

Novel method for a posteriori uncertainty quantification in wildland fire spread simulation

Gautam Dilip Hariharan

Sandeep Bhandari

gautam.hariharan@tu-dortmund.de

sandeep.bhandari@tu-dortmund.de

TU Dortmund

Dortmund, Nordrhein-Westfalen, Germany

Gautam Dilip Hariharan and Sandeep Bhandari. 2022. Novel method for a posteriori uncertainty quantification in wildland fire spread simulation. In *Seminar: Uncertainty Quantification in machine learning*. , 10 pages.

ABSTRACT

Simulation of the wildland fire is essential for making accurate predictions of fire in real-time and information should be provided to the forest manager. Large uncertainty in the input parameters exists which needs to be quantified. An ensemble forecast is used for this purpose and the input probability distribution is applied based on expert knowledge.

A novel method is used to generate calibrated ensembles using input distribution of whose PDF is described by posteriori pseudo-likelihood function. For calculating the distance between simulated and observed burn surface for all the cases of seven fires, Wasserstein distance is used. A gaussian emulator is constructed using the MCMC algorithm in about one day with 8 computing cores to generate the calibrated input distribution due to the high dimensions and computation requirements of the input parameters.

The calibrated ensembles lead to better accuracy and consistency than uncalibrated ensembles. The marginal probability distribution of the input parameters results in a significant difference when the calibration value is increased. This help in limiting over-prediction and restricting the probability to cover more of the observed burn surface.

1 INTRODUCTION

Wildfire prediction is a very challenging task due to high uncertainty in the input parameters and the environmental conditions. Several physical and quasi-physical models[14] have been proposed to understand the fire spread behavior with different complexity. The dynamics of the fire behavior shape are described in the 2D fire simulation model[15], which are better and quicker for generating operational predictions. This model relies on the semi-empirical equation for calculating the rate of spread(ROS) of fire which further depends on wind and slope factors. Instead of using a single deterministic prediction, a probabilistic model for quantifying the uncertainties can be proposed. The Monto Carlo method can be used to propagate the uncertainties where several input sets are sampled independently from a predefined probability distribution that describes the input uncertainty. This leads to the ensemble of the fire spread simulation which can be combined to form a burn probability map showing the probability in observed burn surface as illustrated in Figure 1.

2022-07-29 22:39. Page 1 of 1–10.

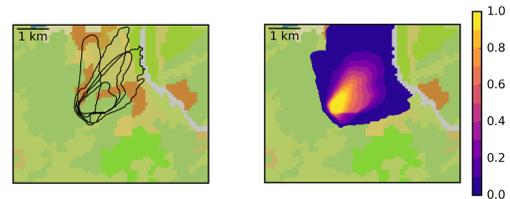


Figure 1: An ensemble result for the simulated fire front. The left shows the five black lines of a firefront. Right is the aggregation of five firefront in to burn probability map.

The main idea of the present paper is to calibrate the probability distribution of the input parameters and to solve the problem of inverse uncertainty quantification[16]. Two major problems were identified: one with a large number of uncertain inputs and other models considered is "black box" whose formula for likelihood cannot be implemented easily. So, a novel score is applied for the calibration of input probability distribution comparing surfaces using Wasserstein distance for various wildland fires. Various uncertain inputs were found in[1] using direct uncertainty quantification. Therefore, the main goal is to find better probability scores by calibration of uncertain distribution on the observed burn surface of seven Corsican fire spreads.

The paper outlines the following sections. In Section 2, the theoretical method for the calibration is discussed. In Section 3, those methods are applied for the real-life application of seven Corsican fire spreads. In Section 4, the results for the emulation energy function, calibrated distribution simulation and ensemble evaluation is shown. In Section 5, conclusion is made and further research is defined.

2 A POSTERIORI UNCERTAINTY QUANTIFICATION

The author explained three sources of uncertainty namely unknown parameters, model inadequacy, and observation error. They proposed a Bayesian approach for the calibration of different forms of input uncertainty. The starting part of this method points to a model updating equation (equation (4) in [9]). Two additive terms that, respectively, take into account model inadequacies and observation mistakes are used to quantify the difference between observations and the associated model outputs at given parameter values. Input uncertainties are approached by perturbation whereas model adequacy and observation errors are not taken into account. Error

59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116

in observed burn surfaces is considered to be measured negligible when compared to the error due to other sources of uncertainty.

