar inputs, namely wind speed, air temperature, relative humidity and rainfall over the previous 24 hours, as well as the values of the indices from the day before. Although one could consider computing the indices at high temporal resolution, they are generally computed to represent peak, or at least high, fire weather conditions

12. As previously mentioned, the notion of risk and danger are related and their definition may differ from a study to another. Most methods used to assess fire danger account for the ignition and/or spread components of risk, whereas identification of exposure is carried out separately.
13. https://www.wfas.net/, last checked 2021.02.01
14. https://cwfis.cfs.nrcan.gc.ca/maps/fw, last checked 2021.02.01
15. https://effis.jrc.ec.europa.eu/, last checked 2021.02.01
16. https://gwis.jrc.ec.europa.eu/, last checked 2021.02.01
17. e.g. for the state of Victoria:http://www.bom.gov.au/vic/forecasts/fire-map.shtml, last checked 2021.02.01.

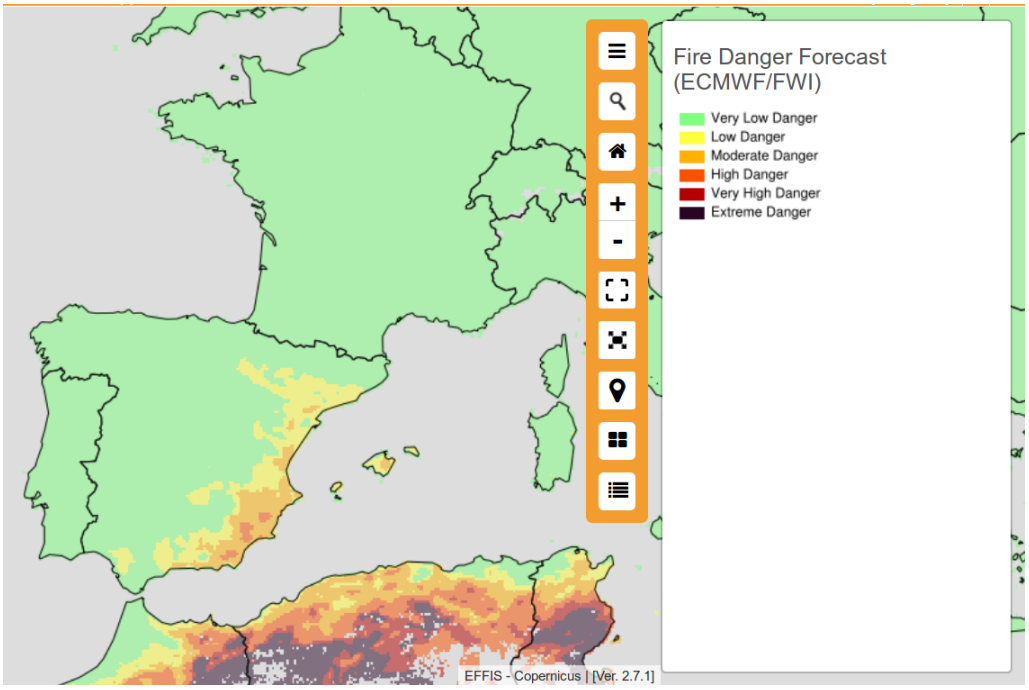
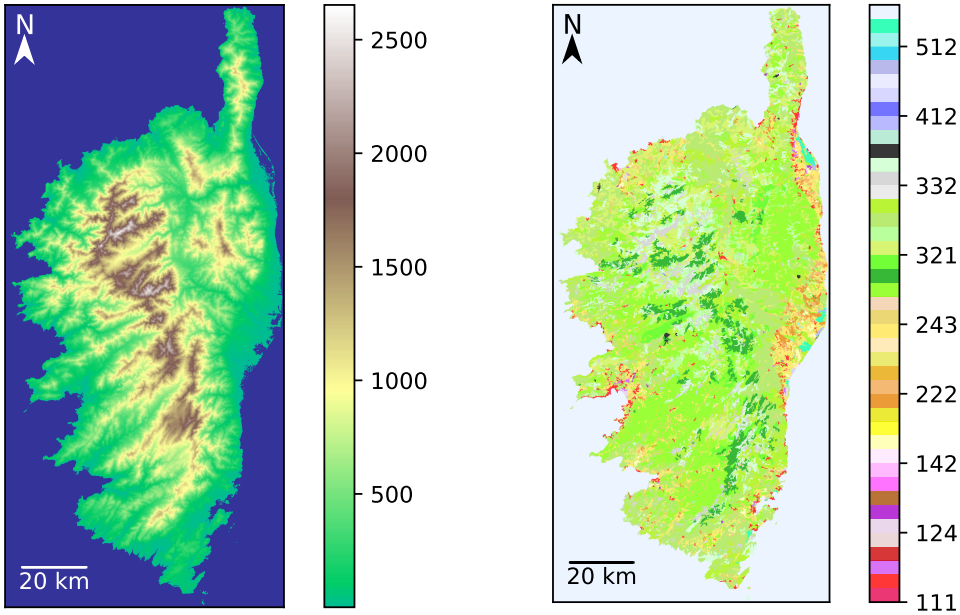


Figure I.9. – Example of daily predicted map of fire danger rating.
The map represents the prediction for 2021.02.04 based on the ECMWF forecast of the FWI and was retrieved from the EFFIS website https://effis.jrc.ec.europa.eu/static/effis_current_situation/public/index.html on 2021.02.03.

in a given day. For instance the FWI is designed to use input weather data at noon (local time) and was originally formulated by correlating these values with the daily maximum fire behavior, whereas the FFDI uses the maximum daily temperature. Also, although the FWI was calibrated to describe fire behavior in a jack pine stands, it has been implemented in many other countries despite the differences in vegetation and climate where it showed reasonable performance (see for instance [56] and references therein). Forecasted maps of such indices generally have a temporal resolution of one day even if the underlying weather forecasts have higher temporal resolution, but the temporal extent and spatial resolution are generally the same. As an example, EFFIS relies on global forecasts of the European Centre for Medium-Range Weather Forecasts (ECMWF, see Figure I.9) and Météo-France which respectively extend from 1 to 9 days with a spatial resolution of 8 km and up to 3 days with a spatial resolution of 10 km.

Weather forecasts covering a smaller area may have higher spatial resolution: for instance the regional AROME limited area model in use at Météo-France has a spatial resolution of 1.3 km covering France and part of neighboring countries and allows the computation of fire danger index at much higher resolution.

In spite of the wide operational use of traditional fire danger indices, they mostly depend on weather information and consist in a unitless quantities that represent potential ignition and/or fire behavior without actually taking into account the influence of slope and fuel type in themselves on fire spread, nor their distribution over the land-



(a) Elevation field.

(b) Land cover field.

Figure I.10. – Data maps of Corsica used to describe the landscape in ForeFire simulations; their spatial resolution is approximately 80 m.
(a) Locations with an altitude of 0 m or less (mostly maritime waters) are represented in blue.
(b) The color scheme corresponds to the classification of the Corine Land Cover

scape. Evaluation of the performance of fire danger indices and design of new ones is an ongoing process, as exemplified by the recently developed Daily Fire Hazard Index [81] that relies on elevation data and satellite observations in complement to weather inputs.

The strategy adopted to quantify wildland fire danger and risk in the present work is to use ROS-based simulations of fire spread over the landscape as a cornerstone. As for traditional fire danger indices, this implies the use of weather forecasts to make day to day predictions, yet opens the possibility to account for elevation and fuel characteristics at high spatial resolution and represent the potential for fire spread in further detail using the characteristics of simulated burned surfaces such as the burned locations or the surface area.

The area of study for application of the methods presented in the following chapters is Corsica island, for which fire spread simulation using ForeFire is carried out using data fields of elevation and land-use in raster format at approximately 80-m spatial resolution (cf. Figure I.10). As part of the FireCaster project, the methods and tools developed in the present project are designed to fit a timeline that satisfies time constraints of operational conditions, which is represented in Figure I.11. Two situations previously mentioned are concerned by the timeline: the “short-term” situation that leads to the computation of predicted fire danger maps with a time extent of 42 hours from the origin time 00:00, all available by 12:00, and the “crisis” situation where a wildfire

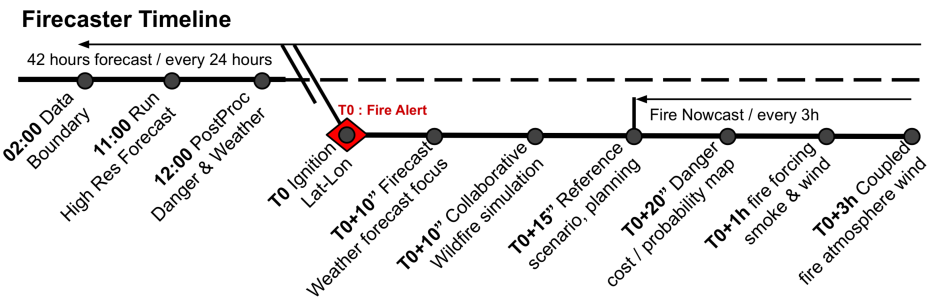


Figure I.11. – Forecasting timeline followed in the FireCaster project. The main goals in the present work are the fire danger maps issued every day for 12:00 and the burn probability map of a specific fire at T0+20", where T0 is the time of fire alert.

starts and the fire alert is issued at time T0 when the ignition location can be estimated, that leads to the computation of an ensemble of fire spread simulations summarized by predicted burn probability maps of the event over time, available ideally by T0+20". The predictions rely on a high resolution (600 m) weather forecast with origin time 00:00 that extends over 42 hours, whose computations can be completed by 11:00 based on a model using boundary and initial conditions available from 02:00. Other products of the FireCaster project include a quick representation of the weather forecast in the area of the fire ("FireCast", cf. T0+10"), a reference simulated scenario of wildfire spread (cf. T0+15"), cost estimation of fires based on simulations (cf. T0+20", see the joint work reported in Appendix A for more details), estimation of smoke dispersion (cf. T0+1h) and computationally intensive simulations using coupled fire-atmosphere model (cf. T0+3h). Computations of the general high-resolution forecast were carried out on 915 days since may 2017, while the specific fire alert forecasting chain was run for 14 fire events since then, 13 of which are analyzed in Chapter V.

Despite this higher temporal and spatial resolution of initial boundary and initial conditions, fire forecasts are not perfectly accurate. There is still considerable uncertainty in models, input data and consequently wildland fire spread simulations. Chapter II addresses this aspect via direct UQ in wildland fire predictions in crisis situations as well as evaluation of burn probability maps stemming from an ensemble of simulations, which can be carried out in a reasonable computational time using ForeFire and distributed computing. The computational requirements of the method proposed for inverse UQ in Chapter III, however, require an unreasonably high number of fire spread simulations, therefore warranting the use of an emulator. For the "short-term" situation, fire danger mapping with a ROS-based fire spread solver requires a high number of simulations to cover a region such as Corsica, especially if one seeks predictions at high spatial and temporal resolution, and an even higher number to also account for uncertainty. This difficulty was addressed by building a NN emulator of wildfire spread predictions in Chapter IV designed to emulate 1-hour burned surface with an input definition domain accounting for any ignition location, a wide range of weather input values and input uncertainty on fuel parameters. Use of this emulator to generate a fire danger index and high-resolution, high-frequency fire danger maps is investigated in Chapter V.

II. Generation and evaluation of an ensemble of wildland fire predictions

In the present chapter, we study direct uncertainty quantification of numerical simulations of wildfire spread of an ongoing fire by attributing probability distributions to the uncertain inputs and propagating them using a Monte Carlo approach. We define several properties and introduce probabilistic scores that are common in meteorological applications to evaluate the probabilistic predictions resulting from the ensemble of simulations. The method is applied to seven fires that occurred in Corsica from mid-2017 to early 2018.

In Section II.7, additional background regarding the notion of consistency presented in Section II.3.4 is provided. We notably present mathematical proofs for the fact that a rank histogram of a consistent ensemble is expected to be flat: first in the usual case of continuous variables, then in the case of binary variables for which we proposed an adapted definition of the rank histogram.

Summary

II.1. Introduction **49**
II.2. Probability distribution of wildfire spread simulations . . . **51**
II.2.1. Fire spread simulation 51
II.2.2. Modeling uncertainty in input data 51
II.2.3. Uncertainty propagation in fire spread simulations 53
II.3. Probabilistic evaluation **54**
II.3.1. Accuracy 54
II.3.2. Reliability and sharpness 55
II.3.3. Probabilistic resolution 55
II.3.4. Consistency 56
II.3.5. Evaluation domain 57
II.4. Application to seven Corsican fires **57**
II.4.1. Data sources 57
II.4.2. Probability distributions 59
II.4.3. Complementary data and error corrections 61
II.4.4. Computational set-up 62
II.5. Results **63**
II.5.1. Detailed case: Calenzana fire 63
II.5.2. Performance for all seven fires 65
II.6. Conclusions **68**
II.7. Consistency and rank histogram **71**
II.7.1. Continuous case 71

II.7.2. Binary case 72

The following published article is reproduced here:
F. Allaire, J.-B. Filippi, and V. Mallet. Generation and evaluation of an ensemble of wildland fire simulations. *International Journal of Wildand Fire*, 29(2):160–173, 2020.

II.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 [128]. 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 wildfire management, there are significant uncertainty sources that may lead to considerable difficulties in determining the most appropriate decisions in an operational context [133]. 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., [119, 28, 87, 86, 40]) to fire spread simulations at regional scale [46, 99, 102, 117, 82]. 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 [114] have been carried out either on one fuel model [119, 87] or more [86] where wind speed norm and direction, slope, fuel moisture content and most (if not all) fuel model parameters are considered as uncertain inputs. In [87, 86], 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 [119], 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.* [24] 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 [46] is commonly used in the United States, and BURN-P3 [99] in Canada. Similar studies were conducted in the Mediterranean region [102, 117] and in California [82]. 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 [49] and other works (e.g., [94, 96]) 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., [37, 43]). More sophisticated methods exist, for instance some use the distance between the vertices of the fire perimeter [52, 36] while others consist in computing scores based on information on the dynamics of the simulated and observed fire surfaces [43]. Evaluation of model performance 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., [35, 118]). Another possibility is to run the simulation based on data that are available at the time of fire start (e.g., [44]). 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 mete-

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

II.2. Probability distribution of wildfire spread simulations

II.2.1. Fire spread simulation

In this study, we used ForeFire [45] 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.

II.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 $15\,\mathrm{m\,s^{-1}}$, a perturbation of $-5\,\mathrm{m\,s^{-1}}$ is admissible, but not if the reference value is $3\,\mathrm{m\,s^{-1}}$. 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 μ and standard deviation σ, 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 σ. 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.

II.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 $(b_1, ..., b_N)$ can be understood as the realization of a random vector $(B_1, ..., B_N)$ 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:

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

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:

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

and

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

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

II.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 [44]) . 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 [140] 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.

II.3.1. Accuracy

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

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

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)

$$BSS = 1 - \frac{BS}{BS_{ref}}. \tag{II.5}$$

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 (II.4) is obtained *a posteriori* with the probability

$$p_c = \frac{1}{N} \sum_{i=1}^{N} o_i. \tag{II.6}$$

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

II.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.

II.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 [97] 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 $i \in \{0, 1, \ldots, n\}$ as $p^j = \frac{j}{n}$, so with the notation introduced in the previous section the partition of the Brier score can be written as follows:

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

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 (II.7) becomes

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

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

II.3.4. Consistency

The idea of consistency for an ensemble was presented by Anderson [9] 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:

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

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

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

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

Finally, to make an overall evaluation based on all the events $A_1, ..., A_N$, 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 $j \in \{0, ..., n\}$.

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.

II.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: $x_1, ..., x_N$. The elementary evaluation event was A_i: 'x_i belongs to the burned surface'.

II.4. Application to seven Corsican fires

II.4.1. Data sources

The fuel layer data are derived from Corine Land Cover (CLC) data [42] coupled with data from the IGN (Institut Géographique National) product BD TOPO® for road and drainage networks. The Digital Elevation Model (DEM) is extracted from another IGN product: BD ALTI® 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 [114]. 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 [10] instead of the one originally proposed by Rothermel. Second, the mineral

CLC	Model	S_v (m^{-1})	e (m)	σ_d ($\mathrm{kg\,m}^{-2}$)	ρ_d ($\mathrm{kg\,m}^{-3}$)	ΔH ($\mathrm{MJ\,kg}^{-1}$)
141	GS2^A	6562	0.46	0.22		
Agricultural	GR2^A	6560	0.30	0.13		
231	GR6^A	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^A	5905	0.46	0.54	512	18.6
322	SH5^A	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 II.1. – Fuel parameterization.
Corresponding Corine Land Cover (CLC) classification and fuel models from [60] (^{A}: with slight alterations). S_v: surface-volume ratio; e: fuel height; σ_d: fuel load; ρ_d: particle density; ΔH: heat of combustion.

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 [60] to determine the values of the fuel parameters, as summarized in Table II.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 [79] 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 [132]. 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 simu-

lations 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 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 [34] 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.

II.4.2. Probability distributions

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

In Table II.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. [64] 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 σ 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 2σ. The chance of sampling

Input	Unit	Perturbation	σ	Range
Wind direction	°	Additive, global	30	$[-60, 60]$
Wind speed norm	$\mathrm{m\,s^{-1}}$	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	$\mathrm{MJ\,kg^{-1}}$	Additive, global	1.0	$[-2.0, -2.0]$
Particle density	$\mathrm{kg\,m^{-3}}$	Additive, global	100	$[-200, 200]$
Fuel height	m	Multiplicative, individual	0.15	$[0.7, 1.3]$
Fuel load	$\mathrm{kg\,m^{-2}}$	Multiplicative, individual	0.15	$[0.7, 1.3]$
Surface-volume ratio	$\mathrm{m^{-1}}$	Multiplicative, individual	0.15	$[0.7, 1.3]$
Ignition point	m	Additive	{50, 150, 250}	In a radius of 2σ
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 II.2. – Properties of the perturbations on scalar inputs.
For the last three inputs, the value of the uncertainty parameter σ 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.

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 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 σ 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.

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

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

We tried to account for most of the uncertainty in the fire simulator inputs, but other approaches that focus on different inputs and use different probability distributions are possible. In [94], 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 [96], 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 simulator. A sophisticated approach regarding input uncertainty quantification is presented in [15] 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.

II.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 σ 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 II.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) [55] (MODIS Collection 6 NRT Hotspot / Active Fire Detections MCD14DL) and VIIRS (Visible Infrared Imaging Radiometer Suite) [121] (NRT VIIRS 375 m Active Fire product VNP14IMGT).

Fire name	Start time	End time	σ_{loc}	σ_{start}	σ_{end}
Olmeta 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 1700 hours	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 II.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 (σ_{loc}), time of fire start (σ_{start}), and time of fire end (σ_{end}).

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°, 9.005°) instead of (42.569°, 8.998°). These modifications came with an increase in the uncertainty parameter σ 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°, 9.164°) were assigned together with a relatively high value of σ.

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

II.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 (∼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.

II.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.

II.5.1. Detailed case: Calenzana fire

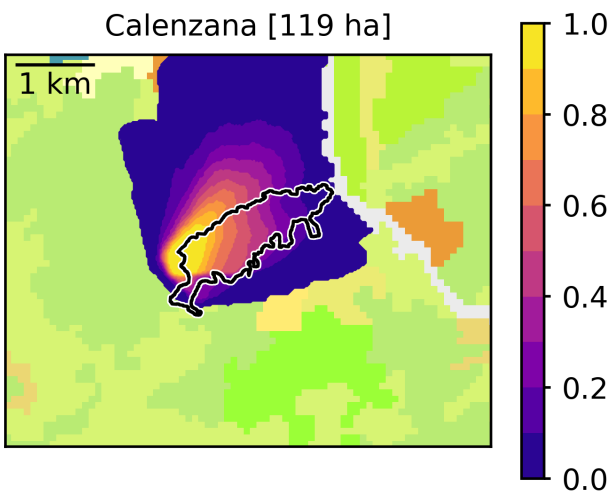
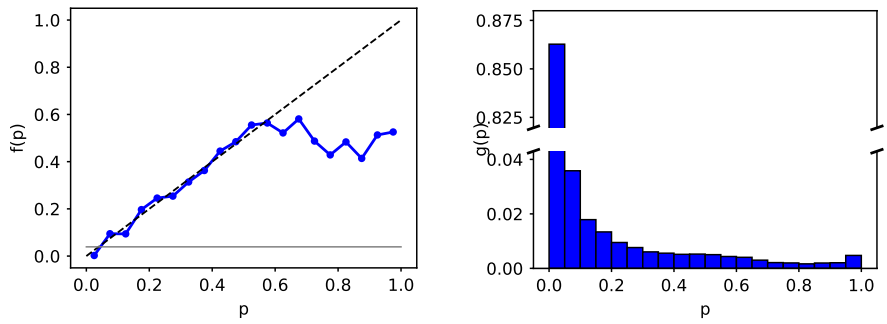


Figure II.1. – Burn probability map obtained in the case of Calenzana fire (5 August 2017).
The colorbar indicates the predicted burn probability ; black and white line is the contour of the observed burned surface ; background colors represent the Corine Land Cover data

The burn probability map in this first case is shown in Figure II.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 II.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, 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 II.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.



(a) Reliability (b) Sharpness

Figure II.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).

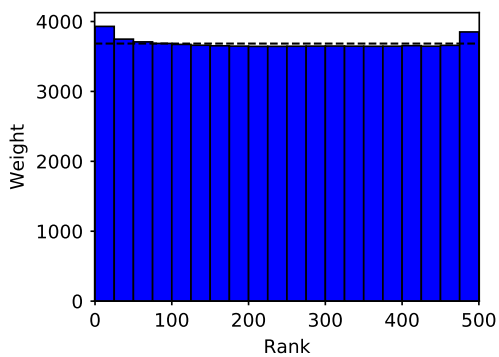


Figure II.3. – Rank diagram obtained in the case of the Calenzana fire (5 August 2017). The 501 ranks are regrouped in 20 bins of 25 consecutive ranks (26 for the first bin). The dashed line indicates the flat histogram that would be obtained with a consistent ensemble. Overall, the ensemble appears fairly consistent.

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 II.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 II.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.

II.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 II.4. Most maps in Figure II.4 do not cover the entire evaluation domain. Characteristics and performance indicators of the ensembles are summarized in Table II.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

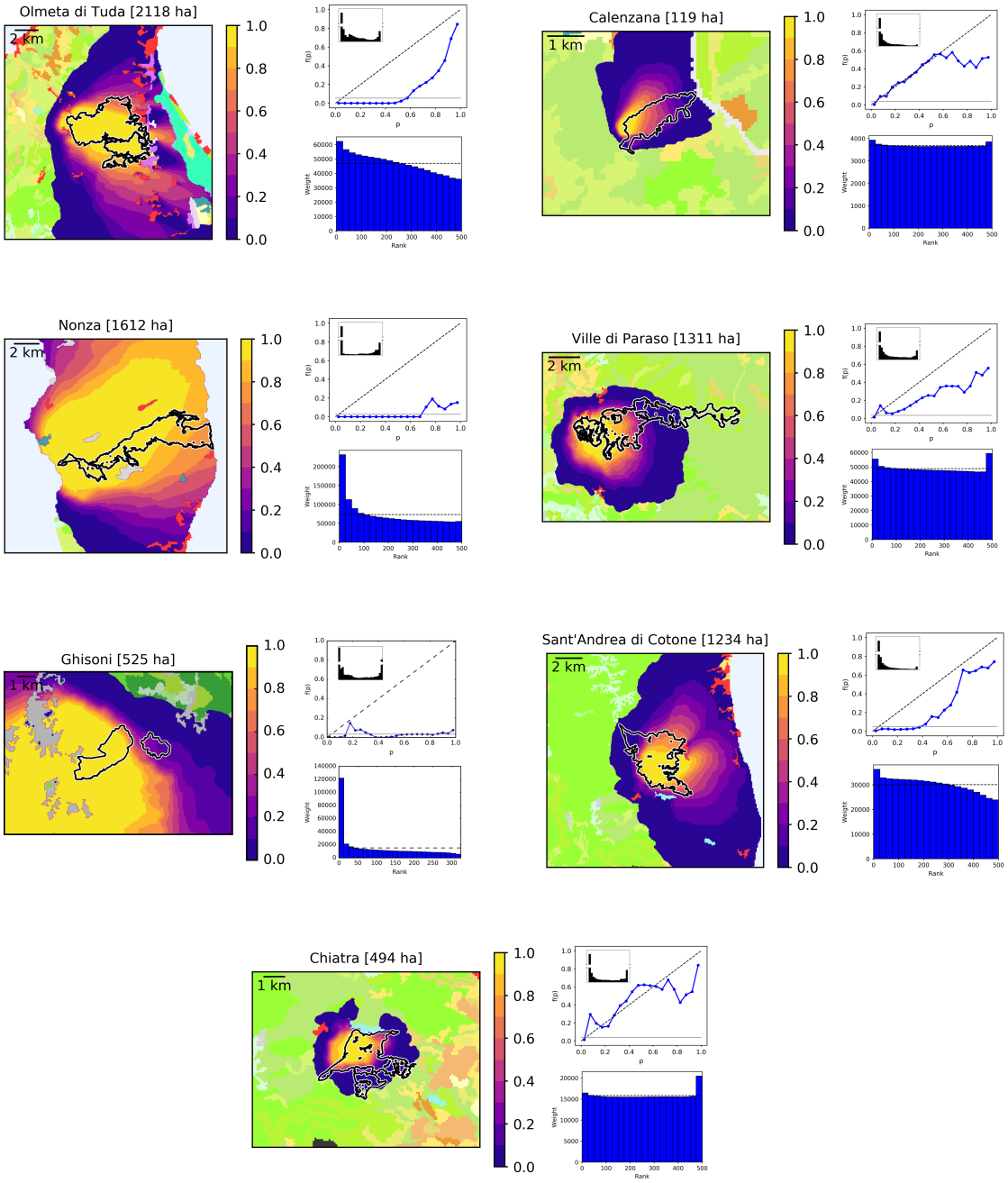


Figure II.4. – Simulations and evaluation for all fires.
For a given fire case, the burn probability map is on the left; the reliability diagram is in the upper right, with the sharpness histogram overlaid in black; the rank histogram is in the lower right. For almost all cases, only a part of the evaluation domain is covered by the map.

Fire name	Duration	p_c	Observed size	Mean size	BS	BSS	Runtime
Olmeta 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 II.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.

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

II.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 wildfire simulations. The probability distributions could be refined by comparing different sources of data [15] 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 [23] (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 [14].

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.

II.7. Consistency and rank histogram

We propose to define consistency in mathematical terms as follows: the observation and the ensemble members are the realization of i.i.d. random variables following a given probability distribution. (NB: This probability distribution may differ from an event to another)

II.7.1. Continuous case

Usually, in a deterministic setting, the rank $R(x)$ of a real number x among a set of n real numbers $x_1, ..., x_n$ is defined as follows: if exactly j of the x_k are lower than x, then $R(x) = j$.
Let us assume that, the scalars $x_1, ..., x_n$ are realizations of the i.i.d. continuous random variables $X_1, ..., X_n$ with PDF g and the corresponding cumulative distribution function (CDF) G. In this case, the rank $R(x)$ of x among $X_1, ..., X_n$ is a random variable and for a given $j \in \{0, ..., n\}$, we have:

$$\begin{aligned}\mathbb{P}[R(x) = j] &= \tbinom{n}{j}\mathbb{P}[X_1 \leq \ldots < X_j < x < X_{j+1} < \ldots < X_n] \\ &= \tbinom{n}{j} \textstyle\prod_{i=1}^{j} \mathbb{P}[X_i < x] \prod_{i=j+1}^{n} \mathbb{P}[x < X_i] \\ &= \tbinom{n}{j} \textstyle\prod_{i=1}^{j} \mathbb{P}[X_1 < x] \prod_{i=j+1}^{n} (1 - \mathbb{P}[X_1 \leq x]),\end{aligned}$$

hence, the probability distribution of $R(x)$ can be expressed as follows:

$$\forall j \in \{0, ..., n\},\ \mathbb{P}[R(x) = j] = \binom{n}{j} G(x)^j (1 - G(x))^{n-j}. \tag{II.10}$$

Now, let us consider a random variable X instead of x. We study $R(X)$, the rank of X among $X_1, ..., X_n$. Under the assumptions that $X, X_1, ..., X_n$ are independent and that X is defined by the pdf f with the corresponding cdf F, it follows from (II.10) that

$$\forall j \in \{0, ..., n\},\ \mathbb{P}[R(X) = j] = \binom{n}{j} \int_{-\infty}^{+\infty} G(x)^j (1 - G(x))^{n-j} f(x) dx. \tag{II.11}$$

(NB: If the independence assumption regarding all $n + 1$ variables is not made, Equation (II.11) does not hold. As an example, if we consider $n = 1$, $X \sim \mathcal{N}(0, 1)$, and $X_1 = X + 1$, we have $\mathbb{P}[R(X) = 0] = 1$.)
If X and the X_k follow the same probability distribution, (i.e., if $f = g$ and $F = G$), we can prove by induction that:

$$\forall j \in \{0, ..., n\}, \int_{-\infty}^{+\infty} F(x)^j (1 - F(x))^{n-j} f(x) dx = \frac{1}{\binom{n}{j}(n + 1)}, \tag{II.12}$$

and, consequently,

$$\forall j \in \{0, ..., n\},\ \mathbb{P}[R(X) = j] = \frac{1}{n + 1}. \tag{II.13}$$

Now let us consider N observed events x_i with the corresponding ensemble predictions $x_{i,1}, x_{i,2}, ..., x_{i,n}$. We assume that $\forall i \in \{1, ..., N\}, x_i, x_{i,1}, ..., x_{i,n}$ is a realization of the i.i.d. continuous random variables $X_i, X_{i,1}, ..., X_{i,n}$ described by PDF f_i with the cor-

responding cdf F_i. This implies that we have a consistent ensemble forecasting system and, from the previous discussion, we have

$$\forall i \in \{1, ..., N\}, \forall j \in \{0, ..., n\},\ \mathbb{P}[R(X_i) = j] = \frac{1}{n+1}. \tag{II.14}$$

The probabilistic perspective put aside, let us define $\forall j \in \{0, ..., n\}, y_j = \sum_{i=1}^{N} \mathbb{1}_{R(x_i=j)}$, where $\mathbb{1}_A$ denotes the indicator function equal to 1 if A, equal to 0 otherwise. The rank histogram is obtained by plotting the y_j against j. Now, back to a probabilistic setting using random variable Y_j instead of it deterministic counterpart, for a consistent ensemble we easily obtain from (II.14) that $\forall j \in \{0, ..., n\}$,

$$\mathbb{E}[Y_j] = \mathbb{E}\left[\sum_{i=1}^{N} \mathbb{1}_{R(X_i=j)}\right] = \sum_{i=1}^{N} \mathbb{E}\left[\mathbb{1}_{R(X_i)=j}\right] = \sum_{i=1}^{N} \mathbb{P}[R(X_i) = j] = \frac{N}{n+1}, \tag{II.15}$$

which indicates that, "overall" (due to the use of the expected value) a consistent system will return a flat rank histogram.

Remark: When the PDFs g_i of the ensemble members are different from f_i, the observations' PDFs, the ensemble is not consistent anymore. It is still possible to compute the shape of the rank histogram under the independance assumption. Considering the i-th event, the substitution $u = F_i(x)$ in (II.11) yields

$$\forall j \in \{0, ..., n\},\ \mathbb{P}[R(X_i) = j] = \int_0^1 \binom{n}{j} (G_i(F_i^{-1}(u)))^j (1 - G_i(F_i^{-1}(u)))^{n-j} du. \tag{II.16}$$

In the absence of a simpler formula, a numerical integral approximation (by the means of a trapezoidal rule, for instance) can be implemented to obtain an estimate of the rank probabilities. The expected shape of the rank histogram is still given by

$$\forall j \in \{0, ..., n\}, \mathbb{E}[Y_j] = \sum_{i=1}^{N} \mathbb{P}[R(X_i) = j], \tag{II.17}$$

which holds for any joint probability distribtion of $(X, X_1, ..., X_n)$.

II.7.2. Binary case

In the situation where we observe *strictly* binary events (1 if an event occurs, 0 otherwise), the rank of the i-th observation is defined as a vector following Equation (II.9).

In this binary context, we assume that each observation x_i is a realization of a random variable X_i following a Bernoulli law of parameter q_i and that each associated observations $x_{i,1}, ..., x_{i,n}$ are realizations of the indentically distributed random variables $X_1, ..., X_n$ that follow a Bernoulli law of parameter p_i. We also assume that $X, X_1, ..., X_n$ are independent. In this case, instead of using the deterministic $w_{i,j}$, we focus on random variable $W_{i,j}$ (its probabilistic counterpart) so as to calculate $\mathbb{E}[W_{i,j}]$ for given $i \in \{1 \dots N\}$ and $j \in \{0 \dots n\}$. Only the cases where $W_{i,j} > 0$ are of interest

in this calculus. It follows that:

$$\mathbb{E}[W_{i,j}] = \sum_{k=0}^{n-j} \frac{1}{n-k+1} \mathbb{P}\Big[X_i = 0 \cap \sum_{j=1}^{n} X_{i,j} = k\Big] + \sum_{k=n-j}^{n} \frac{1}{k+1} \mathbb{P}\Big[X_i = 1 \cap \sum_{j=1}^{n} X_{i,j} = k\Big], \tag{II.18}$$

and, based on the definition of the random variables, it yields:

$$\mathbb{E}[W_{i,j}] = (1-q_i) \sum_{k=0}^{n-j} \frac{1}{n-k+1} \binom{n}{k} (p_i)^k (1-p_i)^{n-k} + q_i \sum_{k=n-j}^{n} \frac{1}{k+1} \binom{n}{k} (p_i)^k (1-p_i)^{n-k}. \tag{II.19}$$

After an index substitution in the second sum of Equation (II.19) and simple developments on the binomial coefficients, we obtain

$$\mathbb{E}[W_{i,j}] = \frac{1}{n+1} \left((1-q_i) \sum_{k=0}^{n-j} \binom{n+1}{k} (p_i)^k (1-p_i)^{n-k} + q_i \sum_{k=n-j+1}^{n+1} \binom{n+1}{k} (p_i)^{k-1} (1-p_i)^{n-k+1} \right). \tag{II.20}$$

If $\forall i \in \{1, ..., N\}, p_i = q_i$, then we have a consistent ensemble forecasting system. Equation (II.20) can be simplified as follows:

$$\mathbb{E}[W_{i,j}] = \frac{1}{n+1} \sum_{k=0}^{n+1} \binom{n+1}{k} (p_i)^k (1-p_i)^{n+1-k} = \frac{1}{n+1} (p_i + 1 - p_i)^{n+1} = \frac{1}{n+1}. \tag{II.21}$$

Finally, let us define $y_j = \sum_{i=1}^{N} w_{i,j}$ defining the height of the bar j in the rank histogram in a deterministic setting. In the probabilistic setting, we employ the random variable $Y_j = \sum_{i=1}^{N} W_{i,j}$ instead. As in the continuous case, with a consistent system, it follows from Equation (II.21) that

$$\forall j \in \{0, 1, ..., n\}, \mathbb{E}[Y_j] = \frac{N}{n+1}. \tag{II.22}$$

In other words, a consistent system is expected to return a flat rank histogram.

III. A posteriori uncertainty quantification for wildland fire simulation

In the present chapter, we build upon the study of the previous chapter and focus on inverse uncertainty quantification of wildland fire spread predictions. We propose a method to provide a relevant input probability distribution a posteriori *by integrating the information provided by the observed burned surfaces that were previously used for evaluation purposes. We make use of the Wasserstein distance as an evaluation metric between surfaces to define a pseudo-likelihood function involved in the posterior PDF. Due to the high dimension and the computational requirements of the pseudo-likelihood function, a Gaussian process emulator is built to obtain a sample of the calibrated input distribution with a MCMC algorithm. We apply the method using the seven Corsican fires presented in the previous chapter and investigate the calibrated input distribution as well as the resulting probabilistic predictions of wildland fire spread.*

Summary

III.1. Introduction **76**
III.2. A posteriori uncertainty quantification **78**
III.2.1. Distribution based on Wasserstein distance 79
III.2.2. Emulation 81
a. Gaussian process modeling 82
b. Design of experiments and error metrics 83
III.2.3. Extension to several fire cases 83
III.2.4. Sampling from the calibrated distribution 84
III.3. Application to wildland fire spread **85**
III.3.1. Fire spread simulation 85
III.3.2. Prior uncertainty in input data 86
III.3.3. Application to seven Corsican wildland fires 88
III.4. Results **89**
III.4.1. Emulation 89
III.4.2. Calibrated distribution 89
III.4.3. Ensemble evaluation 95
III.5. Discussion and conclusions **99**
III.6. Appendix: Bootstrap confidence interval for the Brier skill score **100**

The following published article is reproduced here:
F. Allaire, V. Mallet, and J.-B. Filippi. Novel method for a posteriori uncertainty quantification in wildland fire spread simulation. *Applied Mathematical Modelling*, 90:527–546, 2021.

III.1. Introduction

Modeling wildland fire spread is a challenging task due to the high nonlinearity of the phenomenon and the significant uncertainties in the modeling process. Several models have been developed to describe the dynamics of wildland fire spread [129] with varying degrees of complexity leading to (semi-)physical and (semi-)empirical models. Physical models are complex and the associated simulations are too time-expensive for the computation of a very large scale wildfire in real time. Meanwhile, 2D fire spread simulators [131] that describe the dynamics of the shape of the fire are faster and more suited to making predictions in an operational context. Such simulators typically resolve the spread of the interface between unburned and burned areas (aka fire front) based on an empirical rate of spread (ROS) model. Typical ROS models try to estimate the velocity at which the flames advance locally and express it as a function of local environmental parameters (such as wind speed, slope angle normal to the front, fuel moisture and vegetation properties).

Although the use of empirical models implies a drastic simplification of the physics of wildland fire spread, they are usually non-linear. For instance, most ROS models imply a power-type relation between ROS and wind speed [130]. Model simplifications, together with high variability of the environmental conditions, difficulty of measurements, etc., lead to considerable uncertainty in the modeling process which should be accounted for with appropriate strategies if simulation results were to be used by operators in the field.

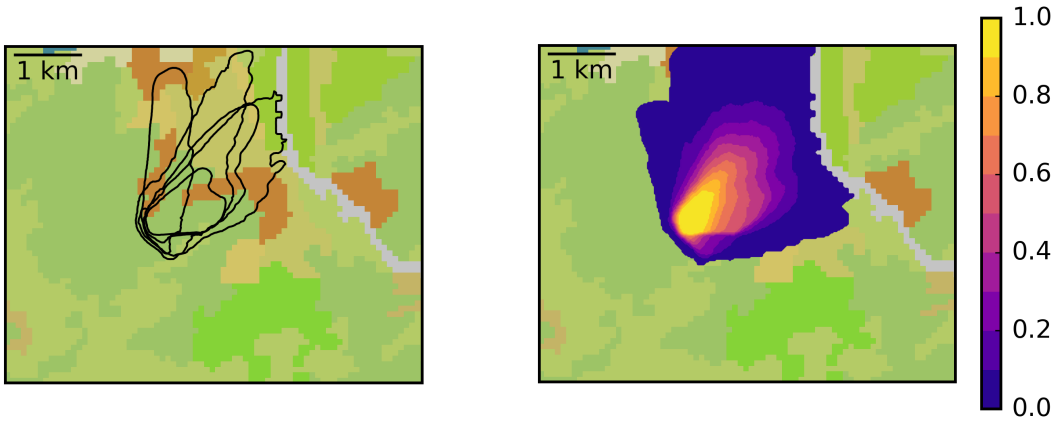


Figure III.1. – Ensemble of simulated fire fronts. Left: the black contours represent five of the simulated fire fronts ; right: burn probability map resulting from the whole ensemble ; background colors represent the Corine Land Cover data [42].

Instead of relying on a single deterministic prediction, an alternative consists in generating a probabilistic prediction of fire spread in order to quantify this uncertainty. In wildland fire predictions, probabilistic methods mostly focus on the uncertainty of the inputs, which is propagated through the fire spread simulators. In [24], the authors investigated the uncertainty in fuel parameters via a sensitivity analysis in fire spread simulations of an actual fire; however, other influential inputs such as wind and fuel

moisture content were fixed in this case study. In the context of a "crisis" situation, that is to say when a wildland fire has just started, there is significant uncertainty regarding the meteorological inputs, which are typically derived from weather forecasts. Other uncertain inputs may include fuel parameters, as well as the exact time and location of fire start. The propagation of uncertainty through the simulator is usually carried out by a Monte Carlo method where several input sets are sampled independently from a given probability distribution that describes input uncertainty. This results in an ensemble of fire spread simulations, as illustrated in Figure III.1. The ensemble members can then be aggregated into a *burn probability map*, which estimates the probability for a zone to be reached by the fire, according to the input uncertainty that was specified for the simulations. Several works proposed a strategy to make a probabilistic prediction of fire spread stemming from an ensemble of simulations (e.g. [49, 94, 96, 15]), taking into account input uncertainty while other sources (notably, model errors) are usually unaccounted for.

The goal of the present study is to provide a method to generate a relevant probabilistic forecast of burned area, based on a set of input data that would be available when a fire occurs. A first step was to generate the ensemble and assess (with several evaluation tools) the performance of such a prediction system [5]. Up to that point, quantitative assessment of the performance of fire spread simulators was mostly limited to deterministic approaches. Typical methods consisted in comparing a simulated burned surface with its observed counterpart, for instance by computing indices that measure how much the two surfaces match (e.g. [43, 37]) or by analyzing the distance between vertices of the fire perimeter (e.g. [52, 36]).

In spite of these advances, the quantification of input uncertainty varies a lot from a study to another and remains quite arbitrary. Input uncertainty is typically quantified *a priori* via a probability distribution that is based on data measurements and expert knowledge. A sophisticated approach is presented in [15] 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. Still, in wildland fire modeling, little efforts have been made to use the information contained in the observed fires by comparing them to the corresponding predictions to improve the input distribution. To the authors' knowledge, this topic has only been investigated in [14] by means of the Generalized Likelihood Uncertainty Estimation (GLUE) method, with a focus on the *a posteriori* distribution of the output stemming from the uncertainty in ROS adjustment factors.

The main contribution of the present study is to propose and apply a method to calibrate the probability distribution of the inputs of the model based on observed fires. This procedure can be seen as a calibration of the input distribution and consists in solving a problem of *inverse uncertainty quantification* (e.g. [141]). The probabilistic predictions generated with the calibrated distributions should lead to better probabilistic scores than those generated with the prior distribution on the input variables. Two major difficulties are encountered: first, the large number of uncertain input variables; second, the model being considered a "black box" whose output is a surface, a formula for the likelihood, fundamental in a Bayesian approach, cannot be written easily. A method inspired by the Bayesian framework is proposed in order to circumvent these two difficulties by making use of a novel score for the comparison of surfaces relying

on the Wasserstein distance for several observed fires. This leads to the definition of a calibrated distribution that can be sampled from via traditional Markov chain Monte Carlo (MCMC) algorithms. An emulator (also called surrogate model or metamodel) of the score is used to drastically decrease computational time, at the cost of relatively low approximation error, so that sufficient MCMC iterations can be carried out in a reasonable amount of time.

The uncertain inputs were identified in [5], with a focus on direct uncertainty quantification. This setting is followed in the present study for inverse uncertainty quantification. Here, the goal is to obtain calibrated probabilistic predictions that are as relevant as possible based on the observation of seven relatively big Corsican fires.

The theory behind the strategy for calibration of input uncertainty is described in Section III.2 and the technical aspects of its application are presented in Section III.3. Section III.4, focuses on the results obtained regarding the emulation, the calibration of the distributions and the evaluation of the resulting ensembles. These results are then discussed in Section III.5.

III.2. A posteriori uncertainty quantification

Several sources of uncertainty can be identified in numerical models, a classification of which was proposed by Kennedy and O'Hagan [77]. The authors identify *unknown parameters*, which are distinguished from other model inputs that are controllable and/or measurable in experiments, *model inadequacy*, and *observation error*, among other sources. They also propose a Bayesian approach for calibration that accounts for the different forms of uncertainty. The starting point of this approach is a model updating equation (see for instance equation (4) in [77]) where the difference between observations and corresponding model outputs at given parameter values is expressed with two additive terms that respectively account for model inadequacy and observation errors. Several studies in engineering applications follow this model updating equation (e.g. [11, 141, 106]) and may include other sources of uncertainty due to interpolation error or model selection, for instance.

In this framework, calibration is understood in a probabilistic sense as it yields distributions, which is also the case in the present study. The present study follows the work presented in [5], which focused on input uncertainty in wildland fire spread predictions by means of perturbations in model inputs, whether these inputs can be seen as model parameters or not. These perturbations allow to account for a fair amount of input uncertainty associated with weather predictions, fuel representation, information regarding fire start, as well as some error of "representativeness" due to the comparison between the final observed burned area and the simulated burned surface computed for a given duration of fire spread. Uncertainty due to model inadequacy is not accounted for. The same goes for observation error, as measurement error in the observed burned surfaces is assumed to be negligible compared to the error due to other sources of uncertainty.

An intuitive way of expressing error in a surface is via simple geometric operations such as translation or rotation. Another way is to consider local "binary" corrections: some unburned locations could belong to the burned surface after correction, and vice versa. However, applying any of these transformations to a surface may lead to "non-

physical" surfaces. For instance, the ignition point could not be located inside the burned surface after transformation, whereas unburnable areas such as lakes could become part of it. Also, classical data assimilation methods applied to front-tracking problems through the assumption of Gaussian errors in the position of a simulated contour can lead to poor performance. This led some authors to consider approaches based on shape-oriented discrepancy measures, such as the Chan-Vese contour fitting functional [112]. A similar approach is followed in the present study since the method used for calibration relies on a similarity measure between surfaces. Arguably, this is better suited to our application than expressing errors between model predictions and observations through usual model updating equations.

In this section, let us first consider the observation of one burned surface. Application to observations of several fires will be introduced in subsection III.2.3. Let $\mathcal{S}_{\text{obs}}$ denote the observed burned surface and $\mathcal{M}$ denote the numerical model of fire spread, whose inputs may vary according to a vector $\boldsymbol{u}$ of d perturbations applied to reference inputs. The model is dynamical and may return burned surfaces at different times but the focus in the present study is on the surface that corresponds to the (estimated) observation time of $\mathcal{S}_{\text{obs}}$. The simulated burned surface being denoted as $\mathcal{S}_{\boldsymbol{u}}$, one may write $\mathcal{S}_{\boldsymbol{u}} = \mathcal{M}(\boldsymbol{u})$. Direct comparison between $\mathcal{S}_{\boldsymbol{u}}$ and $\mathcal{S}_{\text{obs}}$ can be made, but there is uncertainty in the input variables. This uncertainty is modeled by attributing a probability distribution to the perturbation vector, that can be seen as a random vector $\boldsymbol{U}$. Consequently, the output is also stochastic: $\mathcal{S}_{\boldsymbol{U}} = \mathcal{M}(\boldsymbol{U})$. Although $\mathcal{S}_{\boldsymbol{U}}$ is probabilistic and $\mathcal{S}_{\text{obs}}$ is deterministic, they can be compared by means of probabilistic evaluation tools. We seek a distribution that is as suitable as possible for the random vector $\boldsymbol{U}$. It is assumed that the distribution of $\boldsymbol{U}$ is described by the probability density function (PDF) g and that there already exists a *prior* density function f for $\boldsymbol{U}$. In this section, a method is proposed in order to obtain g by making the best possible use of f, $\mathcal{S}_{\text{obs}}$ and $\mathcal{M}$.

III.2.1. Distribution based on Wasserstein distance

A classical choice for g would be the *posterior* density function $p(.|\mathcal{S}_{\text{obs}})$ that is obtained according to Bayes' rule:

$$p(\boldsymbol{u}|\mathcal{S}_{\text{obs}}) = \frac{\mathcal{L}(\mathcal{S}_{\text{obs}}|\boldsymbol{u})f(\boldsymbol{u})}{\int \mathcal{L}(\mathcal{S}_{\text{obs}}|\boldsymbol{u})f(\boldsymbol{u})d\boldsymbol{u}}, \tag{III.1}$$

where $\mathcal{L}(\mathcal{S}_{\text{obs}}|\boldsymbol{u})$ would be the likelihood of the observation $\mathcal{S}_{\text{obs}}$ knowing the perturbation vector $\boldsymbol{u}$. However, defining the likelihood requires to make an appropriate probabilistic hypothesis, where $\mathcal{S}_{\text{obs}}$ is a realization of a 2D stochastic process whose distribution depends on $\boldsymbol{u}$. Making such a hypothesis is not trivial, but a step in this direction would be to use $\mathcal{S}_{\boldsymbol{u}}$ rather than simply $\boldsymbol{u}$, and define a (conditional) probability distribution for $\mathcal{S}_{\text{obs}}$ based on $\mathcal{S}_{\boldsymbol{u}}$. A desirable property of such a probability distribution is that the most likely realizations of $\mathcal{S}_{\text{obs}}$ are the ones that are most similar to $\mathcal{S}_{\boldsymbol{u}}$. Also, similarity should take into account high correlation between two points in a 2D domain when they are close. For instance, if a given location has high probability of being burned, so should have its neighboring locations. Still, while defining a likelihood for a vector is feasible, this might not be the case for a random surface.

Therefore, one may use a calibrated distribution that is inspired by Bayes' rule, where the density function g can be written in the following form:

$$g_{E,\beta}(\boldsymbol{u}) = \frac{e^{-\beta E(\boldsymbol{u})} f(\boldsymbol{u})}{\int e^{-\beta E(\boldsymbol{u})} f(\boldsymbol{u}) d\boldsymbol{u}}, \tag{III.2}$$

where $\beta > 0$ and E is a positive "energy" function that is equal to 0 when $\mathcal{S}_{\boldsymbol{u}} = \mathcal{S}_{\text{obs}}$ and increases with the dissimilarity between $\mathcal{S}_{\boldsymbol{u}}$ and $\mathcal{S}_{\text{obs}}$. Here, a pseudo-likelihood function plays the role of $\mathcal{L}$ in equation (III.1). This calibrated family of functions is inspired by Gibbs measures, but is different because the exponential is multiplied by f, the prior PDF. The higher the parameter β, the more weight is given to the pseudo-likelihood function. Also, when $\beta = 0$, the calibrated PDF is equal to the prior PDF.

Several scores to compare $\mathcal{S}_{\text{obs}}$ and $\mathcal{S}_{\boldsymbol{u}}$ exist and could be used directly or after minor modifications to make suitable choices for E. The present study introduces a novel score that makes use of the Wasserstein distance, which is a metric between probability distributions. The reader may refer to the book [120] for a more extensive review. Following the Kantorovitch formulation of the optimal transport problem, let us consider the Wasserstein distance, denoted as $\mathcal{W}_2(\mu, \nu)$, between two probability measures μ and ν both defined on $\mathbb{R}^q$, whose square can be defined as follows:

$$\mathcal{W}_2^2(\mu, \nu) = \inf \left\{ \int_{\mathbb{R}^q \times \mathbb{R}^q} ||x - y||_2^2 \, d\gamma(x, y) \;\middle|\; \gamma \in \Gamma(\mu, \nu) \right\}, \tag{III.3}$$

where $||.||_2$ is the Euclidean distance (in $\mathbb{R}^q$) and $\Gamma(\mu, \nu)$ is the ensemble of the measures defined on $\mathbb{R}^q \times \mathbb{R}^q$ such that their conditional measure relatively to the first variable is μ and their conditional measure on the second variable is ν. For comparison between surfaces, it is natural to consider $q = 2$ and choose uniform measures whose support is respectively $\mathcal{S}_{\text{obs}}$ and $\mathcal{S}_{\boldsymbol{u}}$ for the probability measures μ and ν. By making these choices, $E(\boldsymbol{u})$ is defined as follows:

$$E(\boldsymbol{u}) = \inf_{\gamma} \left\{ \int_{\mathcal{S}_{\text{obs}} \times \mathcal{S}_{\boldsymbol{u}}} ||x - y||_2^2 \, \gamma(x, y) dx dy \;\middle|\; \begin{array}{l} \displaystyle\int_{\mathcal{S}_{\boldsymbol{u}}} \gamma(x, y) dy = \frac{\mathbb{1}(x \in \mathcal{S}_{\text{obs}})}{|\mathcal{S}_{\text{obs}}|}, \\ \displaystyle\int_{\mathcal{S}_{\text{obs}}} \gamma(x, y) dx = \frac{\mathbb{1}(y \in \mathcal{S}_{\boldsymbol{u}})}{|\mathcal{S}_{\boldsymbol{u}}|} \end{array} \right\}, \tag{III.4}$$

where $\mathbb{1}$ stands for the indicator function, $||.||_2$ is the Euclidean distance (now in $\mathbb{R}^2$), and $|\mathcal{S}|$ is the surface area of $\mathcal{S}$. It can be thought of the minimum energy that is required to move the points contained in $\mathcal{S}_{\text{obs}}$ so as to transform the surface into $\mathcal{S}_{\boldsymbol{u}}$. Also, $E(\boldsymbol{u}) = 0$ when both surfaces are the same.