In this section, we consider only one observed burn surface. Let S_{obs} be an observed burn surface and M represents a numerical model of fire spread in which input depends upon input u of perturbation d based on reference inputs. This simulation is applied on the observed burn surface which results simulated burned surface denoted as S_u (one may define $S_u = M(u)$). To make a direct comparison between S_u and S_{obs} , uncertainty is modeled by attributing probabilistic distribution u to the perturbation vector. The distribution of u is assumed to be described by a new quantity called probability density function denoted as g and its prior distribution by prior density function denoted as f . So, the main idea is to obtain g by making the best possible use of f , S_{obs} and M .

2.1 Distribution based on Wasserstein distance

$p(\cdot | S_{obs})$ is the posterior density function that would be the classical choice for g and is obtained according to Bayes' rule:

$$p(u | S_{obs}) = \frac{\mathcal{L}(S_{obs} | u) f(u)}{\int \mathcal{L}(S_{obs} | u) f(u) du}, \dots\dots\dots(1)$$

where the likelihood of the observation S_{obs} knowing the perturbation vector u is $L(S_{obs} | u)$. But if we define the likelihood, then we need to make an appropriate probabilistic hypothesis, where the realization of a 2D stochastic process whose distribution is dependent on u is S_{obs} . Hypothesizing it like this is not trivial, but to move forward, we would use S_u instead of just u and define a conditional probability distribution for S_{obs} based on S_u . The fact that the most likely realizations of S_{obs} are the ones that are most similar to S_u is a desirable property of such a probability distribution. When two points in a 2D domain are close, similarity should take into account high correlation between these points. For example, if a given location has a high probability of being burned, the neighboring locations should also have a high probability of being burned. Even so, it may not entail that just because defining a likelihood for a vector is feasible, it has to be the same for a random surface as well. Hence, we may use a calibrated distribution where the density g can be written as (inspired by Bayes' rule):

$$g_{E,\beta}(u) = \frac{e^{-\beta E(u)} f(u)}{\int e^{-\beta E(u)} f(u) du}, \dots\dots\dots(2)$$

where $\beta > 0$ and E is a positive "energy" function that is 0 when $S_u = S_{obs}$ and the more dissimilar S_u and S_{obs} become, the more it increases. Here, in equation (1), the role of \mathcal{L} is played by a pseudo-likelihood function. The calibrated family of functions is different from the Gibbs measures although inspired by it as the exponential is multiplied by the prior PDF f . The higher β , the more weight the pseudo-likelihood function holds and when $\beta = 0$, the calibrated PDF and prior PDF are equal.

To make suitable choices for E , there exists several scores to compare S_{obs} and S_u and could be used directly or after minor modifications. A metric between the probability distributions called the Wasserstein distance is used to introduce a novel score in the present study. A reference to [13]. 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 is defined as follows:

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

where $\|\cdot\|_2$ 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 μ is the conditional measure relative to the first variable and ν is the conditional measure on the second variable. It is natural to consider $q = 2$ for comparison between surfaces and choose uniform measures whose support is respectively S_{obs} and S_u for the probability measures μ and ν . After making these choices, $E(u)$ can now be defined as:

$$E(u) = \inf_Y \left\{ \int_{S_{obs} \times S_u} \|x - y\|_2^2 \gamma(x, y) dx dy \mid \int_{S_u} \gamma(x, y) dy = \frac{1(x \in S_{obs})}{|S_{obs}|}, \dots\dots\dots(4) \int_{S_{obs}} \gamma(x, y) dx = \frac{1(y \in S_u)}{|S_u|} \right\}$$

where 1 is the indicator function, $\|\cdot\|_2$ is the Euclidean distance (in \mathbb{R}^2), and $|S|$ is the surface area of S . We can think of it as the minimum energy required to move the points contained in S_{obs} to transform the surface into S_u . When both surfaces are the same, $E(u) = 0$. There is no simple analytic formula for Wasserstein distance except for some particular cases. This leads to us considering a discrete approximation of $E(u)$ instead, which is obtained numerically via a discretization of the PDFs by a sum of Dirac delta distributions. Now, $E(u)$ can be defined as:

$$E(u) = \inf_Y \left\{ \int_{S_{obs} \times S_u} \|x - y\|_2^2 \gamma(x, y) dx dy \mid \int_{S_u} \gamma(x, y) dy = \frac{1}{J} \sum_{j=1}^J \delta_{x_j}(x), \dots\dots\dots(5) \int_{S_{obs}} \gamma(x, y) dx = \frac{1}{K} \sum_{k=1}^K \delta_{y_k}(y) \right\}$$

where δ_x is the Dirac delta distribution at point x belongs to \mathbb{R}^2 and each x_j belongs to S_{obs} , whereas y_k belongs to S_u . The admissible distributions in this discrete setting 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 . Here, the infimum of (5) is arrived at and is the solution of the linear programming problem below:

$$\min_{\gamma_{j,k}} \sum_{j=1}^J \sum_{k=1}^K \gamma_{jk} \|x_j - y_k\|_2^2, \dots\dots\dots(6)$$

subject to

$$\gamma_{jk} \geq 0, \sum_j \gamma_{jk} = \frac{1}{K},$$

and

$$\sum_k \gamma_{jk} = \frac{1}{J}, \dots\dots\dots(7)$$

which is also called the Earth Mover's distance [3]. From graph theory, we know that the optimal γ is a sparse matrix that has at most $J + K - 1$ non-zero cells. There is an issue that we need to address and that is of the denominator of $g_{E,\beta}(u)$ in (2), which is an intractable high-dimensional integral. But this integral does not depend on u , so for a given beta, the PDF is known up to some constant factor. In this case, we can draw samples from that distribution using the Metropolis-Hastings (MH) algorithm. We can compute S_u and the Wasserstein distance to obtain $E(u)$ in a reasonable amount of time. Even still, a large number of iterations (10^5) of the MH algorithm may be required to obtain a sufficiently large sample which is extremely time consuming. To make the MH algorithm faster, we use an emulator \tilde{E} . The emulated value $\tilde{E}(u)$ is much faster to compute and provides a good approximation of $E(u)$. In order to determine how appropriate the sample returned by the MH algorithm is, a multivariate diagnostic metric as proposed in [4] can be used.

2.2 Emulation

2.2.1 Gaussian Process Modelling. We use the Gaussian process (GP) modelling method in this study, also called kriging. Here, $y(\mathbf{u})$ is seen as a realization of a Gaussian process Y_u and is indexed by u . It means that any random vector $[Y_{u^1}, \dots, Y_{u^n}]^T$ with $n < \infty$ components follows a Gaussian multivariate distribution. Let the trend function of the process be denoted by $a: \mathbb{E}[Y_u] = a(\mathbf{u})$. $Z_u = Y_u - a(\mathbf{u})$, which is the centered process, is also a Gaussian and has a covariance function of the form $\text{Cov}(u, u') = \sigma^2 \rho(u - u')$, where $\sigma^2 > 0$ and ρ are the correlation functions between two input points u and u' .

We have a set of training data for use: $(u^i, y(u^i))_{i=1, \dots, n}$. Let us denote $Y^n = [Y_{u^1}, \dots, Y_{u^n}]^T$, $\mathbf{y}_n = [y(u^1), \dots, y(u^n)]^T$ and define the correlation matrix R_n on the inputs of the training data:

$$R_n = \left(\rho(u^i - u^j) \right)_{1 \leq i, j \leq n} \dots \dots \dots (8)$$

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

For a new point u^* (irrespective of whether it is in the training sample), the correlation vector is defined as

$\mathbf{r}^* = [r(u^* - u^1), \dots, r(u^* - u^n)]^T$. As per the assumptions made on Y_u , the joint probability distribution (Y^n and Y_{u^*}) and conditional distribution of Y_{u^*} knowing Y_{u^*} are Gaussian. We can write:

$$Y_{u^*} | Y^n \sim \mathcal{N}(\mathbb{E}[Y_{u^*} | Y^n], \text{Var}[Y_{u^*} | Y^n]) \dots \dots \dots (9)$$

where

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

and

$$\text{Var}[Y_{u^*} | Y^n] = \sigma^2 \left(1 - \mathbf{r}^{*T} R_n^{-1} \mathbf{r}^* \right) \dots \dots \dots (11)$$

We can define an emulator \tilde{y} of y as the mean of the conditional variable given by (10) for any $u^* \in \mathcal{D}$:

$$\tilde{y}(u^*) = a(u^*) + \mathbf{r}^{*T} R_n^{-1} (\mathbf{y}_n - \mathbf{a}_n) \dots \dots \dots (12)$$