Except for some particular cases, there is no simple analytic formula for the Wasserstein distance. This led us to consider a discrete approximation of $E(\boldsymbol{u})$ instead, which can be obtained numerically via a discretization of the PDFs by a sum of Dirac delta

distributions. From this point, $E(\boldsymbol{u})$ is now defined as follows:

$$E(\boldsymbol{u}) = \inf_{\gamma}\Big\{ \int_{\mathcal{S}_{\text{obs}}\times\mathcal{S}_{\boldsymbol{u}}} ||x-y||_2^2\ \gamma(x,y)dxdy \ \Big|\ \int_{\mathcal{S}_{\boldsymbol{u}}} \gamma(x,y)dy = \frac{1}{J}\sum_{j=1}^{J}\delta_{x_j}(x), \\ \int_{\mathcal{S}_{\text{obs}}} \gamma(x,y)dx = \frac{1}{K}\sum_{k=1}^{K}\delta_{y_k}(y)\Big\}, \tag{III.5}$$

where δ_x is the Dirac delta distribution at point $x \in \mathbb{R}^2$, and each x_j belongs to $\mathcal{S}_{\text{obs}}$, whereas y_k belongs to $\mathcal{S}_{\boldsymbol{u}}$. In this discrete setting, the admissible distributions γ can be represented by a matrix of size $J \times K$ where each cell γ_{jk} is positive and indicates the "probability mass" that is transferred from x_j to y_k. In this case, the infimum of (III.5) is reached and is the solution of the following linear programming problem:

$$\min_{\gamma_{j,k}} \sum_{j=1}^{J}\sum_{k=1}^{K} \gamma_{jk}||x_j - y_k||_2^2 \tag{III.6}$$

$$\text{subject to } \gamma_{jk} \geq 0,\ \sum_j \gamma_{jk} = \frac{1}{K},\ \text{and} \sum_k \gamma_{jk} = \frac{1}{J}. \tag{III.7}$$

which is also referred to as the *Earth Mover's distance* (EMD) [16]. It is known from graph theory that the optimal γ is a sparse matrix that has at most $J+K-1$ non-zero cells.

An issue that remains to be addressed is that of the denominator of $g_{E,\beta}(\boldsymbol{u})$ defined in equation (III.2), which is an intractable high-dimensional integral. However, this integral does not depend on the perturbation vector $\boldsymbol{u}$, so for a given β, the PDF is known up to some constant factor. When a distribution is known up to a factor, the Metropolis-Hastings (MH) algorithm allows to draw samples from that distribution. In this paper, the MH algorithm is therefore used to describe the a posteriori distribution of $\boldsymbol{U}$ with a very large sample. Computing $\mathcal{S}_{\boldsymbol{u}}$ and the Wasserstein distance to obtain $E(\boldsymbol{u})$ can be done in a reasonable amount of time. Nonetheless, a lot of iterations ($\sim 10^5$) of the MH algorithm may be required to obtain a sufficiently large sample, which would take too much time. To speed up the MH algorithm, an emulator $\widetilde{E}$ is used. The emulated value $\widetilde{E}(\boldsymbol{u})$ should still provide a good approximation of $E(\boldsymbol{u})$ while being considerably faster to compute. The general design of the emulator is explained in Section III.2.2 and its application to an energy function specific to several fire cases is given in Section III.2.3. To determine whether an appropriate sample is returned by the MH algorithm, one may use the multivariate diagnostic metric proposed by [19], as detailed in Section III.2.4.

III.2.2. Emulation

The focus of this section is the approximation of a function $y : \boldsymbol{u} \in \mathcal{D} \subset \mathbb{R}^d \to \mathbb{R}$. This function can be the previous Wasserstein distance $E(\boldsymbol{u})$, or the extension to several fires presented in Section III.2.3. Note that the following emulation approach is fairly general and can be applied to a wide range of functions.

a. Gaussian process modeling

The emulation method used in this study is Gaussian process (GP) modeling, also called kriging. In this context, $y(\boldsymbol{u})$ is seen as a realization of a Gaussian process $Y_{\boldsymbol{u}}$ indexed by $\boldsymbol{u}$. It means that any random vector $[Y_{\boldsymbol{u}^1}, \ldots, Y_{\boldsymbol{u}^n}]^T$ with $n < \infty$ components follows a Gaussian multivariate distribution. Let a denote the trend function of the process, i.e., $\mathbb{E}[Y_{\boldsymbol{u}}] = a(\boldsymbol{u})$. The centered process $Z_{\boldsymbol{u}} = Y_{\boldsymbol{u}} - a(\boldsymbol{u})$ is also Gaussian, with a covariance function of the form $\mathrm{Cov}(\boldsymbol{u}, \boldsymbol{u}') = \sigma^2 \rho(\boldsymbol{u} - \boldsymbol{u}')$, where $\sigma^2 > 0$ and ρ is the correlation function between two input points $\boldsymbol{u}$ and $\boldsymbol{u}'$.

We have at our disposal a set of training data $\left(\boldsymbol{u}^i, y(\boldsymbol{u}^i)\right)_{i=1,\ldots,n}$. Let us denote $\boldsymbol{Y}^n = [Y_{\boldsymbol{u}^1}, ..., Y_{\boldsymbol{u}^n}]^T$, $\boldsymbol{y}_n = [y(\boldsymbol{u}^1), ..., y(\boldsymbol{u}^n)]^T$ and define R_n, the correlation matrix on the inputs of the training data:

$$R_n = \left(\rho(\boldsymbol{u}^i - \boldsymbol{u}^j)\right)_{1 \geq i,j \geq n}, \tag{III.8}$$

and $\boldsymbol{a}_n = [a(\boldsymbol{u}^1), ..., a(\boldsymbol{u}^n)]^T$, the vector of trends in the training data.

For a new point $\boldsymbol{u}^*$ (outside the training sample or not), we define the correlation vector $\boldsymbol{r}^* = [r(\boldsymbol{u}^* - \boldsymbol{u}^1), ..., r(\boldsymbol{u}^* - \boldsymbol{u}^n)]^T$. Under the assumptions made on $Y_{\boldsymbol{u}}$, the joint probability distribution between $\boldsymbol{Y}^n$ and $Y_{\boldsymbol{u}^*}$ is Gaussian and so is the conditional distribution of $Y_{\boldsymbol{u}^*}$ knowing $\boldsymbol{Y}^n$. One may write

$$Y_{\boldsymbol{u}^*} | \boldsymbol{Y}^n \sim \mathcal{N}\big(\mathbb{E}[Y_{\boldsymbol{u}^*} | \boldsymbol{Y}^n], \mathrm{Var}[Y_{\boldsymbol{u}^*} | \boldsymbol{Y}^n]\big), \tag{III.9}$$

where

$$\mathbb{E}[Y_{\boldsymbol{u}^*} | \boldsymbol{Y}^n] = a(\boldsymbol{u}^*) + \boldsymbol{r}^{*T} {R_n}^{-1} (\boldsymbol{y}_n - \boldsymbol{a}_n) \tag{III.10}$$

and

$$\mathrm{Var}[Y_{\boldsymbol{u}^*} | \boldsymbol{Y}^n] = \sigma^2 (1 - \boldsymbol{r}^{*T} {R_n}^{-1} \boldsymbol{r}^*). \tag{III.11}$$

For any $\boldsymbol{u}^* \in \mathcal{D}$, one may define an emulator $\widetilde{y}$ of y as the mean of the conditional variable given by equation (III.10):

$$\widetilde{y}(\boldsymbol{u}^*) = a(\boldsymbol{u}^*) + \boldsymbol{r}^{*T} {R_n}^{-1} (\boldsymbol{y}_n - \boldsymbol{a}_n). \tag{III.12}$$

In the present case, a linear trend is chosen for $Y_{\boldsymbol{u}}$, i.e., $\mathbb{E}[Y_{\boldsymbol{u}}] = a(\boldsymbol{u}) = \alpha_0 + \boldsymbol{u}^T \boldsymbol{\alpha}$ where $\alpha_0 \in \mathbb{R}$ and $\boldsymbol{\alpha} \in \mathbb{R}^d$. Also, the choice for correlation function is a product of one-dimensional Matérn 5/2 correlation functions, i.e.,

$$\forall \boldsymbol{u}, \boldsymbol{u}' \in \mathcal{D}, \rho(\boldsymbol{u} - \boldsymbol{u}') = \prod_{l=1}^{d} \left(1 + \frac{\sqrt{5}|u_l - u_l'|}{\theta_l} + \frac{5|u_l - u_l'|^2}{3{\theta_l}^2}\right) \exp\left(-\frac{\sqrt{5}|u_l - u_l'|}{\theta_l}\right), \tag{III.13}$$

where $\theta_1, ..., \theta_d > 0$.

For the sake of clarity, the previous presentation focuses on simple kriging, where the coefficients of a and the covariance hyperparameters are known. In this study, universal kriging is used, so the trend is an unknown polynomial (for more information, see for instance [115]). In practice, the $2d + 2$ hyperparameters $\sigma^2, \alpha_0, \alpha_1, ..., \alpha_d, \theta_1, ..., \theta_d$ used to define the Gaussian process are unknown and can be estimated as the maximum likelihood estimators for the training dataset [101].

b. Design of experiments and error metrics

The inputs of the training sample are obtained via a Latin hypersquare sample (LHS) with optimized discrepancy. As the GP emulator is built from the points of the training sample, a complementary test sample is generated to evaluate the approximation error of the emulator far from the training points. The complementary sample is obtained with the algorithm for an optimal validation design described in [70]. It relies on a low-discrepancy sequence (in the present study, a Sobol' sequence is used) whose points are selected in order to keep a low discrepancy when both training and test samples are taken together. This procedure aims at selecting points that are located far from each other but also far from the points of the training sample, where the approximation error is expected to be higher.

Based on the test sample $\left(\boldsymbol{u}^i, y(\boldsymbol{u}^i)\right)_{i=1,\dots,n_{\text{test}}}$ several error metrics can be used to evaluate the emulator $\widetilde{y}$. In this study, the mean absolute error (MAE) is used and defined as follows:

$$\text{MAE} = \frac{1}{n_{\text{test}}} \sum_{i=1}^{n_{\text{test}}} |\widetilde{y}(\boldsymbol{u}^i) - y(\boldsymbol{u}^i)|. \tag{III.14}$$

Let us introduce the standardized mean square error (SMSE), which is defined as follows:

$$\text{SMSE} = \frac{\sum_{i=1}^{n_{\text{test}}} \left(\widetilde{y}(\boldsymbol{u}^i) - y(\boldsymbol{u}^i)\right)^2}{\sum_{i=1}^{n_{\text{test}}} \left(y(\boldsymbol{u}^i) - \bar{y}\right)^2}, \tag{III.15}$$

where $\bar{y} = \frac{1}{n_{\text{test}}} \sum_{i=1}^{n_{\text{test}}} y(\boldsymbol{u}^i)$ is the sample mean of the emulated function based on the test sample. It can be seen as a mean squared error normalized by the variance of the function on the test sample. The Q_2 metric, which is closely related to the SMSE, is more commonly used, and defined as follows:

$$Q_2 = 1 - \frac{\sum_{i=1}^{n_{\text{test}}} \left(\widetilde{y}(\boldsymbol{u}^i) - y(\boldsymbol{u}^i)\right)^2}{\sum_{i=1}^{n_{\text{test}}} \left(y(\boldsymbol{u}^i) - \bar{y}\right)^2} = 1 - \text{SMSE}. \tag{III.16}$$

As the error of the emulator gets lower, the MAE gets closer to 0 and the Q_2 gets closer to 1. Note that a model that would always predict the mean of the training set would have a Q_2 approximately equal to 0.

III.2.3. Extension to several fire cases

Considering K fire cases, it is possible to compute the energy functions $E_1, ..., E_K$ that correspond to each fire. An intuitive choice for the combined energy function is $E : \boldsymbol{u} \mapsto \sum_{k=1}^{K} E_k(\boldsymbol{u})$. However, a concerning issue is when the variations of $E_1(\boldsymbol{u})$ (for instance) are much higher than for the other fires. In this case, the variations of the pseudo-likelihood will be mostly determined by those of $E_1(\boldsymbol{u})$, and the calibrated distribution will be mostly representative of the information from the first fire at the expense of the other observations.

To circumvent this issue, one may weigh each fire depending on the values taken by $E_k(\boldsymbol{u})$ and define the energy function as the weighted sum of squared Wasserstein

distances:

$$E(\boldsymbol{u}) = \sum_{k=1}^{K} w_k E_k(\boldsymbol{u}), \tag{III.17}$$

where the weights are defined using all points from the training dataset:

$$w_k = \frac{n}{\sum_{i=1}^{n} E_k(\boldsymbol{u}^i)}. \tag{III.18}$$

It is possible to emulate $E(\boldsymbol{u})$ directly but while the function is positive, emulation by GP does not guarantee positivity outside the training sample. Alternatively, one may emulate $L(\boldsymbol{u}) = \log E(\boldsymbol{u})$ by the GP procedure described in Section III.2.2, leading to the emulator $\widetilde{L}(\boldsymbol{u})$. Emulation of $E(\boldsymbol{u})$ simply follows from taking the exponential $\widetilde{E}(\boldsymbol{u}) = \exp \widetilde{L}(\boldsymbol{u})$, which ensures positivity. The GP emulation is implemented in the R-package *DiceKriging* [115].

III.2.4. Sampling from the calibrated distribution

Algorithm 2 Metropolis-Hastings algorithm applied to $g_{\widetilde{E},\beta}$ (several chains)

Define m, n, and an instrumental distribution of PDF $q : \boldsymbol{u} \mapsto q(\boldsymbol{u}|\boldsymbol{v})$
for $j = 1, ..., m$ **do**
  Choose a starting point $\boldsymbol{u}_{1,j}$
  for $i = 2, ..., n$ **do**
    Sample a candidate $\boldsymbol{u}_{c,j} \sim q(.|\boldsymbol{u}_{i-1,j})$
    Compute the ratio

$$\tau = \frac{g_{\widetilde{E},\beta}(\boldsymbol{u}_{c,j})\ q(\boldsymbol{u}_{i-1,j}|\boldsymbol{u}_{c,j})}{g_{\widetilde{E},\beta}(\boldsymbol{u}_{i-1,j})\ q(\boldsymbol{u}_{c,j}|\boldsymbol{u}_{i-1,j})} = \frac{e^{-\beta\widetilde{E}(\boldsymbol{u}_{c,j})}\ f(\boldsymbol{u}_{c,j})\ q(\boldsymbol{u}_{i-1,j}|\boldsymbol{u}_{c,j})}{e^{-\beta\widetilde{E}(\boldsymbol{u}_{i-1,j})}\ f(\boldsymbol{u}_{i-1,j})\ q(\boldsymbol{u}_{c,j}|\boldsymbol{u}_{i-1,j})} \tag{III.19}$$

    if $\tau \geq 1$ **then**
      [Accept the candidate]
      $\boldsymbol{u}_{i,j} \leftarrow \boldsymbol{u}_{c,j}$
    else
      [Accept the candidate with probability τ]
      Sample $p \sim \mathcal{U}(0,1)$
      if $p \leq \tau$ **then**
        $\boldsymbol{u}_{i,j} \leftarrow \boldsymbol{u}_{c,j}$
      else
        $\boldsymbol{u}_{i,j} \leftarrow \boldsymbol{u}_{i-1,j}$
      end if
    end if
  end for
end for
return $(\boldsymbol{u}_{1,j}, ..., \boldsymbol{u}_{n,j})_{j=1,...,m}$

This section focuses on the procedure used to obtain a sample following a PDF of

the form $g_{\widetilde{E},\beta}$ described in equation (III.2). To run the algorithm in reasonable computational time, the emulator $\widetilde{E}$ is used in place of the energy function E as explained in section III.2.3, assuming that the target distribution of MH is close enough to the desired distribution whose PDF is $g_{E,\beta}$. The procedure is presented in Algorithm 2. Let us denote the number of chains as m and the number of samples per chain as n, so that the i-th element of the j-th chain is denoted as $\boldsymbol{u}_{i,j}$. The instrumental distribution that is used to sample a candidate $\boldsymbol{u}_{c,j}$ (here, the subscript 'c' stands for "candidate") from element $\boldsymbol{u}_{i-1,j}$ is defined by the PDF $q : \boldsymbol{u} \mapsto q(\boldsymbol{u}|\boldsymbol{v})$.

The use of a version of MH with several chains as presented in Algorithm 2 is motivated by the convergence diagnosis for MCMC algorithms introduced by Brooks and Gelman [19]. It is recommended to choose the starting points $\boldsymbol{u}_{1,1}, ..., \boldsymbol{u}_{1,m}$ quite far from each other. The loop on the m chains can be parallelized easily. Based on the chains returned by the MH algorithm, the between-sequence covariance matrix B/n (of size d) and the within sequence covariance matrix W are computed as follows:

$$B/n = \frac{1}{m-1}\sum_{j=1}^{m}(\bar{\boldsymbol{u}}_j - \bar{\boldsymbol{u}})(\bar{\boldsymbol{u}}_j - \bar{\boldsymbol{u}})^T, \tag{III.20}$$

$$W = \frac{1}{m(n-1)}\sum_{j=1}^{m}\sum_{i=1}^{n}(\boldsymbol{u}_{i,j} - \bar{\boldsymbol{u}}_j)(\boldsymbol{u}_{i,j} - \bar{\boldsymbol{u}}_j)^T, \tag{III.21}$$

where $\bar{\boldsymbol{u}}_j = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{u}_{i,j}$ is the sample mean of the j-th chain, and $\bar{\boldsymbol{u}} = \frac{1}{m}\sum_{j=1}\bar{\boldsymbol{u}}_j$ is the sample mean over all chains. The metric used for analyzing convergence is

$$\hat{R}^d = \frac{n-1}{n} + \left(\frac{m+1}{m}\right)\lambda_1, \tag{III.22}$$

where λ_1 is the largest eigenvalue of the symmetric, positive definite matrix $W^{-1}B/n$. At convergence, $\hat{R}^d$ tends to 1, and following the recommendations of Gelman and Brooks [19], one may consider that a sufficient number of MH iterations has been carried out if $\hat{R}^d < 1.1$ for the second half of the chains. From this, it follows that the set comprising the second half of all m chains constitutes a representative sample of the target distribution when $\hat{R}^d < 1.1$.

III.3. Application to wildland fire spread

III.3.1. Fire spread simulation

In this study, the open source fire spread solver ForeFire [45] is used. ForeFire uses a front-tracking technique to model the spread of the fire front, i.e., the interface between the burned surface and the rest of the simulation domain (not burned). The fire front is discretized by means of Lagrangian markers linked by a dynamic mesh. Each marker is advanced according to the surface geometry and the rate of spread (ROS). Contrary to discrete time simulation methods, ForeFire relies on a discrete event specification. Each marker is therefore advanced according to a given spatial increment and the time at which the marker will reach its next position is deduced from its speed, making the simulation method asynchronous. Advancing a marker in time is considered as an event.

Other events may lead to a new calculation of the future location and time advance of a marker, such as topology checks that determine whether the markers describe a proper burned surface and reshape the fire front when it is not the case.

In this study, the ROS is computed according to the empirical model of Rothermel [114], widely used in wildland fire simulation, and that contains numerous parameters already fitted and fixed through an analysis of a large set of laboratory experiments. The input variables of the ROS model that are subject to perturbation in the present study are m_c, the fuel moisture content of dead fuel, S_v the surface-volume ratio, ΔH, the heat content, σ_f, the fuel load, ρ_p, the particle density, h, the fuel bed depth (denoted as fuel height in the following section), and W_S, the "effective" wind speed in the direction of fire spread, denoted as $\boldsymbol{n}$.

Here, a few additional assumptions are made; first, that the mineral damping coefficient is equal to 1; second, that the fuel mineral content is negligible, which implies that the net initial fuel loading is equal to the fuel load σ_f. In addition, to account for the fact that the wind speed at mid-height of the flame is usually lower than that of the prediction, a 0.4 factor in ROS computations is applied to $\boldsymbol{W}$, the wind speed vector predicted by the meteorological model, so that $W_S = 0.4\ \boldsymbol{W}.\boldsymbol{n}$. The revised wind speed limit function proposed by Andrews et al., which is expressed by equation (21) in [10], is also applied.

The scheme used to advance the markers of the fire front is based on a first-order approximation. Considering a marker that is located at x_i at time t_i, with its normal to the front denoted as $\boldsymbol{n}_i$ (oriented toward the unburned area), its next location is determined by

$$x_{i+1} = x_i + \delta l\ \boldsymbol{n}_i, \tag{III.23}$$

and the advance in time depends on ROS_i, the ROS computed with the values of the environmental inputs at location x_i and time t_i, as follows:

$$t_{i+1} = t_i + \frac{\delta l}{ROS_i}. \tag{III.24}$$

III.3.2. Prior uncertainty in input data

The marginals of the prior distribution for the perturbation variables used in the present study are defined in Table III.1. It is assumed that the inputs are independent. The description of the perturbation variables is presented in a previous study [5], where the distributions were mostly truncated normal (see notably Table 2 in [5]). The value of the standard deviations was chosen based on expert knowledge and the support of the marginals followed values found in the scientific literature. In the present study, a wider support is chosen for some of the distributions, which may be substituted for uniform distributions. This choice results from initial efforts at calibration where evaluation was carried out on ensembles of simulations resulting from a few specific input distributions. Although somewhat arbitrary, this allowed to obtain better probabilistic scores for most of the fires studied.

The simulations may involve up to 13 burnable fuel types that are linked to the Corine Land Cover classification [42]. When the perturbation is "individual" in Table III.1, it means that one perturbation coefficient is sampled for each fuel type. Another difference compared to the previous study is that the prior and posterior input distributions do

Input	Unit	Perturbation	Distribution	Notes
Wind direction	$^{\circ}$	Additive	$\mathcal{N}(0, 60^2)$	Truncated to $[-180, 180]$
Wind speed norm	$\mathrm{m\,s^{-1}}$	Multiplicative	$\mathcal{LN}(0, (0.5 \log 3)^2)$	Truncated to $[1/3, 3]$
Dead fuel moisture		Multiplicative	$\mathcal{U}(0.4, 1.6)$	
Heat of combustion	$\mathrm{MJ\,kg^{-1}}$	Additive	$\mathcal{U}(-5, 5)$	
Particle density	$\mathrm{kg\,m^{-3}}$	Additive	$\mathcal{U}(-300, 300)$	
Fuel height	m	Multiplicative, individual	$\mathcal{U}(0.4, 1.6)$	
Fuel load	$\mathrm{kg\,m^{-2}}$	Multiplicative, individual	$\mathcal{U}(0.4, 1.6)$	
Surface-volume ratio	$\mathrm{m^{-1}}$	Multiplicative, individual	$\mathcal{U}(0.4, 1.6)$	
Direction from ignition point	$^{\circ}$	Additive	$\mathcal{U}(-180, 180)$	
Distance to ignition point	m	Additive	$\mathcal{U}(0, 1) \times \Delta_{\max}$	$\Delta_{\max} \in \{100, 500, 1000\}$
Time of fire start	min	Additive	$\mathcal{U}(-1, 1) \times \Delta_{\max}$	$\Delta_{\max} \in \{10, 15, 30, 60\}$
Time of fire end	min	Additive	$\mathcal{U}(-1, 1) \times \Delta_{\max}$	$\Delta_{\max} \in \{10, 60, 120, 180, 240\}$

Table III.1. – Prior probability distribution of the perturbations in the simulation inputs. Perturbations are independent and details regarding some of these sources of uncertainty can be found in [5].
For the first two inputs, the distribution is a truncated (log-)normal. For each of the last three inputs, a reduced variable in [0,1] or [-1, 1] is used in the calibration procedure, and when it comes to the simulations, it is multiplied by $\Delta_{\max}$, which depends on the fire case.

not include the fuel type transitions, which means that no transition is applied when running the simulations to compute the emulator, and that transitions are not part of the MH algorithm. However, they are still used in the generation of the ensembles, by generating the transitions independently from the scalar perturbations.

For one simulation, $d = 48$ perturbation coefficients are generated. For wind direction and wind speed norm, the distributions are truncated, and the "Distribution" column corresponds to the distribution before truncation is applied. The main reason behind the choice of distributions with finite support is to avoid sampling extreme values that may be unrealistic or that could lead to non-physical parameter values (e.g. negative fuel load).

Uncertainty in the location of the ignition point is specified by the perturbation parameters "direction from ignition point" and "distance to ignition point". To sample a perturbed ignition point, one may first sample a direction and select the new ignition point at an independently sampled distance from the reference ignition point in this direction. The maximum distance $\Delta_{\max}$ depends on the fire case: the perturbed ignition point may therefore be sampled within a radius ranging from 100 m to 1 km around the reference. Similarly, the maximum perturbation $\Delta_{\max}$ for time of fire start and time of fire end depends on the fire case. This varying uncertainty is due to the information available regarding each fire. For the last three variables, $\Delta_{\max}$ is specific to each fire case. However, for the calibration, there is only one "reduced" variable for each of these three inputs, whose support is either $[0, 1]$ or $[-1, 1]$. The actual perturbation used to run the fire spread simulations for a given fire case is obtained from the reduced

variable after multiplication by $\Delta_{\max}$.

III.3.3. Application to seven Corsican wildland fires

The emulation and calibration procedure is applied to $K = 7$ fires that occurred in Corsica in 2017-2018 and that are presented in [5]. The ensembles obtained in this previous study are also used for comparison and referred to as "reference ensembles" hereafter.

An emulator with a training sample of size 4000 is built and evaluated with a test sample of size 2000. For the computation of the Wasserstein distance following equation (III.5), one may consider an orthogonal uniform grid that covers the burned surface, and identify which points belong to the burned surface to approximate the PDF by a sum of Dirac delta distributions at these points. The spatial resolution of the grid depends on the size of the surface because the computational cost increases drastically with the number of points. For the small burned surfaces, the resolution is approximately 20 m, whereas for the largest ones, it is approximately 80 m. Computations are carried out with the package *ot* from the Python toolbox POT [50].

Then, the MH algorithm is applied to several distributions with different values of β ranging in $\{\frac{1}{10}, \frac{1}{7}, \frac{1}{4}, \frac{1}{2}, 1, 2\}$. For each value of β, $n = 150000$ iterations are carried out for $m = 8$ chains. The instrumental distribution described by $q(\boldsymbol{u}|\boldsymbol{v})$ is a product of independent univariate truncated normal distributions. Before truncation, the k-th normal distribution is centered on the k-th component of $\boldsymbol{v}$ and has a standard deviation equal to a twentieth of the width of the perturbation range. The distribution is then truncated to the perturbation range. For the perturbation of wind speed norm, it is the logarithm of the perturbation that follows a truncated normal distribution.

By taking the latter half of the chains obtained with the MH algorithm, this leads to samples of size $m \times n/2 = 600000$ for each value of β. Based on these empirical distributions, ensembles of wildland fire simulations are generated for the seven fire cases, referred to as "calibrated ensembles" hereafter. The size of a calibrated ensemble ranges between 2000 and 10000. The evaluation domain is the same as for the reference ensembles, but contrary to the previous study, no computational time limit is applied to the simulations. The ensemble generation procedure is also carried out based on the prior distributions, which leads to "prior ensembles".

The ensembles are evaluated following the approach presented in [5]. Here, the definition of some of these evaluation tools is provided. Let us consider an evaluation domain $\mathcal{X}$ large enough to contain $\mathcal{S}_{\text{obs}}$, the observed burned surface, and a regular grid on $\mathcal{X}$ that comprises N points $x_1, ..., x_N$. Now defining $p_i = \mathbb{P}[x_i \in \mathcal{S}_{\boldsymbol{U}}]$ and $o_i = 1$ if $x_i \in \mathcal{S}_{\text{obs}}, 0$ otherwise, the Brier score (BS) is defined as follows:

$$BS = \frac{1}{N} \sum_{i=1}^{N} (o_i - p_i)^2. \tag{III.25}$$

This score ranges between 0 and 1 and is negatively oriented. Among the ensembles that forecast a constant probability, the one with the lowest Brier score is obtained with the probability $p_c = \frac{1}{N} \sum_{i=1}^{N} o_i$, and the Brier score of this ensemble is $BS_c = p_c(1-p_c)$.

The Brier skill score (BSS) is defined as follows:

$$\mathrm{BSS} = 1 - \frac{BS}{BS_c}, \tag{III.26}$$

which is positively oriented. When there are several fires, one may summarize the Brier scores by their mean. The corresponding value of BS_c is obtained with the mean of the p_c. With these two global Brier scores, the global Brier skill score is defined as in equation (III.26). Note that in practice, the probabilities p_i are estimated with a Monte Carlo method, so we only have an estimate of the true values of BS and BSS. One may estimate the standard deviation of the estimator with bootstrap [39], where the probabilities p_i are re-estimated by sampling with replacement among the simulated burned surfaces. With a large enough set of bootstrap samples, we obtain σ^b_{BSS}, an estimator of the standard deviation of BSS. Due to the regularity of BS, it can be shown that the estimators of both BS and BSS are asymptotically normal and that the bootstrap estimation is consistent. Therefore, it is possible to provide an approximate confidence interval at level $1-\alpha$ for BSS based on σ^b_{BSS} and the quantile of the standard normal distribution for probability $1-\alpha/2$ (details and justification are given in appendix III.6). Here, bootstrap resampling was carried out to provide 95% confidence intervals of BSS.

To summarize the information given by the other evaluation tools on several fires (rank histogram, reliability and sharpness diagrams), the contribution of each fire case is weighted by the size of the evaluation domain before summing the contributions of the seven fires (otherwise, fires with the largest simulation domains would have the most influence). For the rank histogram, the values of the rank are normalized because of the varying ensemble size.

III.4. Results

III.4.1. Emulation

The prediction of the emulator on the test set is displayed in Figure III.2. The emulator shows good approximation with MAE = 0.73 and $Q_2 = 95.3\%$. The use of the logarithm allowed slightly better approximation than with direct emulation of the energy function (MAE = 0.97 and $Q_2 = 93.2\%$ without using the logarithm). In the latter case, negative values were obtained at 4 points of the test sample, whereas the use of the logarithm ensured the prediction of positive values. The emulated energy function is computed in approximately 0.6 s, therefore the 150000 iterations of the MH algorithm can be carried out in a bit more than a day.

III.4.2. Calibrated distribution

For all values of β, the convergence diagnosis is positive with $\hat{R}^d$ ranging between 1.035 and 1.045 based on the last 75000 values of the chains. The proportion of accepted values ranges between 65% and 71% and decreases with β. The 1D mixing plots of individual chains does not indicate that the chain is stuck in local regions, as illustrated in Figure III.3 that displays the m_c perturbation component of the first of the eight

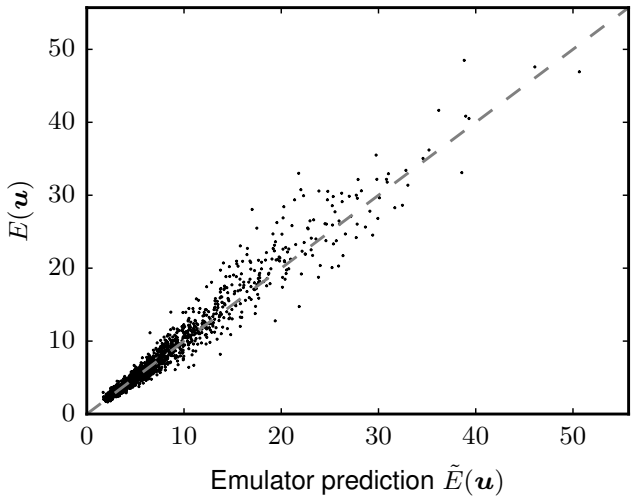


Figure III.2. – Error of the emulator on the test set. The gray line indicates a perfect prediction.

chains obtained with $\beta = 1/2$. For all $m = 8$ chains and all the $d = 48$ perturbations, the autocorrelation coefficient of the second half of the chain was computed with step k varying between 0 and 500. For any value of β, the maximum of the autocorrelation in absolute value tends to decrease with the iteration step k to reach approximately 1.1 at $k \approx 400$, an example of such plot is displayed for $\beta = 1/2$ in Figure III.4. Autocorrelation is fairly high, but chain thinning was not used.

The correlation between different input variables of the calibrated distributions is low, with a maximum absolute value of 0.1046 (for $\beta = 2$). Also, most marginals show very little difference between the prior and the calibrated distribution. Although increasing β leads to more significant difference from the prior distribution, some variables seem to be unaffected by the calibration, namely most individual perturbations in fuel parameters (height, load and surface-volume ratio) and perturbation in the ignition point and the time of fire start.

Histograms of some marginals of the calibrated distribution are shown for $\beta = 1/2$ in Figure III.5. Only the distributions that lead to the most significant change compared to the prior distribution are shown, namely perturbations in ΔH, ρ_p, the three individual fuel parameters (σ_f, S_v, h) specific to fuel type 311 (Broad-leaved forest), wind speed norm and direction, m_c, and time of fire end. Similar results are obtained for other values of β, with an increasing deviation from the prior when β increases. The variables with the most deviation are the same, except for the perturbation in the fuel load σ_f of fuel type 311, which does not differ much with higher β. Meanwhile, for $\beta = 1$ and $\beta = 2$, the marginals of the other individual fuel parameters S_v and h of fuel types 321 (Natural grassland) and 323 (Sclerophyllous vegetation) also show a larger deviation from those of the prior distribution, but not as much as for fuel type 311.

The variables whose marginals change the most after calibration are presumably those that have the most influence on the (emulated) energy function and therefore on

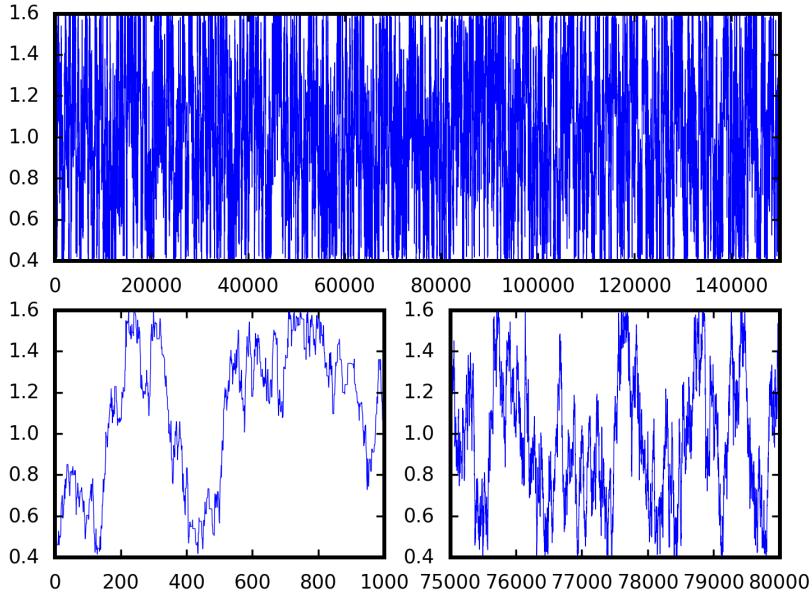


Figure III.3. – Mixing plot obtained for the m_c component of the first chain ran for $\beta = 1/2$. Top: whole chain; bottom left: first 1000 iterations; bottom right: first 5000 iterations of the second half of the chain.

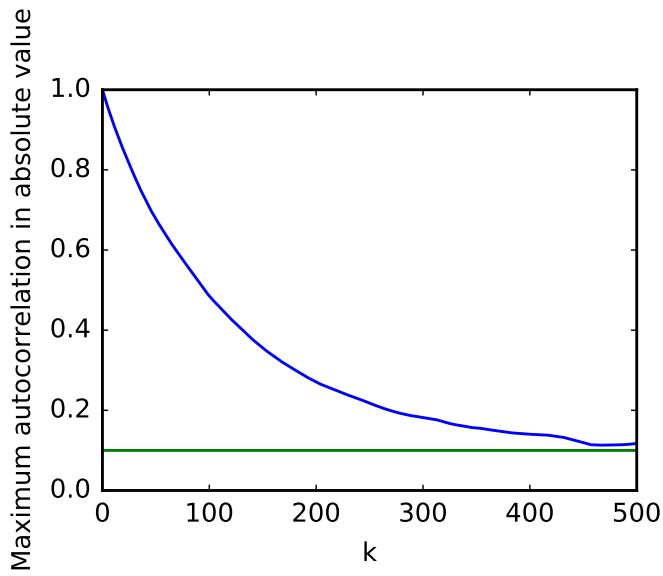
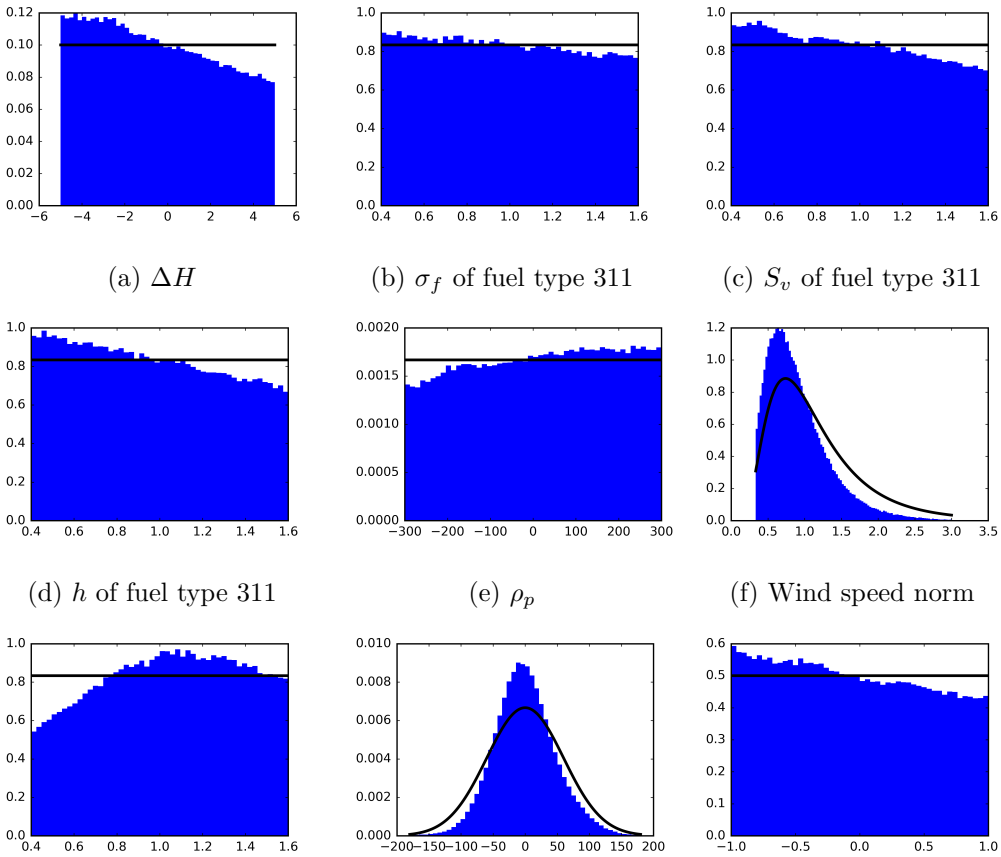


Figure III.4. – Maximum of autocorrelation in absolute value in the latter half of the chains obtained for $\beta = 1/2$ as a function of the iteration step k. The green horizontal line indicates a correlation coefficient of 0.1.



(a) ΔH (b) σ_f of fuel type 311 (c) S_v of fuel type 311

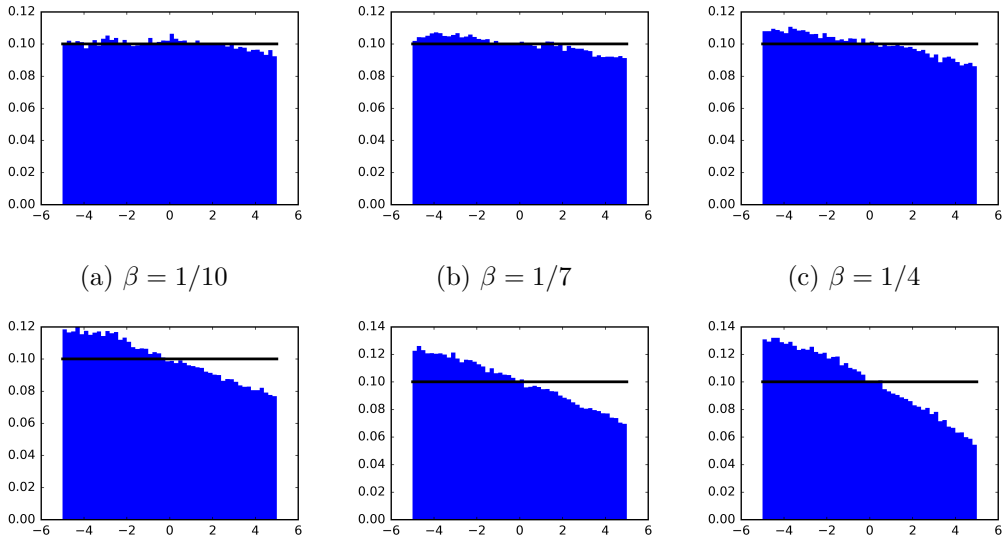
(d) h of fuel type 311 (e) ρ_p (f) Wind speed norm

(g) m_c (h) Wind direction (i) Time of fire end

Figure III.5. – Histograms of some marginals of the calibrated distribution obtained with $\beta = 1/2$. The histograms are normalized to allow comparison with the PDFs of the prior distribution (bold black line). Only the perturbations whose distributions are most different from that of the priors are represented.

the simulations. The ROS is proportional to ΔH and is highly sensitive to wind speed, and the new distributions for these inputs favor a reduced ROS, which gives more probability to smaller simulated burned areas. This tendency to favor a reduced ROS increases with β as can be seen in Figures III.6 and III.7. The calibrated distribution of the perturbation of the time of fire end has the same effect as it favors shorter fire duration. The maximum perturbations of the time of fire start are lower than for the time of fire end for all fire cases. This difference in uncertainty is probably the reason why the marginal distribution for time of fire end is more affected by the calibration than that for time of fire start, although both inputs have an influence on the duration of the simulated fire.

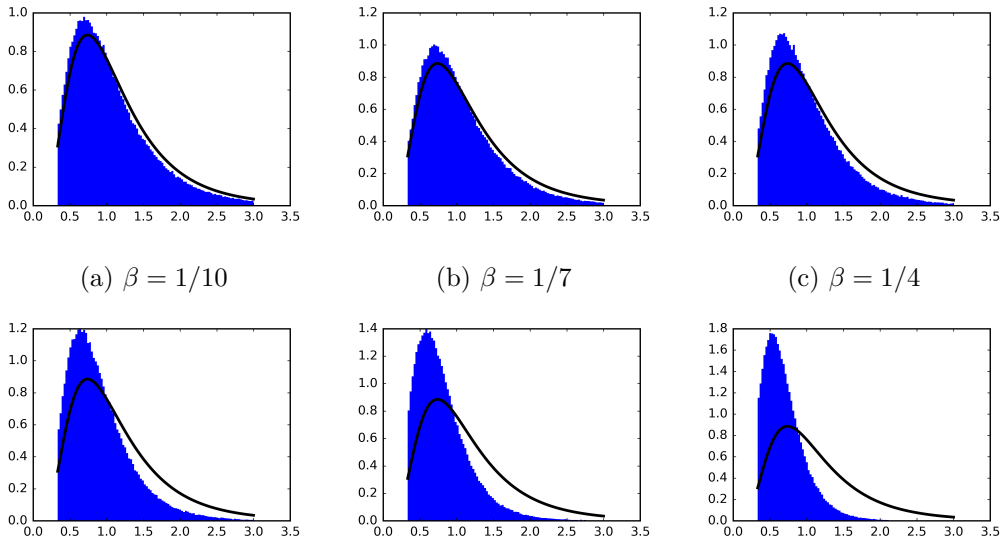
Similarly, it can be assumed that the simulated burned surface is less sensitive to the location of the ignition point compared to the other sources of uncertainty. Individual fuel parameters influence the ROS but only in some regions of the simulation domain,



(a) $\beta = 1/10$ (b) $\beta = 1/7$ (c) $\beta = 1/4$

(d) $\beta = 1/2$ (e) $\beta = 1$ (f) $\beta = 2$

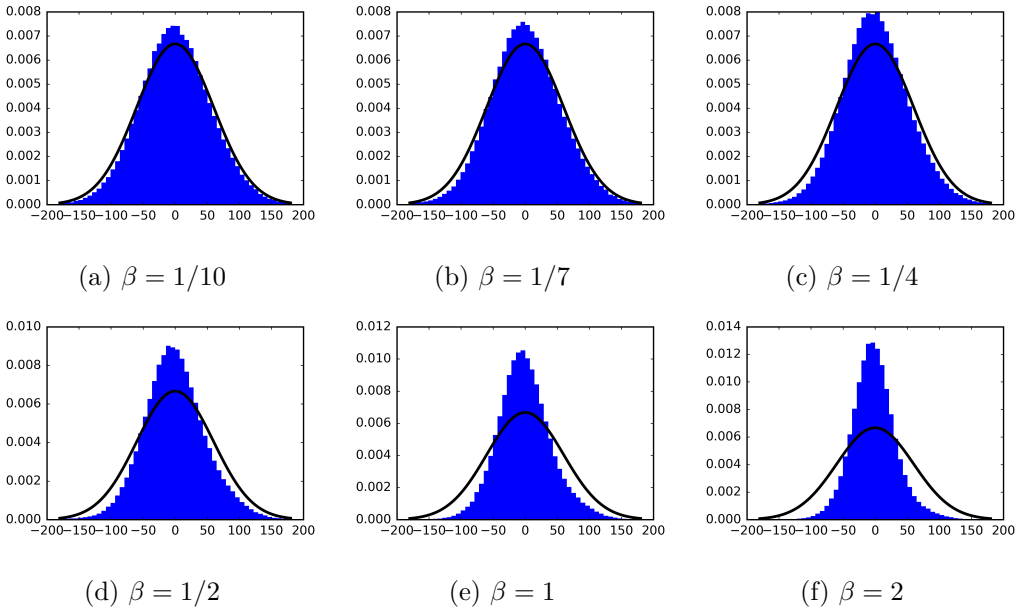
Figure III.6. – Marginal calibrated distribution of ΔH for different values of β. The bold black line indicates the PDF of the prior distribution.



(a) $\beta = 1/10$ (b) $\beta = 1/7$ (c) $\beta = 1/4$

(d) $\beta = 1/2$ (e) $\beta = 1$ (f) $\beta = 2$

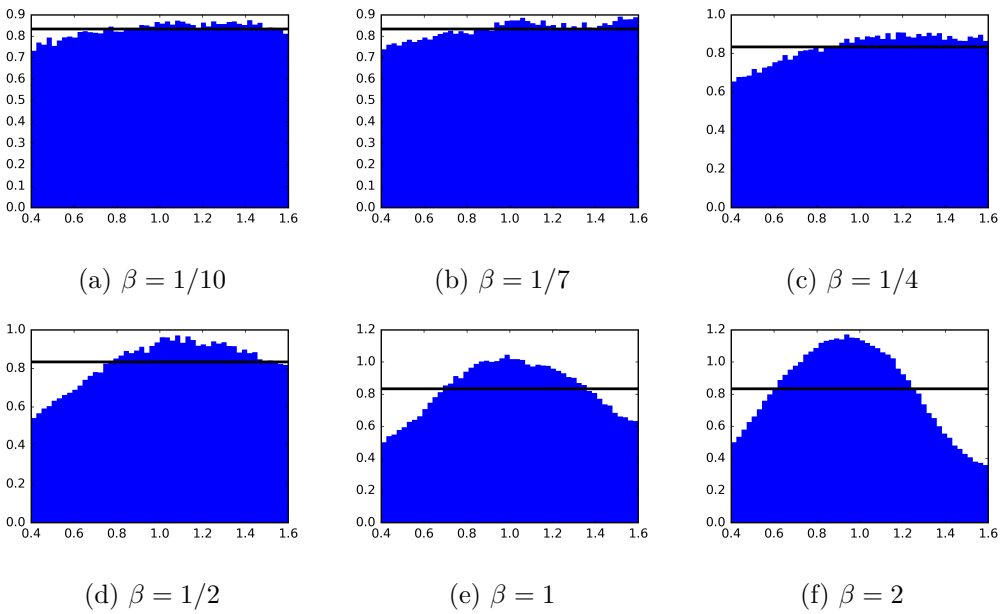
Figure III.7. – Marginal calibrated distribution of wind speed norm for different values of β. The bold black line indicates the PDF of the prior distribution.



(a) $\beta = 1/10$ (b) $\beta = 1/7$ (c) $\beta = 1/4$

(d) $\beta = 1/2$ (e) $\beta = 1$ (f) $\beta = 2$

Figure III.8. – Marginal calibrated distribution of wind direction for different values of β. The bold black line indicates the PDF of the prior distribution.



(a) $\beta = 1/10$ (b) $\beta = 1/7$ (c) $\beta = 1/4$

(d) $\beta = 1/2$ (e) $\beta = 1$ (f) $\beta = 2$

Figure III.9. – Marginal calibrated distribution of the fuel moisture content m_c for different values of β. The bold black line indicates the PDF of the prior distribution.

which comprises two to six major fuel types depending on the fire case. For instance, fuel type 311 is involved in four of the seven fires (Chiatra, Sant'Andrea di Cotone, Nonza and Ghisoni) so its fuel parameters are more likely to be influential than these of fuel type 322 (Moors and heathland), which is almost only involved in Ghisoni fire case. For all values of β, the calibrated distribution of the perturbation of ρ_p, the particle density, favors high values. The mean of the perturbation of wind direction remains close to 0 regardless of the value of β, which is an indication of unbiased meteorological data, and its standard deviation decreases with β (see Figure III.8). Interestingly, the distribution of the perturbation of the fuel moisture m_c at high values of β is dome-shaped, with a mean close to that of the prior (see Figure III.9).

III.4.3. Ensemble evaluation

The Brier skill scores of the calibrated, prior and reference ensembles for all seven fire cases are presented in Table III.2. The size of the reference ensembles is 500 members for all fires other than Ghisoni (324 members) which is smaller than that of the prior and calibrated ensembles. With 10000 bootstrap samples for each ensemble, approximate 95% confidence intervals were estimated and reported in Table III.3.

Fire name (ensemble size)	Reference	Prior	$\beta = 1/10$	$\beta = 1/7$	$\beta = 1/4$	$\beta = 1/2$	$\beta = 1$	$\beta = 2$
Calenzana (10000)	0.269	0.291	0.304	0.308	0.309	**0.314**	0.308	0.284
Chiatra (10000)	0.324	**0.386**	0.385	0.379	0.371	0.358	0.342	0.325
Ville di Paraso (2000)	0.021	0.168	0.179	0.182	**0.189**	0.188	0.176	0.168
Sant'Andrea di Cotone (5000)	0.190	0.408	0.429	0.442	0.454	0.468	0.485	**0.494**
Olmeta di Tuda (2000)	0.063	0.187	0.230	0.219	0.278	0.322	0.378	**0.451**
Nonza (4000)	-5.323	-3.089	-3.124	-3.133	-3.124	**-3.044**	-3.057	-3.053
Ghisoni (2000)	-9.986	-10.273	-9.831	-9.851	-9.333	-9.018	-8.638	**-8.332**
Global	-1.609	-1.332	-1.266	-1.269	-1.191	-1.135	-1.080	**-1.033**

Table III.2. – Brier skill score of the reference, prior and calibrated ensembles for the seven fire cases individually and globally (last line). For a given fire case, the best value of the BSS is shown in bold. The ensemble size applies to all ensembles except the reference ensembles that are of size 500 for all fires other than Ghisoni (324).

There is considerable improvement of the BSS for almost all fire cases with the prior ensembles compared to the reference ensembles. It is also the case with the calibrated ensembles. However, although most calibrated ensembles have better BSS than the reference ensembles, there is not always an improvement compared to the prior ensembles. For the fires of Sant'Andrea di Cotone, Olmeta di Tuda and Ghisoni, there is an overall increase of BSS with β. For the fire of Chiatra, it is the opposite: the Brier skill score decreases with β. For the fires of Calenzana and Ville di Paraso, the variation of BSS with β is relatively low, and there is probably an optimum for intermediate values of β. There might also be an optimum at intermediate β for the fire of Nonza, but the confidence intervals are too large to support this assumption.

Fire name (ensemble size)	Reference	Prior	$\beta = 1/10$	$\beta = 1/7$	$\beta = 1/4$	$\beta = 1/2$	$\beta = 1$	$\beta = 2$
Calenzana (10000)	[0.238, 0.300]	[0.284, 0.297]	[0.298, 0.310]	[0.301, 0.314]	[0.303, 0.315]	[0.308, 0.320]	[0.303, 0.314]	[0.278, 0.290]
Chiatra (10000)	[0.316, 0.332]	[0.383, 0.389]	[0.382, 0.389]	[0.376, 0.383]	[0.367, 0.375]	[0.354, 0.362]	[0.338, 0.346]	[0.322, 0.329]
Ville di Paraso (2000)	[-0.001, 0.042]	[0.156, 0.179]	[0.170, 0.188]	[0.173, 0.192]	[0.181, 0.197]	[0.181, 0.194]	[0.170, 0.182]	[0.162, 0.174]
Sant'Andrea di Cotone (5000)	[0.159, 0.306]	[0.398, 0.419]	[0.419, 0.438]	[0.433, 0.451]	[0.446, 0.462]	[0.461, 0.475]	[0.480, 0.491]	[0.489, 0.499]
Olmeta di Tuda (2000)	[0.011, 0.115]	[0.161, 0.214]	[0.204, 0.256]	[0.193, 0.245]	[0.254, 0.303]	[0.299, 0.346]	[0.356, 0.400]	[0.433, 0.470]
Nonza (4000)	[-5.458, -5.187]	[-3.166, -3.011]	[-3.206, -3.042]	[-3.213, -3.053]	[-3.203, -3.045]	[-3.125, -2.963]	[-3.136, -2.978]	[-3.131, -2.975]
Ghisoni (2000)	[-10.118, -9.854]	[-10.477, -10.069]	[-10.017, -9.645]	[-10.036, -9.665]	[-9.510, -9.156]	[-9.187, -8.850]	[-8.784, -8.492]	[-8.472, -8.192]
Global	[-1.635, -1.586]	[-1.359, -1.305]	[-1.291, -1.241]	[-1.294, -1.245]	[-1.215, -1.168]	[-1.158, -1.113]	[-1.100, -1.060]	[-1.053, -1.014]

Table III.3. – Bootstrap 95% confidence intervals of the BSS values from Table III.2. The ensemble size applies to all ensembles except for the reference ensemble whose size is 500 for all fires other than Ghisoni (324).

There is not an indisputable best value β for which the accuracy is optimal for all fire cases. For $\beta = 2$, we have the best BSS for three fire cases, but for three of the remaining four cases, the BSS is in the lowest among the calibrated ensembles. Still, the global BSS is the best for $\beta = 2$, because the increase in accuracy for some fires is more significant than that of the other calibrated ensembles, while the decrease of accuracy for the other fire cases is relatively low.

In order to determine which value of β might be the most appropriate, one may instead compare the overall ranking of the Brier skill scores. These rankings are reported in Table III.4. According to this method, it results that the best distribution is the one corresponding to $\beta = 1/2$.

Fire name (ensemble size)	Reference	Prior	$\beta = 1/10$	$\beta = 1/7$	$\beta = 1/4$	$\beta = 1/2$	$\beta = 1$	$\beta = 2$
Calenzana (10000)	8	6	5	3	2	1	3	7
Chiatra (10000)	8	1	2	3	4	5	6	7
Ville di Paraso (2000)	8	6	4	3	1	2	5	6
Sant'Andrea di Cotone (5000)	8	7	6	5	4	3	2	1
Olmeta di Tuda (2000)	8	7	5	6	4	3	2	1
Nonza (4000)	8	4	5	7	5	1	3	2
Ghisoni (2000)	7	8	5	6	4	3	2	1
Sum	55	39	32	33	24	18	23	25
Overall ranking	8	7	5	6	3	1	2	4

Table III.4. – Ranking of the Brier skill scores in Table III.2. The ensemble size applies to all ensembles except the reference ensembles that are of size 500 for all fires other than Ghisoni (324).

In Figure III.10 the global (i.e., "average" of the seven fires) rank histogram is shown for four distributions: reference, prior and two calibrated ($\beta = 1/4$ and $\beta = 2$). All distributions lead to higher bars on the left of the rank histogram, which is characteristic of a tendency to overpredict the burn probabilities. Calibration limits overprediction and leads to a histogram that is closer to the ideal uniform histogram.

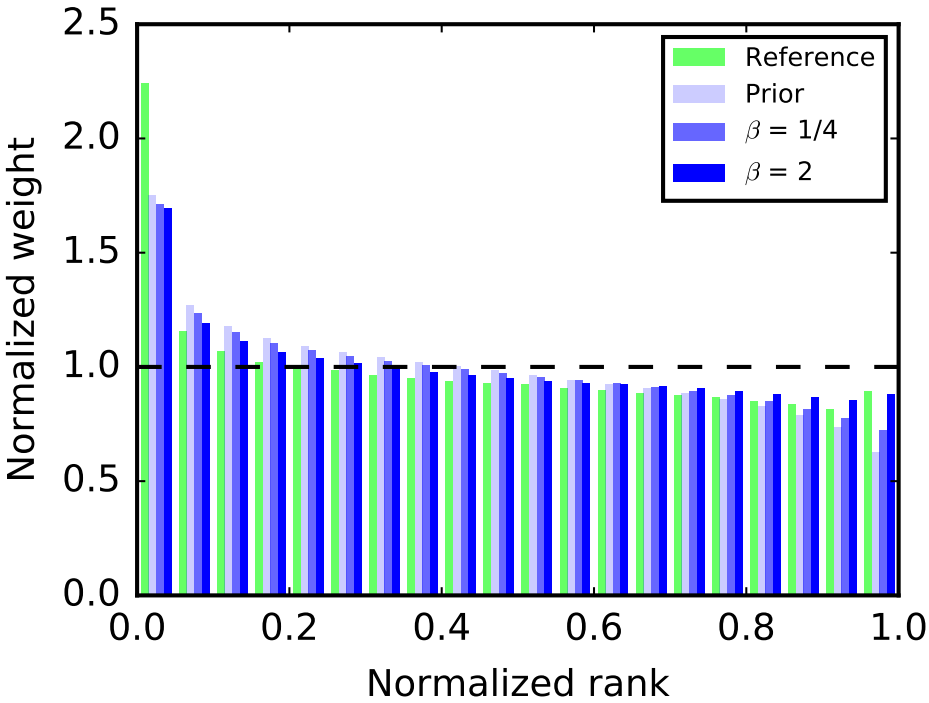
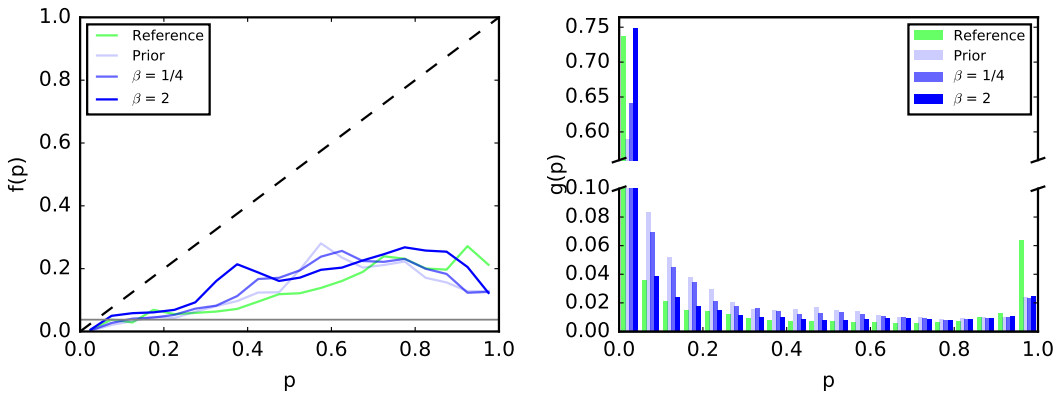
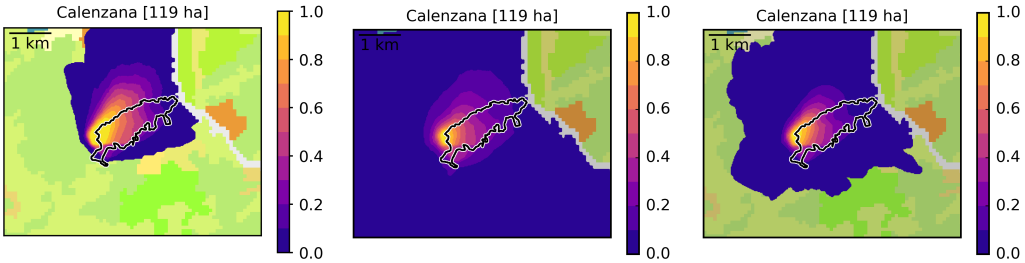


Figure III.10. – Global rank histogram for several ensembles. The dotted line indicates the ideal histogram of a consistent ensemble.



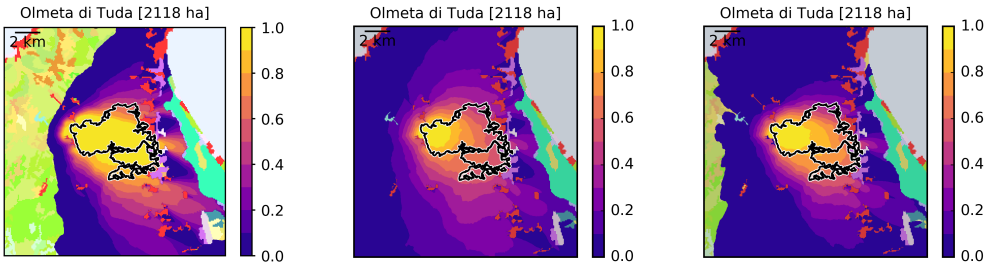
(a) Reliability (b) Sharpness

Figure III.11. – Reliability and sharpness diagrams for several ensembles. (a) The dotted black line indicates a reliable prediction system and the solid gray line the optimal probability p_c of an ensemble with constant predicted probability.



(a) Reference (b) Prior (c) Calibrated, $\beta = 2$

Figure III.12. – Burn probability maps of Calenzana fire for several ensembles. The colorbar indicates the predicted burn probability ; the black and white line is the contour of the observed burned surface ; background colors represent the Corine Land Cover data [42]



(a) Reference (b) Prior (c) Calibrated, $\beta = 2$

Figure III.13. – Same legend as Figure III.12 but for the fire of Olmeta di Tuda

The reliability and sharpness diagrams of these distributions are shown in Figure III.11. Although the calibrated distribution lead to overall slightly more reliable ensembles, there is not an improvement for all predicted probabilities p.

The effect of calibration can also be investigated on individual fires. The maps of burn probability of the fires of Calenzana and Olmeta di Tuda are represented in Figures III.12 and III.13, respectively, with a focus on the reference and prior ensembles as well as the calibrated ensemble corresponding to $\beta = 2$. Non-zero probabilities reach further locations in the prior ensembles than they do in the reference, which is due to the larger uncertainty in the input parameters. Calibration limits this "extension" of the probability field because the calibrated probability distributions mostly favor lower ROS and because the uncertainty in wind direction is much lower. This limitation tends to increase accuracy for some fires, notably that of Ghisoni where the BSS significantly increases with β (from -10.273 for the prior ensemble to -8.332 for $\beta = 2$). The fire of Ghisoni is particularly difficult to predict with simulations because the fire had several starts and spread slowly during several days. As expected, this fact is not altered even after calibration, since accuracy remains poor in spite of its increase.

III.5. Discussion and conclusions

The proposed approach led to the generation of calibrated ensembles whose input distributions are defined by a posterior PDF with a pseudo-likelihood function that involves the Wasserstein distance between simulated and observed burned surfaces of several fire cases. Due to the high dimension and the computational requirements of the pseudo-likelihood function, a Gaussian process emulator was built to obtain a sample of the calibrated input distribution with a MCMC algorithm in about one day of computation on 8 CPU cores.

The emulation with Gaussian process of the pseudo-likelihood of the posterior distribution shows a good accuracy ($Q_2 > 95\%$), and mostly led to calibrated distributions that favored lower ROS and lower uncertainty in wind direction. It resulted in fire spread ensembles with overall lower fire spread and the resulting burn probabilities were less dispersed. Globally on the seven fires studied, the prior ensemble had a tendency to overpredict burn probability. This issue was less and less significant in the calibrated ensembles with increasing β, the "weight" of the pseudo-likelihood function against the prior distribution. The calibration was successful in modifying the probability distribution of the input so that the fire spread predictions have better overall accuracy.

Calibrated distributions with higher values of β were not investigated, but it is reasonable to assume that increasing β will lead to distributions that favor even lower ROS. Although reducing overprediction is a desirable consequence, an adverse effect is that underprediction will become more significant for fires where the prior ensemble already underpredicts burn probability, which was the case for the fire of Chiatra where calibration led to lower accuracy. In the simulations, the firefighting actions are not modeled, so overpredicting of the burn probabilities is preferable to underpredicting. Moreover, underprediction might result in an operationally non acceptable result with areas that might be burned while this is not simulated by any ensemble member.

This raises the question of the choice of the pseudo-likelihood function. It relies on the parameter β to adjust its weight, but the choice of this value is rather arbitrary. Clearly, low values of β will lead to a distribution that is very close to the prior. The rankings of the BSS for each value of β show a best overall for $\beta = 1/2$ that does not result in the best global BSS, but instead a very good one for most fires.

Ideally, β should be representative of the error between simulation and observation: higher error should imply lower β. Yet, the energy function for the comparison between observation and simulation relies on a weighted sum of (squared) Wasserstein distances, whose values are difficult to interpret. The Wasserstein distance remains a suitable metric for surface comparison that allowed us to increase the overall accuracy of the ensemble predictions. Hopefully, given appropriate probabilistic assumptions on the error between observation and simulation, a likelihood function that makes use of the Wasserstein distance could be derived, but modeling this error is not trivial since the model output is a burned surface. Similarly, a perspective is to take into account other sources of uncertainty in the calibration procedure, notably model error, but it is not straightforward due to the nature of model output.

Another open question is the choice of the fire cases used for calibration. Here, all large fires for one season and one region were chosen, which still amounts to a low

number of fires. Even so, all of them were used for calibration and an overall increase in accuracy was obtained. However, there is no guarantee that there will still be an overall improvement if other fires are included and this calibration might be considered valid only for this season and this region. Ideally, there should be a sufficient number of fires that constitute the training basis of the calibration to ensure its application in a wide range of conditions. The evaluation of the ensembles could be carried out not only on the fires used for training but also on other fires that play the role of test sample. More fires in the training sample would provide more information, which should limit over-fitting and potentially lead to calibrated input distributions that are more different from the prior.

Overall, a promising mathematical method was proposed to calibrate the probabilistic predictions of wildland fire spread. Improving prediction accuracy is crucial especially in the field of wildland fires where human lives, infrastructures and ecosystems are endangered. Several points in the method that could be the subject of further work were underlined. Main research perspective is now to combine these calibrated ensembles with models for probability of ignition and values at stake to assess next day wildfire risk, which is relevant to fire managers, and help in the decision of firefighting actions and fire prevention planning.

Acknowledgments

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

III.6. Appendix: Bootstrap confidence interval for the Brier skill score

In this section, justification is given for the use of confidence intervals for the Brier skill score obtained by bootstrap as explained in Section III.3.3. As mentioned previously, let us consider an evaluation domain $\mathcal{X}$ that encompasses $\mathcal{S}_{\text{obs}}$, the observed burned surface, and consider a regular grid on $\mathcal{X}$ that comprises N points $x_1, ..., x_N$. We define $p_i = \mathbb{P}[x_i \in \mathcal{S}_{\boldsymbol{U}}]$, where $\mathcal{S}_{\boldsymbol{U}}$ is the "probabilistic" burned surface corresponding to simulation based on the random input vector $\boldsymbol{U}$. Now defining $o_i = 1$ if $x_i \in \mathcal{S}_{\text{obs}}, 0$ otherwise, the Brier score (BS) is defined as follows:

$$BS = \frac{1}{N} \sum_{i=1}^{N} (o_i - p_i)^2. \tag{III.27}$$

In practice, the probabilities p_i are estimated with a Monte Carlo method. We have an ensemble of n independently sampled input vectors $\boldsymbol{u}_1, ..., \boldsymbol{u}_n$ and their corresponding burned surfaces $\mathcal{S}_{\boldsymbol{u}_1}, ..., \mathcal{S}_{\boldsymbol{u}_n}$. For all i, $\widehat{p_i} = \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}(x_i \in \mathcal{S}_{\boldsymbol{u}_j})$ is the estimate of the burn probability p_i with our finite ensemble, where $\mathbb{1}$ stands for the indicator function. The value of BS is unknown, so in practice, it is estimated by $\widehat{BS}$ which is computed following equation (III.27) using $\widehat{p_i}$ instead of p_i. We are interested in knowing how accurate the estimation $\widehat{BS}$ of BS is. One may estimate the standard deviation of the estimator with bootstrap [39]. For a bootstrap sample, the probabilities

p_i are re-estimated by sampling with replacement among the simulated burned surfaces $\mathcal{S}_{\boldsymbol{u}_1}, ..., \mathcal{S}_{\boldsymbol{u}_n}$. With a large enough set of bootstrap samples, we obtain σ^b_{BS}, an estimator of the standard deviation of $\widehat{BS}$. Provided that the estimator of the Brier score is asymptotically normal and that the bootstrap estimation is consistent, a confidence interval at level $1-\alpha$ for $\widehat{BS}$ can be approximated by:

$$\widehat{BS} \pm z_{1-\alpha/2}\, \sigma^b_{BS}, \tag{III.28}$$

where z_q is the quantile of the standard normal distribution for probability q. The Brier skill score being an affine function of BS, bootstrap can also be carried out to obtain an approximate confidence interval.

Now, let us follow some aspects of bootstrap theory highlighted by Shao [122] to prove the required properties for the confidence interval. The probability measure of the random vector $\boldsymbol{U}$ being denoted as F, one may write the following expression for the simulated burn probabilities: $p_i = \int_{\mathbb{R}^d} \mathbb{1}(x_i \in \mathcal{S}_{\boldsymbol{u}}) dF(\boldsymbol{u})$. The Brier score corresponding to the probability distribution F is denoted as $BS(F)$. First, let us compute the influence function L_F for distribution F. Denoting $\delta_{\boldsymbol{u}}$ as the Dirac delta distribution at point $\boldsymbol{u}$ and setting $\epsilon > 0$, one may write

$$BS((1-\epsilon)F + \epsilon\delta_{\boldsymbol{u}}) = \frac{1}{N}\sum_{i=1}^{N}(o_i - (1-\epsilon)p_i - \epsilon\mathbb{1}(x_i \in \mathcal{S}_{\boldsymbol{u}}))^2, \tag{III.29}$$

which yields

$$\begin{aligned}\frac{BS((1-\epsilon)F + \epsilon\delta_{\boldsymbol{u}}) - BS(F)}{\epsilon} &= \frac{1}{N}\sum_{i=1}^{N}(2o_i - (2-\epsilon)p_i + \epsilon\mathbb{1}(x_i \in \mathcal{S}_{\boldsymbol{u}})))(p_i - \mathbb{1}(x_i \in \mathcal{S}_{\boldsymbol{u}})) \\ &\underset{\epsilon\to 0}{\to} \frac{1}{N}\sum_{i=1}^{N}(2o_i - 2p_i)(p_i - \mathbb{1}(x_i \in \mathcal{S}_{\boldsymbol{u}}))) = L_F(\boldsymbol{u}).\end{aligned}$$

$L_F(\boldsymbol{u})$ is the influence function of the Brier score at point $\boldsymbol{u}$. As expected, we note that $\int L_F(\boldsymbol{u})dF(\boldsymbol{u}) = \frac{1}{N}\sum_{i=1}^{N}(2o_i - 2p_i)(p_i - p_i) = 0$. Now, let us prove the Fréchet differentiability of the Brier score. Using the influence function of the Brier score and defining another probability measure G for which $q_i = \int \mathbb{1}(x_i \in \mathcal{S}_{\boldsymbol{u}})dG(\boldsymbol{u})$, one may write

$$\begin{aligned}\left|BS(G) - BS(F) - \int L_F(\boldsymbol{u})d[G-F](\boldsymbol{u})\right| &= \left|\frac{1}{N}\sum_{i=1}^{N}(o_i - q_i)^2 - (o_i - p_i)^2 - (2o_i - 2p_i)(p_i - q_i)\right| \\ &= \left|\frac{1}{N}\sum_{i=1}^{N}(2o_i - q_i - p_i)(p_i - q_i) - (2o_i - 2p_i)(p_i - q_i)\right| \\ &= \frac{1}{N}\sum_{i=1}^{N}(q_i - p_i)^2\end{aligned}$$

For two probability measures P and Q defined on the same measurable space $(\Omega, \mathcal{F})$,

we define the total variation distance $||P-Q|| = \sup_{A\in\mathcal{F}} |P(A)-Q(A)|$. Since p_i and q_i are probabilities of the same event but according to the distributions F and G respectively, we have $|q_i - p_i| \leq ||G-F||$. It follows that

$$\begin{aligned}\frac{\left|BS(G) - BS(F) - \int L_F(\boldsymbol{u})d[G-F](\boldsymbol{u})\right|}{||G-F||} &\leq \frac{1}{N}\sum_{i=1}^{N}\frac{||G-F||^2}{||G-F||} \\ &= ||G-F|| \\ &\underset{||G-F||\to 0}{\longrightarrow} 0,\end{aligned}$$

and since $\int L_F(\boldsymbol{u})dF(\boldsymbol{u}) = 0$, we conclude that the Brier score is Fréchet differentiable. From [122], the estimator of the Brier score is asymptotically normal with the following distribution:

$$\sqrt{n}(\widehat{BS} - BS) \underset{n\to\infty}{\longrightarrow} \mathcal{N}\left(0, \int L_F(\boldsymbol{u})^2 dF(\boldsymbol{u})\right), \tag{III.30}$$

(provided that the variance in equation (III.30) is finite) and the bootstrap estimator is consistent. All that is left is to calculate

$$\begin{aligned}\int_{R^d} L_F(\boldsymbol{u})^2 dF(\boldsymbol{u}) &= \frac{1}{N^2}\int_{R^d}\sum_{i,j=1}^{N}(2o_i - 2p_i)(p_i - \mathbb{1}(x_i\in\mathcal{S}_{\boldsymbol{u}}))(2o_j - 2p_j)(p_j - \mathbb{1}(x_j\in\mathcal{S}_{\boldsymbol{u}})) \\ &= \frac{4}{N^2}\sum_{i,j=1}^{N}(o_i - p_i)(o_j - p_j)(p_ip_j - p_ip_j - p_jp_i + \int_{R^d}\mathbb{1}(x_i\in\mathcal{S}_{\boldsymbol{u}}\cap x_j\in\mathcal{S}_{\boldsymbol{u}})dF(\boldsymbol{u})) \\ &= \frac{4}{N^2}\sum_{i,j=1}^{N}(o_i - p_i)(o_j - p_j)(p_{ij} - p_ip_j),\end{aligned}$$

which is finite and where p_{ij} is the probability of having both locations x_i and x_j burned according to distribution F. The asymptotic normality of $\widehat{BS}$ is more explicitly given by

$$\sqrt{n}(\widehat{BS} - BS) \underset{n\to\infty}{\longrightarrow} \mathcal{N}\left(0, \frac{4}{N^2}\sum_{i,j=1}^{N}(o_i - p_i)(o_j - p_j)(p_{ij} - p_ip_j)\right). \tag{III.31}$$

IV. Emulation of wildland fire simulations with deep learning

In the present chapter, we focus on the "short-term" situation of a forecasting chain that requires to assess fire risk and danger for the coming day. The aim is to use the size of the simulated burned surface resulting from a fire starting at a given location at any given time and provide a map with high spatial and temporal resolution. The computational time required to carry out the massive amount of simulations for this task is very high, calling for the use of an emulator that can account for all the possible input conditions, including the input uncertainty that was quantified in the previous chapters. We present the design of a deep neural network that was trained to approximate the simulated fire size, and assess the performance of the resulting emulator.

Summary

IV.1. Introduction **104**
IV.2. Simulation of wildland fire spread **107**
IV.3. Emulation with deep learning **111**
IV.3.1. Design of experiments 111
IV.3.2. Neural network architecture 112
IV.3.3. Accuracy metrics and training strategy 115
IV.3.4. Extraction of the actual emulator 116
IV.3.5. Implementation 117
IV.4. Results and discussion **117**
IV.5. Conclusions **127**

The following published article is reproduced here:
F. Allaire, V. Mallet, and J.-B. Filippi. Emulation of wildland fire spread simulation using deep learning. *Neural Networks*, 141:184–198, 2021.

IV.1. Introduction

A major purpose of mathematical modeling and numerical simulation of wildland fire spread across land is to make relevant predictions and support long-term to short-term planning of firefighting actions. Fundamentally, fire spread implies heat transfer at scales of the centimeter, which is too computationally intensive to solve in operational conditions. Alternatively, fire spread modeling can be approached by solving a front-tracking problem where we focus on the propagation of the interface between burned and not burned areas, aka the *fire front*, over a 2D domain that represents the landscape. The growth of the burned surfaces from their initial state is governed by equations involving an model of rate of spread (ROS), that is to say the speed at which the flames advance, which is expressed as a function of local environmental parameters. Among such solvers, marker methods consist in discretizing the fire front by means of markers, which evolve in space and time according to an underlying fire behavior model that determines the speed at which the markers advance as well as other characteristics such as reaction intensity. Notable examples of simulators using this method include FARSITE [47], Prometheus [137], and Phoenix [134], that are commonly used in the US, Canada, and Australia, respectively. Alternatively, level-set methods (e.g. [89, 113]) can be used in simulations to track the fire front, and other approaches were proposed to model fire spread, such as cell-based simulations (e.g. [74]) that adopt a raster representation of the burned surface (see [131] for a detailed review of simulation models). Most of these approaches allow to simulate a fire propagating during more than an hour in a computational time of about a minute or less.

Physical models of wildland fire spread [129], more complex and typically including heat transfer conservation laws, equations describing combustion chemistry, etc. have also been developed. However, their use is generally limited to research purposes, because the computational time for simulations based on such models is prohibitory in an operational context, even more so for large wildfires that may burn during several hours or even days and scale up to thousands of hectares.

Evaluation of simulators of wildland fire spread can be carried out based on the comparison of an observed burned surfaces with its simulated counterpart. Several evaluation metrics have been proposed in wildland fire research, for instance relying on how much the two surfaces intersect (e.g. [37, 43]), on the distance between the vertices of the two fire perimeters [52, 36], or on information regarding the growth of the simulated and observed burned surfaces over time [43]. These metrics can be computed for observed fires, in which case the simulations can make use of data known in hindsight, such as fire suppression actions or observed weather variables (e.g. [35, 118]), or be simply based on data available at the time of fire start (e.g. [44]). ROS models, which may be involved in fire spread simulators, can also be evaluated based on observations obtained from laboratory or outdoor experiments or even from observed wildfires [29].

Given the complexity of wildland fires, there are significant uncertainty sources in modeling that may lead to considerable difficulties in determining the most appropriate decisions in an operational context [133]. In particular, for prediction purposes, there is considerable input uncertainty, which can refer to a range of possible values for a given model parameter or data source, possibly due to the use of weather forecasts or difficulty

to estimate a single "best" value. In control theory, parameter uncertainty can be expressed by means of uncertainty matrices in the model [145] to design robust control laws [145, 127]. The main goal of firefighters, once a wildfire has started (i.e. in a "crisis" situation), is to control the fire by means of suppression actions. These actions are difficult to model, therefore, predictions of wildland fire spread using simulators usually represent free spread (i.e. firefighting actions are not accounted for, but non-burnable areas such as water bodies may halt the progression of the fire front). Uncertainty can still be accounted for in such predictions, usually by propagating the probability distributions of the uncertain inputs following a Monte Carlo (MC) approach, resulting in an ensemble of fire spread simulations (e.g. [49, 96, 5]).

There are several possible applications of simulators of wildland fire spread in an operational context. As previously mentioned, in a crisis situation, they can help in predicting where the fire will spread and optimizing the fire suppression actions and evacuation. Prior to crisis situations, fire spread simulations are a major component of risk assessment frameworks to determine what areas have the highest potential to host a large incident. Wildland fire risk quantification generally involves models describing ignition probability, the probability for a given location to be burned, and the consequences on the objects affected by fire such as properties, timber production, as well as the consequences on human lives, wildlife habitats, etc. Several studies focused on fire risk mapping at the regional or country scale [99, 46, 82], where many fires are simulated to represent a fire season or year according to some probabilistic distribution of ignition and environmental conditions driving fire spread. This process may be repeated hundreds of thousands of times as part of a Monte Carlo method. The purpose of such maps is to help in land management through the reduction of areas at risk in the long-term, by setting up fire breaks and providing more firefighting resources such as reservoirs, etc.

Regarding short-term planning, information for the next day or hours about the areas where a fire is most likely to ignite and how far the resulting fire may spread can be very useful in order to know what locations should be monitored more closely and help in anticipating the distribution of firefighting resources (firefighters, trucks, ...) across the territory. For this purpose, one may focus on the quantification of "fire danger", a term that generally relates to the potential for ignition and spread of a fire at a given location. Traditional fire danger indices, widely used for decision support in an operational context, consist in a unitless value calculated essentially based on weather variables. A notable example of such an index is the Canadian Fire Weather Index (FWI [138]), used for generating maps of predicted fire danger covering Europe and the Mediterranean area as part of the European Forest Fire Information System (EFFIS) [1]. Making use of output burned surfaces of wildland fire spread simulations offers an interesting alternative for quantifying fire danger: for instance, the resulting fire size accounts not only for weather but also for terrain (i.e. elevation and vegetation, which are also influential factors in fire spread) and can be expressed in hectares, a more "concrete" quantity. Numerical simulations of wildland fire spread could be used to generate high-resolution maps of fire spread on the basis of weather forecasts; but this would require numerous computations for different ignition locations, and the constraint on computational time would be too demanding even for simulators used for

1. https://effis.jrc.ec.europa.eu/, last checked 2021.02.01

other operational purposes. As a rough estimate for the region considered in the present study, running one fire spread simulation with a computational of one minute for each hectare of land would amount to a computational time of 872,000 minutes (about 600 days) on a single processor, and even more if an ensemble of simulations is considered for each hectare; which would be too long even after distributing the computations on multiple processors.

In the aforementioned applications, and more particularly in short-term fire danger mapping, a promising approach to reduce computational time is to rely on an *emulator* (aka metamodel or surrogate model) to provide an approximation of some quantity of interest derived from the simulator's output. The idea is to focus on this quantity and compute it much faster with the emulator at the cost of some approximation error that should be as low as possible. Emulation may be used in situations when a fire spread model has high computational time and/or a lot of simulations or calls of a given function are required. Still, emulators are rarely used in wildland fire research even though their potential for reducing computational time of simulations appears desirable in this field. Examples include data assimilation of a fire front via polynomial chaos [113], sensitivity analysis through the computation of Sobol' indices related to the area and shape of the simulated burned surface with emulation by either Gaussian processes (GP) or generalized polynomial chaos [136], uncertainty quantification and computation of Sobol' indices regarding the ROS model of Rothermel [114] using high dimensional model representation methods [85], interpolation in a cell-based wildland fire spread simulator to quickly compute the values of correction factors in the relationship between advection velocity and spread angle on the basis of pre-computed values obtained in a few given configurations using a Radial Basis Function (RBF) approach [54]. Another example outside the scope of fire spread is the emulation of some outputs of a fire emission model with GP [76].

Machine learning (ML) is a very rapidly growing area of study whose methods have been used for prediction and decision-making purposes in a variety of scientific and technical fields [75]. As exemplified by recent reviews, there has been an increasing interest in application of ML to engineering risk assessment [66], emergency management [26], and to a wide variety of topics in wildland fire science [71] as well.

Neural networks, in particular, appear promising to take into account the complexity of wildland fire spread. For instance, an application involving emulation is proposed in [146] where a radial basis function neural network (RBFNN) is trained to emulate the similarity index between an observed burned surface and its simulated counterpart as a function of several ROS adjustment factors; a MC procedure is then applied to the emulator, providing parameter estimation of the adjustment factors for data assimilation of the simulated fire front. Other methods consist in using a convolutional neural network (CNN) as a surrogate for a wildland fire spread simulator to obtain a map of predicted burned areas [107, 68]. Data required to solve wildfire simulations have similarities with these involved in image processing as we are handling gridded maps of elevation and fuel parameters. As deep learning proved to be very appropriate to solve such image processing problems [78], it motivates the use of deep neural networks (DNNs) instead of traditional emulation techniques to approach emulation in wildland fire spread simulations.

In the present study, a method is proposed for the estimation of wildland fire spread in

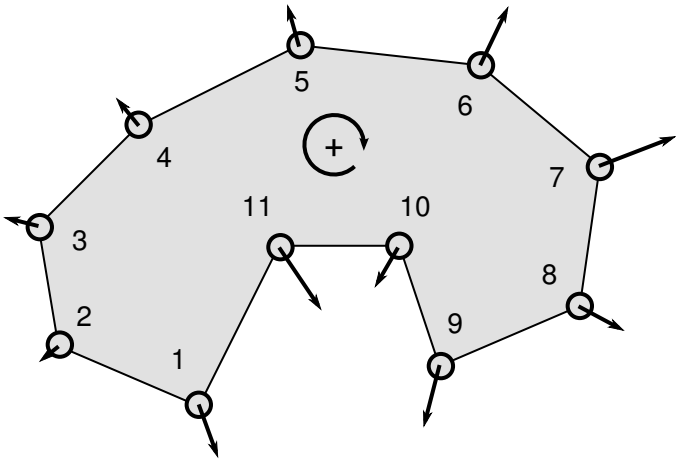


Figure IV.1. – Example of a small fire front discretization with ordered markers. The gray area in the center represents the burned surface and its interface with the unburned locations (white) represents the fire front whose vertices (markers) expand outward along the normal to the front.

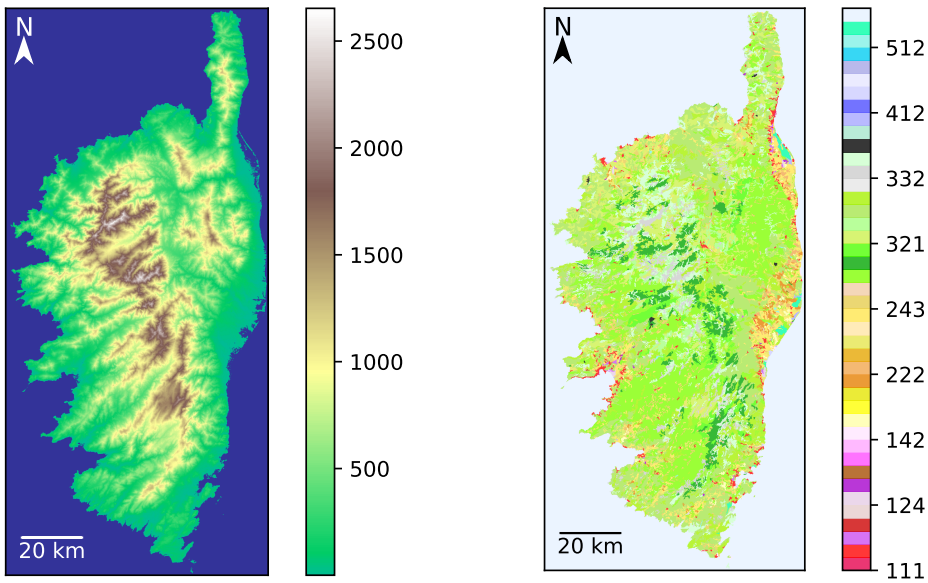
a wide variety of environmental conditions with potential for application to fire danger mapping. The quantity of interest is the burned surface area in hectares provided by a wildland fire simulator and the core of the method consists in the emulation of this output quantity using a deep neural network (DNN) with a hybrid architecture so that both 2D and scalar input data are processed by specific layers. The present study focuses on Corsica island but the method can be extended to other regions.

The numerical simulator of wildland fire spread that is used as basis of the present work is presented in Section IV.2 together with the characteristics of the simulations. The strategy used to obtain the emulator is described in Section IV.3 and the results are provided and discussed in Section IV.4. Conclusions of this work are summarized in Section IV.5, where some perspectives of application of the emulator and possible extensions to the method are also mentioned.

IV.2. Simulation of wildland fire spread

In the present study, wildland fire spread simulations are carried out with the numerical solver ForeFire [45]. ForeFire relies on a front-tracking method where the fire front is represented by Lagrangian markers that are linked to each other by a dynamic mesh. The interface is discretized using an ordered list of Lagrangian markers at given locations on the surface of the Earth. The interface is then tracked by advecting all these markers at the propagation velocity of the front, and by ensuring that the list of markers still holds an accurate representation of the interface. In this ordered list of markers, previous and next are defined by convention in the indirect direction as in Figure IV.1. The outward normal defines the direction of propagation from burning regions toward unburned regions. Although fronts are allowed to contain islands of unburned fuel, they must remain simple polygons (with no self-intersection).

A key aspect of the simulation is the computation of rate of spread (ROS), that is to say the speed at which the flames advance. Several ROS models were proposed in the scientific literature. The model used in present study is the model of Rothermel [114], which is commonly used by fire managers in the US. The ROS is expressed as a function of several environmental properties such as wind speed, terrain slope, fuel moisture content (FMC) and other fuel parameters characterizing the vegetation. A simulation mostly consists in the definition of an initial state of the fire front and the ROS is computed for the markers of the fire front based on underlying 2D fields from which environmental properties are determined. ForeFire relies on a discrete event approach where most computations deal with the determination of the time at which the markers will reach their next destination, this destination being defined by a fixed spatial increment in the outward normal. This discrete event approach includes other types of events such as changes in the values of the layers, notably wind speed and FMC, additions and removals of markers so that the fire front maintains a perimeter resolution in a given range during the simulation, and topology checks that may induce front merging to ensure that the front keeps a physical representation.



(a) Elevation field. (b) Land cover field.

Figure IV.2. – Data maps of Corsica used to describe the landscape in ForeFire simulations; their spatial resolution is approximately 80 m.
(a) Locations with an altitude of 0 m or less (mostly maritime waters) are represented in blue.
(b) The color scheme corresponds to the classification of the Corine Land Cover

The area of study is Corsica island, which is located south-east of France in the Mediterranean sea. For fire simulation on this domain, 2D fields of elevation and land use in raster format at approximately 80-m resolution are used. These two fields are represented in Figures IV.2a and IV.2b, respectively.

Input	Symbol	Unit	Type	Range	Constraint
Ignition point coordinates	(x, y)	m	Raw	Map of Corsica	Initial front in burnable area
Wind speed	(W_x, W_y)	$\mathrm{m\,s^{-1}}$	Raw	$[-35, 35]^2$	Euclidean norm ≤ 35
Fuel moisture content (dead fuel)	m_c		Raw	$[0.04, 0.3]$	
Heat of combustion perturbation	ΔH	$\mathrm{MJ\,kg^{-1}}$	Additive	$[-5, 5]$	
Particle density perturbation	ρ_d	$\mathrm{kg\,m^{-3}}$	Additive	$[-300, 300]$	
Fuel height perturbations	h	m	Multiplicative	$[0.4, 1.6]^{13}$	
Fuel load perturbations	σ_f	$\mathrm{kg\,m^{-2}}$	Multiplicative	$[0.4, 1.6]^{13}$	
Surface-volume ratio perturbations	S_v	$\mathrm{m^{-1}}$	Multiplicative	$[0.4, 1.6]^{13}$	

Table IV.1. – Variable scalar inputs in wildland fire spread simulations. In the case of perturbations, the symbol corresponds to the perturbed quantity, and the perturbation of this quantity can be either additive or multiplicative. The range indicates the boundaries of the domain of definition with two components for the wind and 13 components in the last three rows (one row per fuel type).
The intervals of variation account for uncertainty and variability of weather and ignition locations, as well as for uncertainty.

The land use field comes from Corine Land Cover data [42] coupled with data from the IGN (Institut Géographique National) product BD TOPO® for road and drainage networks. The elevation field is extracted from another IGN product: BD ALTI®, which has originally a 25-m resolution. A fuel parameterization is used to assign reference fuel parameters to each type of vegetation (referred to as "fuel type" in the following) in the land use data for ROS computations. Data used for simulation also include 2D fields of wind speed vectors at a resolution of 200 m that were pre-computed for average wind speed vectors with the mass conserving preconditioner from the atmospheric forecasting system Meso-NH [80] to account for orographic effects. By specifying an average input wind speed vector in the simulations, the underlying 2D wind field is simply obtained from the pre-computed fields corresponding to the closest mean speed vectors.

In the present study, a simulation is always that of a fire with free spread (firefighting actions are not accounted for, but non-burnable areas such as water bodies may halt the progression of the fire front) during one hour. Another fixed input in the simulations is the initial fire front, which is an octagon with a surface area of 0.45 ha, corresponding to an already-propagating fire, that must be located in areas classified as fuel (i.e. burnable vegetation) based on the land cover field.

Several inputs in the simulations may vary from a simulation to another. First are the coordinates of the center of the initial fire front, this point being referred to as the *ignition point*, that may be located in all fuel areas in Corsica. This "high-level" input is of major importance because it determines the location where the fire starts and the part of the spatial fields that will influence how the fire will spread. Next are the zonal and meridional coordinates of the "forcing" wind speed vector, in $\mathrm{m\,s^{-1}}$, that both vary in [-35, 35] on the condition that the wind speed norm be lower than $35\,\mathrm{m\,s^{-1}}$. The FMC of dead fuel varies between 0.04 and 0.3. In contrast to these "raw" inputs, the remaining ones are perturbation coefficients that are applied to reference values of some

fuel parameters. Perturbation in heat of combustion and particle density are additive and applied to a common reference value used for all fuel types, whereas perturbations in fuel height, fuel load or surface-volume ratio are multiplicative coefficients. For all the three latter parameters, each one of the 13 fuel types receives a specific perturbation coefficient. This amounts to 46 variable inputs in the simulations, whose information is summarized in Table IV.1, including the range of each variable.

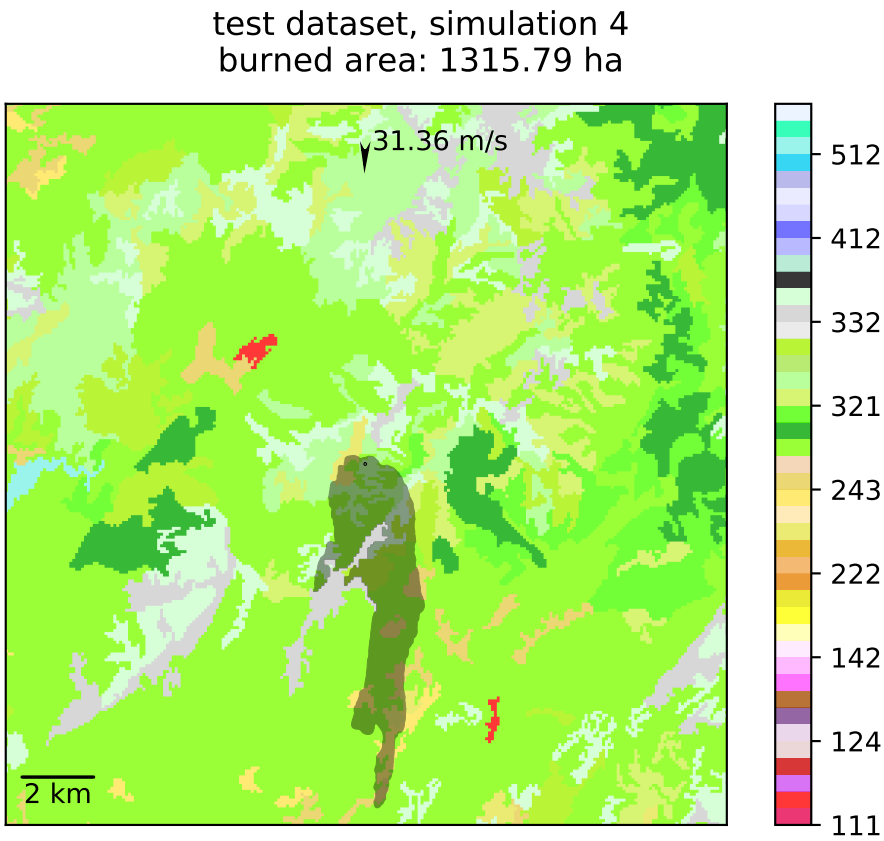


Figure IV.3. – Example of a simulated burned surface after one hour returned by ForeFire.
The initial fire front of 0.45 ha is represented in black at the center of the figure and the final burned surface is the surrounding shaded shape. The input wind speed vector is represented by the arrow at the top. The simulated fire spread to the south, was partly blocked by mountains (in gray), but still burned 1316 ha.
Background colors correspond to the classification of the Corine Land Cover

The simulations are meant to be used for prevision of wildfire spread in Corsica before a fire starts, at any time, so the intervals of variation of the raw inputs were chosen to account for a wide variety of environmental conditions. Moreover, in this context, there is significant uncertainty in the simulations. The weather forecasts used to predict

wind speed and FMC are possible sources of uncertainty; so are model simplifications and the choice of a given fuel parameterization. Therefore, the intervals of variation of both raw inputs and fuel parameters also account for their uncertainty range. Some intervals follow those of a previous study that focused on uncertainty quantification (see notably Table 1 in [7]).

Finally, the quantity of interest in the present study is the area in hectares of the burned surface obtained at the end of the simulation, namely after a free fire spread of one hour. An example of simulated surface is represented in Figure IV.3.

It is possible with ForeFire to simulate any duration of fire and obtain the state of the fire front at any moment between fire start and fire end. Still, the simulated one-hour area alone could be a relevant information for the firefighters as it provides an estimation of the potential of fire growth if a fire that starts at a given location is not contained fast enough, one hour being a typical time for a fire to be detected and firefighters to arrive on-site.

IV.3. Emulation with deep learning

In the context of fire growth prediction mentioned in Section IV.2, the absence of knowledge regarding the location of fire start and the uncertainty in the simulation are considerable difficulties that need to be addressed. An intuitive method consists in running a large number of simulations for ignition points all across the map, where some inputs are determined from weather forecasts. This procedure may or may not include perturbations in the inputs other than ignition point coordinates to account for uncertainty; but in any case, the time required to run all the desired simulations in operational conditions is too high with usual numerical simulators such as ForeFire. This motivates the use of an emulator to compute the area of the output simulated burned surface in a reasonable amount of time, although with some error of approximation. It is desirable to obtain an emulator that approximates this quantity with high accuracy and has a significantly lower computational time than that of the simulator, but it can be quite challenging for an emulator to combine both properties.

IV.3.1. Design of experiments

A common strategy to design an emulator consists in considering the simulator as a “black-box” and build the emulator based on a synthetic dataset of input and corresponding output. The first step of this strategy is to define a design of experiments (DOE) to generate the datasets that will be used to build the emulator and evaluate its approximation error. Given input dimension and model complexity in the present study, we expect a large number of simulations ($\sim 10^5$ at the very least) will be required for an emulator to have good accuracy.

The DOE relies on a Latin Hypersquare Sample (LHS) in $[0,1]^{46}$, which is a popular space-filling design. For all elements of the LHS, we apply an affine transformation from $[0,1]^{46}$ to the hyperrectangle whose boundaries are defined by the ranges in Table IV.1. However, this procedure alone does not account for the restrictions to the definition domain implied by the constraints on ignition point coordinates and wind speed norm. To include these constraints, we generate a LHS with more members than n_{train}, the

desired number of training sample members, and keep only "valid" members, namely those that satisfy the constraints after the affine transformation, so that the resulting sample size is slightly lower than the target. The next step in the constitution of the DOE is to generate a Sobol' sequence in $[0,1]^{46}$, which is a low-discrepancy sequence. We complete the initial LHS (in $[0,1]^{46}$) with members of the Sobol' sequence based on a discrepancy criterion, following the idea proposed in [70] to obtain an optimal complementary design. A notable difference in the present study is that the first elements selected by the algorithm are used to complete the training sample only if they are valid (they are ignored otherwise). Then, when the target size n_{train} is reached, the next valid elements are used to form a test sample of size n_{test}. This procedure aims at selecting the points of the test sample so that they are located far from each other but also far from the points of the training sample, where the approximation error is expected to be higher.

Finally, based on the inputs of the training and test sample, the corresponding fire spread simulations are carried out as described in Section IV.2 and the resulting outputs complete the training and test datasets.

IV.3.2. Neural network architecture

Several techniques can be considered for emulation. Simple statistical methods such as linear regression based on the inputs in Table IV.1 would most likely lead to poor approximation because of the non-linearity of the model. Other methods such as those mentioned in Section IV.1, (i.e., Gaussian processes, polynomial chaos, high dimensional model reduction, radial basis functions) are interesting alternatives, however their computational requirements (regarding time and/or memory space) can become prohibitory when there are both a high dimension ($d = 46$) and a large sample size ($\geq 10^5$).

In this problem, the input variables presented in Table IV.1 can be expressed as a vector of $\mathbb{R}^{46}$, including the coordinates (two scalars) of the ignition point. While these coordinates do locate the origin of the fire, they are not used directly to compute the ROS and simulate how the fire will spread from there. Actually, the restriction of the simulation domain to the surface that is burned after one hour identifies the part of the spatial fields of elevation and fuel parameters that were used in the ROS computations. Therefore, this information could be a better-suited emulator input than the coordinates of the ignition point. Although the simulated burned surface is not known beforehand, the fire will almost never spread further than 10 km in an hour; so *a priori* it will be contained in a 20 km $\times$ 20 km square centered around the ignition point. If one considers the fields of elevation and of fuel parameters h, σ_f, and S_v restricted to this square, given their 80-m spatial resolution, this amounts to four input fields of size 256×256 for emulation. This raises the need for a method that is adapted to handle such high-dimensional data as well as the remaining scalar inputs.

Neural network models appear suitable for emulation of fire spread simulations, not only because they usually perform well when trained on a large dataset, but also because they can handle several types of data. In particular, CNNs proved to be quite successful in the classification of 2D inputs such as images (e.g. [78]), but also for regression (e.g. [142]), which is our target. Here, the simulations are also significantly influenced by the other (scalar) inputs, notably wind speed and FMC, so a network with a hybrid

architecture to process both types of inputs (2D and scalar) seems well suited to our problem. The term "hybrid" may have different meanings when it comes to neural networks. It can refer to the succession of multiple ensembles of layers, with each ensemble appearing like a given type of neural network, as in [105] where DNA sequences are first processed by a convolutional part then by a recurrent part. In the present study, this term is understood as the use of specific types of layers for each type of input, as proposed in [143] where image, sequential, and scalar/categorical inputs are first processed separately by the network.

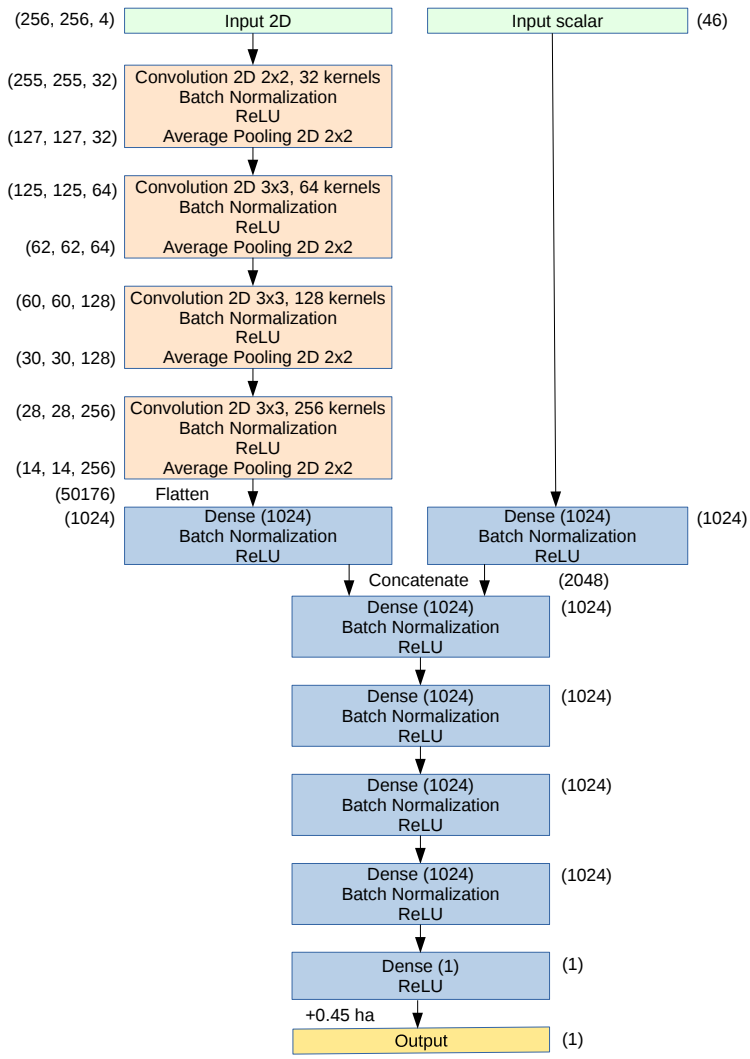


Figure IV.4. – Neural network architecture. The numbers in brackets outside the boxes indicate the shape of the data as they are processed by the network. The architecture is considered "hybrid" because the DNN processes both 2D input corresponding to terrain data and scalar inputs. Processing is first carried out separately until concatenation after which both parts are mixed.

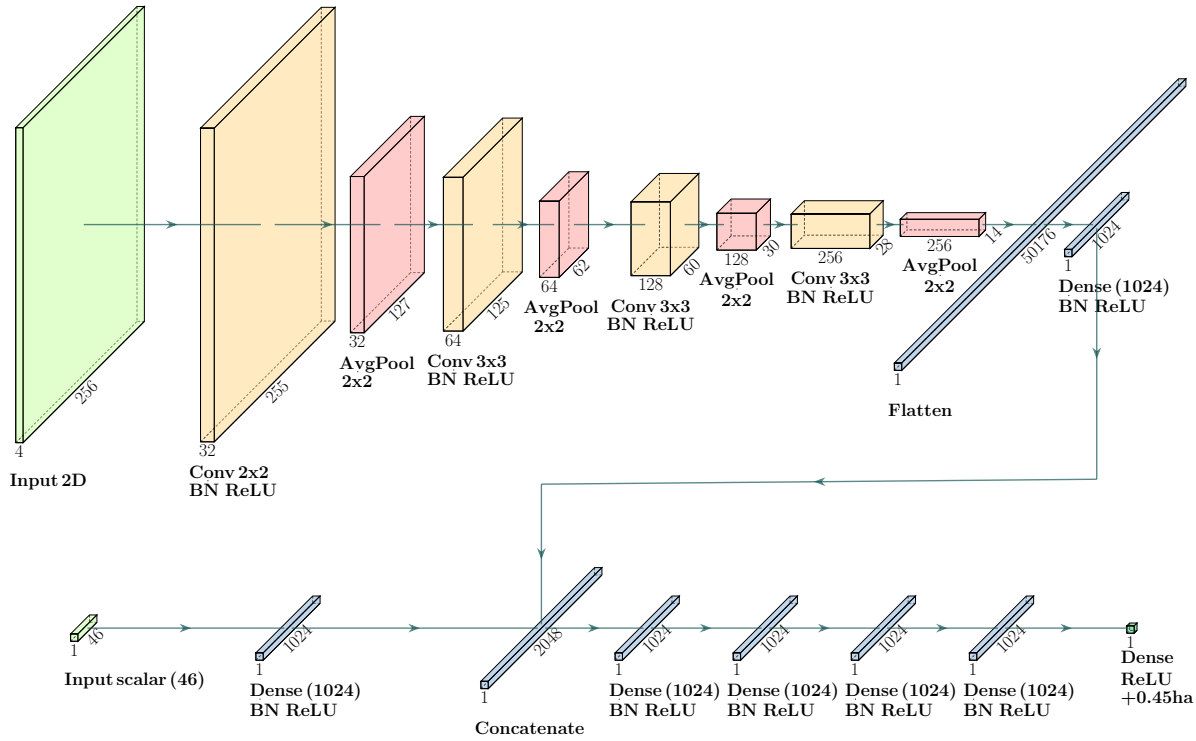


Figure IV.5. – Representation of data processing in the neural network. The blocks indicate the shape of the data. The 2D input is derived from the four fields of elevation, and fuel parameters h, σ_f, and S_v. The 46 scalar inputs are derived from the simulation parameter inputs of Table IV.1. Conv: Convolution 2D; BN: Batch Normalization; AvgPool: Average Pooling 2D.

We propose an emulator based on a DNN with a hybrid architecture. A convolutional part processes the four 2D fields of elevation and fuel parameters (prior to perturbation) h, σ_f and S_v in a square surrounding the ignition point with a side of approximately 20 km, which corresponds to an input of shape (256, 256, 4). Another part of the network processes the vector of size 46 of scalar simulation inputs mentioned in Table IV.1. The "absolute" coordinates (x, y) of the ignition point are replaced by (δ_x, δ_y), which are the coordinates of this point relatively to the center of the surrounding 2D fields. Also, both 2D and scalar inputs are scaled to [-1, 1] through an affine transformation before being processed by the DNN.

The detailed architecture of the DNN is represented in both Figure IV.4 and Figure IV.5. The first figure is more focused on the processing layers (i.e., convolutions, pooling, etc.), while the second represents the successive shapes of the data as they are processed by the network.

First, convolutions with a 2x2 window are applied to the 2D inputs, followed by a batch normalization layer, a Rectified Linear Unit (ReLU) activation and an average pooling layer with a 2x2 window. This succession of layers is repeated three more times,

with a 3x3 window for the convolutions and more and more kernels. Convolutions are carried out without padding nor stride, and the first two average pooling layers result in the edge of the data being cropped, due to the odd input shape. Then, the output of these four blocks of layers is flattened and goes through a block consisting of a fully connected feed forward (aka dense) layer with 1024 output nodes, followed by batch normalization and ReLU activation. As for the scalar input, it goes through a similar block of layers. The output of these two blocks is concatenated and undergoes four similar blocks of layers. The intention behind the application of the dense blocks before concatenation is to concatenate vectors that have the same shape and potentially give similar importance to the 2D part and the scalar part in this mixed architecture. Finally, a dense layer followed by a ReLU activation and an increase of 0.45 ha (the minimum simulated burned surface area, corresponding to a fire that does not spread) are carried out, yielding the output of the network.

IV.3.3. Accuracy metrics and training strategy

Among a dataset of size n, $\boldsymbol{u}^i$ denotes the i-th set of simulation inputs, $y(\boldsymbol{u}^i)$ the resulting output, and $\widetilde{y}(\boldsymbol{u}^i)$ the corresponding value returned by the emulator. Several metrics can be used to evaluate the accuracy of $\widetilde{y}$, the emulator of function y. In this study, we use the mean absolute error (MAE), the mean absolute percentage error (MAPE) and the standardized mean square error (SMSE, cf. [108]), which are respectively defined as follows:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |\widetilde{y}(\boldsymbol{u}^i) - y(\boldsymbol{u}^i)|, \tag{IV.1}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n} \left|\frac{\widetilde{y}(\boldsymbol{u}^i) - y(\boldsymbol{u}^i)}{y(\boldsymbol{u}^i)}\right|, \tag{IV.2}$$

$$\mathrm{SMSE} = \frac{\sum_{i=1}^{n} \left(\widetilde{y}(\boldsymbol{u}^i) - y(\boldsymbol{u}^i)\right)^2}{\sum_{i=1}^{n} \left(y(\boldsymbol{u}^i) - \bar{y}\right)^2}, \tag{IV.3}$$

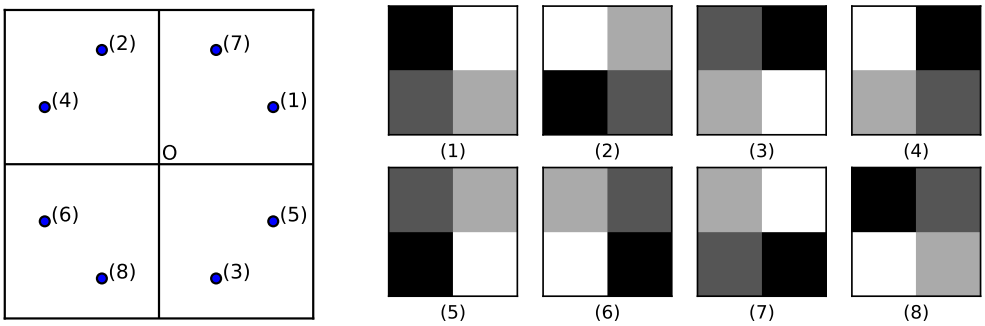
where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y(\boldsymbol{u}^i)$ is the sample mean of the emulated function. The SMSE can be seen as a mean squared error normalized by the sample variance of y, and would be equal to 1 if the emulator was a constant function equal to the sample mean $\bar{y}$. The lower these scores, the more accurate the emulator. The emulator can also be evaluated in terms of mean error, similarly to the MAE but without the absolute value, that will be referred to as "bias" in the following.

The accuracy metrics need to be computed for the test dataset as the error is expected to be much lower for the training dataset, which is used to determine the parameter values of the network. In order to quantify overfitting, the accuracy metrics may also be computed for the training dataset.

The procedure used to train the network's parameters relies on a MAE loss function with an Adadelta optimizer [144], without regularization based on the norm of the layer parameters.

To enrich the train dataset, a form of data augmentation is carried out: over one epoch, each member of the training dataset is used exactly once, but possibly after a ge-

ometric transformation (rotations or axial symmetries). The geometric transformation is applied to the 2D field inputs as well as (W_x, W_y), the wind speed vector, and (δ_x, δ_y), the relative coordinates of the ignition point. There is a 0.5 probability of having no transformation, whereas the other transformations (seven different non-identity applications) each have a 1/14 probability of being applied, all of them being represented in Figure IV.6. We know that in such a configuration, the simulated burned surface would be the same, so this allows us to enrich the dataset (virtually, by a factor of eight) without running additional ForeFire simulations, and might limit overfitting [123] since it allows for more possible configurations than described in Section IV.2. Note that data augmentation is only used during training. Also, with the synthetic datasets, there is no need to split the training dataset to obtain a validation dataset, since the test dataset was designed specifically to evaluate accuracy, as explained in Section IV.3.1. The accuracy metrics of the network are simply computed for the test dataset at the end of each epoch during training.



(a) Transformations applied to a point.

(b) Transformations applied to a matrix.

Figure IV.6. – Geometric transformations used for data augmentation during training. Considering (a) an initial point or vector in the plan with origin O or (b) a matrix in an initial state (1), 8 possible transformations are considered, resulting in state (n), with n $\in \{1, \ldots, 8\}$. Identity transformation, leading to (1), has a 0.5 probability of being applied during training. The probability is 1/14 for all other transformations. (2): 90° rotation, (3): −90° rotation, (4): y axis symmetry, (5): x axis symmetry, (6): 180° rotation, (7): y=x axis symmetry, (8): y=-x axis symmetry. Applying one of these transformations to the input data would result in a similarly transformed simulated burned surface, so the output area in hectares is invariant to these transformations.

IV.3.4. Extraction of the actual emulator

The DNN presented in Section IV.3.2 relies on many convolutions that can be computed much faster with high-performance graphics cards. Even if one is not equipped with such resources, it is possible to compute the output of the DNN even faster once the network has been trained.

To achieve this goal, the final layer of the convolutional part of the network (of size 1024), before concatenation with the scalar part, is pre-computed. Indeed, due to the spatial resolution of the elevation and land cover fields of approximately 80 m, there is a finite amount of possibilities for the 2D input and the subsequent layers up to the end of the convolutional part, which will take the same values as long as the ignition point is located in a given cell of side ∼80 m. In the present case, there are $\sim 1.2 \times 10^6$ possibilities for Corsica.

The actual emulator consists in the remaining part of the DNN and its inputs are the pre-computed final layer of the convolutional part as well as the scalar vector of size 46. This part of the network only involves some dense blocks and a concatenation of the two parts of the network, that can be computed much faster—even on a machine without specific acceleration.

IV.3.5. Implementation

Python scripts are used to process the data, generate the training and test datasets, build and evaluate the DNN. Keras library, which is a high-level neural networks API that is running on top of TensorFlow, is used for building the DNN.

Training and accuracy evaluation of the DNN up to the retrieval of the actual emulator are carried out on a GPU accelerated compute node. The computational time of the actual emulator is evaluated on a machine with 32 CPU.

The size of the datasets are $n_{\text{train}} = 5 \times 10^6$ and $n_{\text{test}} = 10^4$. Training is carried out with data augmentation as explained in IV.3.3 for 100 epochs with batches of size 400, and the hyperparameters of the Adadelta optimizer are a decay rate of 0.95, a conditioning constant ϵ of 10^{-7}, and a learning rate of 0.3, which is an extra factor in the right-hand term of Equation (14) in [144]. The weights of the network are initialized using default TensorFlow arguments, therefore the weights of Dense and Conv2D layers are initialized following a Glorot uniform initializer (cf. Equation (16) in [57]).

The same procedure is also applied to smaller training dataset of size $n_{\text{train}} \in \{10^5, 10^6\}$, each with a specific dataset of size $n_{\text{test}} = 10^4$ generated as explained in Section IV.3.1, to investigate the influence of n_{train} on the approximation error.

IV.4. Results and discussion

The computational time of a simulation (with ForeFire) of wildland fire spread took an average of approximately 25 s. This time highly depends on the input of the simulation and can range from about 0.1 second to more than an hour. Overall, the larger the simulated burned surface, the more computations are carried out during the simulation. Running all the simulations in the training and test datasets would have taken about 4 years if it were not for distributed computing: with several multi-CPU machines, for a total of about 150 CPU cores, the computations were completed in about 10 days. Given the simulation settings presented in Section IV.2, the obtained burned surface areas range from 0.45 ha to 24 804.4 ha among the training dataset. Some statistics of this output in the training dataset are presented in Table IV.2. The high variance of the simulation output is consistent with that of computational time. The minimum output corresponds to the area of the initial burned surface and is obtained in a few

Mean	Std	Minimum	Q1	Median	Q3	Maximum
455.7 ha	782.0 ha	0.45 ha	52.6 ha	181.0 ha	517.7 ha	24 804.4 ha

Table IV.2. – Statistics of the output simulated burned surface area among the training dataset of size 5×10^6.
Std: Standard deviation; Q1: first quartile; Q3: third quartile.
The output has high variance, arguably making "relative" error metrics such as the MAPE and SMSE (cf. Equations (IV.2) and (IV.3), respectively) better suited for expressing the performance of the emulator regarding approximation error.

simulations (approximately half a thousandth) where the FMC is very close to the moisture of extinction (0.3) in the ROS model, leading to a fire that almost does not spread. Similar statistics are obtained with the test dataset, except for the maximum output (14 403.7 ha). The test dataset, having a much lower size than that of the training dataset, is less representative of tail of the output distribution, hence the lower maximum.

Most simulations result in a burned surface of less than 1000 ha, which is realistic for a fire that spreads freely during one hour. Still, a non-negligible amount of simulations result in burned surfaces that are most certainly bigger than what would be observed in reality. This amount would probably be higher were it not for non-burnable zones that significantly contribute to limit fire spread in some cases. This is mostly due to the fact that the simulations rely on simplifying assumptions where wind speed and FMC are constant in time and the DOE allows these inputs to vary in very large intervals. Therefore, it is not surprising to obtain a very large burned surface in a simulation where the wind speed is extremely high, the FMC extremely low, and no unburnable zone is reached during a whole hour of spread. Although somewhat unrealistic, the extremely high values of simulated burned surfaces were not removed from the dataset. This might make the emulation more difficult but the ability to discriminate between a wide range of situations, even extreme ones, is relevant in wildland fire spread.

Carrying out one epoch took a bit less than three hours, so training, which consisted in 100 epochs, lasted for about 10 days. The evolution of the MAE over training of the DNN for these 100 epochs is reported in Figure IV.7. At a given epoch, the predicted values for both test and training datasets result from the model obtained at the epoch's end. Due to high computational time, the MAE was only computed for the training dataset (without applying data augmentation) at the first epoch and every five epoch starting from the fifth. On the one hand, the MAE for the test dataset decreases overall until it reaches 81.5 ha after about 78 epochs after which it oscillates around that value. On the other hand, the MAE for the training dataset decreases overall, faster than the MAE of the test dataset. Therefore, while both scores are almost identical at the start the gap between the two increases with the number of epochs.

The main objective is to have low generalization error, which is measured here using the error metrics for the test dataset. In high-dimensional cases, it is possible to observe a significant gap in error between the training and test datasets when training neural networks (see for instance [2]). It is the case in this study with the original model

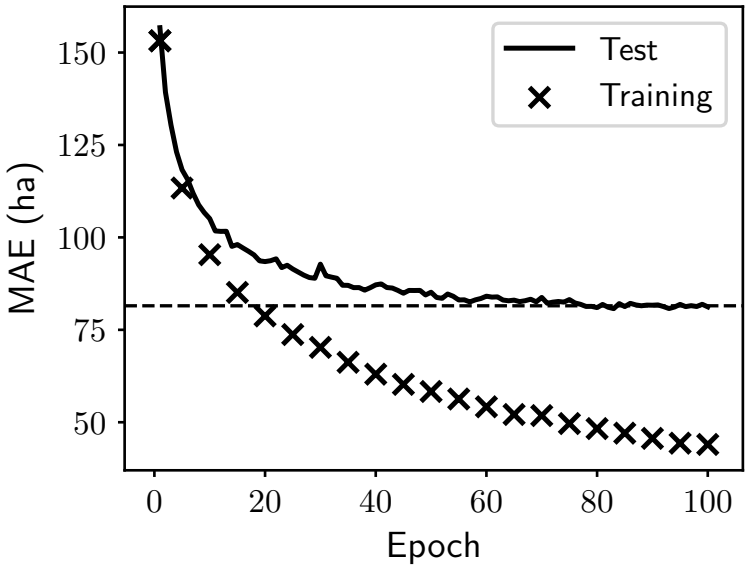


Figure IV.7. – MAE over training. The solid curve represents the MAE for the test dataset, while the crosses represent the MAE computed for the training dataset at the end of the first epoch and after every five epochs starting from the fifth. The horizontal dotted line corresponds to MAE=81.5 ha. Both metrics decrease and this decrease is faster for the training dataset, yet the MAE for the test dataset does not increase and seems to keep around 81.5 ha after about 75 epochs. No significant decrease in the test error is expected after more epochs, so the network after 94 epochs, which has a SMSE on the test dataset of 6.0% (the best over all 100 epochs) is selected for emulation.

(ForeFire) relying on high-dimensional input data and being highly non-linear. Note that the neural network can be interpreted here as a substitute for an interpolator in high dimension, without the constraint to coincide with the training dataset at the training points.

It is unlikely that carrying out more training epochs would result in a significant decrease of the error metrics for the test dataset. Consequently, the model with the best SMSE over the test set, which was obtained at the end of the 94-th epoch, was selected to define the emulator. The emulator with the best MAE was not selected because its MAE was only slightly lower (80.7 ha instead of 81.2 ha), while the other scores were all better for the model with the best SMSE. Even though our loss function (the MAE) may seem high, an absolute error of about 80 ha is at the same time very high for a small simulated burned surface of about 10 ha and very small for larger ones of about 1000 ha. Consequently, "relative" error metrics such as the MAPE and SMSE (cf. Equations (IV.2) and (IV.3), respectively) are arguably better suited for expressing the performance of the emulator regarding approximation error.

The error metrics of the emulator are reported in Table IV.3 and Table IV.4, re-

Model \ Metric	MAE	MAPE	SMSE	Bias
Mean of training	461.9 ha	2266.0%	100.0%	2.2 ha
Linear regression without threshold	387.9 ha	1239%	73.9%	−4.8 ha
Linear regression with 0.45 ha threshold	361.1 ha	493.3%	72.3%	21.9 ha
DNN after 100 epochs	81.2 ha	33.5%	6.2%	−13.1 ha
Emulator (from DNN after 94 epochs)	81.2 ha	32.8%	6.0%	−6.5 ha

Table IV.3. – Model error on test dataset of size 10^4.
The approximation is poor using the three most simple models (constant and linear regression with or without threshold), whereas the DNN trained using a large training dataset shows good approximation.

Model \ Metric	MAE	MAPE	SMSE	Bias
Mean of training	461.5 ha	2139%	100.0%	0 ha
Linear regression without threshold	389.9 ha	1185%	73.7%	0 ha
Linear regression with 0.45 ha threshold	365.6 ha	493.4%	72.3%	24.3 ha
DNN after 100 epochs	44.0 ha	23.8%	1.2%	−7.6 ha
Emulator (from DNN after 94 epochs)	45.1 ha	23.2%	1.2%	−0.9 ha

Table IV.4. – Model error on training dataset of size 5×10^6.
For the three most simple models, the approximation metrics are almost the same to the ones computed on the test dataset (cf. Table IV.3), whereas the DNN has lower error for the training dataset than for the test dataset.

spectively relating to the test dataset and the training dataset. These metrics are also computed for three simple models for comparison: 1) a model that consists in always predicting the mean simulated burned surface of the training dataset (455.7 ha), 2) a linear regression model fitted using the training dataset based on the 46 inputs of Table IV.1, 3) same as the previous model, but applying a 0.45 ha minimum threshold (the minimum simulated area) to avoid non-physical output. The metrics for the DNN with the parameters obtained at the end of training are also reported. Although a MAE of 81.2 ha might seem high, it is much lower compared to that of the three simple models (461.9 ha with the mean, 361.1 ha for the linear regression with threshold). The SMSE of 6.0% means that 94.0% of the variance in the test dataset output is explained by the emulator, which is very good given the range of variation in simulation inputs. The relative error is also satisfactory with a MAPE of 32.8% on the test dataset, especially when compared to that of the simple models (2266.0% using the mean, 493.3% using linear regression with threshold). As for computational time on a 32-CPU machine, the outputs for the test dataset are obtained in about half a second with the emulator against 56 s with the whole DNN, which corresponds to a speed-up by a factor of about 100. Also, the corresponding ForeFire simulations would have been obtained in about two hours with parallel computations on the 32-CPU machine, meaning that the emulator allows a speed-up by about 15,000 times. For a dataset where the simulated burned surface tends to be higher, the average computational time with ForeFire could be higher. This is not the case for the emulator, for which computational time does

not depend on the output fire size, meaning that the resulting speed-up factor would be higher.

The influence of n_{train} on the resulting approximation error of the DNN based on the error metrics for the test dataset is presented in Table IV.5. The MAE in the test dataset did not seem to decrease after a few tens of epochs for smaller training datasets (this was also the case for $n_{\text{train}} = 5 \times 10^6$), and the model with best SMSE over 100 epochs, which had a MAE close to the best value obtained over the 100 epochs, was selected. Figure IV.8 shows the MAE, MAPE, and SMSE from Table IV.5 plotted against n_{train}. Although there are only three points for each metric, linear regression of the logarithms suggests that the metrics decrease following a trend of the form $(n_{\text{train}})^{-\alpha}$ with $\alpha \in [0.2, 0.37]$, depending on the metric, which is quite slow. For instance, assuming this trend using the values of the slopes reported in Figure IV.8 to specify α, reducing the MAE from 81.2 ha to 50 ha (resp. the MAPE from 32.8% to 20% and the SMSE from 6.0% to 2.0%) would require to increase n_{train} by approximately a factor 10 (resp. 7 and 20).

n_{train} \ Metric	MAE	MAPE	SMSE	Bias
10^5 (best SMSE after 34 epochs)	182.0 ha	89.1%	25.4%	−22.6 ha
10^6 (best SMSE after 26 epochs)	127.5 ha	49.9%	13.3%	−16.3 ha
5×10^6 (best SMSE after 94 epochs)	81.2 ha	32.8%	6.0%	−6.5 ha

Table IV.5. – DNN error on complementary test dataset (always of size $n_{\text{test}} = 10^4$) with variable training dataset size n_{train}.
The larger the training dataset of the DNN, the better the approximation.

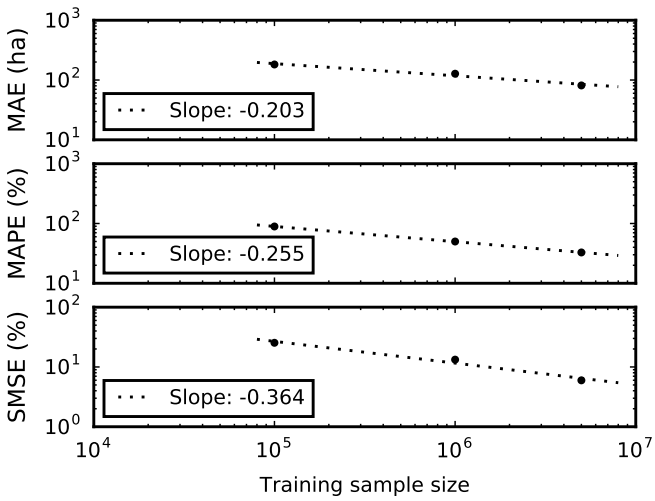
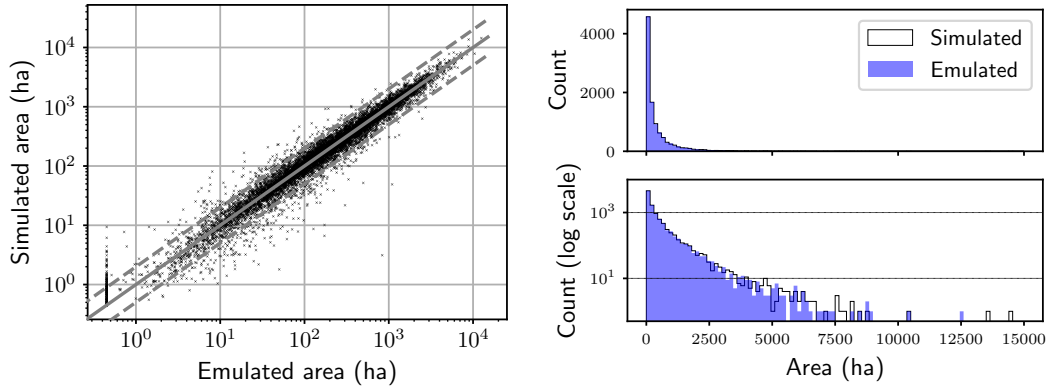


Figure IV.8. – Plot of the error metrics against n_{train} the size of the training dataset (cf. values reported in Table IV.5), in log scale.
Linear regression of the logarithms (dotted lines) results in a good fit with the data, suggesting a slowly decreasing trend of the form $(n_{\text{train}})^{-\alpha}$ with $\alpha \in [0.2, 0.37]$, depending on the metric.



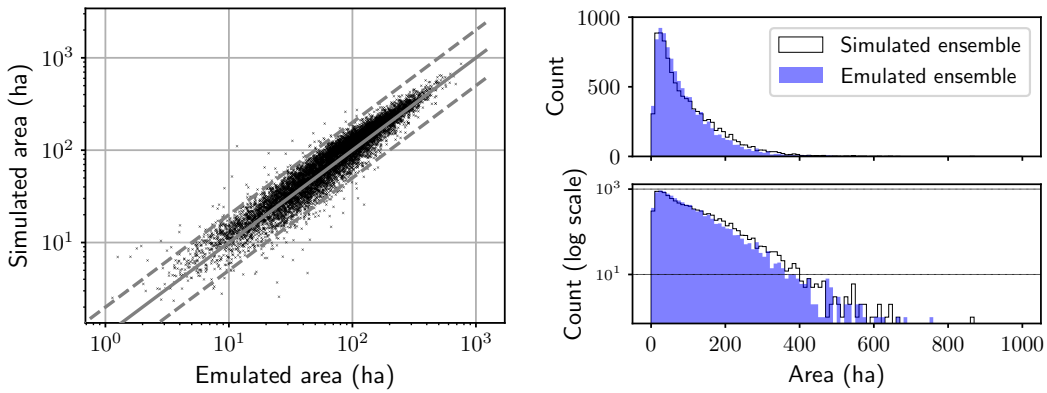
(a) Individual burned areas in log scale.

(b) Histograms of burned areas.

Figure IV.9. – Comparison between the burned area simulated by ForeFire and the corresponding emulator output over the test dataset of size 10^4.
(a) The solid oblique gray line corresponds to a perfect match and the dotted lines correspond to an error by a factor of 0.5 and 2.
(b) Black contour: histogram of simulated areas; blue surface: histogram of emulated areas. Both top and bottom figures represent the same distributions, they share the same abscissa axis but the bottom figure has its ordinate in log scale.
Most of the emulated values are at most either twice higher or half lower, resulting in good error metrics (MAE=81.2 ha, MAPE=32.8%), yet the individual error is quite high for a few members. The distributions of both samples are close.

For more insight regarding the approximation of the selected model, the emulator output for each member of the test dataset is plotted against the actual values of simulated burned area in Figure IV.9. The vast majority of the emulated values are close to their simulated counterparts and 9,332 out of 10,000 are at most either twice higher or half lower. In 157 cases, the emulator returns the minimum value of 0.45 ha, while the actual simulated value may go up to 10 ha. This corresponds to the apparent "black vertical bar" at the lower left of the graph in Figure IV.9a. There are 29 simulations for which the emulated burned area is at least five times lower (11 of them being equal to 0.45 ha) and 43 simulations for which the emulated value is at least five times higher. In the latter cases, most of the simulated burned surfaces are small ($\leq$10 ha in 32 simulations out of 43), which usually contributes to a higher relative error; but not all of them. In some of these cases of overprediction by the emulator, there is a relatively small area close to the ignition point in the main direction of fire spread that seems to considerably slow down the fire. The emulator probably has difficulty when it comes to accounting for some particular configurations of the underlying fuel and altitude fields, especially small non-burnable areas, given that the convolutional part of the DNN reduces the size of inputs by a factor of 256 when processing it for the emulator (from 262,144 to 1024). Overall, the individual errors lead to similar distributions of burned area. The emulator has a small bias of -6.5 ha and, as shown in Figure IV.9b,

the histogram of emulated burned areas is slightly less dispersed (standard deviation of 752.9 ha against 782.5 ha).



(a) Individual burned areas in log scale.

(b) Histograms of burned areas.

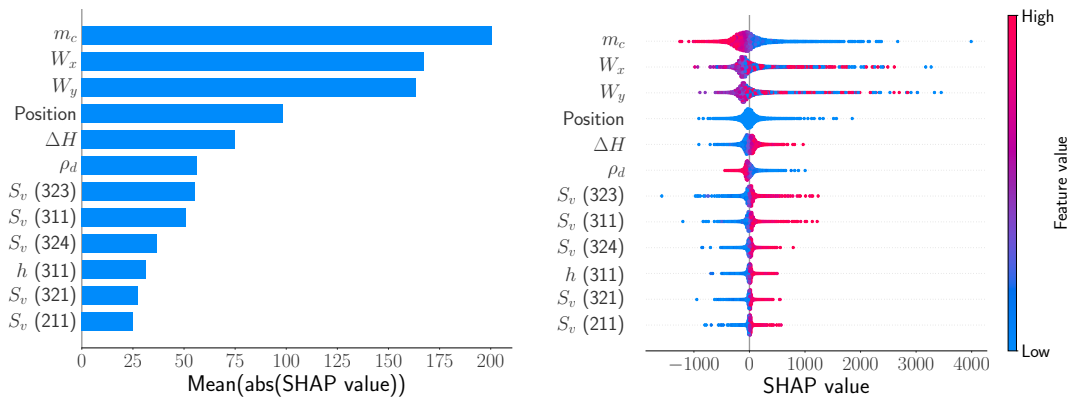
Figure IV.10. – Comparison between the ensemble of burned areas simulated by ForeFire for the fire case of Calenzana and their emulated counterparts.
(a) The solid gray line corresponds to a perfect match and the dotted lines correspond to an error by a factor of 0.5 and 2.
(b) Black contour: histogram of simulated areas; blue surface: histogram of emulated areas. Both top and bottom figures represent the same distributions, they share the same abscissa axis but the bottom figure has its ordinate in log scale.
The inputs have smaller variations than for the test dataset, yet most emulated values fall into the range of half to twice the simulated value as it was the case for the test dataset (cf. Figure IV.9), leading to good error metrics (MAE=18.7 ha, MAPE=22.7%).

The emulator is also evaluated with an ensemble of ForeFire simulations that correspond to a real Corsican fire that occurred near Calenzana during summer 2017 and burned about 120 ha. Most of the spread for this fire took place during the first hour after ignition. For this case, some reference inputs are defined from weather predictions and a presumed ignition point is identified, as explained in [5]. Then, an ensemble of perturbed simulations is generated, where the inputs presented in Table IV.1 follow a calibrated distribution that was obtained in a previous study [7] with $\beta = 1/2$. It should be noted that the resulting ensemble of burned surface areas in the present study is not the same as in [7] because supplementary inputs were variable in the previous study (such as perturbations in the times of fire start and fire end, which could make the simulated fire duration different from one hour). The 10,000 simulated burned surface areas of the ensemble are compared to their emulated counterparts in Figure IV.10. Similarly to the test dataset, most emulated values fall into the range of half to twice the simulated value, leading to a MAPE of 22.7%. A MAE of 18.7 ha is obtained and individual errors result in a distribution of the emulator output that is less dispersed than that of the simulated output, as shown in Figure IV.10b, with a bias of −9.6 ha and a standard deviation of 77.7 ha against 86.1 ha.

The overall agreement between simulation and emulation is good for this simulated fire case, and the simulations were computed in 20 minutes, whereas the emulator predictions only took a bit more than a second. The speed-up factor is about 1000 this time, which is lower than the several thousands obtained for the test dataset. This is explained by the lower simulation time for this fire case (20 min instead of about two hours for the test dataset, while both datasets have the same size). This performance is quite promising for application to ensemble forecasting, but care should be taken as propagation of uncertainty leads to different output distributions according to the model (either ForeFire or its emulator) used.

Linked to the approximation error of the emulator is the influence of the inputs on the output. A desirable property of the emulator is the ability to behave in a similar way as ForeFire so that it keeps the main characteristics of the fire spread model, namely a burned area that, overall, increases with wind speed and decreases with FMC, while the surrounding 2D fields of altitude and fuel can either favor or block fire spread. Perturbations of fuel parameters are expected to have less influence, especially those of fuel parameters that are applied to a specific fuel type (h, σ_f, S_v). Also, the ROS is proportional to heat of combustion ΔH, which is a global parameter, so positive perturbations of this quantity will increase the burned area and negative ones will decrease it.

Given the complexity of the emulator, one may approach it as a black-box and estimate the overall influence of its inputs with Shapley additive explanations (SHAP, cf. [88]), a feature attribution method. The features we focus on are the inputs of the emulator, namely the 1024 "position" scalars linked to the 2D fields surrounding the ignition point stemming from the convolutions and the remaining 46 scalar inputs. Approximate SHAP values are computed for each member of the test dataset by means of expected gradients. This procedure leads to exact SHAP values when the model to explain is linear and the features are independent. While these assumptions are not verified with the emulator, the original algorithm for the computation of the exact SHAP values is too computationally expensive, whereas this method allows to compute approximate values in a reasonable amount of time. Although these results should be taken with care, they can still be used for a qualitative analysis and should provide some insight on the overall input influence over a whole dataset. For each member of the test dataset, the expected gradient is estimated based on a subset of size 50,000 sampled randomly from the training dataset. Given that the 1024 position scalars are difficult to interpret and expected to have little individual influence on the output due to their correlation, we consider the sum of their SHAP values, which is identified via a fictitious variable named "Position". The approximate SHAP values obtained for 12 of the 47 resulting variables are summarized in Figure IV.11. The values obtained for each of the 10,000 test members represented in Figure IV.11b indicate a good overall agreement with the main characteristics of the fire behavior model. High FMC (m_c) tends to decrease the output while low FMC tends to increase it. High positive SHAP values for the coordinates of wind speed (W_x and W_y) are obtained for extreme values of these inputs, i.e. close to either $-35\,\mathrm{m\,s^{-1}}$ or $35\,\mathrm{m\,s^{-1}}$ (in blue and red, respectively) while the negative values are obtained for intermediate values (close to $0\,\mathrm{m\,s^{-1}}$). SHAP values associated to the perturbation of ΔH are also consistent with our expectations. Regarding the rankings of the inputs when looking at the absolute



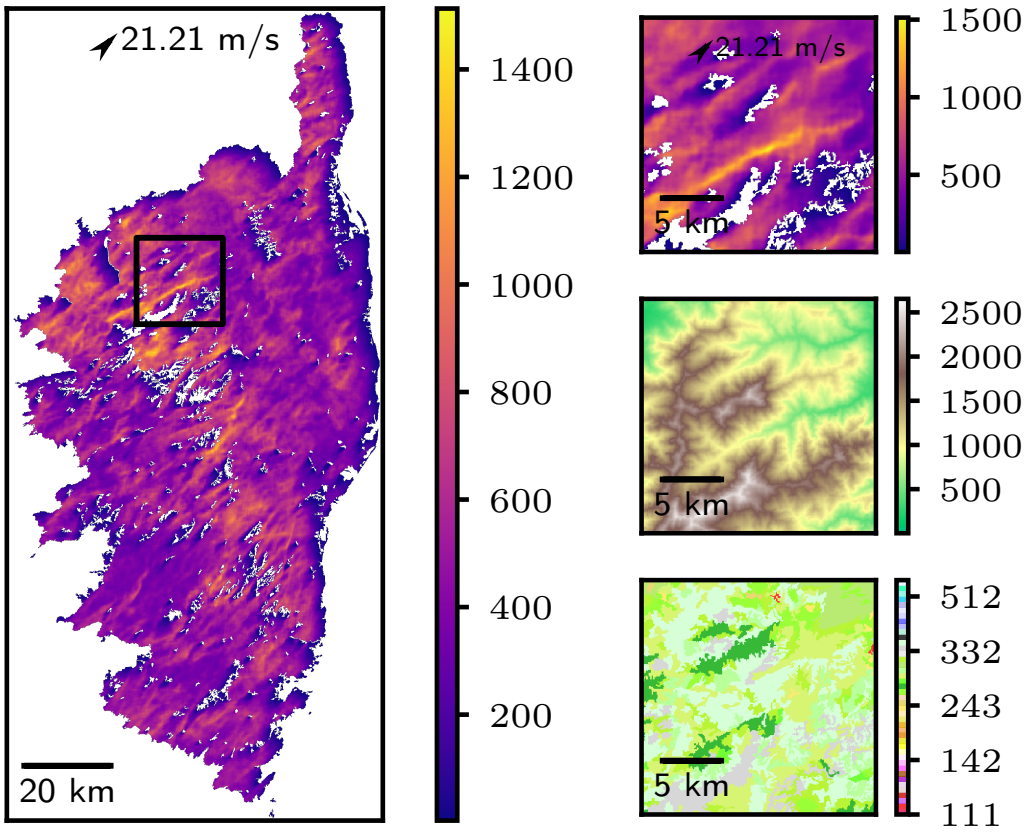
(a) Mean of the absolute value over the test dataset.

(b) Individual values for each member of the test dataset.

Figure IV.11. – Approximate SHAP values associated with the emulator computed for the test dataset, using the training dataset as basis. The SHAP values corresponding to the 1024 inputs resulting from the convolutional part of the DNN are summed up and this sum is identified as "Position" in the figure. Only the 12 most overall influential inputs, as ranked in (a), are represented.
(b) The color indicates the value of the input for each member, while the SHAP value is read in abscissa.
Qualitatively, the influence of the inputs expressed by the SHAP values corresponds to typical behavior of fire spread, both in terms of ranking and in terms of values for individual members (e.g. overall, low FMC m_c leads to a high SHAP value and high FMC leads to a low SHAP value). This suggests that the input-output relationship of the emulator is similar to that of the simulator.

SHAP values averaged over the test dataset in Figure IV.11a, the three most influential inputs are the FMC and the two coordinates of wind speed. Position is ranked fourth, perturbations on fuel parameters that affect all fuel types (ΔH and ρ_d) are ranked fifth and sixth, and the remaining ranks are attributed to the other perturbations of fuel parameters, as well as δ_x and δ_y (ranked last). Interestingly, when the positional inputs are not summed, their individual influence is quite low: the 54th scalar of the vector of size 1024 is the highest ranked at rank 32 only. Although we only have an approximation of SHAP values, these results are qualitatively the ones we would expect from fire spread simulations and indicate that the emulator has an overall relationship between inputs and output that is fairly consistent with typical behavior of wildland fire spread.

The "physical" behavior of the emulator is also analyzed through the lens of fire danger mapping in Figure IV.12, that represents the response surface of the emulator where the ignition point varies in Corsica on grid of the altitude field, whereas the other inputs are fixed to $m_c = 0.13$, $(W_x, W_y) = (15, 15)\ \mathrm{m\,s^{-1}}$, and no perturbation on fuel parameters. This mapping involves $\sim 1.2 \times 10^6$ emulator computations, which are carried out in about 40 s only. Values lower than 200 ha can be observed toward the south-west



(a) Map of emulated burned area on the entire Corsica island.

(b) Zoom in the black square represented in (a).

Figure IV.12. – Map of the area (in hectares) of the burned surface predicted by the emulator with variable ignition point in Corsica. The other inputs are a wind speed vector of (15, 15) $\mathrm{m\,s^{-1}}$ represented with a black arrow, a FMC of 0.13, and no perturbation on fuel parameters. The spatial resolution is approximately 80 m; white pixels correspond to non-burnable locations in the simulations.
(b) From top to bottom: burned area (ha), altitude (m), land cover. The "potential" area at a given ignition point is clearly lower when unburnable locations are close to the ignition point in the direction of the wind speed vector. This is consistent with "physical" behavior of wildland fire spread.

of non-burnable areas (mostly water bodies, rocky mountain tops over 1800 m with no vegetation, and urban areas), while most of the other ignition points are associated to values higher than 300 ha. This is consistent with the input wind speed vector pointing to the north-east. Also, there is a fairly high spatial variation of the emulated burned area that goes up to about 1500 ha. The smaller region shown in Figure IV.12b presents some of the highest emulated values. Comparison with the underlying 2D fields of al-

titude and fuel used in the simulations does not reveal clear influence of either one of these fields on the emulated output (except for the ignition points to the south-west of non-burnable locations). An animated version of Figure IV.12a with varying wind is available as Supplementary material. Considering that the approximation errors of the emulator are relatively low, it appears that, overall, the map generated using the emulator highlights locations where ignition would induce larger burned areas.

IV.5. Conclusions

The basis for the present study was simulations of wildland fire spread with the numerical solver ForeFire using the underlying ROS model of Rothermel. These simulations represented free fire spread during one hour from a small initial burned surface located at all possible areas in Corsica island. The terrain was represented by 2D fields of fuel and altitude at approximately 80-m resolution in the simulations. Some environmental input parameters, namely FMC, wind speed, and perturbation of fuel parameters, were also allowed to vary in a wide range. ForeFire simulations can be computed in a reasonable amount of time, yet too high for applications that require a large number of simulations on a daily basis. This motivated the use of an emulator in order to faster compute an approximation of the output simulated burned area (in hectares).

The proposed approach consisted in training a DNN used for regression. The network has a hybrid architecture to deal with 2D fields of environmental parameters and with scalar inputs. On the one hand, the 2D fields are restricted to a square of 20 km side centered around the ignition point to filter out information that is, for the most part, not used during the simulation, and these fields go through convolutional blocks due to their similarity to images. On the other hand, the remaining scalar inputs go through a dense block and are concatenated with last layer of the convolutional part. Then, the rest of the network consists in more dense blocks. Training was carried out with a large dataset of size 5×10^6 obtained from a LHS sample, which could be augmented during training, and a complementary test sample of size 10^4 was obtained from a low-discrepancy sequence.

The DNN achieved good approximation of burned surface area simulated by ForeFire. The last layer of the convolutional part of the DNN for all fuel cells ($\sim 1.2 \times 10^6$) of the map of Corsica for which ignition is possible in the simulation is pre-computed. This allows to reduce computational time since the resulting positional information can be used together with the scalar inputs to run computations with only the remaining part of the DNN, which was chosen as emulator of burned surface area. The emulator showed satisfactory performance. In the test dataset, it explains 94.0% of the variance of the output, it has a MAPE of 32.8%. Also, compared to the ForeFire simulations for fire danger mapping, the emulator computations are carried out thousands of times faster on a 32-CPU machine. Finally, the overall influence of the inputs on emulator output seems consistent with typical behavior of wildland fire spread.

Preliminary results suggest that the emulator is suited to ensemble predictions and fire danger mapping, notably due to the considerable speed-up factor. For instance, 1.2 million ForeFire simulations requiring 25 s on average would be computed in more than 10 days on a 32-CPU machine, while this took about 40 s with the emulator, that

is to say more than 20,000 times faster.

Even though the DNN was trained for Corsica using ForeFire, the method presented in this work could be applied to other regions and/or similar fire spread simulators. In this method, the two most computationally expensive steps are 1) running the simulations required for the training dataset and 2) training the DNN. Thanks to high-performance computing resources (multi-CPU machines for the simulations and a GPU-accelerated core for training), both steps only took about 10 days in the present study. In other cases with bigger territories (e.g. at the scale of a country), one can expect that an even larger training dataset will be required to obtain comparable approximation error. Also, if the spatial resolution of the data maps is different, one may consider adapting the convolutional part. Regarding data size, the available RAM also poses a constraint on the batch size used during training. For these reasons, the authors strongly recommend using high-performance computing resource to apply the method and starting with a relatively small training dataset before moving on to a larger dataset. In spite of the high computational time for those two steps, once the DNN is trained and the emulator is obtained, the resulting speed-up factor should be worthwhile.

A major research perspective consists in evaluating the emulator for use in ensemble predictions and fire danger mapping, but now in a more extensive manner. In particular, actual weather forecasts that cover the whole island will be used to generate fire danger maps for every hour (at least) of a given day. This process can be carried by considering several real fire cases or an entire fire season. Depending on the ability of the emulator to quickly identify the locations with higher fire danger ahead of time, it could provide valuable help in an operational context.

Another perspective is to investigate how the DNN compares with other approximation techniques (regression or interpolation), but their application can be quite computationally expensive with large training datasets and may require to carry out data reduction on high-dimensional inputs. One may also focus on the neural network architecture to either increase its performance or extend its application to more scenarios of wildland fire spread simulations. A first extension could be to consider more simulation outputs, for instance the burned surface area every ten minutes after ignition. In this case, the DNN could yield a vector output that represents burned areas at different forecast times, instead of a single scalar, where each component could be expressed as the sum of the previous component plus a positive quantity. Similarly, inputs such as wind speed vector and FMC could vary during the simulation time. This would entail more possibilities in simulated scenarios, making the emulator more relevant for simulations of fires spreading during 1 hour or more, provided that it is trained with realistic weather time series, the definition of which is not obvious. As for network architecture, upsampling layers could be considered, hoping that they would re-constitute a good raster approximation of the burned surface. This burned surface either could be used directly as output (as in [68]) or as the layer previous to the final output node estimating the number of hectares burned. Also, multi-dimensional recurrent neural networks [58] could be considered as substitute for the convolutional part of the DNN. Regardless of the complexity of the emulator, the main properties to pursue remain the same: low approximation error and reduction in computational time.

Acknowledgments

Funding: This work was supported by the Agence Nationale de la Recherche, France [grant number ANR-16-CE04-0006 FIRECASTER].

This work was carried out using HPC resources from GENCI-IDRIS (Grant 2020-AD011011630).

V. High-resolution fire danger mapping

In the present chapter, we use the emulator of simulated fire size presented in the previous chapter (referred to as DeepFire in the following) and study its application to fire danger mapping using actual weather forecasts. Predictions based on DeepFire are generated for 13 relatively big fires that occurred in Corsica and compared with an operational fire danger index used in France. Due to the different nature of DeepFire's output (i.e. a fire size expressed in hectares), we then investigate the insights provided by such predictions characterized by high spatial resolution and frequency that can account for input uncertainty in the form of ensembles. We then address how to best summarize such forecasts and how to derive new daily fire danger ratings.

Summary

V.1. Introduction **132**
V.2. Method **135**
V.2.1. DeepFire: fire size emulation using a deep neural network . . 135
V.2.2. Deterministic predictions of fire danger maps 137
V.2.3. Probabilistic predictions representing input uncertainty . . . 139
V.2.4. Application to weather forecasts corresponding to 13 Corsican fires 140
V.3. Analysis of predictions for specific fires **143**
V.3.1. A detailed fire case: Calenzana 2017 143
V.3.2. Overview of predictions for all 13 fires 152
V.4. Possible designs for daily fire danger mapping **156**
V.4.1. Focus on the day of Calenzana 2017 fire 156
V.4.2. Overview of the predictions for the 12 other fires 161
V.5. Conclusions and perspectives **162**

The preprint of the paper "Simulation-based high resolution fire danger mapping using deep learning", written by Frédéric Allaire (INRIA), Jean-Baptiste Filippi (CNRS), Vivien Mallet (INRIA), and Florence Vaysse (Météo-France), submitted to *International Journal of Wildland Fire*, is reproduced here.

V.1. Introduction

Wildfire occurrence and behavior is difficult to predict accurately. The environmental conditions such as weather and the type(s) of vegetation involved may be known to some extent, but the time, location, and cause of fire occurrence are rarely known ahead of time. A common approach in wildfire prediction consists in assessing fire danger, a general term that expresses both fixed and variable factors of the fire environment that influence the ease of ignition, rate of spread and difficulty of control.

Fire danger rating systems include assessment of one or several fire danger indices, which are used among other information to provide a rating, i.e. a class whose possible values may include "low", "moderate", "severe", "extreme", etc., but the notion of "rating" may also refer to the scalar indices composing the system. Regardless of the type of quantity used to assess it, fire danger relates to the proneness for ignition, spread and/or intensity of a wildfire according to the state of the vegetation and its environment at a given time, therefore reflecting how difficult it may be to control fire. Development of fire danger rating methods has led to the implementation of systems at the national scale about 50 years ago in Canada and in the US, with the Canadian Forest Fire Danger Rating System (CFFDRS [83], whose development started in 1968) and the National Fire Danger Rating System (NFDRS [17], first published in 1972, but whose development as well as fire danger rating methods dates further back, see for instance [65]).

Maps of fire danger ratings are usually generated every day to assess the situation of the current day and forecast the situation for the day(s) to come. Such maps may be available among other data via internet-based information systems; for instance, covering the US as part of the Wildland Fire Assessment System (WFAS [20]) [1], covering Canada as part of the Canadian Wildland Fire Information System (CFWIS) [2], covering Europe and the Mediterranean area as part of the European Forest Fire Information System (EFFIS) [3], or even covering the globe as part of the Global Wildfire Information System (GWIS) [4], which builds on activities of EFFIS.

Calculation of the aforementioned fire danger indices mostly depends on weather inputs and, to some extent, fuel moisture, which is generally derived from weather information. In the CFFDRS, the final output is a composite Fire Weather Index (FWI), that depends on such inputs, more precisely wind speed, air temperature, relative humidity and rainfall over the previous 24 hours, as well as the values of the indices from the day before. Although one could consider computing the indices at high temporal resolution, they are generally computed to represent peak, or at least high, fire weather conditions in a given day. Also, although the FWI was calibrated to describe fire behavior in a jack pine stands, it has been implemented in many other countries despite the differences in vegetation and climate where it showed reasonable performance (see for instance [56] and references therein). Forecasted maps of such indices generally have a temporal resolution of one day even if the underlying weather forecasts have higher temporal resolution, but the temporal extent and spatial resolution are generally the same.

1. https://www.wfas.net/, last checked 2021.02.01
2. https://cwfis.cfs.nrcan.gc.ca/maps/fw, last checked 2021.02.01
3. https://effis.jrc.ec.europa.eu/, last checked 2021.02.01
4. https://gwis.jrc.ec.europa.eu/, last checked 2021.02.01

Weather forecasts covering a smaller area may have higher spatial resolution: for instance the regional AROME limited area model in use at Météo-France has a spatial resolution of 1.3 km, covers France and part of neighboring countries, and allows for the computation of fire danger index at much higher resolution. AROME is also an operational system used as a reference in this study. Meteorological assistance for forest fires started in the 60s and the Météo-France methodology has been gradually evolving over the years, depending on the research but also on the feedback of the various departments of Civil Protection and the ONF who fight against forest fires.

In summer, the Mediterranean vegetation is evolving between a growth phase in spring to a dormancy phase between mid-summer and early fall. When the scrubland's shrubs are falling over to this vegetative state, there is a very small quantity of water in leaves and thus the Mediterranean vegetation is becoming extremely flammable by dry weather. The expertise of weather fire danger requires, firstly, the monitoring of drought to evaluate whether or not the vegetation is prone to big summer fires.

In this configuration, at Météo-France, the forecaster assesses weather fire danger with 3 main indices; FWI: for vegetation in low or moderate drought, IPse: a local rate of spread index for very dry vegetation and IEP (Indicateur d'éclosion et Propagation, French for "Ignition and spread indicator"): for dead or dormancy vegetation for which the water content is directly linked with air moisture and therefore Fine Fuel Moisture Content (FFMC). All those indices are calculated at the maximum of the day (in which case, they are post-fixed with an x, leading to FWIx, IPsex, and IEPx). Weather fire danger depends on wind speed and highest temperatures/lowest air moisture, which can happen at different moments during the day depending on the local meteorological context; for instance, strong and dry winds can happen at any moment in the day and even in the night in Mediterranean areas.

Instead of relying on empirically calibrated formulas to compute indices that are, for the most part, unitless, a promising strategy is to rely on a large number of simulations of fire spread using weather forecasts as input, therefore providing a more "concrete" representation of fire spread via simulated burned surfaces. Two interesting methods following such a strategy, although applied at different time scales than day to day predictions, are burn probability (BP) modeling and ensembles of simulations for specific fire cases.

BP modeling consists in propagating input distributions of environmental characteristics (weather and fuels) and fire characteristics (ignition, fire duration, number of fires in a year) via several simulations of wildland fire ignition and spread, this process being repeated independently several times via a MC method to represent potential scenarios of wildland fire activity in a region and/or a country. In this context, the output BP maps represent the probability of local areas to be burned by a fire (typically over the course of a year). Notable implementations include Burn-P3 [99] which relies on the Prometheus solver [137] and FSim [46] which relies on FARSITE [47]. A large number of studies have proposed and/or applied BP modeling (see for instance [100] and references therein).

Similarly, ensembles of simulations for specific fire cases allow to assess "burn probabilities", following a Monte Carlo approach that accounts for input uncertainty in fire spread simulations, notably regarding weather variables. A notable implementation of this method is the FSPro system [49], which relies on an ensemble of FARSITE simu-

lations. Due to knowledge of the ignition location, the burn probability maps resulting from such ensembles cover a much smaller area than maps of BP modeling.

From an operational perspective, BP modeling can be seen as a decision support tool for "long-term" planning, whereas ensembles of simulations for specific fire cases apply to "crisis" situations. In "short-term" predictions, however, the input distributions should be more representative of the expected weather conditions as in ensembles of simulations for specific fire cases, but should also account for the lack of knowledge regarding future ignitions over a vast territory, as in BP modeling. Although BP modeling methods could be adapted to account for daily weather forecasts, their computational requirement due to the large number of simulations involved is too high for use in an operational context.

The approach developed in this work proposes to use wildland fire simulations as basis for short-term assessment of fire danger, while relying on deep learning to reduce the computational cost of running a high number of simulations. A deep neural network (DNN) referred to as *DeepFire* in the following is used to approximate the potential fire size returned by a fire spread simulator.

DeepFire is able to compute a high number of fire size estimations in a short amount of time, and could generate fire danger maps at both high spatial resolution (approximately 80 m) and high frequency (10-minute time step) in a operational context. The fire size computations could be obtained using a fire spread simulator, but would require several days in this context. Compared to traditional fire danger indices, two major differences of DeepFire are that it accounts not only for weather but also for the influence of terrain on fire spread, and that fire danger is expressed as a physical quantity, namely a surface area (in hectares). Its output and overall design are therefore quite different from the CFFDRS, for instance. It is also possible to run an ensemble of DeepFire predictions to quantify uncertainty, although at lower spatial resolution (640 m). Compared to BP modeling, a noteworthy difference is that DeepFire focuses on "sources" of fire spread (i.e. how large will be the burned surface, assuming that an ignition will occur at a specific location), while burn probabilities represent "sinks" of fire spread (i.e. how often a specific location is likely to be burned, as a result of several scenarios of fire spread each with variable ignition location).

A key aspect of this work is the the availability of high-resolution weather forecasts, which is arguably a major limitation for application of a BP approach to short-term predictions. In the frame of this study, a limited-area model running at 600-m resolution is used to generate a forecast of surface wind, fuel moisture, and air temperature with a time extent of 42 hours. In order to be representative of operational applications, this high-resolution weather forecast was run daily (at midnight) from 10/05/2017 and was available around 11:00. To account for rapidly changing conditions, a high-frequency simulation output with a time step of 10 minutes is provided, resulting in time-series of 253 spatial fields of weather variables, which are forecasted every day.

Section V.2 details the method that is used to obtain predictions of fire danger with DeepFire and with IEP. Analysis of forecasts for days were relatively large fires occurred is presented in Section V.3. A detailed analysis is first provided for one fire, then more synthetically for 12 other fires. Section V.4 addresses the question of generating daily maps that indicate the danger for the coming day in order to summarize the high number of maps resulting from the high forecast frequency. Finally, the main

conclusions of this study and some perspectives are given in Section V.5.

V.2. Method

V.2.1. DeepFire: fire size emulation using a deep neural network

The new approach proposed to quantify fire danger consists in predicting, for a given time and location, the size of the burned surface that would result from one hour of free wildfire spread after an early stage where the fire has already ignited and spread over about one acre. Ignition probability is assumed homogeneous, except for locations without vegetation where it is considered null. A duration of one hour is generally more than the time necessary for the first attack on the fire to be carried out, even more so if one assumes that the fire has been detected in the early stage. These simplifying assumptions imply that fire danger mostly expresses potential for fire spread if it is not attacked rather than potential for ignition.

Input	Symbol	Unit	Type	Range	Constraint
Wind speed	(W_x, W_y)	$\mathrm{m\,s^{-1}}$	Raw	$[-35, 35]^2$	Euclidean norm ≤ 35
Fuel moisture content (dead fuel)	m_c		Raw	$[0.04, 0.3]$	
Heat of combustion perturbation	ΔH	$\mathrm{MJ\,kg^{-1}}$	Additive	$[-5, 5]$	
Particle density perturbation	ρ_p	$\mathrm{kg\,m^{-3}}$	Additive	$[-300, 300]$	
Fuel height perturbations	h	m	Multiplicative	$[0.4, 1.6]^{13}$	
Fuel load perturbations	σ_f	$\mathrm{kg\,m^{-2}}$	Multiplicative	$[0.4, 1.6]^{13}$	
Surface-volume ratio perturbations	S_v	$\mathrm{m^{-1}}$	Multiplicative	$[0.4, 1.6]^{13}$	
Relative ignition point coordinates	$(\delta x, \delta y)$	m	Raw	In a fuel cell	

Table V.1. – Scalar inputs of DeepFire. In the case of perturbations, the symbol corresponds to the perturbed quantity, and the perturbation of this quantity can be either additive or multiplicative. The range indicates the boundaries of the domain of definition with two components for the wind and 13 components in the last three rows (one row per fuel type).

In the present study, the one-hour fire size is estimated using the DNN *DeepFire*. Its architecture is "hybrid" in the sense that it has two types of input data. The first type is similar to an image: it is composed of four rasters describing the landscape with a spatial resolution of about 80 m in a square of 20 km × 20 km centered around the ignition location. One raster represents elevation and the other three are derived from the same land cover data. Aside from non-burnable areas, each type of land use is matched with a fuel behavior model, for a total of 13 different fuel models. The three fields other than elevation therefore represent the variable parameters of the fuel models, namely, the height h, the load σ_f, and the surface-volume ratio S_v. The second type or inputs are scalars, listed in Table V.1, that describe physical quantities influencing fire behavior: the wind speed vector (W_x, W_y), the FMC m_c, and coefficients used to perturb the default values of fuel model parameters. Two fuel parameters, namely the heat of combustion ΔH and the particle density ρ_p, have the same default value for all fuel

models and only have one additive perturbation coefficient each. The other parameters (h, σ_f, and S_v) depend on the fuel model and each have one multiplicative perturbation coefficient per fuel type, which, including the previous two coefficients, amounts to 41 perturbation coefficients. The remaining scalar inputs are two coordinates that locate the ignition point inside the fuel cell of the land cover raster where fire starts, but given the small size of the fuel cells, these two inputs barely have any influence on fire size estimation.

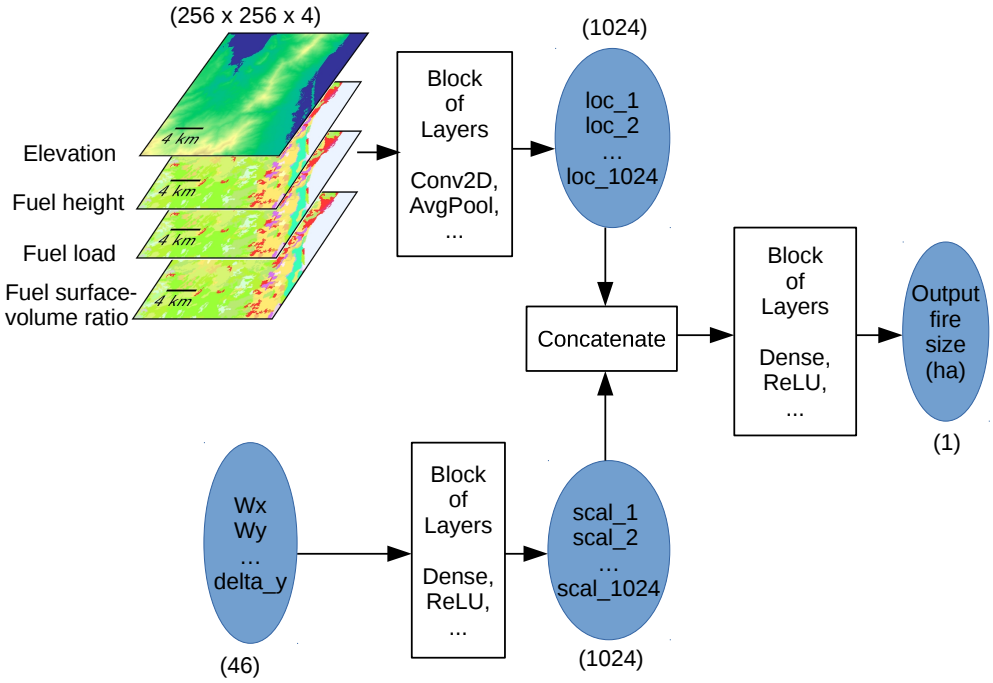


Figure V.1. – Architecture of the DeepFire neural network.

Each type of input is first processed differently by DeepFire, a summarized representation of the neural network architecture being shown in Figure V.1. Several convolutional layers and average pooling layers (among others) are applied to the rasters while the scalars mostly go through a dense layer and a rectified linear unit (ReLU) layer. This results in two intermediate vectors of size 1024: the first vector only depends on the terrain around ignition point, whereas the second vector depends on the scalar inputs. The rest of the network consists in concatenating both vectors, and applying several dense and ReLU layers, for finally providing an estimate of the fire size. The parameters of the network were fitted so that DeepFire returns a good approximation of the one-hour fire size that would be simulated by the fire spread solver ForeFire [45] using the rate of spread model of Rothermel [114] with the same inputs. In the simulations, the initial burned surface is a 0.45-ha octagon, which is also the minimum returned by both ForeFire and DeepFire (this generally happens when the FMC is close to the moisture of extinction). DeepFire is an emulator (aka, metamodel or surrogate model) of the one-hour fire size returned by ForeFire, that is to say a model that computes this output considerably faster. The first intermediate vector of "location" is pre-computed for each fuel cell in the region studied (about 1.2 million

fuel cells for Corsica island). Based on these pre-computed vectors, the one-hour fire size can be computed for all cells in about one minute for any ignition location, which is thousands of times faster than with ForeFire and makes daily fire danger mapping in operational conditions technically possible. More details on the neural network and its use for emulation of fire spread simulation can be found in Chapter IV.

V.2.2. Deterministic predictions of fire danger maps

The design of DeepFire allows for fire danger mapping of Corsica island at high spatial resolution. First, for a given fuel cell, the intermediate location vector (that was pre-computed) is used as input of the concatenate block. The ignition point is assumed to be located at the center of the fuel cell, which defines δ_x and δ_y. Then, scalar fuel parameters perturbations are set to 0 and 1 for additive and multiplicative coefficients, respectively, which corresponds to no alteration of the default values of the fuel model parameters. Finally, wind speed and FMC are quantities that vary during the day, so they are estimated based on a weather forecast. A spatial interpolation can be carried out to determine the value of these three inputs at the center of a given fuel cell to address differences in spatial resolution and/or coordinate system (in the present study, a barycentric interpolation is used). In extreme cases where the wind speed norm is higher than $35\,\mathrm{m\,s^{-1}}$, wind speed is reduced to this threshold because DeepFire was not trained based on simulations outside this scope. Similarly, when the FMC is lower than 0.04, this lower bound is used as input of DeepFire. Also, the minimum output value (0.45 ha) is automatically returned when the FMC is higher than 0.3, which is the moisture of extinction used in ForeFire simulations. The computation of the DeepFire output with the aforementioned inputs corresponding to each fuel cell of the map and each forecast time of the weather forecast results in a time sequence of fire danger maps at high spatial resolution. Given that no perturbations are applied to the scalar inputs, this is referred to as a *deterministic prediction* of fire danger in the following.

Using the same weather forecast as for DeepFire, a time sequence of fire danger maps at high spatial resolution based on the IEP can be obtained similarly. IEP is one of the operational danger rating indicator in France that serves for both summer and winter. The design of IEP is mostly empirical and based on phenological assumptions. In the autumn, as soon as days become shorter and temperatures lower, the vegetation transforms itself again. Deciduous trees lose their leaves which accumulate on soil and the forest litter thicken. Other trees, shrubs and bushes 'hibernate' and become in dormancy with a very low quantity of sap/water in their leaves. At the same time, herbaceous are at the end of their cycle and die. In addition to this traditional and seasonal dormancy, with climate warming, more and more vegetation dies with huge and record summer drought. The dead or dormancy vegetation water content is directly linked with air humidity (as a sponge effect). In this case, the IEPx (daily maximum) is the most relevant indicator. In the 2010s, some huge fires occurred with moderate or just severe weather fire conditions in places where dead vegetation were numerous. By studying indices and conditions in a file constituted by more than 5000 summer alerts from 2001, Météo-France set up an indicator by crossing FFMC, correlated with outbreak potential, and wind. Since then, by empiricism and following winter and summer situations, with Fire Brigades and National Forest officers, it was concluded that this indicator was the most relevant indicator for winter fires and a good complement in

summer for dead or dormancy vegetation in very dry conditions (as it was the case in summer 2017 in Corsica).

Wind speed \ FFMC	≤ 80	]80, 85]	]85, 89]	]89, 93]	]93, 96]	> 96
≤ 10 kt	low (1)	low (1)	minor (2)	moderate (3)	severe (4)	severe (4)
]10 kt, 20 kt]	low (1)	minor (2)	moderate (3)	severe (4)	severe (4)	very severe (5)
]20 kt, 30 kt]	low (1)	minor (2)	moderate (3)	severe (4)	very severe (5)	very severe (5)
> 30 kt	minor (2)	moderate (3)	severe (4)	very severe (5)	very severe (5)	very severe (5)

Table V.2. – Computation of the IEP based on the norm of wind speed (in knots) and the FFMC.

The IEP has five possible categories ranging from "low" to "very severe" that only depend on the norm of wind speed and the FFMC as determined from Table V.2. Let us underline that both FMC and FFMC are diagnostic variables of the weather forecast, i.e. they are deduced from the other variables. The FFMC is generally based on the forecast for 12:00 in local time, but this method does not account for variations of fuel moisture during the day, whereas the forecasted FMC used as input of DeepFire may vary over time. For consistency, we decided not to compute the input FFMC of the IEP in the "standard" way but based on the value of FMC m_c, according to Equation (V.1):

$$\text{FFMC} = 59.5 \times \frac{250 - m_c}{147.2 + m_c}, \tag{V.1}$$

which is the final equation intervening in calculation of the FFMC in the Canadian system (cf. Equation (10) in [138], where the other equations are used to provide an estimate of m_c).

Categories of DeepFire values were determined based on thresholds of burned surface area to provide maps with a categorical color scale for easier interpretation. For simplicity, the thresholds of the US classification of fire size [5] were chosen, providing seven categories ranging from A to G, as reported in Table V.3. This choice is not perfectly suited to DeepFire which estimates the one-hour fire size of a freely spreading wildfire, whereas the US classification accounts for all observed fires, even if there were firefighting actions or the fire did not last one hour. The minimum value returned by ForeFire is 0.45 ha, which implies that class A will not be obtained in the present study at all. Similarly, class G should be quite rare as it is only returned by DeepFire given extreme input conditions (very high wind speed and very low FMC). Still, classes B to F should all contain at least a fair proportion of DeepFire predictions so this choice, although arbitrary, makes sense as a first approach. In practice, the thresholds should be calibrated in order to provide a fire danger indicator that is as relevant as possible. Finally, for qualitative comparison of danger class predicted by DeepFire and IEP, a correspondence reported in Table V.3 was made.

5. cf. https://www.nwcg.gov/term/glossary/size-class-of-fire, last accessed on 2020.12.11

Fire size (ha)	< 0.1	[0.1, 4.0[	[4.0, 40.5[	[40.5, 121.4[	[121.4, 404.7[	[404.7, 2023.4[	≥ 2023.4
Category	A	B	C	D	E	F	G
Corresponding IEP	None	low (1)	minor (2)	moderate (3)	severe (4)	very severe (5)	very severe (5)

Table V.3. – Values of fire sizes used to determine a class of fire danger. The fire size classes follow the US classification and a correspondence of fire danger classes was made between DeepFire and IEP in the present study.

V.2.3. Probabilistic predictions representing input uncertainty

A major aspect that is not accounted for by the deterministic predictions is the uncertainty associated to the modelling process, that could lead to a difference between the "actual" one-hour fire size that would be observed and its estimate. Although it may not be possible to determine the actual value (firefighting actions usually occur, it can be difficult to measure, etc.), several sources of uncertainty can be identified such as the error of approximation between DeepFire and ForeFire and simplifying model assumptions. Arguably, the main source of uncertainty regarding DeepFire predictions stems from the weather predictions and values of fuel model parameters, that is to say input uncertainty.

Previous studies [5, 7] focused on the quantification of input uncertainty in simulations of wildland fire spread. To quantify the uncertainty in the inputs of DeepFire, we use a *calibrated* probability distribution associated to the inputs of ForeFire simulations. The calibration method is detailed in [7]; its key aspect consists in using an *a priori* distribution where the marginals of each individual input have higher variance than in [5] and including the information of observed burned surfaces of seven Corsican fires. The *probability density function* (PDF) g of the calibrated distribution is intended to be higher than the prior PDF f for a given input vector $\boldsymbol{u}$ when there is good overall agreement between observed burned surface $\mathcal{S}_{\text{obs}}$ and corresponding simulated burned surface $\mathcal{S}_{\boldsymbol{u}}$. More precisely, the PDF g can be expressed as follows:

$$g(\boldsymbol{u}) = \frac{e^{-\beta E(\boldsymbol{u})} f(\boldsymbol{u})}{\int e^{-\beta E(\boldsymbol{u})} f(\boldsymbol{u}) d\boldsymbol{u}}, \tag{V.2}$$

where $\beta > 0$ and E is a positive "energy" function that is equal to 0 when $\mathcal{S}_{\boldsymbol{u}} = \mathcal{S}_{\text{obs}}$ for each fire and increases with the dissimilarity between simulated and observed burned surfaces. Measurement of shape dissimilarity relies on the Wasserstein distance, which is a metric that appears in the field of optimal transport (see, for instance, [120] for an extensive review), and $E(\boldsymbol{u})$ can be understood as an energy that is required to transform $\mathcal{S}_{\boldsymbol{u}}$ into $\mathcal{S}_{\text{obs}}$. The definition of g is inspired from Bayes' rule:

$$p(\boldsymbol{u}|\mathcal{S}_{\text{obs}}) = \frac{\mathcal{L}(\mathcal{S}_{\text{obs}}|\boldsymbol{u}) f(\boldsymbol{u})}{\int \mathcal{L}(\mathcal{S}_{\text{obs}}|\boldsymbol{u}) f(\boldsymbol{u}) d\boldsymbol{u}}, \tag{V.3}$$

where $p(.|\mathcal{S}_{\text{obs}})$ is the posterior PDF that would be obtained from $\mathcal{L}(\mathcal{S}_{\text{obs}}|.)$, the likelihood of the observation, and f, the prior PDF. The exponential in Equation (V.2) can therefore be seen as a pseudo-likelihood function, whose weight increases with the parameter β. When $\beta = 0$, the calibrated PDF g is equal to the prior PDF f.

In the present study, the calibrated input distribution obtained with $\beta = 1/2$ is used. The components of the input vector $\boldsymbol{u}$ are similar to the scalar inputs of DeepFire listed in Table V.1. The perturbations in fuel parameters are exactly the same and can be used directly as input of DeepFire instead of the values 0 and 1 used for the deterministic prediction. Regarding wind speed and FMC, however, the corresponding components of $\boldsymbol{u}$ are perturbation coefficients so that each value of W_x, W_y, and m_c in the weather forecast is perturbed using the corresponding components of $\boldsymbol{u}$. As for the deterministic prediction, a threshold is applied so that the perturbed value of wind speed norm and FMC fall in the range used to train DeepFire. By sampling n independent sets of perturbation coefficients $(\boldsymbol{u}_1, ..., \boldsymbol{u}_n)$ following the calibrated distribution, one may obtain n replicas of the deterministic DeepFire predictions but using perturbed inputs so that each of these n predictions is different. The resulting ensemble of n DeepFire predictions is referred to as a *probabilistic prediction* in the following.

A probabilistic prediction based on a large ensemble can be fairly easily obtained when one considers a single ignition location. However, computing an ensemble of M high-resolution maps requires M times more computational time, making the generation of a representative ensemble too long for operational context conditions. To circumvent this issue, the spatial resolution of the fire danger maps can be reduced. For each large cell and each forecast time, we define a "representative cell": the small cell which has the highest DeepFire value according to the deterministic prediction. For each large cell, the computations of the probabilistic prediction is only carried out for the small representative cell. In case the large cell contains a high number of non-burnable cells, it may be considered as non-burnable as well (here, it is considered to be the case when at least 90% of the cells are non-burnable). It must also be noted that these ensembles are distinct from atmospheric forecast ensembles, while the later may also be used as inputs for these simulations to account for weather uncertainties, all fire danger predictions for a given day in the present study, whether deterministic or probabilistic, were based on a single deterministic weather forecast.

V.2.4. Application to weather forecasts corresponding to 13 Corsican fires

Daily forecasts are generated with a high-resolution run of the limited area model Meso-NH version 5.4 [80]. While being the support of coupled fire/weather simulations, Meso-NH is also the atmospheric research model of the French community, maintained and developed by two laboratories associated on the research program (Centre National de la Recherche Scientifique and Laboratoire d'Aérologie) that provided access to operational runs providing the boundary and initial conditions. Boundary and initial conditions are gathered from the Météo-France archive of the AROME model [8]. The 00Z run is downloaded daily from the system and is available at around 3AM. The daily computation is performed by Meso-NH on a Cartesian regular grid at 600 meter resolution and 46 atmospheric levels (up to an altitude of 7000 meters). In the application over Corsica this resulted in a horizontal grid of 300 by 180 points. Overall computation of the 42 hours of forecast takes an average of 2 hours on 360 CPU. In these runs, Meso-NH is run coupled to ForeFire[45] code in order to provide high frequency (2 minutes) model state outputs, resulting in 252 output files containing water, temperature, cloud fraction, turbulence and U,V and W wind fields. Such high frequency outputs is an important point to, at the same time, display the complexity of

local situations that can locally be very perturbed by orography and rapidly changing situations that may be critical on the field.

Still, as these forecasts cover more than one day, we focused more particularly on the prediction between T+6 h and T+30 h. The spatial resolution of the probabilistic DeepFire predictions is 640 m to reduce computational time, instead of 80 m for the deterministic predictions (cf. Section V.2.3). This allows to compute an ensemble of 100 maps in a time that is about twice as long as for computing a deterministic map (presented computed duration for non-emulated maps are estimations based on a limited number of simulations, full computations were not run). For specific ignition locations, a larger ensemble of size 10,000 is computed as well. The characteristics of both deterministic and probabilistic predictions are summarized in Table V.4.

Type of prediction	Ignition points	Spatial resolution	Total number of time steps	Number of time steps in 24 hours	Ensemble members
Deterministic	1.2×10^6	80 m	253	145	1
Probabilistic	2.0×10^4	640 m	253	145	100 (10,000 locally)
Type of prediction	**Total number of fire danger estimations**	**Computational time DeepFire, total**	**Computational time ForeFire, total**	**Computational time DeepFire, 1 forecast step**	**Computational time ForeFire, 1 forecast step**
Deterministic	3.0×10^8	2 hours	4 months	30 seconds	half a day
Probabilistic	5.1×10^8	3.4 hours	8 months	1 minute	one day

Table V.4. – Characteristics of deterministic and probabilistic predictions and estimation of computational time.

The predictions are carried out and analyzed for 13 daily weather forecasts, each corresponding to the occurrence of a wildfire of 100 ha or more in Corsica island. Data on these 13 fires are available in the French database Prométhée [6] and some are reported in Table V.5. Regarding the "time of fire alert" reported in Prométhée, one may assume that, before this time, the fire has ignited and started spreading and, although its size at the time is unknown, it may be close to the initial fire size of 0.45 ha assumed in DeepFire estimations. The fire size in Table V.5 corresponds to the final burned surface, which results from firefighting actions and fire spread for generally more than one hour, so it does not make much sense to compare it with DeepFire predictions. Depending on the fire, the coordinates of the ignition point may be available, and in any case, its estimated location is at least identified by a square of 2 km × 2 km. When only the latter information is available, it will be assumed that the ignition point is located at the center of the square. In either situation, the ignition point coordinates may not be perfectly accurate and one may assume that there could be an error ranging between 100 m and 1 km. For the vil_2017 fire case, the fire started on 20-10-2017 (the date indicated in Prométhée) and only spread a little then it started spreading again two days later from which it burned most of the final observed burned surface; therefore, corrected estimations of time of fire start and ignition location are reported for this fire instead of the ones available in Prométhée.

The 13 fires studied all occurred between May 2017, date from which the weather forecasts were running, and late 2019. This relatively long data collection period ensures

6. cf. https://www.promethee.com/, last accessed 2021.01.21

Fire id	Town	Date and time of first alert (UTC)	Fire size (ha)	Fire season
bon_05-2017	Bonifacio	30-05-2017 11h43	380.0	summer
bon_07-2017	Bonifacio	17-07-2017 11h50	121.8	summer
olm_2017	Olmeta di Tuda	24-07-2017 10h49	2118.0	summer
pal_2017	Palneca	02-08-2017 13h48	185.5	summer
cal_2017	Calenzana	05-08-2017 15h42	124.6	summer
non_2017	Nonza	10-08-2017 22h45	1617.0	summer
man_2017	Manso	11-08-2017 13h10	109.0	summer
sai_2017	Sainte Lucie de Tallano	15-08-2017 09h42	130.0	summer
vil_2017	Ville di Paraso	22-10-2017 09h00	1517.8	summer
ghi_2017	Ghisoni	26-10-2017 13h37	526.0	winter
san_2018	Saint'Andrea di Cotone	01-02-2018 04h13	1317.2	winter
chi_2018	Chiatra	03-01-2018 18h44	565.7	winter
cal_2019	Calenzana	23-02-2019 19h03	1110.2	winter

Table V.5. – Information on the 13 fires studied, sorted in chronological order. The "Fire id" is used to identify the fires in the present study, the one in bold being the most detailed in the following sections.

a relative representativeness of high fire danger meteorological conditions. Among these 13 fires, nine of them are considered to have occurred during hot, windy and dry events in the summer fire season, and the remaining four during windy events in winter fire season. In the present study, no weather forecast corresponds to a day without at least one big fire. In these situations, one may consider that fire danger was high at least at some point during the event for a fire size of at least 100 ha to be observed, especially during the summer fire season during for which more firefighting forces are mobilized to carry out the initial attack as early as possible. Availability of firefighting forces are obviously a strong factor in having a reduced burned area, a notable example being bon_05-2017, a fire happening on a low-danger day in May (no fire brigades were deployed in the territory), yet in a rather remote area so that the fire propagated freely for several hours and was barely fought.

Among those 13 fires, olm_2017, non_2017, vil_2017, chi_2018, san_2018 and cal_2019 are wind-driven fires, but with notable differences concerning the last 3 winter fires that were the result of extremely high (>150 km h^{-}1) wind creating intense hot and dry down slope conditions. Ghi_2017 is a different situation, with a rather high wind and dry day (not an unusual situation in autumn) but a fire started in large areas of grounded deciduous leaves that were not yet wet as no rain had fallen for a month. This created a not very intense fire, but with a lot of spotting in a mountainous area that were difficult to fight for several days. Other situations were typical of hot dry day with some winds with less danger and resulted in smaller burned area due to the efficiency of fire fighting.

This database is yet representative of 3 years of active fire seasons with unusual, mild

and severe cases, chosen thereof as a first approach to evaluate the ability to predict high danger and analyze predictions obtained using DeepFire. Although this implies that the database is somewhat biased, both winter and summer seasons are represented and the weather conditions are quite diverse from a case to another. Also, given the time range of the forecasts, there are also moments when the predicted weather conditions are less prone to fire spread.

V.3. Analysis of predictions for specific fires

Although IEP and DeepFire serve similar purposes, results are not directly comparable due to their different design. This section tries to tackle this issue by analyzing fire danger predictions, illustrated by figures, that address specific questions. Given the ability of the fire danger prediction system to provide very local values, we first try analyze what is the predicted fire danger in the neighborhood of the ignition location (cf. Figure V.5). Then, by leveraging the high temporal resolution, we investigate how the predicted fire danger evolves over time, locally (cf. Figure V.4 & V.5), and for the whole simulation domain Figure V.6. Using ensembles it is also possible to estimate how uncertain the prediction is (cf. Figure V.8 at the ignition location and time of fire alert, Figure V.7 at the same location over time, Figure V.9. and on the whole domain, cf. Figure V.11, V.12, V.13 at the time of fire start).

Comparison with the IEP as a reference danger rating system is also of importance but not only as a pure numerical output. As being part of a differently designed system, the IEP does not address the same issues. While investigating the ability to forecast local danger, the following comparative analysis also tries to illustrate the interest of having model outputs that are both probabilistic and quantifiable in hectares.

The first part of this section illustrates this difference in designs with a detailed analysis a single case (Calenzana 2017) before a mode synthetic comparison is presented for the other 12 cases.

V.3.1. A detailed fire case: Calenzana 2017

Calenzana 2017 fire took place at 15h42 (UTC) on August 5th, and burned a modest 119 ha but the most part was burned in less than two hours. Figure V.2 presents the meteorological situation that rapidly changed in the area, thus creating a high weather fire danger. At 12:00 a sea breeze was keeping temperatures and air relative humidity within a typical summer day, nevertheless, south-westerly winds strengthened to more than $10\,\mathrm{m\,s^{-1}}$ at 12:30, generating a down slope Foehn effect that took over the sea breeze with higher temperatures and air moisture under 30% at 13:00, which persisted up to the time of reported ignition at 15:40. The phenomenon was also observed at the nearby weather station of Calvi Airport Figure V.3 although a little later and more quickly than in forecast.

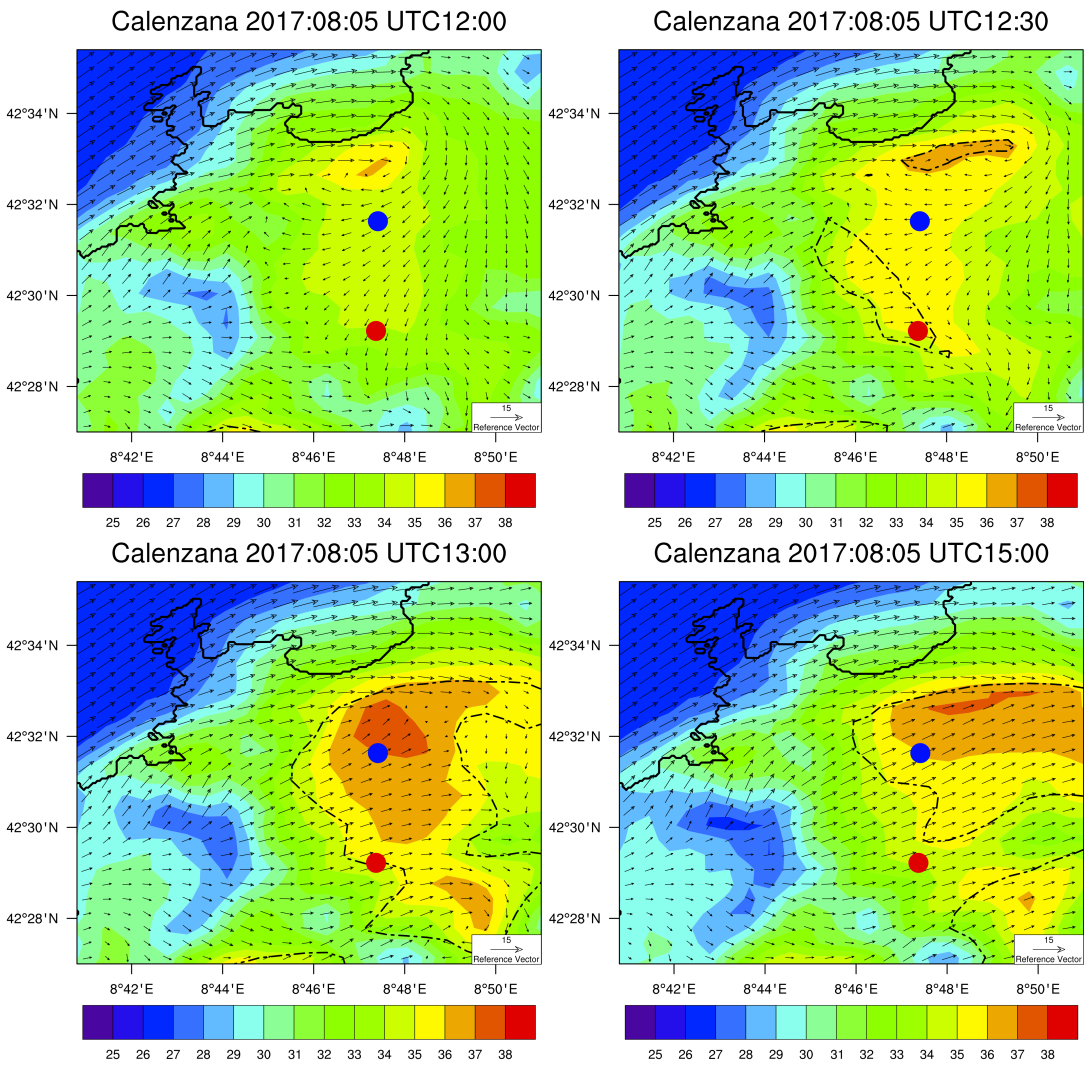


Figure V.2. – Model forecast at 12:00, 12:30, 13:00 and 15:00 on 08-05-2017 with surface temperature (color), wind (arrows) and area under 30% of relative humidity (dashed lines). Blue dot is the location of Calvi Airport, red dot is the fire ignition location.

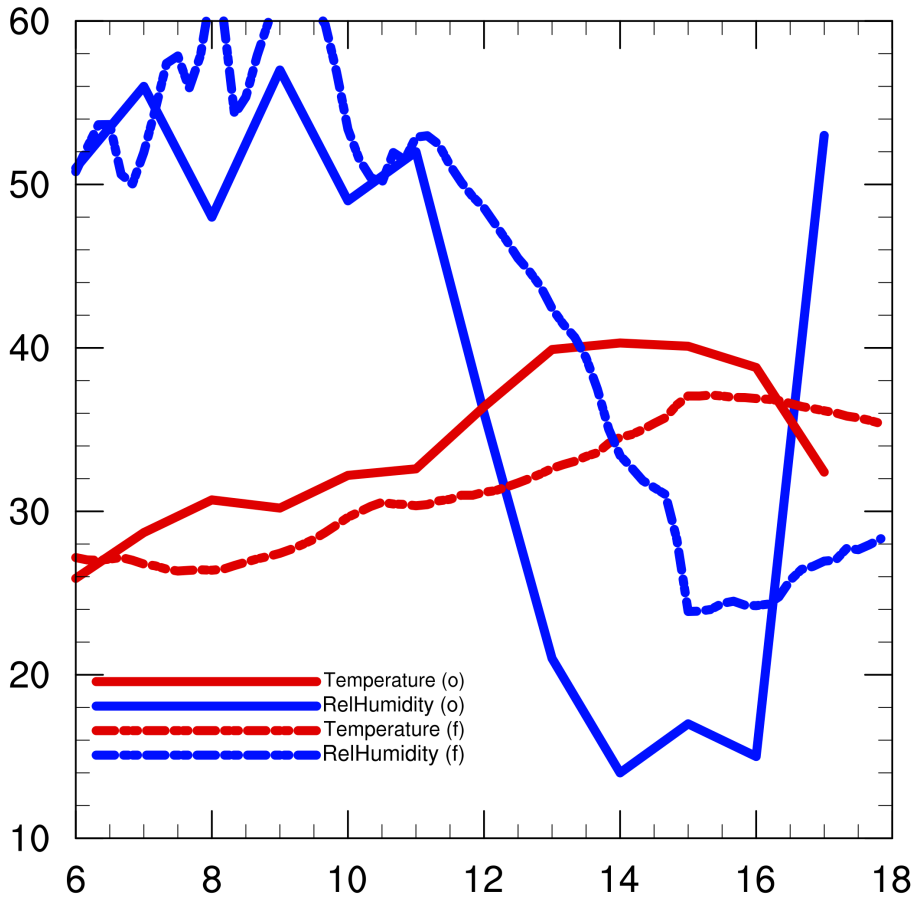
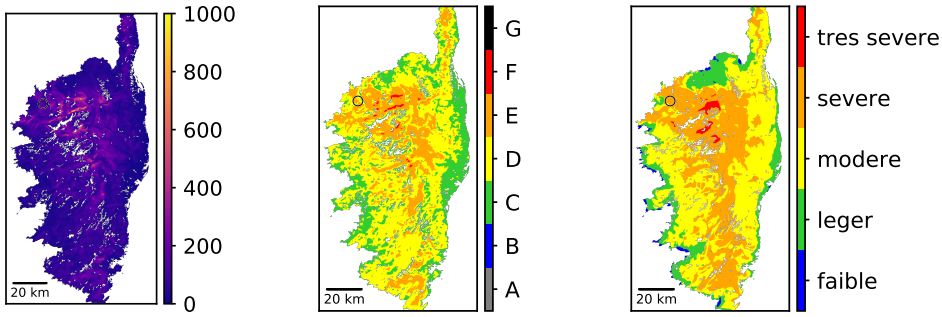
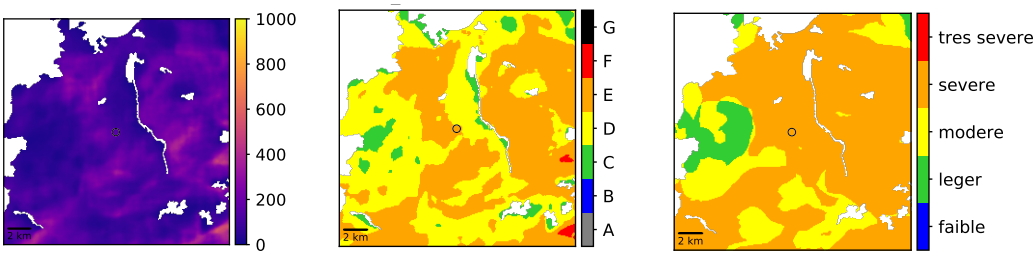


Figure V.3. – Relative humidity (blue) and temperature (red) observations (plain lines) and forecasts (dashed lines) at Calvi airport (blue dot in Figure V.2) on 08-05-2017.



(a) DeepFire (continuous) (b) DeepFire (categorical) (c) IEP

Figure V.4. – Deterministic maps predicted based on the weather data of forecast time T+15:40 UTC (right before presumed fire start). In white are the locations that are not vegetation, where it is assumed that neither ignition nor spread is possible.
The black circle indicates the presumed location of the ignition point.

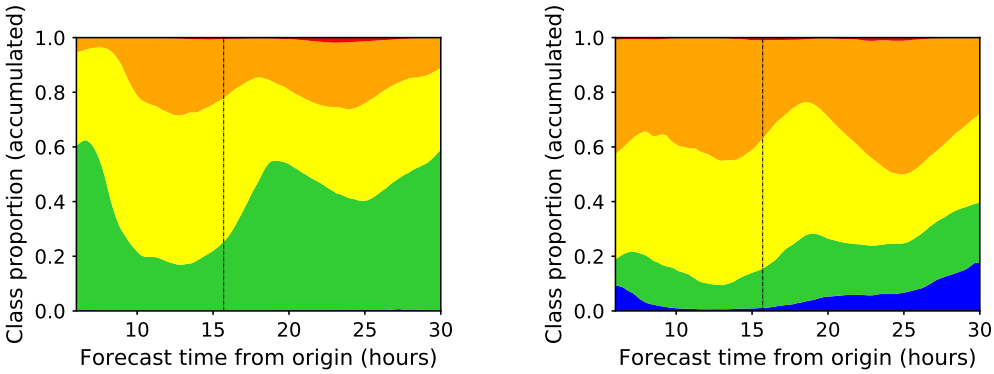


(a) DeepFire (continuous) (b) DeepFire (categorical) (c) IEP

Figure V.5. – Same as Figure V.4 but zoomed around the ignition point

The deterministic fire danger maps of both DeepFire and IEP predicted at the time of fire start, are represented in Figure V.4. The DeepFire value in hectares is shown in Figure V.4a with a continuous color scale. The locations in white indicate the absence of vegetation, where it is assumed that neither ignition nor spread is possible. Among the locations where fuel is present, the prediction ranges in three orders of magnitude in this specific map (from less than 1 ha to several hundreds), making it difficult to highlight the differences in fire danger with a continuous color scale. This appears more clearly with the categorical color scale in Figure V.4b, where most of the island is in class D or E. Based on the correspondence between danger classes, the same can be seen on the IEP map shown in Figure V.4c. Although these two maps do not perfectly match, their respective categories were determined independently, so we could expect more blatant differences. This overall similarity is probably due to the fact both indices are very sensitive to wind speed and FMC. A notable difference is that the map of IEP seems more regular than that of DeepFire. This is most likely due to the fact that the "driving" input field of IEP is the Meso-NH forecast whose original spatial resolution is

600 m, while DeepFire also accounts for a terrain field at an original spatial resolution of 80 m. Another consequence of the sensitivity of DeepFire to the terrain is more visible in Figure V.5 which is zoomed around the ignition point. In Figure V.5b, the ignition point is in class E, but slightly to the north-east of this point, some locations are in class D or even C. This is due to the forecasted wind in this region, which is oriented to the north-east where there is a long area from north to south with no vegetation where the fire cannot spread, hence the lower potential fire size at the aforementioned locations. In Figure V.5c, it clearly appears that the IEP does not account for this characteristic, as these locations are in the same danger class as that of the ignition point (4: "severe").



(a) DeepFire (b) IEP

Figure V.6. – Proportion of points on the island predicted in a given area/danger category as a function of forecast time between T+6 and T+30. The vertical line indicates the time of fire start.
(a) Categories of burned surface area range from A to G, cf. colorbar in Figure V.4c.
(b) Categories of IEP range from "faible" to "très sévère", cf. colorbar in Figure V.4b.

To summarize the overall of fire danger on the island, one may compute the proportion of locations in each danger category. This "spatial" distribution is represented between T+6 and T+30 in Figure V.6 for both DeepFire and IEP categories. The proportion of locations in DeepFire class E is lower than the proportion for IEP class 4 at all times, but interestingly, they increase or decrease at almost the same times. For both indices, fire danger seems to decrease on most of the island around the time of fire start, whereas the evolution of both deterministic and probabilistic DeepFire predictions at the ignition location represented in Figure V.7 shows the opposite trend. At this location, the DeepFire deterministic prediction is fairly low until T+13 where an abrupt increase by about 100 ha occurs in only one hour and stays in class E until it decreases at the end of the night. This sudden increase is yet well-connected with the local conditions and the sudden increase in temperature and lower relative humidity occurring locally near the ignition area as exposed in Figure V.2.

Focusing on the distribution of DeepFire values, the histogram of the probabilistic prediction obtained at the ignition location at the time of fire start is represented

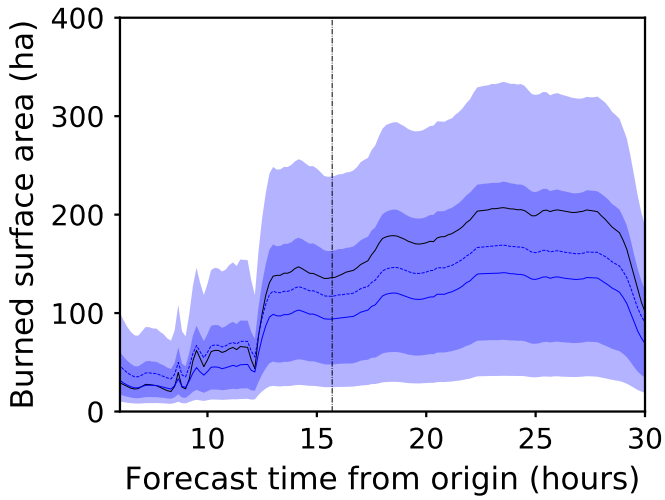


Figure V.7. – Ensemble of 10,000 emulated burned area (in hectares) predicted at the ignition point as a function of forecast time between T+6 and T+30. The vertical line indicates the time of fire start.
Solid black line: deterministic prediction; solid blue line: mean; dotted blue line: median.
Shaded areas delineate quantiles: first and third quartiles (dark blue); and the first and ninth deciles (light blue).

in Figure V.8. The raw histogram is right-skewed and has a single mode; it may resemble a log-normal distribution at first glance, but the empirical distribution of its logarithm is fairly left-skewed instead of being symmetrical. At several other time steps and/or locations, however, a log-normal distribution fits the predicted DeepFire ensemble quite well, as shown in Figure V.9a. A good agreement was obtained for most DeepFire ensembles when fitting the parameters of a Johnson's S_U distribution to the logarithm of DeepFire values by maximum likelihood estimation. This family includes normal distribution, hence the good fit in several occasions, but is also well-suited to the left-skewed distribution obtained at the ignition location and time of fire start. Still, although a Johnson's S_U distribution generally allows for a decent agreement, the shape of the fitted PDF associated to either the raw DeepFire value or its logarithm could match the histogram better around the PDF's mode when the distribution is skewed as in Figure V.8. Also, the ensemble in Figure V.9b shows that the minimum DeepFire value (0.45 ha) may be predicted for several ensemble members, resulting in a Dirac at this value that a Johnson's S_U distribution (which is continuous) cannot account for.

The distribution of DeepFire ensembles cannot be simply summarized by a simple distribution such as a log-normal. However, focusing on the ensemble mean and standard deviation reveals that over the time range of the forecast, the coefficient of variation of the DeepFire ensemble predicted at the ignition location remains between 0.7 and 1.1. For other times and/or locations, the standard deviation of the ensemble

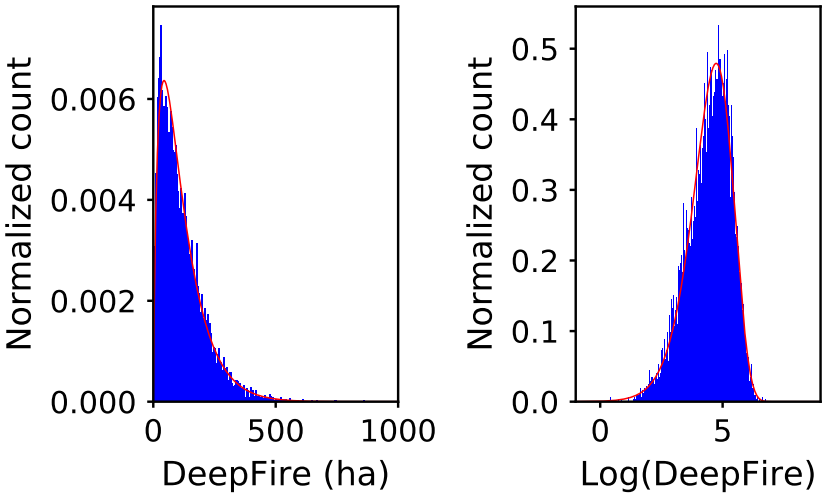
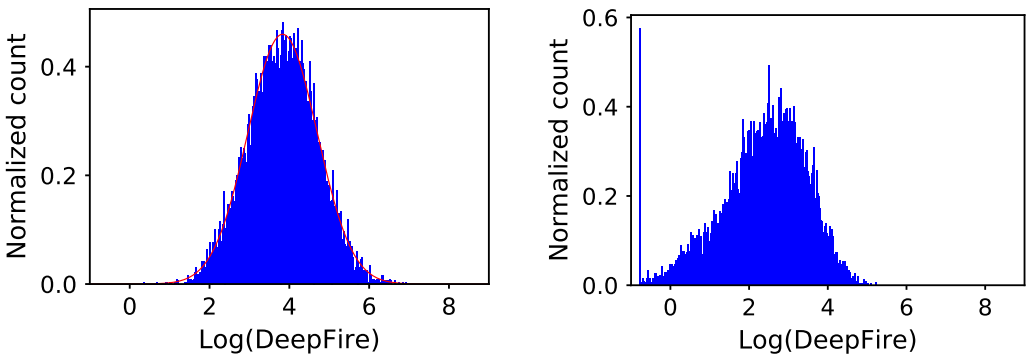


Figure V.8. – Distribution of the ensemble of 10,000 DeepFire values predicted at the time of fire start at the ignition location.
The red curves are the PDFs obtained from fitting the parameters of a Johnson's S_U distribution to the logarithm of DeepFire values by maximum likelihood estimation.



(a) Forecast time: T+18h20

(b) Forecast time: T+3h20

Figure V.9. – Distribution of the ensemble of 10,000 logarithm of DeepFire values predicted for the cal_2017 fire case. Contrary to Figure V.8, the location is not the ignition point.
(a) The red curve is the PDF obtained from fitting the parameters of a normal distribution to the logarithm of DeepFire values by maximum likelihood estimation.

is in the same order of magnitude as its mean as well, so one may focus on the latter for simplicity. The evolution of the ensemble mean at the ignition location is represented in Figure V.10 with the deterministic prediction, showing their respective fire danger categories, as well as that of the deterministic IEP for comparison. This "concise" plot

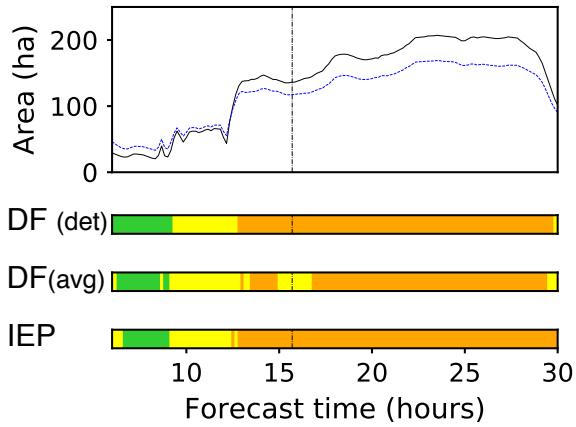


Figure V.10. – Predictions at the ignition point as a function of forecast time between T+6 and T+30. The vertical line indicates the time of fire start.
Top most: DeepFire both deterministic (solid black) and 10,000-size ensemble mean (dotted blue).
Horizontal colorbars, from top to bottom: category of the DeepFire deterministic area; category of the DeepFire mean area; IEP

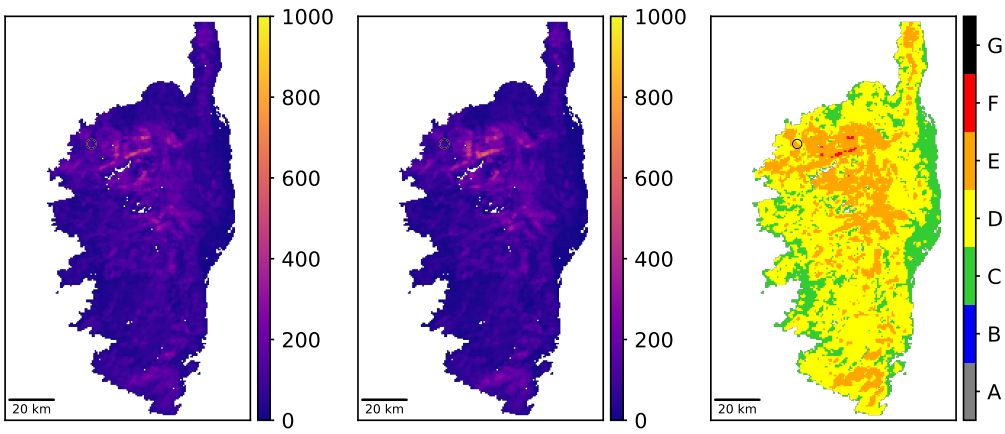


Figure V.11. – "Probabilistic" predicted maps at forecast time 15h40 UTC (right before fire start). From left to right: mean (continuous), standard deviation, mean (categorical).

reveals that the deterministic value and the ensemble mean are predicted in the same category and match with those of IEP for most of the forecast.

The probabilistic DeepFire prediction on the whole island consists in a smaller ensemble (100 members) than the local one that was analyzed previously (10,000 members). The ensemble of DeepFire maps predicted for the time of fire start is summarized in Figure V.11 by the ensemble mean and standard deviation computed for each location. Overall, the higher the mean, the higher the standard deviation, which was also the

case for the large ensemble at the ignition location. More precisely, the coefficient of variation is close to unity and for almost all locations and forecast times, it is contained in [1/5, 5] in this case. Focusing on the map of ensemble mean represented with a categorical color scale, it appears similar to its deterministic counterpart represented in Figure V.4b and does not provide much additional information.

In order to show the uncertainty represented by the probabilistic prediction by means of maps, one may consider other quantities than the ensemble mean and standard deviation. For instance, the fire size category with highest probability (aka, "most likely category") is computed for each location and shown in Figure V.12 together with the probability of this category. Like the ensemble mean, the map of most likely category is similar to its deterministic counterpart in Figure V.4b. Regarding the associated probability, most locations have a probability ranging in [0.3, 0.7]. Locations whose most likely category is C tend to have probabilities closer to unity, so despite providing an intuitive representation of uncertainty, these maps do not seem to provide useful information.

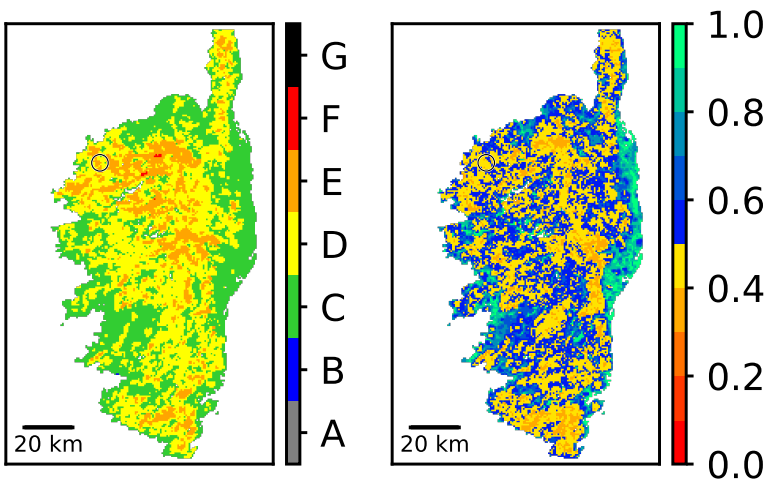


Figure V.12. – Most likely DeepFire size category in the ensemble (left) and associated probability (right)

Arguably, the probabilistic prediction is best represented by maps of quantiles computed for each location. Using the same quantiles as in Figure V.7 as well as the ensemble minimum and maximum, the resulting map are shown in Figure V.13. These maps can be understood as predicted scenarios that range from most optimistic (ensemble minimum at the left) to most pessimistic (ensemble maximum at the right). It is arguably a more intuitive way to represent the uncertainty in the prediction than with the standard deviation or the probability of the most likely category. In this case, it seems that the map of minimum and, to some extent, the map of first decile show scenarios where fire danger categories are underestimated. Likewise, the map of maximum overestimates fire danger, but the other maps where fire danger is not as extreme seem fairly relevant.

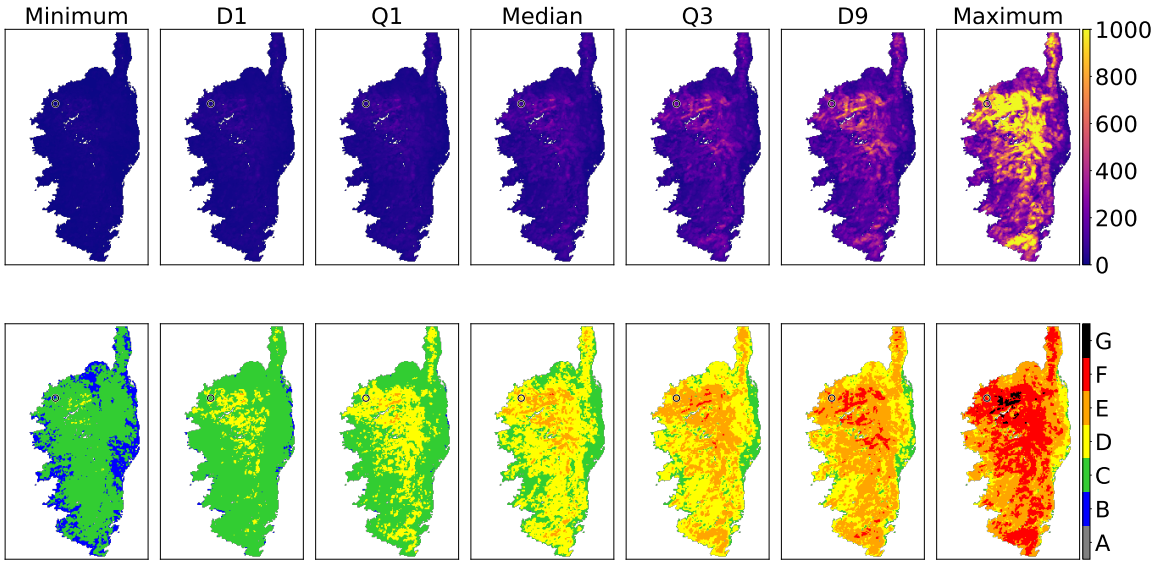


Figure V.13. – Statistics of the ensemble of DeepFire values predicted for the time of ignition mapped over Corsica island (Calenzana 2017.08.05)
Upper row: continuous; lower row: categorical
Q1 and Q3: first and third quartiles; D1 and D9: first and ninth deciles.

V.3.2. Overview of predictions for all 13 fires

The fire danger predictions for fire case cal_2017 that were presented in detail in Section V.3.1 were also carried out for the 12 other fire cases listed in Table V.5.The present section summarizes results for all 13 fire cases.

The deterministic predictions of IEP and DeepFire at the ignition location and time of fire start for each fire are listed in Table V.6, together with some statistics of the associated ensemble of 10,000 DeepFire predictions. The DeepFire ensemble mean is generally a bit lower than the deterministic value, except for two fires (ghi_2017 and cal_2019). Both ensemble mean and deterministic value are generally slightly lower than the ensemble's third quantile, except for man_2017 fire where the deterministic value is slightly higher. In terms of fire size categories, whether we consider the deterministic DeepFire value or the ensemble mean, the most likely category is either the same or one level lower. The most likely category is either C, D, or E, and in almost all cases, probabilities of being in category B and G are very low. Comparing the deterministic predictions of IEP and DeepFire, there is an agreement between fire danger categories for 9 out of 13 cases. Due to the categorical nature of these ratings, difference of one category are expected and require a more detailed analysis proposed thereafter.

Focusing on deterministic prediction at the time of fire start for each fire, the maps of both DeepFire and IEP are summarized in Table V.7 by the proportions of locations on the island predicted in either category of both fire danger indices. In all fire cases, the proportion of the island that is in class G is null or quasi-null and in most cases, the respective proportions of class F and class 5 are also quite low. Proportion of class

Fire id	IEP	DeepFire (ha)	Mean (ha)	[Q1, Q3] (ha)	$\mathbb{P}$[B]	$\mathbb{P}$[C]	$\mathbb{P}$[D]	$\mathbb{P}$[E]	$\mathbb{P}$[F]	$\mathbb{P}$[G]
bon_05-2017	3	72.6 (D)	67.2 (D)	[26.3, 86.2]	0.008	0.397	**0.464**	0.127	0.004	0
bon_07-2017	4	159.8 (E)	143.0 (E)	[60.5, 193.3]	0	0.126	**0.422**	0.415	0.037	0
olm_2017	4	368.3 (E)	272.5 (E)	[103.3, 362.5]	0	0.048	0.260	**0.478**	0.214	0.001
pal_2017	4	194.8 (E)	137.0 (E)	[54.5, 159.8]	0	0.139	**0.489**	0.324	0.048	$< 10^{-3}$
cal_2017	4	136.0 (E)	117.1 (D)	[47.8, 163.1]	$< 10^{-3}$	0.207	**0.403**	0.377	0.012	0
non_2017	3	118.9 (D)	106.7 (D)	[35.3, 130.8]	0.001	0.292	**0.433**	0.246	0.028	0
man_2017	2	38.7 (C)	27.9 (C)	[3.45, 32.8]	0.268	**0.531**	0.163	0.037	0.002	0
sai_2017	4	71.7 (D)	66.7 (D)	[25.3, 83.2]	$< 10^{-3}$	**0.439**	0.431	0.126	0.004	0
vil_2017	4	453.6 (F)	309.6 (E)	[59.6, 370.5]	$< 10^{-3}$	0.167	0.263	**0.344**	0.213	0.013
ghi_2017	3	55.9 (D)	63.6 (D)	[32.1, 77.6]	0	0.384	**0.518**	0.097	0.001	0
san_2018	3	172.3 (E)	144.2 (E)	[49.6, 182.1]	0.003	0.198	0.369	**0.378**	0.052	$< 10^{-3}$
chi_2018	4	483.0 (F)	312.5 (E)	[79.1, 405.6]	0	0.099	0.272	**0.378**	0.245	0.005
cal_2019	5	431.9 (F)	450.1 (F)	[181.5, 578.9]	0	0.007	0.121	**0.467**	0.395	0.010

Table V.6. – Summary of the fire danger predictions on all 13 fire cases at the presumed location of fire start.
Q1 and Q3: first and third quantiles.
$\mathbb{P}$[X]: predicted probability of being into class X of burned surface area; in bold is the highest probability among all 6 classes.

Fire id	B	C	D	E	F	G	1	2	3	4	5	E+	4+
bon_05-2017	4.6%	39.0%	**52.0%**	4.3%	< 0.01%	0%	32.1%	**40.5%**	23.1%	4.3%	0%	4.3%	4.3%
bon_07-2017	< 0.01%	10.0%	**52.8%**	37.0%	0.2%	0%	0.1%	3.3%	33.0%	**63.3%**	0.2%	37.2%	63.5%
olm_2017	5.1%	13.3%	39.7%	**40.6%**	1.3%	0%	9.1%	22.4%	**39.8%**	28.4%	0.3%	41.9%	28.7%
pal_2017	< 0.001%	13.1%	**47.5%**	38.9%	0.5%	0%	< 0.1%	2.1%	18.3%	**76.1%**	3.4%	39.4%	79.5%
cal_2017	< 0.01%	25.2%	**52.6%**	21.7%	0.5%	0%	1.1%	14.4%	**47.5%**	36.0%	1.0%	22.5%	37.0%
non_2017	0.8%	29.9%	**34.5%**	33.2%	1.6%	0%	14.2%	28.8%	**33.9%**	22.3%	0.8%	34.8%	23.1%
man_2017	7.9%	8.4%	32.1%	**46.2%**	5.4%	0%	8.8%	14.8%	**41.8%**	31.6%	3.0%	51.6%	34.6%
sai_2017	< 0.01%	23.6%	**67.8%**	8.7%	< 0.001%	0%	0%	5.6%	**64.5%**	29.9%	0%	8.7%	29.9%
vil_2017	28.6%	17.4%	**33.1%**	18.8%	2.0%	0%	**36.4%**	31.2%	26.7%	5.6%	< 0.1%	20.8%	5.6%
ghi_2017	< 0.01%	41.8%	**52.6%**	5.6%	0%	0%	4.9%	32.0%	**53.3%**	9.8%	0%	5.6%	9.8%
san_2018	**38.4%**	11.6%	27.7%	22.3%	< 0.1%	0%	**43.4%**	39.6%	15.6%	1.5%	0%	22.3%	1.5%
chi_2018	**35.0%**	6.7%	16.4%	32.8%	9.0%	0%	26.4%	**36.4%**	18.4%	15.0%	3.8%	41.8%	18.8%
cal_2019	1.1%	9.7%	31.2%	**43.2%**	14.8%	< 0.1%	1.6%	4.9%	19.6%	**55.6%**	18.2%	58.0%	73.8%

Table V.7. – Summary of the fire danger predictions on all 13 fire cases over the whole island at the time of fire start: proportion of each class for both DeepFire (B to G) and IEP (1 to 5). In bold is the highest proportion among all classes.
Last two columns: proportion of locations in DeepFire class E or higher (E+) and proportion in IEP class 4 or higher (4+).

B is low in many cases too, but there are 3 of them where it is higher than 25% because a non-negligible part of the island is predicted with a FMC higher than 0.3. This, together with the predicted probabilities listed in Table V.6 reveals a limit of using the fire size classes of the US system to quantify fire danger with DeepFire. The thresholds of class B and G are very restrictive (upper bound of 4 ha and lower bound of about 2000 ha, respectively) considering that DeepFire estimates the potential fire size of a fire that spreads freely during one hour. The same holds to some extent for class F (lower bound of about 400 ha), explaining the lower representativeness of the extreme classes in most cases.

Among the 13 cases, the fire danger category that has the highest proportion is generally class D for DeepFire and 3 for IEP. Comparison with the local deterministic predictions in Table V.6 reveals that, in almost all cases, the deterministic prediction is in equal or higher danger class than the category with the highest proportion over the island for both IEP and DeepFire, the only exception being man_2017 fire case. This means that using either index, the ignition point belongs to the areas that are "highlighted" at the time of fire start, which is a desirable property of a fire danger map. It would be even better to observe the same property at the sub-regional scale because the environmental conditions and values at stake can be quite different from one to another, and make more sense from an operational perspective than the average over the island.

The previous tables provide data regarding the prediction at the time of fire start, but the evolution of fire danger over time needs to be assessed as well. For simplicity, we decided to present only two of the figures that were shown in the previous section regarding cal_2017 fire case, for the 12 other cases. The zoom around ignition point of the DeepFire deterministic map at the time of fire start with categorical color scale (cf. Figure V.5c) provides a prediction of fire danger at the sub-regional scale, while the "concise" plot in Figure V.10 shows the evolution at the ignition point of the deterministic DeepFire prediction, the ensemble mean, and the IEP. These two figures are shown for the 12 fire cases in Figure V.14. Similarly to the cal_2017 fire case, lower DeepFire values are obtained close to areas with no vegetation that block fire spread in several instances, but this mainly depends on the direction of wind speed.

In the case of bon_07-2017 for instance, the wind comes from the east, so when getting closer to areas without vegetation, lower values (classes D and C) are obtained to the west of the ignition location, but higher values (class E) are obtained eastward. Regarding the DeepFire predictions over time at the ignition location, the deterministic value and the ensemble mean are generally close and in the same fire danger category. The ensemble mean seems to follow the same variations with equal or lower magnitude, resulting in predictions in a smaller ranger than its deterministic counterpart, especially when the latter is very high ($>$ 200 ha). The category of IEP generally matches that of DeepFire, but can be higher or lower by one level for a few hours. There is a notable exception for Sainte Lucie de Tallano where IEP is in the highest class (5) while DeepFire is in class D (2 levels lower), but it can be observed as well that area nearby is only 1 level under (class E).

Looking at the DeepFire deterministic prediction more globally, one may notice that close to the ignition point and/or in a few hours after fire start, higher fire danger classes can be obtained. Five fire cases are concerned: olm_2017 (eastward), non_2017

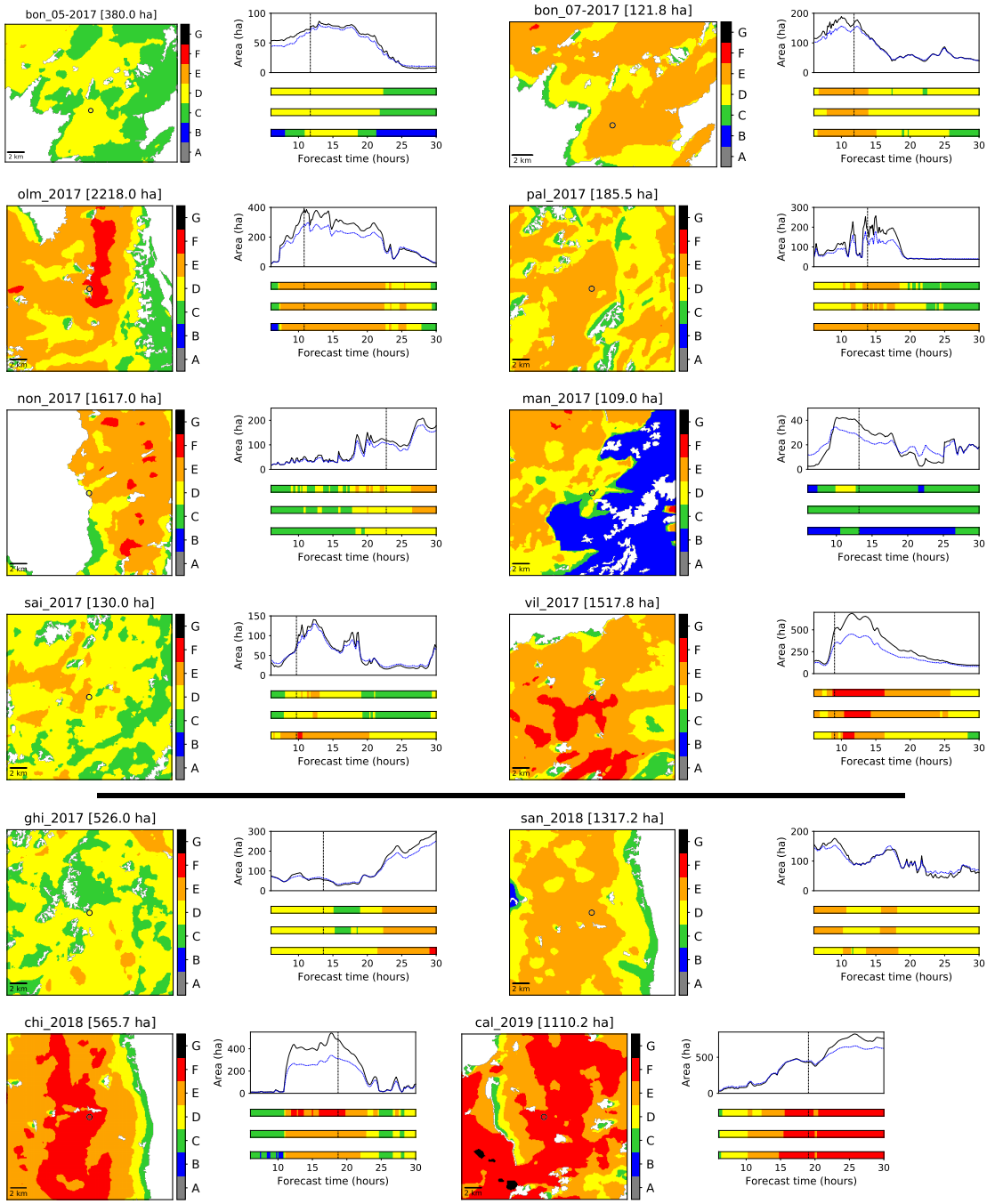


Figure V.14. – For all 12 complementary fire cases: map of DeepFire predicted at the ignition time zoomed around the ignition location (left, cf. Figure V.5c), as well as the "concise" plot for the ignition location (right, cf. Figure V.10).
Eight top-most: summer fires; four bottom-most: winter fires.

(eastward and about five hours after fire start), man_2017 (to the north-west), sai_2017 (eastward and after about two hours), and ghi_2017 (about eight hours after fire start). Therefore, considering the vicinity of the ignition point and/or a few hours after fire start, class E or higher is predicted in most cases. This does not happen in two cases: ghi_2017 (class D) where the fire burned slowly and several fire starts occurred, and bon_05-2017 (class D), whose observed fire size (109 ha) is the smallest among the 13 fires studied.

It may make more sense to focus on local predictions in time and space to evaluate predictions of ignition, but DeepFire estimates potential fire spread, so it should be analyzed over a larger range in time and space. For instance, some fires such as non_2017 may spread slowly at first, yet become difficult to stop after some time due to a change in weather conditions and/or when they reach locations with high slope or specific fuel type that are more favorable to fire spread.

V.4. Possible designs for daily fire danger mapping

Focusing on the vicinity of the ignition location and the few hours after fire start, the DeepFire predictions proved to be satisfactory. However, in practice, information on the fires used in the previous section is not known before making fire danger predictions. For the fire cases studied, we knew what were the areas and times to focus on because the information was known *a posteriori*. One could analyze the maps predicted for each time step, but in an operational context one or a few maps that summarize the daily predictions of fire danger. In the present section, we propose several ways of summarizing the predicted fire danger maps and discuss how relevant the resulting daily maps could be in an operational context.

The maps proposed in the present section attempt to address several questions. First, what are the locations and time where/when fire danger is highest for the day to come? (cf. deterministic: Figure V.15 & V.16, and probabilistic: Figure V.17 & V.18) Then, starting from when and for how long is there high fire danger? (cf. Figure V.19 and Figure V.20.) Regarding these aspects, the question of uncertainty in the prediction is also addressed to some extent by comparing the deterministic maps to their probabilistic counterparts.

Similarly to Section V.3, several maps are first proposed and detailed for cal_2017 fire case, then a summarized analysis is provided for all 13 fire cases.

V.4.1. Focus on the day of Calenzana 2017 fire

From the deterministic prediction of either DeepFire or IEP, the maximum over the day can be computed easily for each location. The maps of the maximum between T+6 and T+30 of DeepFire (resp., IEP) and of the associated time of maximum are shown in Figure V.15 (resp., Figure V.16). On this fire day, relatively strong south-westerly wind was definitely the driving factor, but this wind also brought some humidity, leaving area of high danger either where a downslope effect was strong. By looking at the time of the maximum of DeepFire on the right of Figure V.15 it can be seen that although there is a high danger potential (some areas in red in map on the left), there is actually much contrast regarding time of highest danger in the area of Calvi, suggesting that if

there is an event, a detailed local analysis is required. On the IEP map in Figure V.16 such requirement is less obvious, with large areas marked in orange and less contrast, indicating a dangerous but more general situation (less discriminant) with an event having strong probabilities of occurring in the morning.

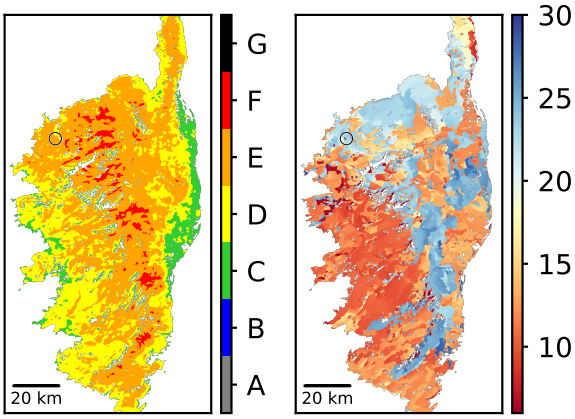


Figure V.15. – Maximum over the forecast between T+6 and T+30 (left) and time thereof (right) of the DeepFire prediction.

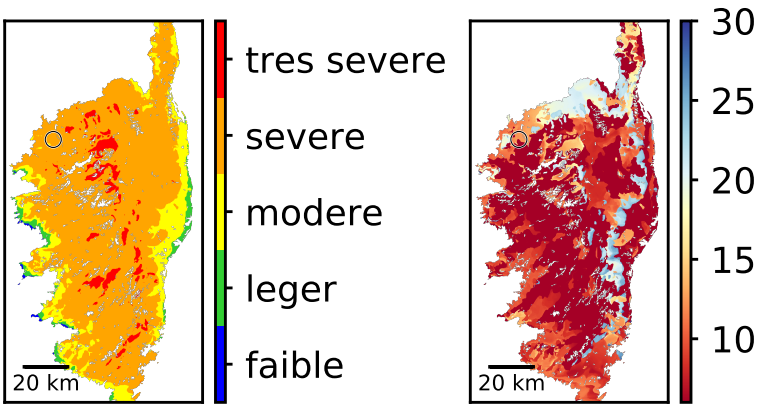


Figure V.16. – Same as Figure V.15 but for IEP prediction

If the maximum is obtained at several forecast times, we show the earliest one. Because the IEP has only 5 categories, the maximum value will most likely be predicted at several times. Figure V.16 shows that the time of maximum is T+6 for most locations, making it hard to tell when fire danger is highest. In practice, this issue can be avoided by identifying the time of maximum of another a continuous quantity, such as the FFMC which is one of the two components of IEP. In Figure V.15, although the maximum is represented by a categorical color scale, it corresponds to a continuous index, so this issue does not occur. The time of maximum for DeepFire predictions is between T+8 and T+18 on the majority of the island, but for some locations (even among these predicted in class F) the time of maximum is after T+18, notably around

the ignition location where the fire occurred at T+16 while the maximum is more around T+22.

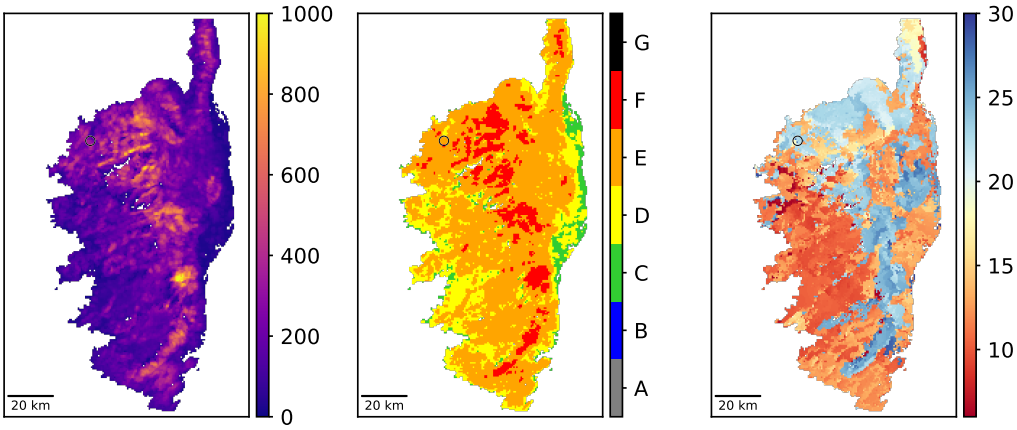


Figure V.17. – Maximum over the forecast between T+6 and T+30 of the quantile for probability 0.8 in the ensemble of DeepFire predictions.
From left to right: continuous scale; categorical scale; time of the maximum.

From the probabilistic forecast of DeepFire, the computation of a counterpart to Figure V.15 is less direct. The ensemble can be summarized by a statistic such as the mean or a quantile. For either statistic, it makes more sense to first compute it for all locations and forecast times, then to identify the time of maximum. In the case of a quantile, for instance, the time of maximum can therefore be interpreted to that of a more or less optimistic predicted scenario. In Figure V.17, the quantile for probability 0.8 was chosen. According to the Prométhée database, about 80% of the fires in Corsica have a final burned surface of 1 ha or less, which is quite low considering the range of DeepFire predictions. Although it is intuitive to define quantile from a meaningful fire size derived from a database of observations, it does not seem relevant here, and it makes more sense to interpret the chosen quantile as a quantity that represents a quite "pessimistic" scenario, yet not too extreme given the previous analysis of Figure V.13. However, as can be seen in Figure V.17 (middle map) there are still many locations where fire danger is at least in class D. It might be possible to make a distinction among the high-danger areas by looking at the continuous value (left map) less intuitive. Regarding the time of maximum, it is similar to that of the deterministic counterpart in Figure V.15.

Alternatively, one may consider another maximum over time: that of the probability (based on the ensemble) of being into class E or higher. The resulting map for cal_2017 fire case is shown in Figure V.18 together with the associated time of maximum. Compared to Figure V.17, the maximum probability seems better suited to discriminate among locations with high fire danger during the day, whereas the time of maximum appears similar overall, except for some locations where the predicted probability is 0 over all 24 hours, resulting in a time of maximum at the value of T+6 by default.

Among all the maps presented in the present section up to this point, the maps of

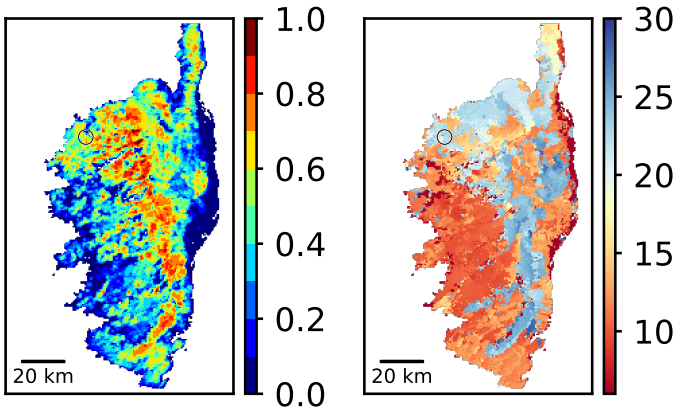
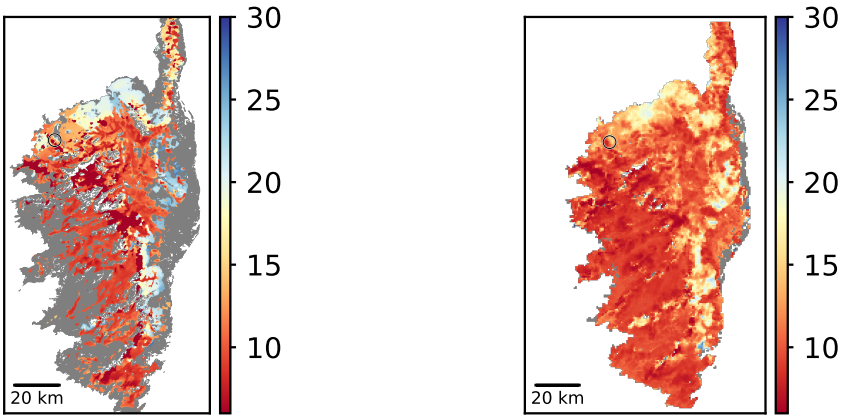


Figure V.18. – Maximum over the forecast between T+6 and T+30 of the probability of being into class E or higher (left) and time thereof (right).



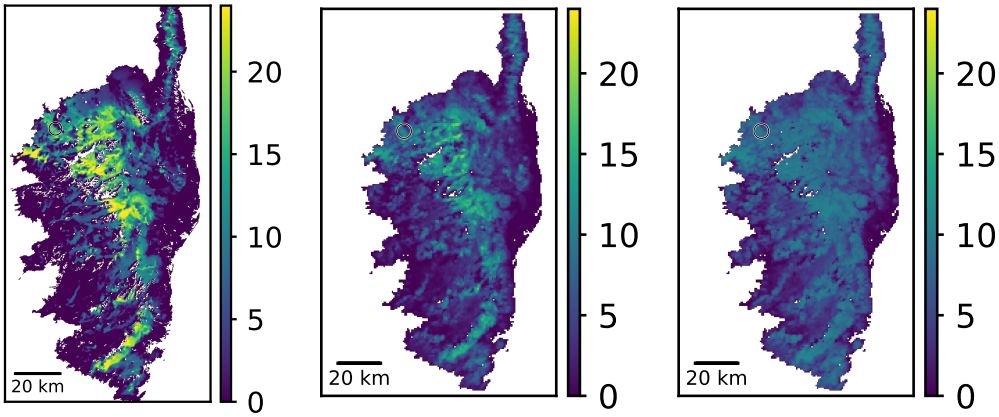
(a) Deterministic (b) Probabilistic

Figure V.19. – Earliest time between T+6 and T+30 of a location being predicted into class E or higher (in gray if it does not happen).
The probabilistic version is the conditional mean among the ensemble of the time when class E is reached on the condition that it is reached (a gray area is shown otherwise)

time of maximum to point out that the evolution of fire danger over time does not necessarily result in a peak at a single given time (e.g., midday) everywhere. Still, the time of maximum fire danger might not be the most relevant information for daily predictions as there can be high fire danger for an extended period of time.

Regarding the evolution of DeepFire predictions during the day, a more relevant quantity could be the earliest time of being into class E or higher if it happens at all. The associated maps resulting from the deterministic and probabilistic predictions are represented in Figure V.19. In the case of the probabilistic version, the earliest time when class E is predicted may differ from a member of the ensemble to another

and the class may not be reached for some of them, so the conditional mean among the sub-ensemble of members for which the class is reached is provided. According to the deterministic version in Figure V.19a, there is about a quarter of the island where class E is never reached. However, according to its probabilistic counterpart in Figure V.19b, almost all locations have at least one member in the ensemble for which class E is predicted at some point (but the amount of these members may differ between locations), and the map is more homogeneous with fewer extreme values.



(a) Deterministic (b) Probabilistic prediction

Figure V.20. – Total duration a location is predicted in class E or higher over the forecast between T+6 and T+30. The colorbar indicates the duration in hours.
(b) Mean (left) and standard deviation (right).

To complete the information provided by the previous maps, one may compute the total duration between T+6 and T+30 when a given location has a DeepFire prediction of class E or higher. In the case of the probabilistic prediction, there is no need to consider a conditional mean as the duration is null if class E is never reached. The deterministic version is shown in Figure V.20a, while both mean and standard deviation of the ensemble are shown in Figure V.20b to represent the probabilistic forecast. Once again, the map of ensemble mean is more homogeneous than its deterministic counterpart and the standard deviation seems higher in locations where the mean is higher.

The analysis of cal_2017 fire case reveals that predicted time of maximum fire danger is similar even when the input uncertainty is accounted for using the calibrated distribution. However, based on the calibrated distribution, the predicted duration of high fire danger is quite uncertain. Such quantity may not serve as a daily fire danger index, but it appears interesting as a complement, because the longer high fire danger conditions last, the more severe the situation may be considered, even if the maximum fire danger category over the day is the same. Still, in the case of DeepFire predictions, not only is the prediction uncertain, but it is also quite sensitive to the choice of the fire size threshold used to define high fire danger categories.

V.4.2. Overview of the predictions for the 12 other fires

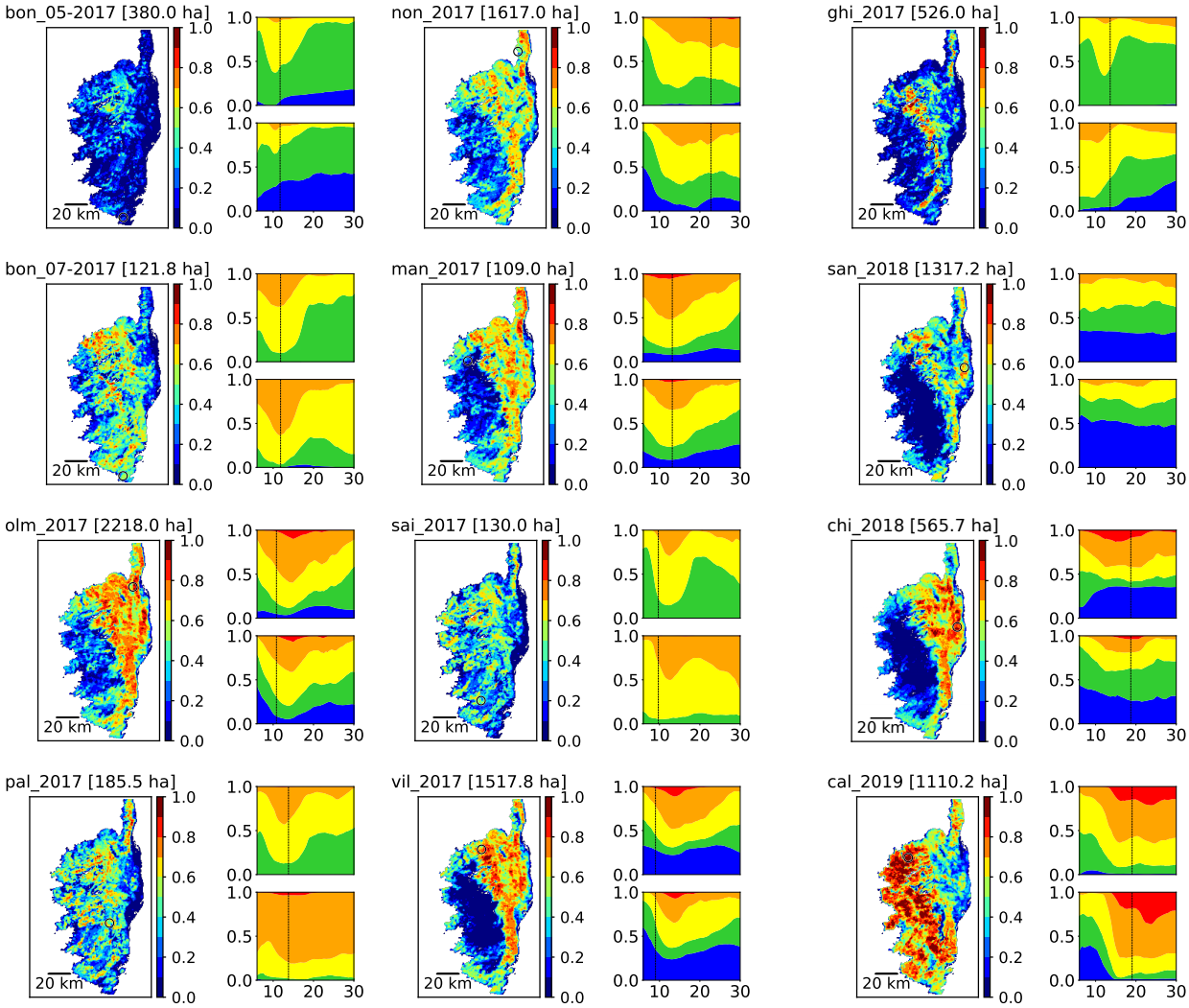


Figure V.21. – For all 12 complementary fire cases: maximum between T+6 and T+30 of the probability of being into class E or higher (left, cf. Figure V.18), together with the evolution of the proportions over the island of DeepFire (top right, cf. Figure V.6a) and IEP (bottom right, cf. Figure V.6b) classes.
First two columns: summer fires; last column: winter fires.

A number of numerical quantities that could serve as a daily indicator were proposed. Due to its ability to discriminate between locations with high fire danger in the cal_2017 fire case, the maximum over the forecast between T+6 and T+30 of the probability of being into class E or higher is chosen as a daily fire danger map for the 12 other fire cases. This map is shown for all 12 cases in Figure V.21, together with the evolution of the proportion in the island of each class for both DeepFire and IEP according to the deterministic prediction (cf. Figure V.6) to have an idea of how the overall spatial

distribution of fire danger evolves over time.

Overall there is a general agreement in variations between T+6 and T+30 between IEP and DeepFire, with a general trend of more "contrasted" DeepFire predictions, indicating a better ability to pinpoint high danger locations (i.e. it is more discriminant).

As in Section V.3.2, one could present some scalar results regarding the overall spatial distribution of fire danger over time, but considering an average in time and space does not reflect the variations of fire danger and, therefore, does not seem very relevant. Arguably, a daily rating of fire danger based on DeepFire should be computed on a relatively large area, at the sub-regional level for Corsica island and the maximum value over the day can be considered. At this spatial scale, a representative fire danger rating could be the one associated to the DeepFire value that separates the 80% lower values from the 20% highest in a given area. As a complement to fire danger ratings, that are associated to a low spatial resolution, it makes sense to use the predictions with high spatial resolution and high frequency to analyze the situation in more detail.

V.5. Conclusions and perspectives

We have shown a new method to compute fire danger predictions by estimating a potential fire size using the deep neural network DeepFire in a reasonable computational time. This allows for the generation of a sequence of high resolution maps of fire danger with high frequency, and an ensemble of such predictions that account for input uncertainty, although at lower spatial resolution. Several quantities derived from predictions of potential fire size can be computed such as maximum fire danger over time, duration of high fire danger, and probabilistic versions thereof using the ensemble. Analysis of the predictions on 13 fire cases showed that, overall, DeepFire predicts higher fire danger in the neighborhood of the ignition location and around the time of fire start. The temporal variation over the day make it difficult to summarize the predictions by a few daily fire danger indices. A few quantities were proposed to serve the purpose of providing a daily indicator, but the evolution in time, the granularity, and the uncertainty of the predictions can provide useful information to a forecaster. It is also important to note that all these computations were actually run in a batch processing mode to ensures its potential to run in an operational context.

A major strength of the prediction using DeepFire seems to be its spatial granularity allowing to be more discriminant. Compared to traditional fire danger indices that mostly rely on weather forecasts, the potential fire size estimated by DeepFire accounts for the influence of terrain on fire spread at via the variability over space in type of vegetation, presence of non-burnable areas, and slope. The high-resolution maps could be used as complement of fire danger ratings, that generally attribute a single value to a large area, for better anticipation but potentially to help to decide firefighting actions after a fire has started spreading. For instance, the maps can be used to finely identify locations that, if reached at a some point, the fire will spread even faster and become harder to control. Moreover, another strength regarding its design, compared to other fire danger rating systems, is that it is not based on empirical knowledge, except for the actual choices of fire size for each class, these results are not based on experience of past fires, nor on expert analysis.

The present study allowed identifying several aspects that could be investigated fur-

ther or improved. Regarding the DeepFire ensemble at a given time and location, it seems that the logarithm follows a Johnson's S_U distribution in several instances, although a more complex family of distribution could lead to a better fit in more instances. Statistical inference of the ensembles could also be analyzed at a larger scale, by considering several forecast times and locations, and even the predictions for several days. The analysis of the distribution of numerous potential fire sized predicted by DeepFire could lead to the definition of meaningful thresholds to define categories of fire danger that are better suited than the US classifications which pertains to all observed fires. Defining the thresholds based on observed fire sizes appears difficult in Corsica where a vast majority of fires are attacked early enough to spread far, even when fire danger is high. Information on the intermediate sizes of the fires, rather than that of the final burned surface, would be more relevant for comparison with DeepFire, but is difficult to measure. If DeepFire predictions were to be used in operational conditions, the feedback would be very valuable to adjust the thresholds, but also to evaluate the usefulness of such a prediction system. Regarding uncertainty, only a deterministic weather forecasts were used, but probabilistic weather forecasts could be used as an alternative or as a complement to represent the probabilistic distribution of wind and FMC inputs of DeepFire.

DeepFire focuses on wildfire spread, and could be associated to predictions representing ignition probability for a better estimation of fire danger. A model of ignition probability should account at least for weather predictions, and other data sources such as proximity to roads or cities could be included to design such a model, which would result in a distribution that is not uniform among burnable locations. Ignition probability is also a major component of wildfire risk quantification frameworks, where DeepFire could be used to estimate fire spread. Although the computational time constraints of fire risk assessment are not as tight as for those of daily fire danger predictions, DeepFire could estimate a very large number of potential fire sizes in a reasonable amount of time, much larger than it would be using traditional fire spread simulators, allowing for better and/or more detailed risk quantification. For a complete assessment of risk, however, the values at stake are not just hectares of land burned and should be accounted for more explicitly than with DeepFire. A step in this direction could therefore be to train a DNN in a similar fashion as DeepFire but based on an estimated cost (cf. Appendix A).

Whether it be for fire danger or risk, DeepFire predictions should be also considered and evaluated over a long time period (e.g. a whole year) that includes both days with and without fires occurrence. Due to their ability to account for several types of data, and to carry out fast computations, deep neural networks appear promising to address the complexity of wildfires and design better prediction systems.

Acknowledgments

Funding: This work was supported by the Agence Nationale de la Recherche, France [grant number ANR-16-CE04-0006 FIRECASTER].

VI. Conclusion

VI.1. Thesis summary

In the present thesis, we studied the assessment of wildland fire risk using simulations of wildland fire spread to provide probability distributions of burned surface and fire size that would result from a wildfire ignition. The general purpose of this study is to design decision support tools that can be used in an operational context. The methods presented in this work were carried out based on simulations using the fire spread solver ForeFire, and relied on the underlying rate of spread (ROS) model of Rothermel, data maps of the highly contrasted Mediterranean Corsica island for elevation and land-use, as well as weather forecasts covering the region; but the approach is generic and may very well be applied to other models and/or other regions.

We first focused on the errors associated to predictions in "crisis" (when there is an ongoing fire) situations pertaining to the input uncertainty in fire spread simulations by means of direct Uncertainty Quantification (UQ) in Chapter II. The input probability distributions were expressed by the PDFs of perturbation coefficients in reference values used in the simulations on the basis of weather forecasts (wind speed, fuel moisture content (FMC)), fuel model parameters, information regarding the fire (time of fire start, time of fire end, and ignition location), and by "transition" probabilities to account for error in land-use type. Propagation of these distributions in the simulations relied on a Monte Carlo (MC) approach, which resulted in maps of burn probabilities covering the vicinity of the presumed ignition location.

Evaluation required the introduction of several properties and tools suited to probabilistic predictions, including accuracy using a "spatialized" Brier score for comparison with an observed burned surface, reliability with a reliability diagram, probabilistic resolution using a sharpness graph and consistency using a rank diagram designed for binary variables. The calibration of the input probability distributions was then undertaken by means of inverse UQ in Chapter III. Inspired from Bayesian approaches, we defined a posterior input PDF proportional to the product of a pseudo-likelihood function by the prior input PDF. Assuming independence for events such as "location x belongs to the burned surface" is not satisfactory due to high correlation between close locations, so we proposed to employ the Wasserstein distance, involved in the field of optimal transport, as a deterministic evaluation score between simulated and observed burned surface to address this difficulty. The pseudo-likelihood function involved a weighted average of this score calculated for several wildfires. Investigation of the posterior input distribution relied on a Metropolis-Hasting (MH) algorithm for which a Gaussian process (GP) emulator was used to approximate the weighted average to reduce computational time, so that the high number of iterations desired for the MH algorithm could be carried out in a reasonable amount of time. The resulting MCMC chains can be used to draw samples and easily perform direct UQ as in the previous

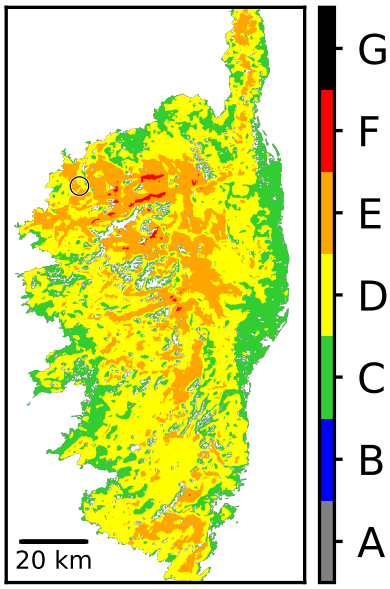
chapter but this time with a calibrated input distribution. The resulting a posteriori output distributions were evaluated in the same fashion as previously mentioned.

Both direct and inverse UQ methods were applied to seven relatively big fires that occurred in Corsica island from mid-2017 to early 2018, and for which final observed burned surfaces were available. In Chapter II, evaluation of the predicted burn probability maps based on 500-member ensembles resulted in fair performance in 5 cases, but lower for the two cases of highest fire duration where fire growth was overpredicted (cases of Nonza and Ghisoni) indicating an expected uncertainty building up with time. An individual fire spread simulation was typically completed in less than 10 minutes so the MC method can fit the time constraints of an operational context by distributing the simulations, provided that high-performance resources are available.

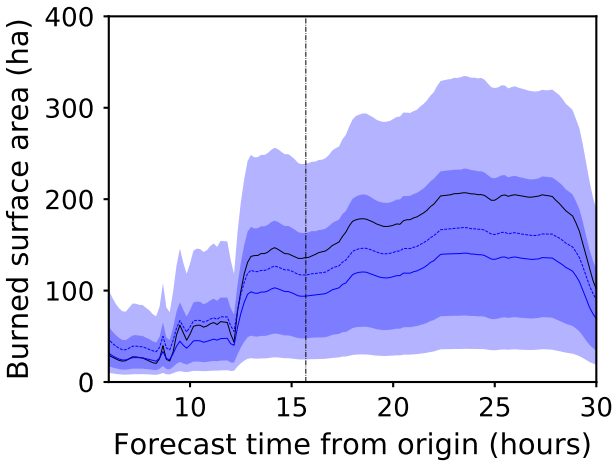
In Chapter II, compared to the prior input PDF, the marginals of the posterior input PDF mostly favored input values leading to lower ROS, and the most significant changes were observed in perturbations in wind speed norm, fuel moisture content and heat of combustion rather than perturbations in other fuel parameters. Another significant difference was that the posterior distribution had lower standard deviation for perturbation in wind direction. Overall, for the seven fire cases, the burn probability maps representing the posterior distribution showed less dispersion than with the prior distribution and a lower tendency to overpredict burn probabilities. Increasing the weight β of the pseudo-likelihood relatively to the prior PDF improved overall accuracy, but not that of all fire cases. We selected the posterior input distribution obtained using $\beta = 1/2$ as a calibrated input distribution for the rest of this work, due to its overall good accuracy ranking out of all the values of β tested.

The calibrated input distribution was employed in an interdisciplinary work led by Antoine Belgodère, reported in Appendix A. This study focused on estimating the economical cost of wildfires and marginal damage as a function of time using simulations of fire spread, and application to our seven Corsican fires resulted in a surge of the marginal cost five hours after ignition. Applications of fire spread simulation to economical models allows to assess the evolution of fire risk in "crisis" situation using either the mean cost or based on the certainty equivalent cost, which involves a convex disutility function, to account for risk aversion.

We then focused on wildfire risk assessment in anticipation of fire starts. We proposed to deal with the "spread" component of risk by estimating the potential fire size of a fire starting from any ignition location at any given time. Considering the number of ignition locations required for high-resolution mapping alone (about 1.2 million for Corsica island with a target resolution of 80 m), the time to compute all required fire spread simulations is too high for "short-term" situations, and is even higher when uncertainty is accounted for using a MC approach. In Chapter IV, we proposed to emulate the size of a 1-hour simulated fire for a large range of input values, notably wind speed vector, FMC, and ignition location, but also perturbation in fuel parameters (consistently with the calibrated distribution). For this purpose, a deep neural network (DNN) with a hybrid architecture accounting for both raster inputs (data maps of elevation and fuel parameters for Corsica island) and scalar inputs was trained based on a training dataset of size 5×10^6. Considering that the fire sizes ranged across 4 orders of magnitude for this complex problem, the achieved approximation error of the DNN was satisfactory with a mean absolute percentage error (MAPE) $\approx 33\%$ on our



(a) Map of predicted fire danger



(b) Evolution of the prediction of local potential fire size (10-min resolution). The shaded blue areas represent quantiles of the output fire size distribution resulting from propagation of input uncertainty in DeepFire.

Figure VI.1. – Illustration of fire danger predictions obtained with the DeepFire emulator. For more details regarding the figures, see (a) Figure V.4b and (b) Figure V.7

test dataset. It required pre-computing the part of the network pertaining to the data maps for all ignition locations to obtain an emulator, called DeepFire, able to compute a large amount of potential burned surfaces in a very short amount of time. As an example, 1.2 million values of potential fire size were obtained in less than a minute on a 32-CPU machine using DeepFire, whereas this would take several days on average with non-emulated ForeFire simulations.

Application of DeepFire to day to day fire danger mapping was studied in further detail in Chapter V. Using a 32-CPU machine, it was possible to generate maps of 1-hour potential fire size at a spatial resolution of 80 m and a temporal resolution of 10 min (an example of results is represented in Figure VI.1) in only a few hours. Uncertainty can also be accounted for by MC to generate maps of quantiles of the calibrated distribution of potential fire size, for instance, although base on ensembles of maps at lower spatial resolution (600 m) to keep computational time reasonably low. Compared to traditional fire danger indices that depend mostly on weather inputs, DeepFire also accounts for other influential factors in fire spread, notably local elevation and fuel cover data in the vicinity of the ignition location at high spatial resolution (80 m). Threshold values of fire size were used to derive discriminant classes of fire danger ratings. Using the high-resolution maps, more elaborate quantities of fire danger can be derived, such as the duration in a given class or the maximum over 24 hours of the probability (estimated based on the ensemble of maps generated via MC) of being in a high-danger class. The fire danger maps were computed a posteriori for 13 days when relatively big fires occurred in Corsica, on the basis of the 42-hour weather

forecasts available at 10:00 AM on the fire day. For most of these case studies, DeepFire predicted high fire danger in the neighborhood of the ignition location around the time of fire start, which was also higher than the average prediction over the territory and the course of the day.

Overall, we adapted mathematical methods in an innovative manner suited to wildland fire spread predictions, while striving to account for modeling uncertainties. All these developments were designed for a "urgent high performance computing" computational system that could realistically be used in an operational context. Arguably, the main achievement of this thesis is the design of the DeepFire emulator. Compared to traditional fire danger indices expressed as unitless quantities based on weather variables such as the FWI, DeepFire can provide a potential fire size in hectares, which provides a more concrete representation of fire spread; and training of the underlying DNN based on simulations following a systematic optimization process to be as exempt of empirical parameter fitting as possible. Another major outcome of these new application of mathematical and numerical methods is the opening of research perspectives.

VI.2. Research perspectives

A promising direction is the application of neural networks in wildland fire science. Although DeepFire was designed for emulation of simulated fire size and fire danger rating, its ability to account for the variety of input data is encouraging and the idea of using a hybrid NN architecture could be extended to other detection/prediction systems. As for network architecture, upsampling layers could be considered, hoping that they would re-constitute a good raster approximation of the burned surface which could be used directly as output as in [68], from which ensembles of fire spread predictions, BP modeling, or estimations of cost could be carried out.

Regarding UQ, the uncertainty in weather inputs has been expressed via perturbation coefficients. Probabilistic weather forecasts is another way of accounting for this uncertainty; for instance, the ensemble forecasts of the ECMWF are used in EFFIS for predictions of fire danger. It would be insightful to evaluate the burn probability maps of specific fires resulting from propagating such an ensemble (e.g. for Corsica, this could be based on the ensemble version of AROME forecasts) in wildland fire simulations and compare this with the calibrated distribution. Another idea is to combine both sources of uncertainty (e.g. by applying perturbation coefficients to any member of the ensemble of weather forecasts) and carry out the comparison. Similarly, this approach could be applied to propagation of uncertainty in DeepFire to compare the probabilistic fire danger predictions resulting from different input distributions.

Another source of uncertainty can be attributed to the approximation error in metamodeling. We assumed that this error could be neglected compared to other sources of uncertainty, yet accounting for this source could allow to bypass or verify this simplifying assumption. In Chapter III, for inverse UQ, we emulated the energy function E intervening in the pseudo-likelihood function by the mean $\widetilde{E}$ of a Gaussian process. The standard deviation $\widetilde{\sigma}$ of the said GP was not used but could account for approximation error. Informally, instead of simply using the approximated value of the pseudo-likelihood function $e^{-\beta\widetilde{E}(\boldsymbol{u})}$, we could imagine using the Gaussian assumption to

derive an updated pseudo-likelihood function of the form $e^{-\beta\widetilde{E}(\boldsymbol{u})/(1+\widetilde{\sigma}(\boldsymbol{u}))^2}$; this would still guarantee equality to $e^{-\beta E(\boldsymbol{u})}$ for $\boldsymbol{u}$ in the training dataset and lower likelihood would be obtained for higher values of $\widetilde{\sigma}(\boldsymbol{u})$ (i.e. higher uncertainty due to approximation). As for DeepFire, although it consists in a neural network and not a GP, the plot of emulated fire size against simulated fire size among the test dataset (cf. Figure IV.9a) suggests that a linear regression model applied to the logarithms of these two values would yield a regression coefficient close to 1. Using such a model, which typically assumes i.i.d. Gaussian noise, could yield an output distribution of potential fire sizes to compare to and/or use in combination with the probabilistic DeepFire predictions resulting from the calibrated input distribution. Another way to account for uncertainty would be to add noise to ROS estimations based on empirical models involved in the fire spread simulations, which could lead to the design of a stochastic version of fire spread solvers relying on markers methods, comparably to stochastic cellular automata models (e.g. [135]).

Related to UQ is the evaluation of probabilistic predictions. The Brier score, which we used as a measure of accuracy, can be understood as an estimation of the expected value between a observation and the associated predicted probability (cf. Equation (I.15)). This estimation assumes independence, whereas for burned surfaces, the observations and the predicted probabilities of locations that are close to one another are correlated. It would be interesting to have a probabilistic score that accounts for correlation. The Wasserstein distance, which we used to provide a deterministic accuracy score between burned surfaces, can very well be adapted to compare a burn probability map. Outside of "crisis" situations, the principles and tools presented for probabilistic evaluation could be applied to "long-term" situations, notably to BP modeling, where evaluation and calibration usually rely on qualitative assessment and comparison of the predicted distribution of fire sizes to the one observed over several years.

A notable concern regarding the evaluation of fire spread predictions whether it be deterministic or probabilistic is the use of final observed burned surfaces. Although such data have the advantage of being quite accurate and available after most fire events, the final burned surface are the result of firefighting actions and only represent a static final state, whereas the simulations represent a dynamic process. Observations of burned surfaces at intermediate times of fire spread over the whole event would be more relevant for evaluation; given the recent rise in the use of unmanned aerial vehicles (UAVs, aka drones) in firefighting operations, including fire mapping, one could expect such data to become increasingly available in the near future. This raises the need for methods suited to evaluating use a sequence of observed burned surfaces for evaluation and calibration purposes, due to the presence of both spatial and temporal correlations. In this regard, the extension of the Wasserstein distance to distributions with values in $\mathbb{R}^2 \times \mathbb{R}$ (i.e. space and time) could be investigated; more generally, this raises the question of connections between front-tracking and optimal transport.

Out of the three components of wildfire risk assessment, we mostly focused on spread and, to some extent, exposure. A way to combine these two aspects in short-term and long-term situations is to use the estimated cost of the simulated burned surface of a 1-hour fire, instead of its size. To do so, an emulator that computes the potential cost of a 1-hour fire could be built based on a DNN, in a similar fashion as DeepFire was built. Both potential cost and potential fire size should be used in combination with

an ignition probability model to provide a more comprehensive assessment of wildfire risk and danger, respectively. A simple approach is to compute the product of ignition probability by potential fire size (resp. cost) for all locations and provide an updated fire danger (resp. risk) map. Regardless of the approach, it is likely that the locations with higher fire danger would be different than these with higher fire risk, which could also differ if ignition probability is accounted for or not. In any case, this strategy focuses on potential consequences of a "source" of ignition. It is a notable difference from BP modeling, where the burn probabilities can be seen as representing "sinks" which result from a multitude of possible ignition scenarios. One can therefore wonder if fuel management and allocation of firefighting resources should focus more on the sources or on the sinks. This could depend on the situation (i.e. "short-term" or "long-term"), and a notion that addresses this issue is the so-called source-sink ratio [3], quantified by the logarithm of the ratio between fire size originating from a given point by the burn probability. Maps of potential fire size could be used in combination with BP maps to compute this ratio, to provide additional information for decision support.

From a modeling perspective, planning of fire-fighting allocation in the short-term could consist in solving an optimization problem, which is a possible extension of the fire danger maps in decision support. The idea is to minimize the average / maximum fire danger over the territory based on a finite number of resources. Along these lines, one could consider that there are I types of resources (water bombers, UAVs, fire trucks, etc.) each available in a given number $n_1, \ldots, n_N$. By denoting $n_{i,j} \in \{0, \ldots, n_i\}$ the amount of resources of type i present at location $x_j \in \mathbb{R}^2$ (where j could be an integer equal to $\sum_i n_i$, or lower), we could consider a local danger-reduction function $f : x \in \mathbb{R}^2 \mapsto f(x, x_j, n_{1,j}, \ldots, n_{I,j})$ that reduces fire danger for locations x in a neighborhood around x_j depending on the resources allocated. The problem would therefore be to find the values of $n_{i,j}$ and x_j that minimize the maximum fire danger over the territory based on the application of f to the predicted map of fire danger. Such a problem could be solved using stochastic methods such as simulated annealing, for instance. Alternative formulations of the problem include minimizing the spatial average of fire danger, or using values of costs instead of fire danger. A possible problem extension when using an economical cost is to include the cost of mobilizing resources in the objective function, then the problem could express that expensive resources should be mobilized if they can significantly reduce the maximum cost. Throughout this thesis, we carried out simulations of fire spread in the absence of fighting actions, which provide "what-if" scenarios; in this context, cost estimation can more simply help to quantify the benefits of firefighting actions. Another step in this direction would be to account for the intensity of fire in the simulations of fire spread, which is correlated to the damage caused. Optimization problems similar to those previously mentioned could be formulated and solved in order to help with the use of firefighting resources during "crisis" situations. Even though this work has addressed the question of the performance of prediction systems, one may wonder how well they perform in terms of decision-making, i.e. does the system help the decision makers in taking good decisions? are there biases or misinterpretations of the tools? etc.

Finally, all the predictive systems considered throughout this thesis would benefit from a large database of observed fires. In this regard, fire danger maps that provide potential fire sizes could undergo evaluation and calibration using databases of fire sizes

such as the Prométhée database. Although most of the observed fires are small and may not last 1 hour, we could investigate how their distribution compares with the output distribution of DeepFire values resulting from an input distribution representative of a large number of weather forecasts. This input distribution could also be used to provide "long-term" fire danger maps.

A. Hurry! There is a fire

Avoiding catastrophic wildfires is a natural rationale for fighting fires in their early stage. Beside this benefit, may a marginal decrease in the duration of smaller wildfires be worthwhile? The present work addresses this topic by estimating the marginal damage of forest fires as a function of time. We simulated 3500 scenarios of fire growth based on 7 real events that took place in Corsica between 2017 and 2018 and estimated the damage based on the type of land use in burned areas. Results suggest that the marginal cost surges 300 minutes after ignition. This pattern is magnified when wind speed is above median weather conditions. Using a conservative calibration, we find important marginal costs, reaching more than €1M per hour after 300 minutes of propagation in the 50% windiest conditions. These results corroborate the principle of early initial attack already in use in countries with sufficient fire fighting forces, but subject to debate because of its cost.

Summary

A.1. Introduction **174**
A.2. Numerical methods **176**
A.2.1. Simulation of wildland fire spread 176
A.2.2. Generation of an ensemble of simulations 178
A.3. Evaluating the cost of fire **180**
A.3.1. Land Use in Corsica 180
A.3.2. Cost of a single fire 181
A.3.3. Modeling risk aversion 184
A.4. Evolution of the damage **184**
A.5. Discussion **192**
A.6. Conclusion **193**

The preprint of the paper "Hurry! There is a fire" (to be submitted), written by Antoine Belgodère (UMR CNRS LISA 6240), Frédéric Allaire (INRIA), Jean-Baptiste Filippi (CNRS), Vivien Mallet (INRIA), and Florian Guéniot (UMR CNRS LISA 6240), is reproduced here. The work presented in this chapter is the result of a joint work among the FireCaster project, where the calibrated input distribution from Chapter III is employed.

A.1. Introduction

Gigantic wildfires, such as the ones that recently devastated large areas of Australia, California, Portugal, Brazil, shed light on the potential dramatic outcomes of an unrestrained fire. In those countries Mediterranean climate and landscapes are among those that suffers the most dramatic consequences, but some countries are barely hit by such massive events. France, for instance, is regularly struck by wildfires, but they remain of size that is manageable by the fire fighting forces. Left panel of Figure A.1 presents the maximum fire size in the Mediterranean Departments of France (MDF) [1], which are the most exposed to fires, between 1973 and 2019. Only one year (in 1990) did a single fire burn slightly more than 10 000 hectares. Right panel presents the yearly total area burnt in MDF for the same period. The maximum is reached in 2003 with 600 000 hectares. These are low numbers compared to the 20 000 000 hectares that burnt in 2019-2020 in Australia for instance. France have more than twice (40.000) the amount of career firefighters than Australia (17.000) [2], for a much reduced area to protect, and can apply a principle of early initial attack with the benefit of avoiding some of these gigantic events. Is it the only benefit? This paper want to assess the marginal benefit of reducing the duration of already small fires. Small fires are the submerged parts of the iceberg. 47% of the burnt area in MDF since 1973 have gone up in smoke in fires of less than 400 hectares [3], which represent 99.6% of all the events. Even if the marginal cost of fires' size is increasing, it is safe to say that a significant part of the overall cost is due to these small fires. In the context of the present article, the largest simulated fire reaches 13 155 hectares after 8 hours, whereas the largest fire in MDF since 1973 reached 11 580 hectares. Even these extreme values are far from the catastrophic events aforementioned.

The economics of wildfires highlights the needs to balance two types of costs ([109]): i- the expected damage from the fire itself, ii- the different public expenditures aimed at reducing the expected damage, such as prevention, fuel management and suppression. Reducing the duration of fires would decrease the first one and increase the second one. This paper is focused on the damage side of the equation. In other words, if the cost of reducing the duration of a fire by one hour was given, the question we ask in the paper is: "is it worthwhile?"

A large ensemble of wildfire is simulated in order to address this question, coupled with land-use data on the island of Corsica, which is a subset of MDF, and represents 36% of the total burnt area since 1973. The land-use data allows to estimate the cost of each fire for each of the 24 20-minute time steps of use for the simulations. If fire ignition was purely random, ignition cells could just be uniformly drawn over the territory and rely on them for the simulations. However, since most wildfires are a direct or indirect (such as presence of power lines) result of humans presence, there are reasons to believe that the ignition points and the ignition dates are selected as actual plausible candidates.

1. Alpes de Haute Provence, Hautes Alpes, Alpes Maritimes, Ardèche, Aude, Bouches du Rhône, Drôme, Corse-du-Sud, Haute-Corse, Gard, Hérault, Lozère, Pyrénées Orientales, Var, Vaucluse

2. source: European Commission (European Forest Fire Information System), CTIF, Eurostat, FAO

3. This number is set in [21] as the limit between small and big fires. It also corresponds to the fires of categories F and G by the National Wildfire Coordination Group in the USA https://www.nwcg.gov/term/glossary/size-class-of-fire

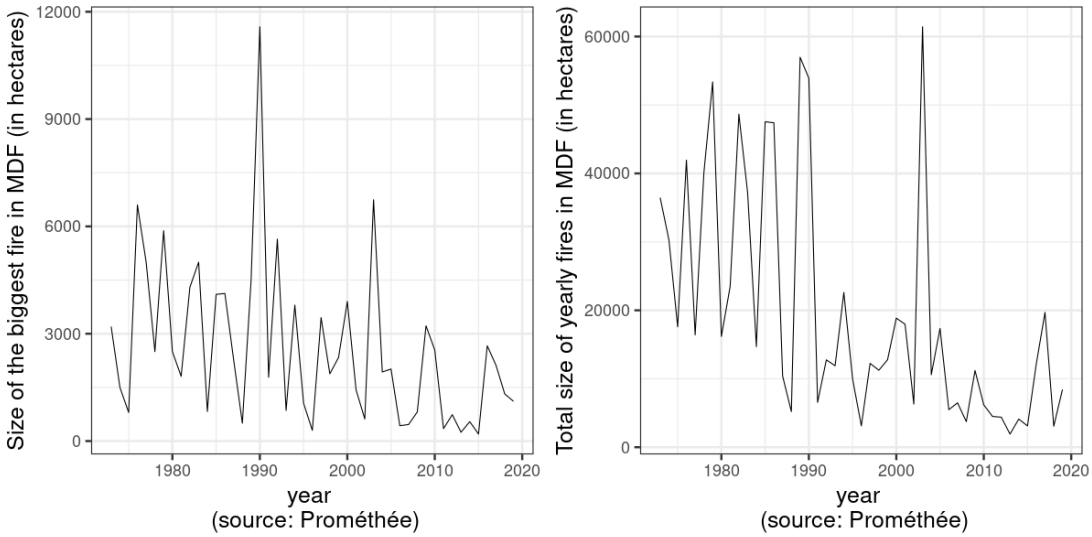


Figure A.1. – Size of the largest versus total size of wildfires in France between 1973 and 2019

Our approach to account for ignition is the following: meteorological conditions and ignition points are retrieved for each of the 7 biggest real fires that took place in Corsica between 2017 and 2018 to account for the simulation of realistic fires. For each of these 7 fires, an ensemble of 500 simulations are run, that each differ due to random variations in the meteorological parameters, in the precise ignition point, as well as in fuel parameters. These variations were calibrated [5] to that they maximize the likelihood of the simulated events and are sufficient to provide significantly variable outcomes.

The number of 7 real cases we use to build our ensembles might seem weak. However, the official database Prométhée [4] that lists all the fires in MDF since 1973 does not allow does not provide sufficiently precise data to apply the method described in this paper to a larger number of cases. Indeed, the ignition point is given in DFSI coordinate system, with 2km × 2km resolution, and the ignition time is often arbitrarily set to midnight. On the other hand, the 7 cases we use in this paper happened after the start of this research, which allowed us to retrieve precise data on: i- the ignition points and times, via a close relationship with the fire fighting services, and ii- on the meteorological conditions, via Météo France.

First the 3500 simulations [5] are performed to constitute the base data. Then, the cost of the fire is estimated at a fixed time step for each simulation by using reasonable cost parameters for each type of land use. If no firefighting actions are provided, the simulated fire will continue to spread until it reaches nonburnable land, requiring an external criterion to stop the computation. In the current set-up, simulation is set to stop at a given fire duration. For each simulation, the fire shape is sampled every 20 minutes until it reaches 8 hours. The resulting 84 000 [6] are pooled to perform cost

4. https://www.promethee.com/
5. 7 fires × 500 simulations
6. 3500 simulations × 24 time steps

estimates in order to assess the marginal cost of time, controlling for moisture and wind speed. Our main results suggests that:

1. The marginal cost of fire increases sharply after 300 minutes.
2. The marginal burnt area increases sharply after 300 minutes, following an evolution very similar to the marginal cost.
3. The cost per hectare increases sharply in the 200 first minutes, then roughly stabilizes around €750 per hectare.
4. After 5 hours of burning, the total cost is nearly 20 times higher when the speed of wind is higher than the when median wind is lower.
5. The cost per hectare increases much faster in the presence of wind, which is much less obvious than the previous point.

The paper is organized as follows: Section A.2 presents the numerical simulation methods for the evolution of a wildfire and to generate the ensemble of simulations; Section A.3 describes the method and data of use to predict the cost of a fire with a given contour; Section A.4 presents the main results of simulations; Section A.5 is a discussion section and the last section concludes.

A.2. Numerical methods

A.2.1. Simulation of wildland fire spread

Knowing how a wildland fire will spread is a key aspect of firefighting as it helps in deciding the most effective actions to stop the fire and/or mitigate its impacts on life and property. With the availability of powerful computational resources, numerical simulation has become an invaluable tool to study and predict wildland fire spread. Models from varying complexity can be used to describe the dynamics of fire propagation and can be classified as (semi-)physical or (semi-)empirical [129]. The former are more comprehensive in the description of the physics of combustion but the resulting simulations are too time-expensive for the computation of very large fires in real time. Meanwhile, fire spread simulators [131] that describe the dynamics of the shape of the fire are faster and more suited to make predictions in an operational context. Such simulators typically make use of an empirical model where the rate of spread (ROS), i.e., the speed at which the flames advance, is expressed as a function of local environmental parameters. Most ROS models take into account wind speed V, slope α, fuel moisture content m (= mass of water / mass of dry fuel), as well as parameters that describe the properties of the vegetation (which is also the fuel), and can be summarized as follows:

$$\frac{dx}{dt} = r(V, \alpha, m, \ldots), \tag{A.1}$$

where x denotes the location of the flames, t denotes time, and r is the ROS function. Usually, the ROS increases with wind speed V and slope α, while higher fuel moisture content m leads to reduced ROS. Equation (A.1) is inherently 1-dimensional whereas numerical simulators of wildland fire are inherently 2D. In these models, a fire most commonly follows a binary representation where any location in the the landscape is either burned or not burned at a given time. Simulating fire spread therefore amounts

to determining the evolution of the burned part of the landscape over time. The interface between burned and not-burned areas is referred to as the *fire front*, and the evolution of a fire can be viewed by displaying the successive states of this interface across the landscape, as depicted in Figure A.2.

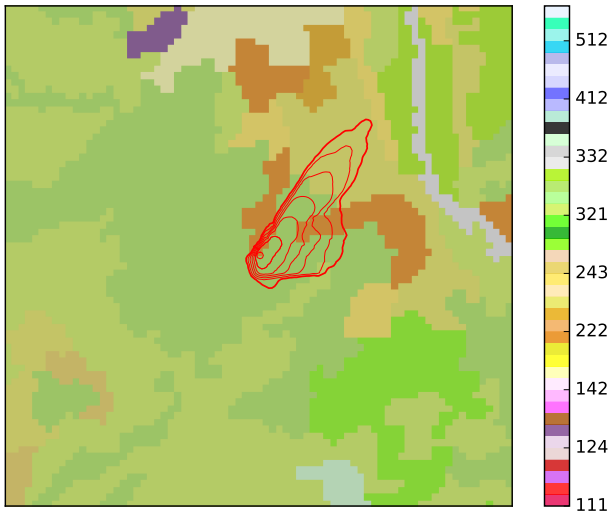


Figure A.2. – Representation of simulated fire spread. The red contours are isochrones of a simulated fire front separated by ten minutes of spread, for a total duration of one hour. The fire started at the center of the figure. Background colors correspond to a land use raster at 80-m resolution derived from Corine Land Cover [42] data

Fire propagation solver ForeFire [45] is performing the simulations. ForeFire uses a front-tracking technique to model the propagation of the fire front. The fire front is discretized by the means of Lagrangian markers linked by a dynamic mesh. Each marker is advanced according to the surface geometry and the ROS. Here, the latter is computed according to the empirical model of Rothermel [114], widely used in wildland fire simulation, and that contains numerous parameters already fitted and fixed through an analysis of a large set of laboratory experiments.

Running a simulation relies on several data fields that determine the local environmental parameters from which the ROS will be computed wherever the markers of the fire front are positioned. Slope α is computed from an altitude layer. Wind speed V varies locally due to the spatial variations in the altitude layer but is determined by a single "triggering" value of wind speed vector. This triggering value is obtained from a weather forecast and may be modified at some points in a simulation to account for changes in weather. Fuel moisture content m is derived from the weather forecast and may vary over time in the same fashion as wind speed, but it is constant over the landscape. Another layer is the land cover map, which identifies different fuel types over the landscape. For each of these fuel types exists a corresponding parameterization (a "fuel model") which determines the remaining input parameters of the ROS used to characterize the vegetation. Areas that are not vegetation (cities, bodies of water, bare

rocks, etc.) are not considered as fuel and, therefore, cannot be burned. The remaining inputs that are necessary to run a simulation are a few spatial and temporal pieces of information on the fire, namely the coordinates of the presumed ignition point, which will define a small initial burned area, as well as the time of ignition, and the duration of the fire. The characteristics of the seven fire cases mentioned in the present study can be found in [5], and the interested reader will find more details regarding the data sources used to run the simulations. A major difference to note in is that the present study is focusing on the potential evolution of the cost and are not attempting to compare the simulations to the observed burned surfaces. Here, fires were simulated for a 8 hours duration, regardless of the actual duration of the fire.

A.2.2. Generation of an ensemble of simulations

In spite of the recent advances in wildland fire spread modeling, making accurate predictions of large fires is still a considerable challenge. Complex physical phenomena among which convection, heat transfer, pyrolysis and combustion, come into play at possibly very small spatial and temporal scales, whereas a fire may spread during several hours or even days, resulting in burned surfaces that can amount to thousands of hectares. This also implies that fire spread may occur in highly variable environmental conditions, as weather, vegetation state, and topography can be considerably different across the landscape and over time. Simulators that are used in an operational context often rely on an incomplete or over-simplified formulation of the underlying physical and numerical model, resulting sometimes in inaccurate predictions.

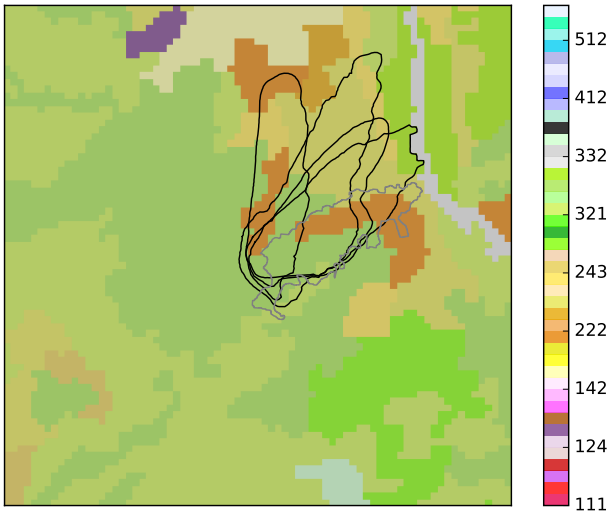


Figure A.3. – Ensemble of wildland fire simulations. The black contours indicate some of the simulated fire fronts after 1h of propagation. The gray contour delineates the burned surface observed at the end of the actual wildfire. Background colors correspond to a land use raster at 80-m resolution derived from Corine Land Cover [42] data

A major source of uncertainty in wildfire spread prediction relates to the input variables of the simulation. Wind speed, fuel moisture content, fuel parameterization, time

and location of fire ignition, as well as fire duration are not known perfectly. The best estimates of these variables may result in a simulation that significantly differs form the observed spread of fire. Instead of relying on a single deterministic simulation, a promising alternative consists in adopting a probabilistic approach where several simulations are run based on different likely sets of inputs, which results in an ensemble of simulations traducing possible scenarios, as exemplified in Figure A.3.

More formally, $\mathcal{S}_{\boldsymbol{u}}(t)$ is introduced as the burned surface at time t returned by the simulator based on the set of inputs $\boldsymbol{u} \in \mathbb{R}^d$. Input uncertainty is modeled by attributing a probability distribution $\mathbb{P}_{\boldsymbol{U}}$ to the inputs, which is now perceived as a random vector $\boldsymbol{U}$. The resulting burned surface $\mathcal{S}_{\boldsymbol{U}}(t)$ is now probabilistic and traduces the probability for any location in the landscape to be burned at time t, according to the simulator and the input uncertainty. In practice, the *burn probability* at time t is estimated by the means of an ensemble of M simulated burned surfaces $\mathcal{S}_{\boldsymbol{u_1}}(t), \ldots, \mathcal{S}_{\boldsymbol{u_M}}(t)$, where $\boldsymbol{u_1}, \ldots, \boldsymbol{u_M}$ are sampled independently from the same probability distribution $\mathbb{P}_{\boldsymbol{U}}$.

Several approaches for ensemble generation are available in the literature [49, 94, 96]. Depending on the study, the uncertain inputs may not be the same and for a commonly identified uncertain input, its marginal probability distribution may also differ. In the present study, the uncertain inputs are those identified in [5], with the exception of time of fire start and time of fire end: here, simulations of fires are systematically run with a fixed duration of 8 hours and fixed time of fire start. Another notable difference is that a *calibrated* probability distribution is used instead of the *a priori* distribution proposed in the aforementioned study. The calibration method is detailed in [7]. The key elements of this calibration procedure consist in using an *a priori* distribution where the marginals have higher variance than in [5] and including the information of observed burned surfaces of the seven large fires mentioned in the present study. The *probability density function* (PDF) g of the calibrated distribution is intended to be higher than the prior PDF f for a given $\boldsymbol{u}$ when there is good overall agreement between observed burned surface $\mathcal{S}_{\text{obs}}$ and corresponding simulated burned surface $\mathcal{S}_{\boldsymbol{u}}$. More precisely, the PDF g can be expressed as follows:

$$g(\boldsymbol{u}) = \frac{e^{-\beta E(\boldsymbol{u})} f(\boldsymbol{u})}{\int e^{-\beta E(\boldsymbol{u})} f(\boldsymbol{u}) d\boldsymbol{u}}, \tag{A.2}$$

where $\beta > 0$ and E is a positive "energy" function that is equal to 0 when $\mathcal{S}_{\boldsymbol{u}} = \mathcal{S}_{\text{obs}}$ for each fire and increases with the dissimilarity between simulated and observed burned surfaces. Measurement of shape dissimilarity relies on the Wasserstein distance, which is a metric that appears in the field of optimal transport (see, for instance, [120] for an extensive review), and $E(\boldsymbol{u})$ can be understood as an energy that is required to transform $\mathcal{S}_{\boldsymbol{u}}$ into $\mathcal{S}_{\text{obs}}$. The definition of g is inspired from Bayes' rule:

$$p(\boldsymbol{u}|\mathcal{S}_{\text{obs}}) = \frac{\mathcal{L}(\mathcal{S}_{\text{obs}}|\boldsymbol{u}) f(\boldsymbol{u})}{\int \mathcal{L}(\mathcal{S}_{\text{obs}}|\boldsymbol{u}) f(\boldsymbol{u}) d\boldsymbol{u}}, \tag{A.3}$$

where $p(.|\mathcal{S}_{\text{obs}})$ is the posterior PDF that would be obtained from $\mathcal{L}(\mathcal{S}_{\text{obs}}|.)$, the likelihood of the observation, and f, the prior PDF. The exponential in Equation (A.2) can therefore be seen as a pseudo-likelihood function, whose weight increases with the

parameter β. When $\beta = 0$, the calibrated PDF g is equal to the prior PDF f and $\beta = 1/2$ in the present study.

A.3. Evaluating the cost of fire

This section describes the procedure to assess the impact of a fire. Assessing the cost of a fire requires collecting data on land use in endangered areas. Especially, it is required to identify the areas containing buildings, agricultural lands, and human beings in the middle of the wild. Then to attribute a value that is likely to be lost in case the fire reaches a given place. Finally, adding a risk premium to the expected cost of a given fire in order to account for risk aversion.

A.3.1. Land Use in Corsica

Representations of land use in Corsica is decomposed into 6 rasters of 85.5km $\times$ 183.3km with a resolution of 50 meters with:

1. Buildings. *BD-topo* by the French *Institut National de Géographie* (IGN) [7] provides vector data on all identified buildings in Corsica. Each cell's value is the sum of the areas it contains expressed in square meter.
2. Remarkable buildings. Some of the buildings are deemed remarkable by IGN. These include churches, castles, city halls... These buildings are already included in the building raster. The remarkable building raster assigns to each cell the number of remarkable buildings it contains.
3. Agricultural lands. The database *Registre Parcellaire Graphique* (RPG) by IGN and the *Agence de Service et de Paiement* (ASP) provides vector data on agricultural lands, with distinct types of cultures (wheat, corn, olives, orchards...). The raster of agricultural lands simply assigns to each cell the corresponding type of culture.
4. Forests. The database *BD-Forêts* by IGN provides vector data on forests. Forests is considered as homogeneous, with a procedure similar as the agriculture class: in each cell indicates the type of forest (deciduous trees, chestnut trees, pines...)
5. Campsites: Corsica is an important tourist destination, with about 3 million tourists each year [8] for less than 400 000 inhabitants [9]. OpenStreetMap provides a list of campsites represented by dots. With no information on the size of these sites, a campsite is assumed to expand over a $250 \times 250m^2$ square, which corresponds to the cell where the dot is located plus the 24 cells around this one. The number assigned to each cell in the raster file corresponds to the share of campsites it contains. If a single site expands over a given cell, the cell will be assigned the value 0.04 (1/25). In the case of two sites that are close neighbors, cell that contains both sites will be assigned the value 0.08.

7. IGN changed its name in 2012 for *Institut national de l'information géographique et forestière* commonly called IGN

8. source: Agence du Tourisme de la Corse https://www.corsica-pro.com/fr/observatoire/chiffres-cles

9. 335 000 in 2017 according to INSEE https://www.insee.fr/fr/statistiques/fichier/version-html/4271414/co_inf_46.pdf

6. Hiking paths: we represent hiking paths listed by the Corsican natural reserve [10] as poly-lines. When such a poly-line crosses a given cell, value of 50 linear meters of path is assigned, which corresponds to the length of the square it represents [11].

For each of these rasters, a value must be assigned to each cell in order to assess the economic value at risk in case the corresponding area is burnt.

A.3.2. Cost of a single fire

For all categories of cost, the actual cost of a fire strongly depends on fire intensity. If a house is completely destroyed by a fire, then it sounds wise to assess the cost of the fire as being equal to the cost of rebuilding the house. This represents the maximum cost that can be born by the householder [12]. However, damage will, hopefully, often be much lower than this maximum value (for instance, only the roof can be burnt). The strategy followed to estimate the cost of a fire consists in: i- estimating the maximum cost based on the type of land use, ii- multiplying this maximum by a scaling factor $s \in]0;1[$ chosen so as to produce correct orders of magnitude. Identifying correct orders of magnitude proves to be complex for two reasons:

First, there is a large variability of estimates in the literature. For instance, [1] estimate the costs of 4 fires that took place around 2000 and find costs that range from \$300 to nearly \$18 000 per hectare. [103] have narrower estimates for 3 other fires (between \$1000 and \$5600 per hectare). The tiny literature on the topic does not allow to perform multivariate analysis in order to identify the determinants of this variability.

Second, the literature based on real cases presumably focuses on costly fires, whereas this paper aims at accounting for fires of all degrees of severeness.

With this unavoidable limitations in mind, $s = 0.01$ was chosen, which results in a median cost of €139.98, a mean cost of €668.75 and a standard deviation of €2217.33 [13]. Clearly, the level of the estimated cost that calculated hereafter are very sensitive to this somehow arbitrary scaling parameter. However, these results fall into the (wide) range of the estimated costs per hectare of real-life fires comforting this choice. Moreover, it is important to stress that the evolution of the marginal cost of duration over time, and especially the doubling of this marginal cost after 5 hours, is independent from this choice.

Table A.1 presents the coefficients of use to evaluate the endangered economic value of the fires. Here is a brief justification of these choices:

Forests

Forests cover 58% of Corsican landscape. Yet, commercial use of local timber is limited due to transportation difficulties ([62]). Of course, that doesn't mean that the forest is without value. It certainly have a patrimonial value for Corsican people, and a recreational value, both for local people and for tourists. [4] estimate the net commercial value of the dense forests in the region of the Blue Mountain ecoregion of eastern Oregon and southeastern Washington in the United States at \$650 per hectare. Same value was chosen here, rounded to €500 per hectare.

10. Parc Naturel Régional de Corse
11. This length is assumed as an average, since intersections cannot be evaluated at a lower resolution than the raster.
12. except for psychological costs.
13. those numbers include each of the 24 time steps of the 3500 simulations.

Land use	Unit	Parameter value
Forests	€ per hectare	€500
Agriculture	€ per hectare	from €5 to €1.12M
Buildings	€ per built square meter	€1000
Remarkable buildings	additional € per building	€50 000
Value of human life (campsites and hiking paths)	€ per life	€3 000 000

Table A.1. – List of parameters used to estimate the maximum value at risk for the different land uses.

Agriculture

Some 20% of Corsican land is devoted to agriculture (191 104 hectares out of 872 200) according to [69]. However, a very large share can produce only fodder: in 2018, 165 180 hectares were permanent grassland, 7320 were artificial or temporary grassland, and 1400 where fallow-land. A large share of the 17 204 remaining hectares are devoted to chestnut, olive and clementine groves and vines. [69] provides data on the precise surface occupied by different types of agricultural products, the production by product, and the unitary price. These data are reproduced in Table A.2 for the main products of Corsican agriculture. These data are used to estimate the value of the yearly production per hectare. However, as explained in detail in [103], burning some agricultural parcels such as vines or olive groves can have dynamic impacts. In order to estimate the maximum value that can be destroyed by a fire, a gross assumption was made so that the yearly production value represents 5% of the value of the trees, which is used to parameterize the potential cost (or value at risk) per hectare (last column of Table A.2).

Product	Prod. in 2018	Unit	Surface (ha)	Value of prod.	Yearly value per ha	Dyn.	VAR per hectare
Olives (oil)	276	t	2150	€3500	€449	1	€8986
Chestnut (flour)	15	t	1370	€15 000	€164	1	€3280
Vines (wine)	374 649	hl	5884	€400	€25 469	1	€0.5×10^6
Clementines	37 336	t	1332	€2000	€56 060	1	€1.2×10^6
Fodder	403 993	t	285 370	€20	€28	0	€28

Table A.2. – Calculus for the cost per hectare for the main agricultural land-use. VAR: Value at risk

Buildings

The maximum value at risk in a building is the reconstruction cost. €1000 was chosen as the value per square meter, which might be slightly below the average construction cost for new buildings, in order to account for the older ones. In addition,

a buffer is defined around the cells containing buildings, that replicates a share of the cost of these building, which is an admittedly quite ad-hoc way to account for all the inconveniences due to the proximity of a fire from inhabited areas. When a building is labeled "remarkable" in IGN data, it adds an arbitrary €50 000 to the value of the cell containing the building.

Human lives

Wildfires kill. Hopefully, only a minority of them do. Pricing the risk of fatalities is not an easy task in the context of this article. Indeed, most fatalities occur during exceptionally long events. In this work, assumes the fire can be stopped after a given duration, ranging from one hour to the simulated fire duration of eight hours. The probability to have a massive number of fatalities in less than eight hours is too low to account satisfactorily for those extreme cases. With that limitation in mind, estimation of human lives especially endangered in the case of wildfires are those of people in campsites or hiking paths. It is not taking account for the endangering of firefighters because this pertains to the intervention cost, whereas this paper only focuses on the proper cost of fires.

For campsites, the number of available beds is taken in all campsites (65 200) [14] and assume they are evenly shared among the 204 registered campsites. This implicitly assumes that the campsites capacity is always full, which is not a realistic assumption, since tourism in Corsica is a very seasonal activity. However, it is also true that most wildfires happen in the touristic season.

For hiking paths, it is assumed that 1000 hikers are permanently present over the 1563 kilometers of registered paths, which roughly corresponds to 0.2% of the population present on the island during the touristic season.

For both campsites and hiking paths is used a Value of Human Life (VHL) of €3 000 000 in order to give a price to the endangered lives. This value is within the wide range of values identified by [111] from $1 million to $10 millions.

Health

While health costs due to particle matters present in the smoke are an important part of the economics of wildfires, they were not taken into account in this paper. The reason is that in the case of small-medium scale fires that take place in a small density region like Corsica (less than 40 inhabitants per km^2), health costs are likely to be marginal.

Estimating the catastrophic wildfires that took place in Florida in 1998, and that burnt some 200 000 hectares in 6 weeks, [22] estimate the health costs to be less than $1 million, whereas the global cost is estimated between $600 million and $800 million. This would represent less than $5 per burnt hectare. [110] estimate the health cost of the 2001 fire season in Chilsom, Alberta to be CAN$4.1 $\times 10^6$ for 116 000 hectares burnt, which amounts to CAN$35 per burnt hectare, a higher number than the one for Florida. Yet, this estimate relies on the assumption that more than 1 million people where exposed to smoke, which in unlikely to happen in the case of Corsica, where the whole population is less than 400 000, and spread along the coastal regions that are separated from each other by mountain chains.

14. source ATC, the Corsican Tourism Agency.

A.3.3. Modeling risk aversion

The previous section described how to assign a cost to one specific event. But since 500 simulations are run for each fire, how to synthesize the different costs? A simple average cost appears insufficient, because it tends to attribute an unsatisfactory low weight to the rare but very costly outcomes. In other words, an average cost neglects risk aversion, that is the cornerstone of insurance theory.

The theory of expected utility is the most classical way to model risk aversion in economics. The basic idea is that an agent with a wealth W enjoys a utility level of $U(W)$ where $U()$ is a strictly increasing and strictly concave function. In those conditions, a rational agent having to choose between several options, each yielding random gains $\tilde{g}$, will pick the one that maximizes the expected utility $E(U(W_0 + \tilde{g}))$ (where W_0 is the agent's wealth before earning the gain). Given this expected utility $\bar{U}$, it is possible to define a certainty equivalent gain as $\bar{g} \equiv U^{-1}(\bar{U}) - W_0$. Then the risk premium, $E(\tilde{g}) - \bar{g}$ would denote the increase in expected gain, compared to the risk-free hypothetical situation, that is required in order to compensate for the risk.

This paper slightly deviates from this ideal framework in two respects. The first is not a very substantial one: it is thought in terms of costs and not in terms of gains, for the obvious reason that wildfires are unlikely to generate positive gains. The second is that when if the aim is to estimate the social cost of a fire, there is no relevant equivalent for the initial wealth W_0: should it be the wealth of the people directly hurt by the fire? Or the mean income in the municipality where the fire took place? Or the mean income at the Corsican level? There is no obvious answer to this question. What is at stake here is to determine whether W_0 should vary from one place to another. The most convincing solution was to assume it is constant.

Thus, risk is accounted by computing a "Certainty Equivalent Cost" (CEC) defined as follows: let $c_{i,j}$ be the cost of fire i (in 1...7) in simulation j (in 1...500). Expected disutility of fire i is defined as $\bar{D}_i \equiv \frac{\sum_{j=1}^{500} c_{i,j}^{\alpha}}{500}$ (with $\alpha \geq 1$). Then, the CEC is: $cec_i \equiv \bar{D}_i^{1/\alpha}$. In the special case where $\alpha = 1$, the CEC simply corresponds to the average cost calculated over the 500 simulations. For $\alpha > 1$, the CEC will be higher than the average cost. Thus, α can be interpreted as a measure of the policy-maker's risk-aversion.

A.4. Evolution of the damage

Ensembles of 500 simulations described in Section A.2.2 are used to feed the model described in Section A.2.1. For each simulation is retrieved successive polygons representing the contour of the fire every 20 minutes. The different rasters described in Section A.3 are aggregated into a cost map of Corsica, where each cell representing the value at risk of a 50 m × 50 m area. An intersection is done for each polygon of the burnt area with this map, and then a sum the values at risk of for each intersection. Finally this value at risk is multiplied by the scaling factor s to obtain an estimate of the cost of the fire.

Figure A.4 shows the distribution of the final loss (the loss after 480 minutes) for the 7 fires. The 7 fires correspond to 7 different geographical and meteorological situations, which explain the differences both in terms of median loss and in terms of variability.

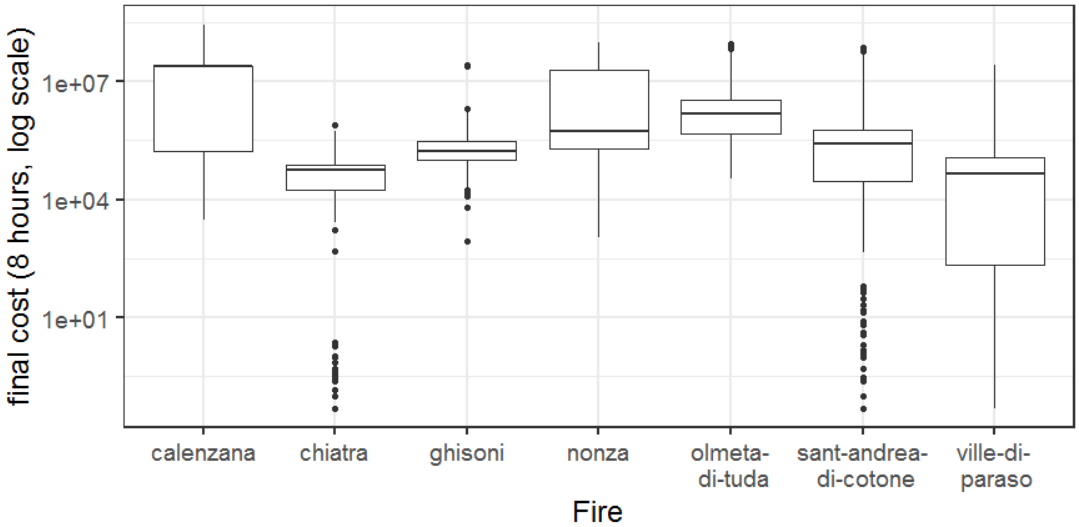


Figure A.4. – The distribution of the cost after 480 minutes for the 7 fires

Figure A.5 shows the evolution of the mean loss for the 7 fires. A general pattern shows up when looking to this figure: a sharp increase in the cost within the first minutes (around 100 minutes), followed by a growth at a smaller rate.

The evolution of CEC [15] follows the same pattern (Figure A.6), but tends to magnify the small occasional surges in the mean cost, which indicates that those surges are born by extreme values.

It is possible to decompose the evolution of the cost as the product of the evolution of the cost per hectare times the number of burned hectares. In 6 fires out of 7, the cost per hectare (averaged over the 500 simulations) sharply increases at the early stage. Then it evolves either slowly or becomes flat (Figure A.7). An interpretation could be that fires tend to start in places with small economic value, probably because supervision is weak in those places. The burnt area evolves in a similar way, with a smoother decline in the growth rate (Figure A.8)

To sum up, the trend towards increasing costs at a decreasing rate seems to stem from a similar evolution of both the burnt areas and their economic value per hectare.

However, just looking at these curves can be confusing, since meteorological conditions might explain changes in marginal costs as well as time. Moreover, the log scale of use for convenience for these graphs does not allow to capture the evolution of marginal cost. An approach to overcome these drawbacks is to estimate the cost of the simulated fires using meteorological variables as regressors, with both fire and time fixed effects:

$$c_{i,j,t} = \alpha * W_{i,j,t} + \beta * M_{i,j,t} + \mu_i + \nu_t + \epsilon_{i,j,t} \tag{A.4}$$

Where $c_{i,j,t}$ is the cost of fire $i \in [1;7]$ calculated at simulation $j \in [1;500]$ t minutes after ignition. $W_{i,j,t}$ and $M_{i,j,t}$ are the corresponding speed of wind and moisture. μ_i and ν_t are fire and time fixed effects, and $\epsilon_{i,j,t}$ is an idiosyncratic shock.

15. calculated with $\alpha = 2$

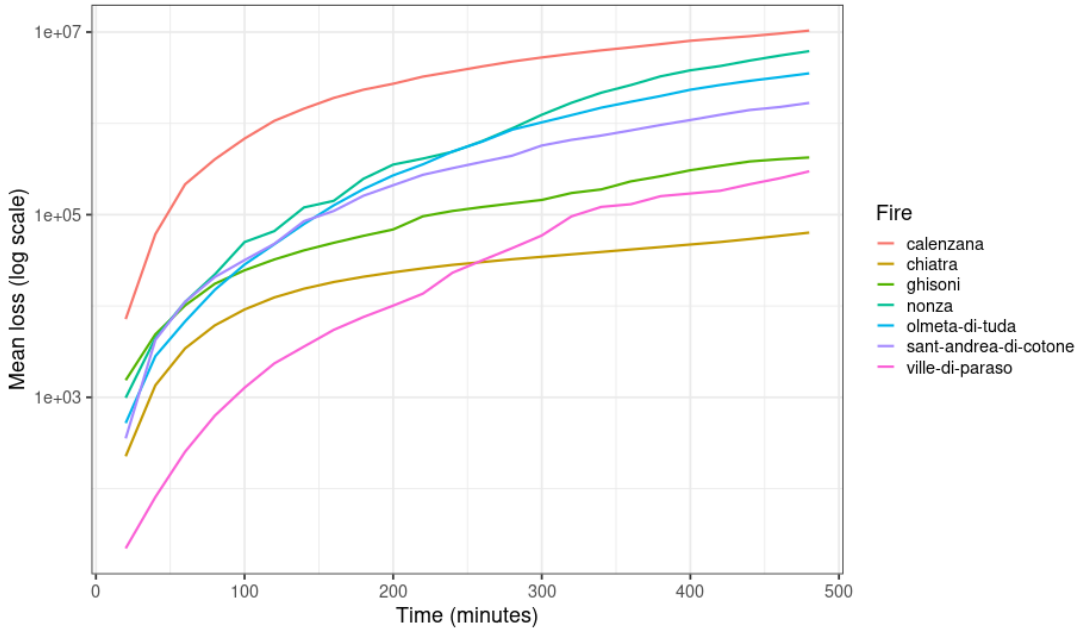


Figure A.5. – Evolution of the mean cost for the 7 fires

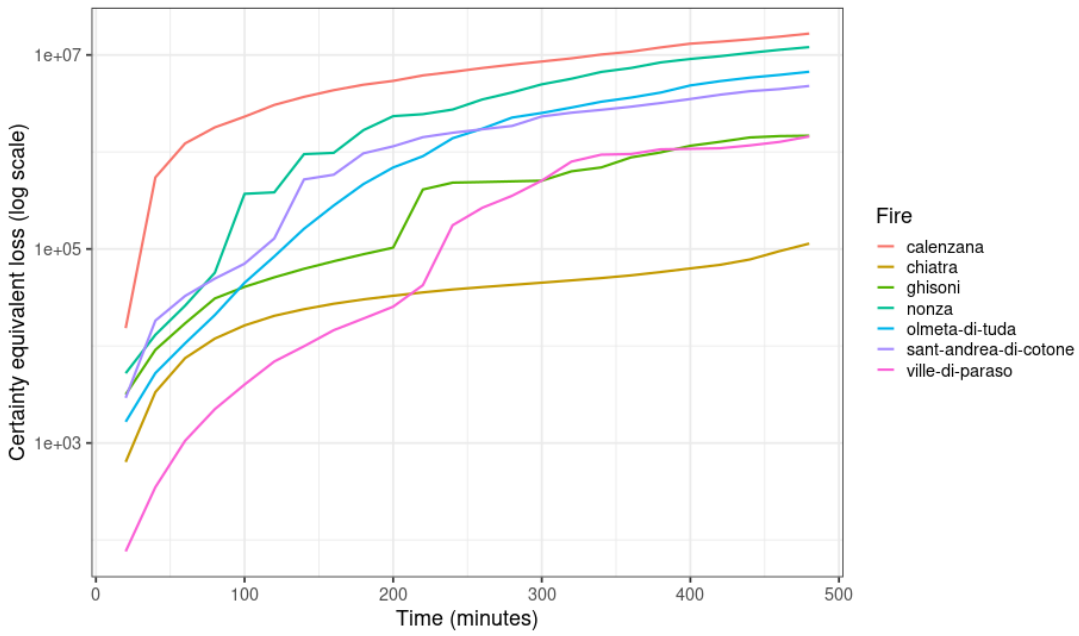


Figure A.6. – Evolution of certainty-equivalent cost for the 7 fires

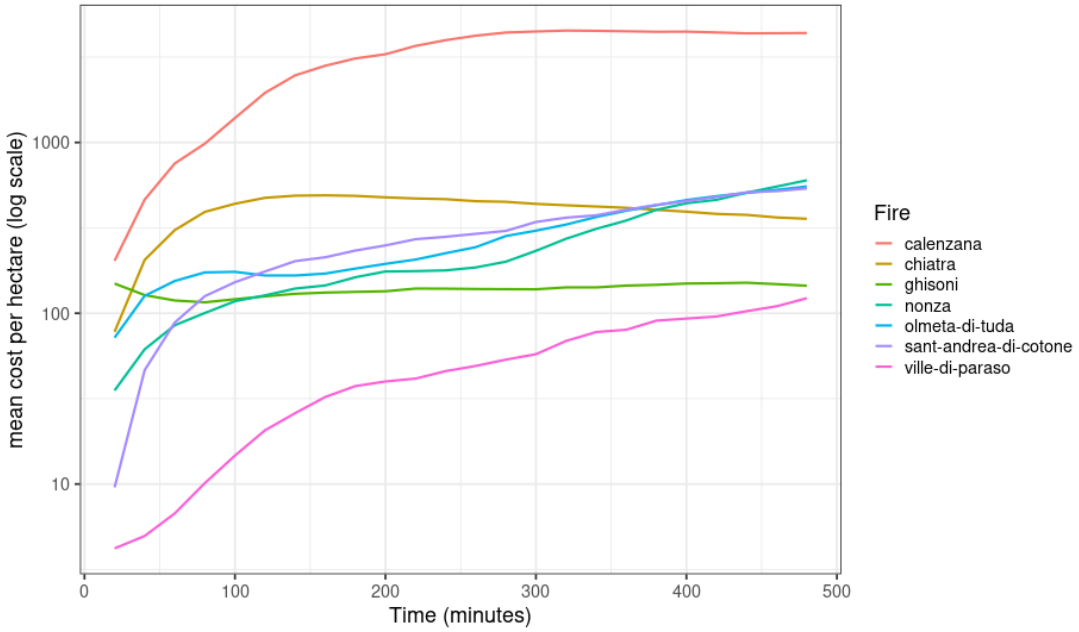


Figure A.7. – Evolution of the mean cost per hectare for the 7 fires

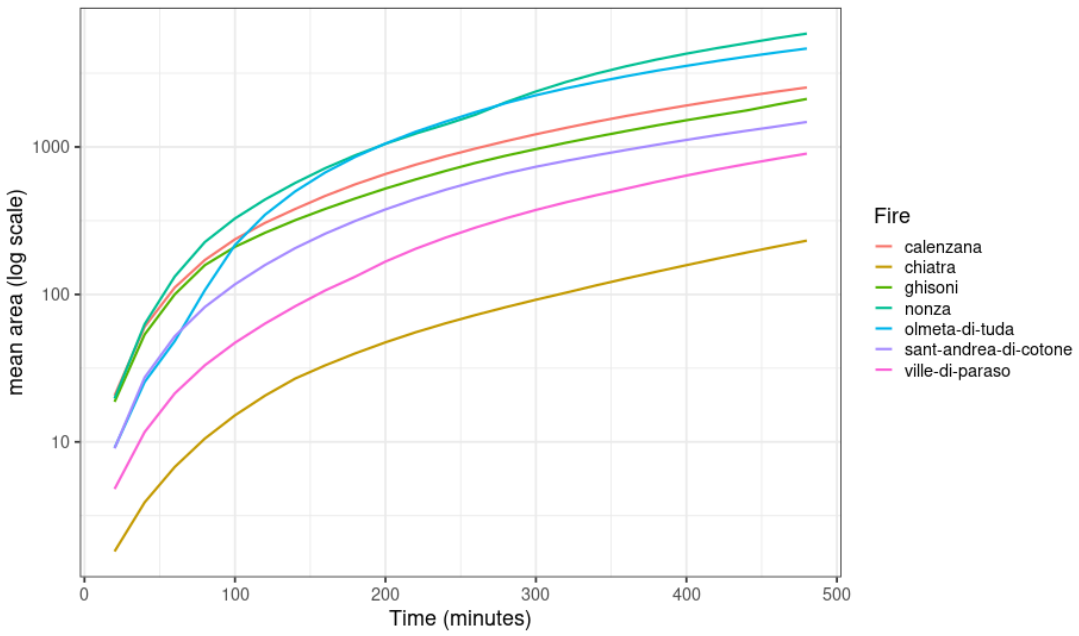


Figure A.8. – Evolution of the burnt area for the 7 fires

The results of this regression are presented in Table A.3, that shows the predictable impact of wind (positive) and moisture (negative). Figure A.9 presents the time fixed-effect coefficients (left panel) with 95% confidence interval. This figure confirms the sharp increase in the total cost after 300 minutes. The right panel presents the difference between coefficients with 1 hour lag, to make the increase clearer.

Independent variable	coefficient	SE	P-value
Intercept	1 657 282	80 511	$< 2e^{-16} * **$
moisture	−6 694 084	234 678	$< 2e^{-16} * **$
speed of wind	420 351	5053	$< 2e^{-16} * **$
Number of observations	84 000		
R^2	0.2435		

Table A.3. – Dependent variable: cost of the fire. The 84 000 observations correspond to 7 fires times 500 simulations times 24 20-minute time steps. Fire and time fixed-effects are not reported in this table.

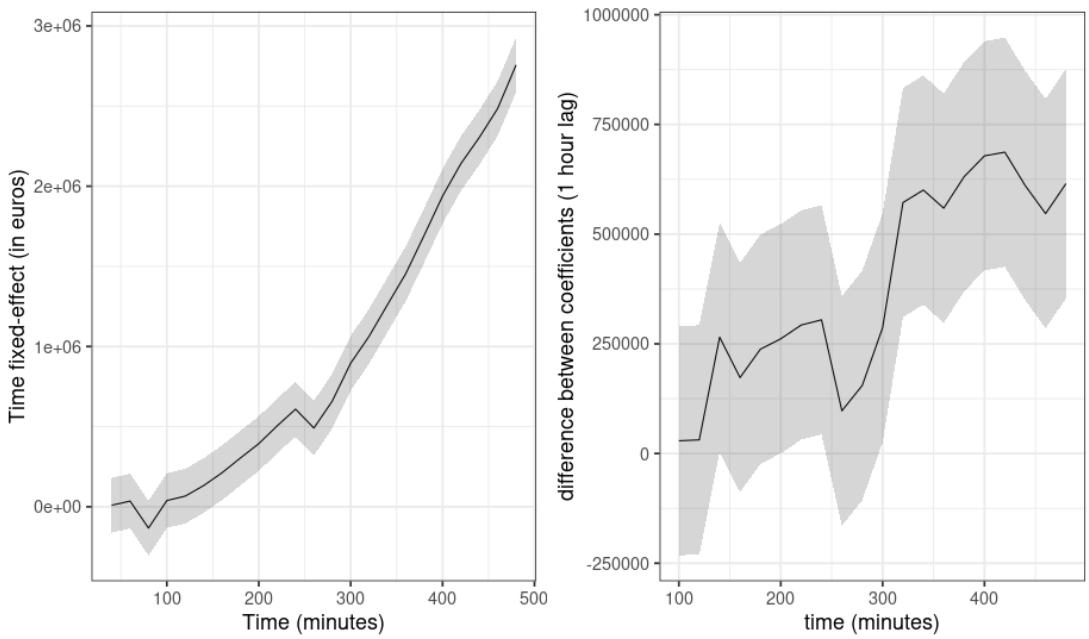


Figure A.9. – Evolution of the time fixed effect for the regression of global cost.

Right panel represents the 60 minute difference between coefficients. The regression also includes fire fixed-effects, the moisture and the speed of wind. Shaded areas represent 95% confidence intervals.

Table A.4 and Figure A.10 present the same exercise for the cost per hectare, whereas Table A.5 and Figure A.11 present the same exercise for the burned area.

As for the global cost, wind and moisture have a significant impact on both the cost per hectare and on the burnt area. Clearly, the significant impact of meteorological

Independent variable	coefficient	SE	P-value
Intercept	2398	42.3	$< 2e^{-16} * **$
moisture	−1778	123.4	$< 2e^{-16} * **$
speed of wind	85.8	2.657	$< 2e^{-16} * **$
Number of observations	84 000		
R^2	0.2557		

Table A.4. – Dependent variable: cost of the fire per hectare. The 84 000 observations correspond to 7 fires times 500 simulations times 24 20-minute time steps. Fire and time fixed-effects are not reported in this table.

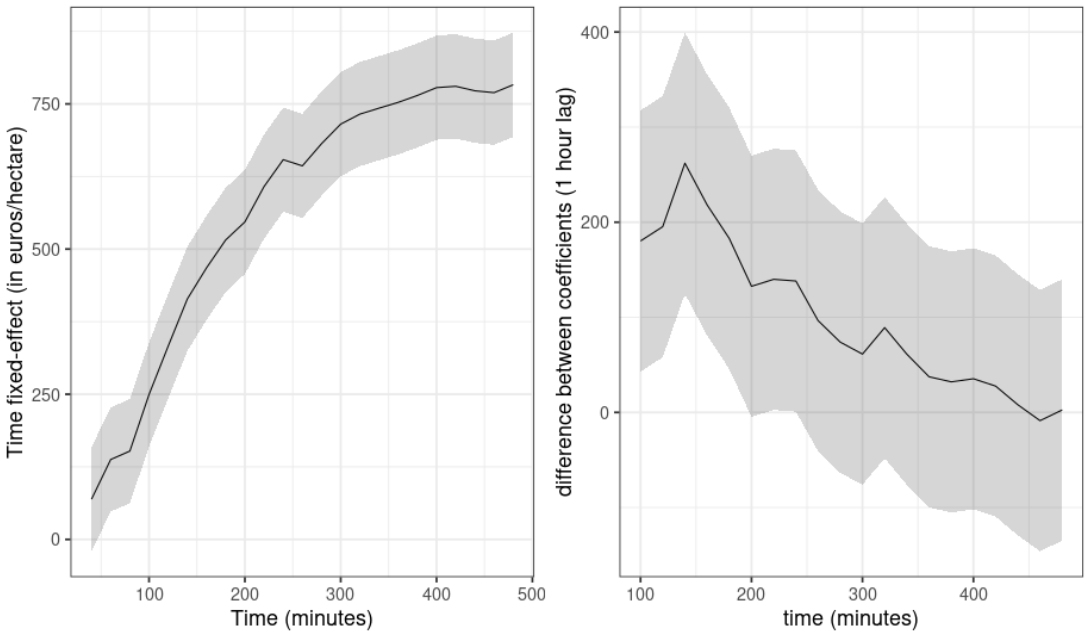


Figure A.10. – Evolution of the time fixed effect for the regression of the cost per hectare.

conditions on the cost per hectare is human-made phenomenon. It is due to the seven fires having their ignition points away from the most costly area. It might be because arsonists want to avoid big wastes, or because the less sensitive areas are less watched by firefighters. As a consequence, when the fires evolve quickly, they tend to burn lands with higher values than the lands close to the ignition point. This finding strengthens the view that taking real-life ignition points is more relevant than just randomly drawing them.

Right panel of Figure A.9 can be interpreted as the marginal cost of the duration of a fire, which is the main focus of this paper. It suggests that the marginal cost increases sharply after 300 minutes, from roughly €250 000 per additional hour to €600 000, more than the double. Two things happen after 300 minutes: first, the cost per hectare seems to reach a plateau after a sharp increase in the 5 first hours, to stabilize around

Independent variable	coefficient	SE	P-value
Intercept	−646.6	29.3	$< 2e^{-16} * **$
moisture	−5946	85.4	$< 2e^{-16} * **$
speed of wind	256.5	1.839	$< 2e^{-16} * **$
Number of observations	84 000		
R^2	0.4855		

Table A.5. – Dependent variable: surface of burned area. The 84′000 observations correspond to 7 fires times 500 simulations times 24 20-minute time steps. Fire and time fixed-effects are not reported in this table.

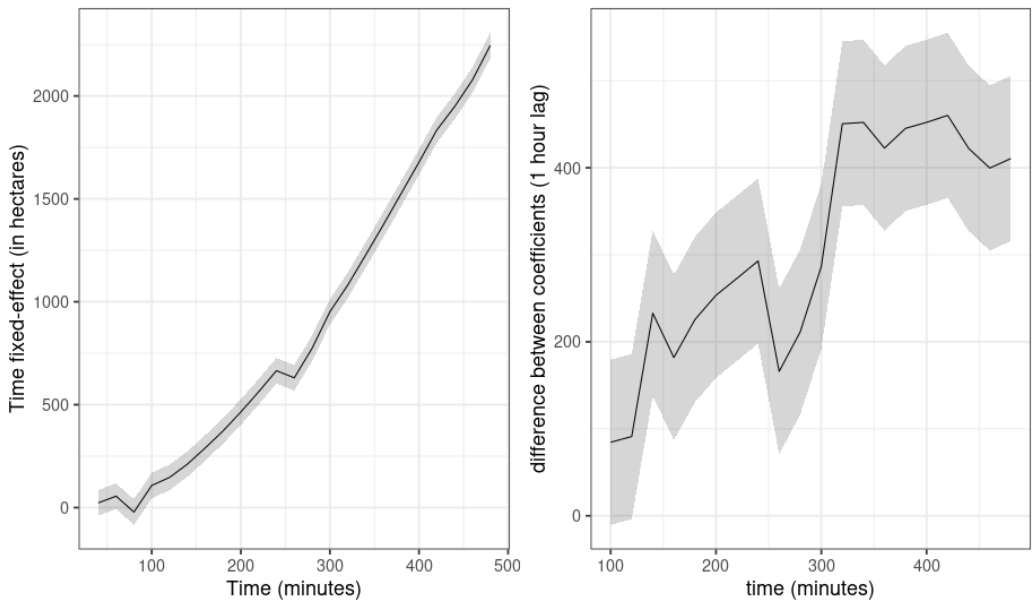


Figure A.11. – Evolution of the time fixed effect for the regression of the burned area.

€750 per hectare. Second, the marginal area doubles from 200 hectares per additional hour to 400 hectares.

Equation (A.4) assumes that the marginal cost of duration is independent on the meteorological conditions. This is admittedly a limitation of the exercise, since it is expected that high-speed wind increases the marginal cost of duration. In order to take this impact into account, a modified version of Equation (A.4) is used where the wind is removed as regressor, but limit the observations to the 50% fires with the highest speed of wind, and then re-run it for the remaining 50%. The results of this modified regression procedure are represented in Figure A.12 for the global cost, Figure A.13 for the cost per hectare, and Figure A.14 for the size of the fire.

The shapes of marginal effects in the case of the windy fires looks closely like the shapes from Equation (A.4), but with higher impacts. The increase in marginal cost of duration after 300 minutes is still visible in Figure A.12, from roughly €500 000 to more than €1×10^6.

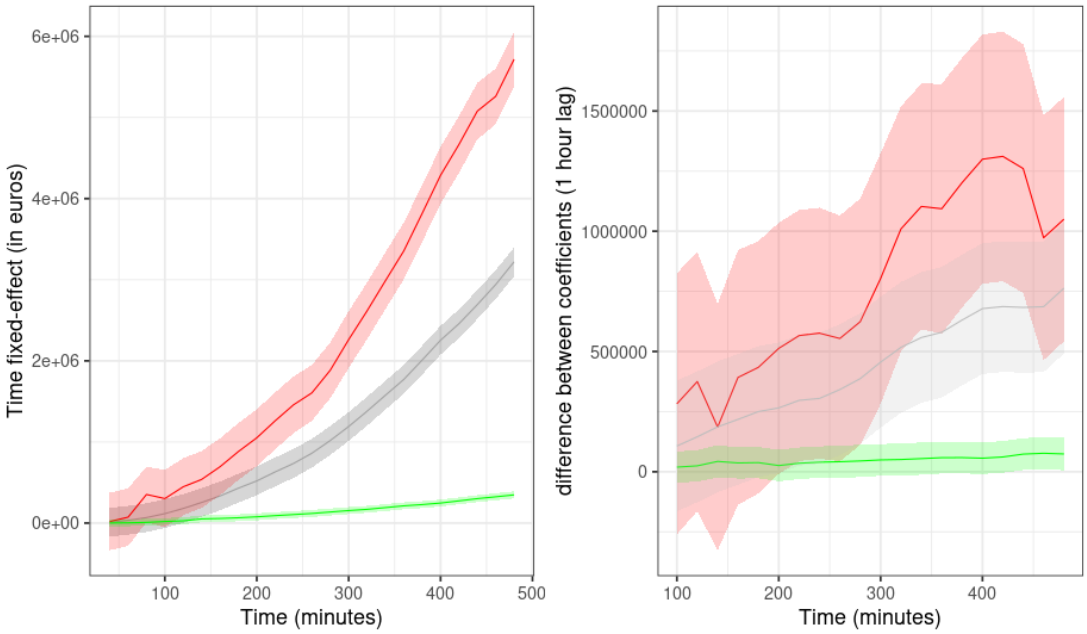


Figure A.12. – Evolution of the time fixed effect for the regression of global cost depending on the speed of wind.
Red line is for fast wind, green for slow wind, grey for the whole sample. Shaded areas represent 95% confidence intervals.

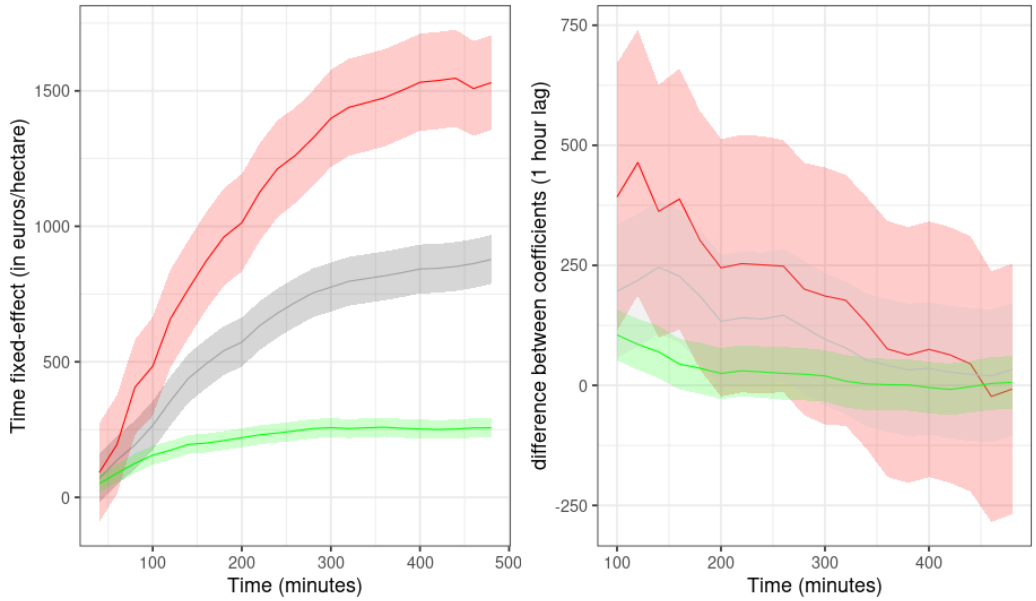


Figure A.13. – Evolution of the time fixed effect for the regression of cost per hectare depending on the speed of wind.
Red line is for fast wind, green for slow wind, grey for the whole sample. Shaded areas represent 95% confidence intervals.

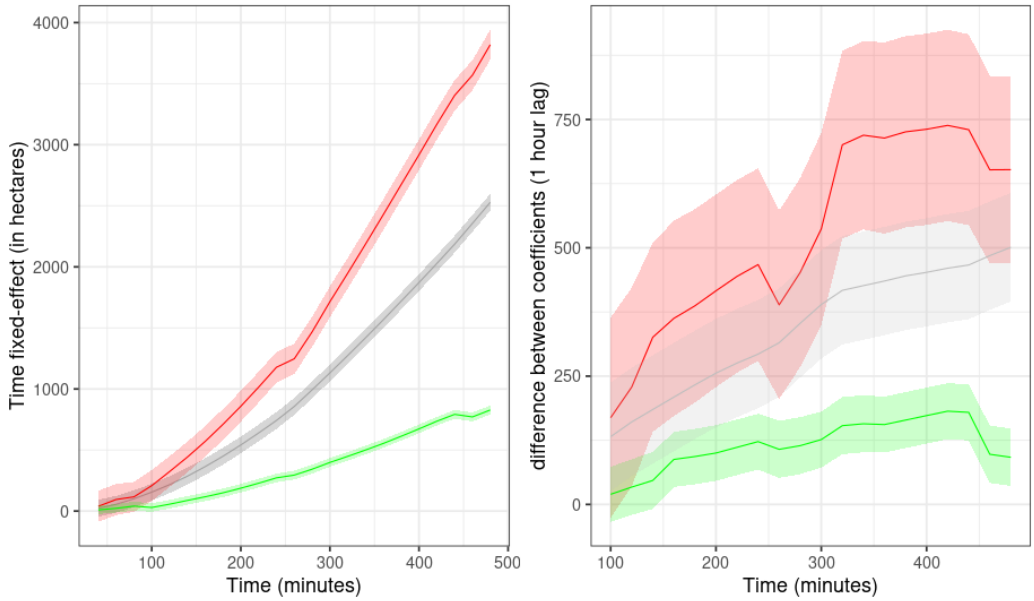


Figure A.14. – Evolution of the time fixed effect for the regression of the burned area depending on the speed of wind.
Red line is for fast wind, green for slow wind, grey for the whole sample. Shaded areas represent 95% confidence intervals.

Figure A.12 also highlights the impact of wind. The curve for slow-wind fires appears low and flat compared to the one for high-speed fires. Indeed, the time coefficient for low-wind fires is 154 431, some 14.6 times lower than the same coefficient for high-speed wind (2 261 840). In terms of marginal cost, the slow-wind fires exhibit a much smoother increase over time, from €19 024 to €73 845, with no particular break after 5 hours. It is important to note that the median speed, that separates our subsamples, is only 11.7km/h, which is quite slow in absolute terms (a typical average wind in a breeze), whereas the mean speed of the 50% windiest fires is 24.26km/h. It follows that even a very modest wind should alert about the marginal cost of a fire's duration.

A.5. Discussion

The results presented in the previous section rely on a model where firefighting actions are totally absent. Thus, they should not be considered as a realistic estimate of the actual costs of the fires but more of what-if simulated scenari. For instance, the endangerment of human life is probably overestimated, because in real life situations, firefighters generally prioritize saving human lives over other immediate objectives. However, those simulations can make sense since they pinpoint the hot-spots that are the worthiest saving. In other words, the costs presented here should be interpreted as the maximum cost of a worthwhile intervention aiming at stopping a fire. Anyway, it is hoped that the coupled fire-economy model can help firefighters sizing their interventions and directing them toward the most sensitive areas or to the least use the 5 hours threshold to take fire fighting planning decisions. Corsica is also very typical of

a Mediterranean climate and landscape that my be found in parts of Australia, California, and Europe, so this model may be adapted to other regions as an operational tool for firefighting management.

Beside these practical matters, it was not expected to find how large the gross gain from rapid intervention seems to be. Needless to say, a complete economic analysis of wildfire duration would require a careful account for the firefighting cost, with its complex balance between fixed cost and variable cost, fuel management policy, suppression and prevention. Again, that would go beyond the scope of the present paper. But as a of discussion, by roughly estimating a reduction of the duration of a given fire by one hour would require 3 hours of use of 5 water-bomber aircraft, which sounds like a comfortable support. If one hour of aircraft costs €10 000, then this extra support would be barely worthwhile in the case of slow-wind events. In fact, in these situations, aircraft might not even be necessary to achieve rapid containment of the fire. In the presence of wind, on the other hand, such massive support would more than pay for itself, at least from a social point of view, even before the break at 5 hours. And even if these estimates strongly depend on the choice of a scaling parameter, the mean cost per hectare falls in the range of cost estimates found in the literature. Moreover, the difference between the roughly €500 000 marginal cost of duration and the €150 000 of this hypothetical support is so massive that it would probably resist a quite substantial fall in the scaling parameter. To be clear, saying that a massive intervention would pay for itself does not mean it would be an optimal policy: other actions such as prioritizing the protection of the most sensitive parcels might be a less costly way to mitigate the cost of a fire. But a €500 000 to €1×10^6 marginal cost for the windiest events leaves a lot of room for even costly interventions to be worthwhile.

A.6. Conclusion

Wildfires are obviously not a *normal* phenomenon whose upper tail can be neglected. But even the central part of the distribution is worth studying. Reducing the duration of small scale fires could well pay for itself. A wildfire simulator was used to run 3500 virtual 8-hour long events. The ignition points and meteorological conditions were chosen from 7 real fires that took place in Corsica between 2017 and 2018. Assessing the cost of real-life wildfires requires a very careful investigation to observe the extent to which the burnt areas are impacted. There is no known equivalent of such investigations in the case of simulated fires, with results may be taken with caution, but parameterization of the simulations were made so as to generate a mean cost per hectare (€669) that is in line with values that are reported in the literature. By focusing on the 50% windiest events, the marginal cost of the duration of a fire is around €500 000 per hour before 5 hours, and more than €1×10^6 after 5 hours. This finding is an argument for trying to reduce the duration of already small fires, even if this assertion should be supported by a careful analysis of the fighting cost side of the equation that is not provide here.

We hope that the integrated tool presented in this article will contribute to feed the fire-fighting policy making, in line with the 10th recommendation of French Senator Vogel's report on the bush fire policy[139]: "in order to improve knowledge of the costs of damages related to forest fires, initiate work on the development of an assessment

tool economic loss and damage caused or likely to be caused."

Eventually, this tool aims at finding other purposes than just estimating the marginal cost of fire duration. These other purposes include the real-time crisis management and long term planing, in terms of fuel and urban-wild interface management.

Bibliography

[1] K. Abt, R. Huggett, and T. Holmes. Designing economic impact assessments for usfs wildfire programs. In T. P. Holmes, J. P. Prestemon, and K. L. Abt, editors, *The Economics of Forest Disturbances: Wildfires, Storms, and Invasive Species*, pages 151–166. Springer Science & Business Media, 2008.

[2] M. S. Advani, A. M. Saxe, and H. Sompolinsky. High-dimensional dynamics of generalization error in neural networks. *Neural Networks*, 132:428–446, 2020.

[3] A. A. Ager, N. M. Vaillant, M. A. Finney, and H. K. Preisler. Analyzing wildfire exposure and source–sink relationships on a fire prone forest landscape. *Forest Ecology and Management*, 267:271–283, 2012.

[4] A. A. Ager, K. C. Vogler, M. A. Day, and J. D. Bailey. Economic opportunities and trade-offs in collaborative forest landscape restoration. *Ecological Economics*, 136:226–239, 2017.

[5] F. Allaire, J.-B. Filippi, and V. Mallet. Generation and evaluation of an ensemble of wildland fire simulations. *International Journal of Wildand Fire*, 29(2):160–173, 2020.

[6] F. Allaire, V. Mallet, and J.-B. Filippi. Emulation of wildland fire spread simulation using deep learning. *Neural Networks*, 141:184–198, 2021.

[7] F. Allaire, V. Mallet, and J.-B. Filippi. Novel method for a posteriori uncertainty quantification in wildland fire spread simulation. *Applied Mathematical Modelling*, 90:527–546, 2021.

[8] M. Amodei, S. Isabelle, and J. Stein. Verification of the french operational high-resolution model arome with the regional brier probability score. *Meteorological Applications*, 22, 10 2015.

[9] J. L. Anderson. The impact of dynamical constraints on the selection of initial conditions for ensemble predictions: Low-order perfect model results. *Monthly Weather Review*, 125(11):2969–2983, 1997.

[10] P. L. Andrews, M. G. Cruz, and R. C. Rothermel. Examination of the wind speed limit function in the Rothermel surface fire spread model. *International Journal of Wildland Fire*, 22:959–969, 2013.

[11] P. D. Arendt, D. W. Apley, and W. Chen. Quantification of model uncertainty: Calibration, model discrepancy, and identifiability. *Journal of Mechanical Design*, 134(10), 09 2012. 100908.

[12] A. Bachmann and B. Allgöwer. A consistent wildland fire risk terminology is needed! *Fire Management Today 61*, pages 28–33, 2001.

[13] R. Baïle, J.-F. Muzy, and X. Silvani. Multifractal point processes and the spatial distribution of wildfires in french mediterranean regions. *Physica A: Statistical Mechanics and its Applications*, 568, Apr. 2021.

[14] A. Benali, A. C.L. Sá, A. R. Ervilha, R. M. Trigo, P. M. Fernandes, and J. M.C. Pereira. Fire spread predictions: Sweeping uncertainty under the rug. *Science of The Total Environment*, 592:187 – 196, 2017.

[15] A. Benali, A. R. Ervilha, A. C.L. Sá, P. M. Fernandes, R. M.S. Pinto, R. M. Trigo, and J. M.C. Pereira. Deciphering the impact of uncertainty on the accuracy of large wildfire spread simulations. *Science of The Total Environment*, 569-570:73 – 85, 2016.

[16] N. Bonneel, M. van de Panne, S. Paris, and W. Heidrich. Displacement interpolation using Lagrangian mass transport. *ACM Trans. Graph.*, 30(6):1–12, Dec. 2011.

[17] L. S. Bradshaw, J. E. Deeming, R. E. Burgan, and J. D. Cohen. The 1978 national fire-danger rating system: technical documentation. Gen. Tech. Rep. INT-169. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station. 44 pp., 1984.

[18] G. W. Brier. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1):1–3, 1950.

[19] S. P. Brooks and A. Gelman. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7(4):434–455, 1998.

[20] R. E. Burgan, P. L. Andrews, L. S. Bradshaw, C. H. Chase, R. A. Hartford, and D. J. Latham. Current status of the wildland fire assessment system (wfas). *Fire Management Notes*, 57(2):14–17, 1997.

[21] D. Butry, M. L. Grumpertz, and M. G. Genton. he production of large and small wildfires. In T. P. Holmes, J. P. Prestemon, and K. L. Abt, editors, *The Economics of Forest Disturbances: Wildfires, Storms, and Invasive Species*, pages 79–106. Springer Science & Business Media, 2008.

[22] D. T. Butry, D. E. Mercer, J. P. Prestemon, J. M. Pye, and T. P. Holmes. What is the price of catastrophic wildfire? *Journal of Forestry*, pages 9–17, 2001.

[23] R. E. Caflisch. Monte carlo and quasi-monte carlo methods. *Acta Numerica*, 7:1–49, 1998.

[24] L. Cai, H. S. He, Y. Liang, Z. Wu, and C. Huang. 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, 01 2019.

[25] D. E. Calkin, M. P. Thompson, M. A. Finney, and K. D. Hyde. A real-time risk assessment tool supporting wildland fire decisionmaking. *Journal of Forestry*, 109(5):274–280, 2011.

[26] N. Chen, W. Liu, R. Bai, and A. Chen. Application of computational intelligence technologies in emergency management: a literature review. *Artificial Intelligence Review*, 52(3):2131–2168, 2017.

[27] J. D. Cohen and M. A. Finney. An examination of fuel particle heating during fire spread. In D. X. Viegas, editor, *Proceedings of the VI International Conference on Forest Fire Research*, 15-18 November 2010 2010. 13 p.

[28] M. G. Cruz. Monte carlo-based ensemble method for prediction of grassland fire spread. *International Journal of Wildland Fire*, 19:521–530, 07 2010.

[29] M. G. Cruz and M. E. Alexander. Uncertainty in model predictions of wildland fire rate of spread. In *Advances in forest fire research*, pages 466–477. Imprensa da Universidade de Coimbra, 2014.

[30] G. Damblin, M. Couplet, and B. Iooss. Numerical studies of space filling designs: optimization of Latin Hypercube Samples and subprojection properties, 2013.

[31] L. V. del Hoyo, M. P. M. Isabel, and F. J. M. Vega. Logistic regression models for human-caused wildfire risk estimation: analysing the effect of the spatial accuracy in fire occurrence data. *European Journal of Forest Research*, 130:983–996, 2011.

[32] M. D'Este, A. Ganga, M. Elia, R. Lovreglio, V. Giannico, G. Spano, G. Colangelo, R. Lafortezza, and G. Sanesi. Modeling fire ignition probability and frequency using hurdle models: a cross-regional study in southern europe. *Ecological Processes*, 9, 2020.

[33] A. J. Dowdy, G. A. Mills, K. Finkele, and W. de Groot. Australian fire weather as represented by the mcarthur forest fire danger index and the canadian forest fire weather index. *Centre for Australian Weather and Climate Research Tech. Rep*, 10, 2009.

[34] M. Drusch, F. Gascon, and M. Berger. Gmes sentinel-2 mission requirements document. EOP-SM/1163/MR-dr, 2010. issue 2.1.

[35] T. J. Duff, J. G. Cawson, B. Cirulis, P. Nyman, G. J. Sheridan, and K. G. Tolhurst. Conditional performance evaluation: Using wildfire observations for systematic fire simulator development. *Forests*, 9(4), 2018.

[36] T. J. Duff, D. M. Chong, P. Taylor, and K. G. Tolhurst. 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, 2012.

[37] T. J. Duff, D. M. Chong, and K. G. Tolhurst. Indices for the evaluation of wildfire spread simulations using contemporaneous predictions and observations of burnt area. *Environmental Modelling & Software*, 83:276 – 285, 2016.

[38] D. Dupuy, C. Helbert, and J. Franco. DiceDesign and DiceEval: Two R packages for design and analysis of computer experiments. *Journal of Statistical Software*, 65(11), 2015.

[39] B. Efron and R. Tibshirani. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statist. Sci.*, 1(1):54–75, 02 1986.

[40] A. Ervilha, J. Pereira, and J. Pereira. On the parametric uncertainty quantification of the rothermel's rate of spread model. *Applied Mathematical Modelling*, 41:37 – 53, 2017.

[41] H. Faure. Discrépance de suites associées à un système de numération (en dimension s) [discrepancy of sequences associated with a number nystem (in dimension s)]. *Acta Arithmetica*, 41(4):337–351, 1982.

[42] J. Feranec, T. Soukup, G. Hazeu, and G. Jaffrain, editors. *European Landscape Dynamics: CORINE Land Cover Data.* CRC Press, Boca Raton, USA, 2016.

[43] J.-B. Filippi, V. Mallet, and B. Nader. Representation and evaluation of wildfire propagation simulations. *International Journal of Wildland Fire*, 23:46–57, 2013.

[44] J.-B. Filippi, V. Mallet, and B. Nader. Evaluation of forest fire models on a large observation database. *Natural Hazards and Earth System Sciences Discussions*, 2(11):3077–3091, 2014.

[45] J.-B. Filippi, F. Morandini, J. H. Balbi, and D. R.C. Hill. Discrete event front-tracking simulation of a physical fire-spread model. *SIMULATION*, 86(10):629–646, 2010.

[46] M. Finney, C. McHugh, I. Grenfell, K. Riley, and K. Short. A simulation of probabilistic wildfire risk components for the continental united states. *Stochastic Environmental Research and Risk Assessment*, 25:973–1000, 10 2011.

[47] M. A. Finney. Farsite: Fire area simulator-model development and evaluation. Res. Pap. RMRS-RP-4, Revised 2004, Ogden, UT: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 47 p., 1998.

[48] M. A. Finney, J. D. Cohen, S. S. McAllister, and W. M. Jolly. On the need for a theory of wildland fire spread. *International Journal of Wildland Fire*, 22(1):25–36, 2013.

[49] M. A. Finney, I. C. Grenfell, C. W. McHugh, R. C. Seli, D. Trethewey, R. D. Stratton, and S. Brittain. A method for ensemble wildland fire simulation. *Environmental Modeling & Assessment*, 16(2):153–167, Apr 2011.

[50] R. Flamary and N. Courty. POT Python Optimal Transport library, 2017.

[51] W. L. Fons. Analysis of fire spread in light forest fuels. *Journal of Agricultural Research*, 72(3):92–121, 1946.

[52] F. M. Fujioka. A new method for the analysis of fire spread modeling errors. *International Journal of Wildland Fire*, 11:193–203, 01 2002.

[53] A. Ganteaume, A. Camia, M. Jappiot, J. San-Miguel-Ayanz, M. Long-Fournel, and C. Lampin. A review of the main driving factors of forest fire ignition over europe. *Environmental Management*, 51(3):651–662, 2013.

[54] T. Ghisu, B. Arca, G. Pellizzaro, and P. Duce. An optimal cellular automata algorithm for simulating wildfire spread. *Environmental Modelling & Software*, 71:1–14, 2015.

[55] L. Giglio, W. Schroeder, and C. O. Justice. The collection 6 modis active fire detection algorithm and fire products. *Remote Sensing of Environment*, 178:31–41, 06 2016.

[56] F. D. Giuseppe, C. Vitolo, B. Krzeminski, C. Barnard, P. Maciel, and J. San-Miguel. Fire weather index: the skill provided by the european centre for medium-range weather forecasts ensemble prediction system. *Natural Hazards and Earth System Sciences*, 20(8):2365–2378, Aug. 2020.

[57] X. Glorot and Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. In Y. W. Teh and M. Titterington, editors, *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, volume 9 of *Proceedings of Machine Learning Research*, pages 249–256, Chia Laguna Resort, Sardinia, Italy, 13–15 May 2010. JMLR Workshop and Conference Proceedings.

[58] A. Graves, S. Fernández, and J. Schmidhuber. Multi-dimensional recurrent neural networks. arXiv preprint, arXiv:0705.2011, 2007.

[59] F. Guo, Z. Su, G. Wang, L. Sun, F. Lin, and A. Liu. Wildfire ignition in the forests of southeast china: Identifying drivers and spatial distribution to predict wildfire likelihood. *Applied Geography*, 66:12–21, 2016.

[60] J. H Scott and R. E Burgan. Standard fire behavior fuel models : A comprehensive set for use with rothermel's surface fire spread model. *The Bark Beetles, Fuels, and Fire Bibliography*, 06 2005.

[61] J. H. Halton. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik*, 2(1):84–90, 1960.

[62] F. Hamann and S. Bazer. En corse, une filière bois peu développée. *INSEE Analyse*, 10, may 2016.

[63] J. M. Hammersley. Monte-carlo methods for solving multivariable problems. *Annals of the New York Academy of Sciences*, 86(3):844–874, 1960.

[64] S. R. Hanna, J. C. Chang, and M. E. Fernau. 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, 1998.

[65] C. C. Hardy and C. E. Hardy. Fire danger rating in the united states of america: an evolution since 1916. *International Journal of Wildland Fire*, 16(2):217–231, 2007.

[66] J. Hegde and B. Rokseth. Applications of machine learning methods for engineering risk assessment – a review. *Safety Science*, 122:104492, Feb. 2020.

[67] F. J. Hickernell. Lattice rules: How well do they measure up? In P. Hellekalek and G. Larcher, editors, *Random and Quasi-Random Point Sets*, pages 109–166. Springer New York, New York, NY, 1998.

[68] J. L. Hodges and B. Y. Lattimer. Wildland fire spread modeling using convolutional neural networks. *Fire Technology*, 55(6):2115–2142, 2019.

[69] S. Hofferer. Chiffres clés de l'agriculture corse. Technical report, DRAAF de Corse. Service Régional de l'Information Statistique et Economique (SRISE), 2020.

[70] B. Iooss, L. Boussouf, V. Feuillard, and A. Marrel. Numerical studies of the metamodel fitting and validation processes. *International Journal On Advances in Systems and Measurements*, 3:11–21, 2010.

[71] P. Jain, S. C. Coogan, S. G. Subramanian, M. Crowley, S. W. Taylor, and M. D. Flannigan. A review of machine learning applications in wildfire science and management. *Environmental Reviews*, 2020.

[72] S. Joe and F. Y. Kuo. Remark on algorithm 659. *ACM Transactions on Mathematical Software*, 29(1):49–57, 2003.

[73] S. Joe and F. Y. Kuo. Constructing sobol sequences with better two-dimensional projections. *SIAM Journal on Scientific Computing*, 30(5):2635–2654, 2008.

[74] P. Johnston, J. Kelso, and G. J. Milne. Efficient simulation of wildfire spread on an irregular grid. *International Journal of Wildland Fire*, 17:614–627, 2008.

[75] M. I. Jordan and T. M. Mitchell. Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260, 2015.

[76] M. Katurji, J. Nikolic, S. Zhong, S. Pratt, L. Yu, and W. E. Heilman. Application of a statistical emulator to fire emission modeling. *Environmental Modelling & Software*, 73:254–259, 2015.

[77] M. C. Kennedy and A. O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, 2001.

[78] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.

[79] C. Lac, J.-P. Chaboureau, V. Masson, J.-P. Pinty, P. Tulet, J. Escobar, M. Leriche, C. Barthe, B. Aouizerats, C. Augros, P. Aumond, F. Auguste, P. Bechtold, S. Berthet, S. Bieilli, F. Bosseur, O. Caumont, J.-M. Cohard, J. Colin, F. Couvreux, J. Cuxart, G. Delautier, T. Dauhut, V. Ducrocq, J.-B. Filippi, D. Gazen, O. Geoffroy, F. Gheusi, R. Honnert, J.-P. Lafore, C. Lebeaupin Brossier, Q. Libois, T. Lunet, C. Mari, T. Maric, P. Mascart, M. Mogé, G. Molinié, O. Nuissier, F. Pantillon, P. Peyrillé, J. Pergaud, E. Perraud, J. Pianezze, J.-L. Redelsperger, D. Ricard, E. Richard, S. Riette, Q. Rodier, R. Schoetter, L. Seyfried, J. Stein, K. Suhre, M. Taufour, O. Thouron, S. Turner, A. Verrelle, B. Vié, F. Visentin, V. Vionnet, and P. Wautelet. Overview of the meso-nh model version 5.4 and its applications. *Geoscientific Model Development Discussions*, 11(5):1929–1969, 2018.

[80] C. Lac, J.-P. Chaboureau, V. Masson, J.-P. Pinty, P. Tulet, J. Escobar, M. Leriche, C. Barthe, B. Aouizerats, C. Augros, P. Aumond, F. Auguste, P. Bechtold, S. Berthet, S. Bielli, F. Bosseur, O. Caumont, J.-M. Cohard, J. Colin, F. Couvreux, J. Cuxart, G. Delautier, T. Dauhut, V. Ducrocq, J.-B. Filippi, D. Gazen, O. Geoffroy, F. Gheusi, R. Honnert, J.-P. Lafore, C. L. Brossier, Q. Libois, T. Lunet, C. Mari, T. Maric, P. Mascart, M. Mogé, G. Molinié, O. Nuissier, F. Pantillon, P. Peyrillé, J. Pergaud, E. Perraud, J. Pianezze, J.-L. Redelsperger, D. Ricard, E. Richard, S. Riette, Q. Rodier, R. Schoetter, L. Seyfried, J. Stein, K. Suhre, M. Taufour, O. Thouron, S. Turner, A. Verrelle, B. Vié, F. Visentin, V. Vionnet, and P. Wautelet. Overview of the meso-NH model version 5.4 and its applications. *Geoscientific Model Development*, 11(5):1929–1969, 2018.

[81] G. Laneve, V. Pampanoni, and R. U. Shaik. The daily fire hazard index: A fire danger rating method for mediterranean areas. *Remote Sensing*, 12(15):2356, 2020.

[82] C. Lautenberger. Mapping areas at elevated risk of large-scale structure loss using monte carlo simulation and wildland fire modeling. *Fire Safety Journal*, 91:768 – 775, 2017. Fire Safety Science: Proceedings of the 12th International Symposium.

[83] B. Lawson and O. Armitage. Weather guide for the canadian forest fire danger rating system. Natural Resources Canada, Canadian Forest Service, Northern Forestry Centre, Edmonton, Alberta. 84 p., 2008.

[84] N. Liu, J. Lei, W. Gao, H. Chen, and X. Xie. Combustion dynamics of large-scale wildfires. *Proceedings of the Combustion Institute*, Jan. 2021.

[85] Y. Liu, M. Y. Hussaini, and G. Ökten. Accurate construction of high dimensional model representation with applications to uncertainty quantification. *Reliability Engineering & System Safety*, 152:281–295, 2016.

[86] Y. Liu, M. Y. Hussaini, and G. Ökten. Global sensitivity analysis for the rothermel model based on high-dimensional model representation. *Canadian Journal of Forest Research*, 45(11):1474–1479, 2015.

[87] Y. Liu, E. Jimenez, M. Hussaini, G. Ökten, and S. Goodrick. Parametric uncertainty quantification in the rothermel model with randomised quasi-monte carlo methods. *International Journal of Wildland Fire*, 24:307–316, 05 2015.

[88] S. M. Lundberg and S.-I. Lee. A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 4765–4774. Curran Associates, Inc., 2017.

[89] V. Mallet, D. Keyes, and F. Fendell. Modeling wildland fire propagation with level set methods. *Computers & Mathematics with Applications*, 57(7):1089–1101, 2009.

[90] V. Mallet, A. Tilloy, D. Poulet, S. Girard, and F. Brocheton. Meta-modeling of ADMS-urban by dimension reduction and emulation. *Atmospheric Environment*, 184:37–46, 2018.

[91] J. Mandel, J. D. Beezley, and A. K. Kochanski. Coupled atmosphere-wildland fire modeling with WRF 3.3 and SFIRE 2011. *Geoscientific Model Development*, 4(3):591–610, 2011.

[92] W. Mell, M. A. Jenkins, J. Gould, and P. Cheney. A physics-based approach to modelling grassland fires. *International Journal of Wildland Fire*, 16(1), 2007.

[93] C. Miller and A. A. Ager. A review of recent advances in risk analysis for wildfire management. *International Journal of Wildland Fire*, 22(1), 2013.

[94] C. Miller, J. Hilton, A. Sullivan, and M. Prakash. SPARK – A bushfire spread prediction tool. In R. Denzer, R. M. Argent, G. Schimak, and J. Hřebíček, editors, *Environmental Software Systems. Infrastructures, Services and Applications*, pages 262–271, Cham, 2015. Springer International Publishing.

[95] C. Miller, P. Landres, and P. Alaback. Evaluating risks and benefits of wildland fire at landscape scales. In: Neuenschwander, Leon F.; Ryan, Kevin C., tech. eds. Crossing the millenium: integrating spatial technologies and ecological principles for a new age in fire management: Proceedings, Joint Fire Sciences conference and workshop; June 15-17, 1999, Boise, ID. Boise, ID: University of Idaho. 10 p, 2000.

[96] R. M.S. Pinto, A. Benali, A. C.L. Sá, P. M. Fernandes, P. M.M. Soares, R. M. Cardoso, R. M. Trigo, and J. M.C. Pereira. Probabilistic fire spread forecast as a management tool in an operational setting. *SpringerPlus*, 5:1205, 07 2016.

[97] A. H. Murphy. A new vector partition of the probability score. *Journal of Applied Meteorology (1962-1982)*, 12(4):595–600, 1973.

[98] H. Niederreiter. Low-discrepancy and low-dispersion sequences. *Journal of Number Theory*, 30(1):51–70, 1988.

[99] M. Parisien, V. Kafka, K. Hirsch, J. Todd, S. Lavoie, and P. Maczek. Mapping wildfire susceptibility with the burn-p3 simulation model. Canadian Forest Service, Northern Forestry Centre, 2005.

[100] M.-A. Parisien, D. A. Dawe, C. Miller, C. A. Stockdale, and O. B. Armitage. Applications of simulation-based burn probability modelling: a review. *International Journal of Wildland Fire*, 28(12):913–926, 2019.

[101] J.-S. Park and J. Baek. Efficient computation of maximum likelihood estimators in a spatial linear model with power exponential covariogram. *Computers & Geosciences*, 27(1):1 – 7, 2001.

[102] S. Paz, Y. Carmel, F. Jahshan, and M. Shoshany. Post-fire analysis of pre-fire mapping of fire-risk: A recent case study from mt. carmel (israel). *Forest Ecology and Management*, 262(7):1184 – 1188, 2011.

[103] D. Pettenella, M. Marchetti, D. Marino, A. Marucci, M. Ottaviano, and B. Lasserre. Proposal for a harmonized methodology to assess socio-economic damages from forest fires in europe. Technical report, Laboratorio di Ecologia e Geomatica Forestale, Università Degli Studi Del Molise, 2008.

[104] J. P. Prestemon, T. J. Hawbaker, M. Bowden, J. Carpenter, M. T. Brooks, K. L. Abt, R. Sutphen, and S. Scranton. Wildfire ignitions: A review of the science and recommendations for empirical modeling. Gen. Tech. Rep. SRS-GTR-171. Asheville, NC: USDA-Forest Service, Southern Research Station. 20 p., 2013.

[105] D. Quang and X. Xie. DanQ: a hybrid convolutional and recurrent deep neural network for quantifying the function of DNA sequences. *Nucleic Acids Research*, 44(11):e107–e107, 2016.

[106] M. I. Radaideh, K. Borowiec, and T. Kozlowski. Integrated framework for model assessment and advanced uncertainty quantification of nuclear computer codes under Bayesian statistics. *Reliability Engineering & System Safety*, 189:357 – 377, 2019.

[107] D. Radke, A. Hessler, and D. Ellsworth. FireCast: Leveraging deep learning to predict wildfire spread. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pages 4575–4581. International Joint Conferences on Artificial Intelligence Organization, 2019.

[108] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. the MIT Press, 2006.

[109] D. B. Rideout, Y. Wei, A. G. Kirsch, , and S. J. Botti. Toward a unified economic theory of fire program analysis with strategies for empirical modeling. In T. P. Holmes, J. P. Prestemon, and K. L. Abt, editors, *The Economics of Forest Disturbances: Wildfires, Storms, and Invasive Species*, pages 361–380. Springer Science & Business Media, 2008.

[110] R. Rittmaster, W. Adamowicz, B. Amiro, and R. Pelletier. Economic analysis of health effects from forest fires. *Canadian Journal of Forest Research*, 36:866–877, 2006.

[111] L. A. Robinson. How us government agencies value mortality risk reductions. *Review of Environmental Economics and Policy*, 1:283–299, 2007.

[112] M. C. Rochoux, A. Collin, C. Zhang, A. Trouvé, D. Lucor, and P. Moireau. Front shape similarity measure for shape-oriented sensitivity analysis and data assimilation for eikonal equation. *ESAIM: ProcS*, 63:258–279, 2018.

[113] M. C. Rochoux, S. Ricci, D. Lucor, B. Cuenot, and A. Trouvé. Towards predictive data-driven simulations of wildfire spread – part i: Reduced-cost ensemble kalman filter based on a polynomial chaos surrogate model for parameter estimation. *Natural Hazards and Earth System Sciences*, 14(11):2951–2973, 2014.

[114] R. C. Rothermel. 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, 1972.

[115] O. Roustant, D. Ginsbourger, and Y. Deville. DiceKriging, DiceOptim: Two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization. *Journal of Statistical Software*, 51(1):1–55, 2012.

[116] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn. Design and analysis of computer experiments. *Statistical Science*, 4(4):409–423, 1989.

[117] M. Salis, A. Ager, B. Arca, M. Finney, V. Bacciu, P. Duce, and D. Spano. Assessing exposure of human and ecological values to wildfire in sardinia, italy. *International Journal of Wildland Fire*, 22:549–565, 06 2013.

[118] M. Salis, B. Arca, F. Alcasena, M. Arianoutsou, V. Bacciu, P. Duce, B. Duguy, N. Koutsias, G. Mallinis, I. Mitsopoulos, J. Moreno, J. Perez, I. Urbieta, F. Xystrakis, G. Zavala, and D. Spano. Predicting wildfire spread and behavior in mediterranean landscapes. *International Journal of Wildland Fire*, 25:1015–1032, 06 2016.

[119] R. Salvador, J. Piñol, S. Tarantola, and E. Pla. Global sensitivity analysis and scale effects of a fire propagation model used over mediterranean shrublands. *Ecological Modelling*, 136(2):175 – 189, 2001.

[120] F. Santambrogio. *Optimal Transport for Applied Mathematicians*, volume 87 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser / Springer, Cham, 2015.

[121] W. Schroeder, P. Oliva, L. Giglio, and I. Csiszar. The new viirs 375 m active fire detection data product: Algorithm description and initial assessment. *Remote Sensing of Environment*, 143:85–96, 03 2014.

[122] J. Shao. Bootstrap estimation of the asymptotic variances of statistical functionals. *Annals of the Institute of Statistical Mathematics*, 42(4):737–752, Dec 1990.

[123] C. Shorten and T. M. Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1), 2019.

[124] D. Silver, A. Huang, C. Maddison, A. Guez, L. Sifre, G. Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis. Mastering the game of go with deep neural networks and tree search. *Nature*, 529:484–489, 01 2016.

[125] I. Sobol'. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics*, 7(4):86–112, 1967.

[126] I. Sobol'. Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Mathematics and Computers in Simulation*, 55(1-3):271–280, 2001.

[127] V. Stojanovic, S. He, and B. Zhang. State and parameter joint estimation of linear stochastic systems in presence of faults and non-Gaussian noises. *International Journal of Robust and Nonlinear Control*, 30(16):6683–6700, 2020.

[128] A. L. Sullivan. Wildland surface fire spread modelling, 1990–2007. 1: Physical and quasi-physical models. *International Journal of Wildland Fire*, 18:349–368, 07 2009.

[129] A. L. Sullivan. Wildland surface fire spread modelling, 1990–2007. 1: Physical and quasi-physical models. *International Journal of Wildland Fire*, 18:349–368, 07 2009.

[130] A. L. Sullivan. Wildland surface fire spread modelling, 1990–2007. 2: Empirical and quasi-empirical models. *International Journal of Wildland Fire*, 18:369–386, 07 2009.

[131] A. L. Sullivan. Wildland surface fire spread modelling, 1990–2007. 3: Simulation and mathematical analogue models. *International Journal of Wildland Fire*, 18:387–403, 07 2009.

[132] P. Termonia, C. Fischer, E. Bazile, F. Bouyssel, R. Brožková, P. Bénard, B. Bochenek, D. Degrauwe, M. Derková, R. El Khatib, R. Hamdi, J. Mašek, P. Pottier, N. Pristov, Y. Seity, P. Smolíková, O. Španiel, M. Tudor, Y. Wang, C. Wittmann, and A. Joly. The aladin system and its canonical model configurations arome cy41t1 and alaro cy40t1. *Geoscientific Model Development*, 11(1):257–281, 2018.

[133] M. Thompson, D. Calkin, J. H. Scott, and M. Hand. Uncertainty and probability in wildfire management decision support: An example from the united states. In K. Riley, P. Webley, and M. Thompson, editors, *Natural Hazard Uncertainty*

Assessment: Modelling and Decision Support, volume 223 of *Geophysical Monograph*, chapter 4, pages 31–41. American Geophysical Union, first edition, 2017.

[134] K. Tolhurst, B. Shields, and D. Chong. Phoenix: Development and application of a bushfire risk management tool. *Australian Journal of Emergency Management*, 23(4):47–54, 2008.

[135] A. Trucchia, M. D'Andrea, F. Baghino, P. Fiorucci, L. Ferraris, D. Negro, A. Gollini, and M. Severino. PROPAGATOR: An operational cellular-automata based wildfire simulator. *Fire*, 3(3), 2020.

[136] A. Trucchia, V. Egorova, G. Pagnini, and M. Rochoux. On the merits of sparse surrogates for global sensitivity analysis of multi-scale nonlinear problems: Application to turbulence and fire-spotting model in wildland fire simulators. *Communications in Nonlinear Science and Numerical Simulation*, 73:120–145, 2019.

[137] C. Tymstra, R. Bryce, B. Wotton, S. Taylor, and O. Armitage. Development and structure of prometheus: the canadian wildland fire growth simulation model. Natural Resources Canada, Information Report NOR-X-417, Canadian Forest Service, Northern Forestry Centre, Edmonton, Alberta. 88 p., 2010.

[138] C. E. Van Wagner and T. L. Pickett. Equations and fortran program for the canadian forest fire weather index system. Forestry Technical Report No. 33. Ottawa, Environment Canada, Canadian Forestry Service, Petawawa National Forestry Institute, 1985.

[139] J.-P. Vogel. Rapport d'information fait au nom de la commission des finances sur la lutte contre les feux de forêt. Technical report, French Senate, 2019.

[140] D. Wilks. Chapter 7 - statistical forecasting. In D. S. Wilks, editor, *Statistical Methods in the Atmospheric Sciences*, volume 100 of *International Geophysics*, pages 215 – 300. Academic Press, third edition, 2011.

[141] X. Wu, T. Kozlowski, H. Meidani, and K. Shirvan. Inverse uncertainty quantification using the modular Bayesian approach based on Gaussian process, Part 1: Theory. *Nuclear Engineering and Design*, 335:339–355, Aug 2018.

[142] Y. Xie, F. Xing, X. Kong, H. Su, and L. Yang. Beyond classification: Structured regression for robust cell detection using convolutional neural network. In N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 358–365, Cham, 2015. Springer International Publishing.

[143] Z. Yuan, Y. Jiang, J. Li, and H. Huang. Hybrid-dnns: Hybrid deep neural networks for mixed inputs. arXiv preprint arXiv:2005.08419, 2020.

[144] M. D. Zeiler. Adadelta: An adaptive learning rate method. arXiv preprint arXiv:1212.5701, 2012.

[145] L. Zhou, H. Tao, W. Paszke, V. Stojanovic, and H. Yang. PD-type iterative learning control for uncertain spatially interconnected systems. *Mathematics*, 8(9):1528, Sept. 2020.

[146] T. Zhou, L. Ding, J. Ji, L. Yu, and Z. Wang. Combined estimation of fire perimeters and fuel adjustment factors in FARSITE for forecasting wildland fire propagation. *Fire Safety Journal*, 116:103167, 2020.

Abstract

This work addresses the quantification of wildfire risk by relying on simulations of fire spread. The objectives are to compute the probability distribution of burned surfaces that could result from wildfire ignition and quickly generate maps to assess which areas should receive focused protection against wildfires. This probability distribution should represent the uncertainty in the simulations. First, an ensemble of wildland fire spread simulations accounting for sources of uncertainty is generated following a Monte Carlo approach, and probabilistic evaluation of the predictions with observations is carried out. Then, the underlying probability distributions are calibrated based on the observations by adapting the Wasserstein distance to the comparison of burned surfaces to improve prediction performance in presence of uncertainty. Subsequently, a deep learning approach is followed to train a "hybrid" neural network with a convolutional part, thus building an emulator of "potential" fire size simulated by the fire spread model allowing to considerably reduce the computational time implied by the large amount of simulations required for high-resolution maps. Eventually, this emulator is applied to derive fire danger mapping from daily weather forecasts and applied to assess relatively large fire events.

Keywords: uncertainty quantification, probabilistic prediction, metamodeling, deep learning, wildland fire spread, fire danger mapping

Résumé

Ce travail porte sur la quantification du risque incendie en se fondant sur des simulations de propagation de feux de forêt. Les objectifs sont de calculer la distribution de probabilité des surfaces brûlées pouvant résulter d'un départ de feu et de générer des cartes permettant d'estimer quelles zones doivent être protégées en priorité. Les simulations pouvant donner lieu à des erreurs de prévision, la distribution de probabilité en question doit représenter l'incertitude associée aux simulations. Dans un premier temps, un ensemble de simulations de propagation de feux de forêt prenant en compte les sources d'incertitude est généré selon une approche Monte Carlo, et les prévisions, probabilistes, sont comparées à des observations selon des critères adaptés. Ensuite, les distributions de probabilité sous-jacentes sont calibrées à partir des observations en adaptant la distance de Wasserstein à la comparaison de surfaces brûlées afin d'améliorer la qualité des prévisions, tout en tenant compte de l'incertitude. Par la suite, une approche d'apprentissage profond est mise en œuvre pour entraîner un réseau de neurones "hybride" avec une partie convolutionnelle, élaborant ainsi un émulateur de taille de feu "potentielle" simulée par le modèle de propagation afin de diminuer considérablement le temps de calcul associé au grand nombre de simulations nécessaires à l'élaboration de cartes à haute résolution. Enfin, l'émulateur est utilisé pour générer des cartes de danger incendie à partir de vraies prévisions météorologiques générées pour des jours où des feux relativement grands ont eu lieu.

Mots clés: quantification d'incertitude, simulation d'ensemble, métamodélisation, apprentissage profond, propagation de feux de forêt, cartes de danger incendie