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

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

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

In this study, universal kriging is used which signifies that the trend is a unknown polynomial (as seen in [12]). Practically, the $2d + 2$ hyperparameters $\sigma^2, \alpha_0, \alpha_1, \dots, \alpha_d, \theta_1, \dots, \theta_d$ that the Gaussian process is defined by are unknown and can be estimated as the maximum likelihood estimators for the training data set.[10]

2.2.2 Design of experiments and error metrics. We use a Latin hypersquare sample (LHS) to obtain the the inputs of the training sample with optimized discrepancy. Because of the fact that the GP emulator is built from the points of the training sample, to evaluate he approximation of the emulator far from the training points, a complementary test sample is generated (obtained with algorithm for an optimal validation design [?]) which relies on a low discrepancy sequence whose points are chosen to maintain a low discrepancy when both training and test samples are taken together. The goal of this process is to select points that are located far from each other but it is also expected to be far from the points displayed in the training sample where the expectations for approximation error is higher. As per the test sample $(u^i, y(u^i))_{i=1, \dots, n_{\text{test}}}$, we can use several error metrics to evaluate the emulator \tilde{y} . In this study, we use the mean absolute error (MAE) which is defined as:

$$\text{MAE} = \frac{1}{n_{\text{test}}} \sum_{i=1}^{n_{\text{test}}} |\tilde{y}(u^i) - y(u^i)| \dots \dots \dots (14)$$

We then introduce the standardized mean square error (SMSE), which is defined as:

$$\text{SMSE} = \frac{\sum_{i=1}^{n_{\text{test}}} (\tilde{y}(u^i) - y(u^i))^2}{\sum_{i=1}^{n_{\text{test}}} (y(u^i) - \bar{y})^2} \dots \dots \dots (15)$$

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

$$Q_2 = 1 - \frac{\sum_{i=1}^{n_{\text{test}}} (\tilde{y}(u^i) - y(u^i))^2}{\sum_{i=1}^{n_{\text{test}}} (y(u^i) - \bar{y})^2} = 1 - \text{SMSE} \dots \dots \dots (16)$$

As the error of the emulator decreases, the MAE advances towards 0 and the Q_2 advances towards 1. If a model always predicts the mean of the training set, the Q_2 would approximately be equal to 0.

2.3 Extension to several fire cases

If we consider K fire cases, we can compute the energy functions E_1, \dots, E_K that correspond to each fire. Intuitively, a choice for the combined energy function is $E : \mathbf{u} \mapsto \sum_{k=1}^K E_k(\mathbf{u})$. But an issue arises when the variations of $E_1(\mathbf{u})$ (for example) are much higher than for the rest of the fires. If this happens, then $E_1(\mathbf{u})$ will mostly dictate the variations of the pseudo-likelihood, and the calibrated distribution will be mostly represented by the information of the first fire at the cost of the rest of the fires.

To go around this issue, we can weigh each fire depending on the values taken by $E_k(\mathbf{u})$ and the energy function is then defined as the weighted sum of squared Wasserstein distances:

$$E(\mathbf{u}) = \sum_{k=1}^K w_k E_k(\mathbf{u}), \dots \dots \dots (17)$$

where we define the weights using all points from the training dataset:

$$w_k = \frac{n}{\sum_{i=1}^n E_k(\mathbf{u}^i)} \dots \dots \dots (18)$$

We can emulate $E(\mathbf{u})$ directly but even though the function is positive, emulation by Gaussian Process does not ensure positivity outside the training sample. As an alternate, we may emulate $L(\mathbf{u}) = \log E(\mathbf{u})$ by the GP procedure, which leads to the emulator $\tilde{L}(\mathbf{u})$. Emulation of $E(\mathbf{u})$ is derived from taking the exponential $\tilde{E}(\mathbf{u}) = \exp \tilde{L}(\mathbf{u})$, which guarantees positivity. The R-package DiceKriging is used to implement the Gaussian Process emulation. [12]

2.4 Sampling from the calibrated distribution

In this section, we talk about the procedure we use to obtain a sample following a PDF of the form $g_{\tilde{E}, \beta}$ as defined in equation (2).

We use the emulator \tilde{E} instead of the energy function E to run the algorithm in reasonable computational time with the assumption that the target distribution of MH is close enough to the desired distribution which has a PDF $g_{E, \beta}$. This procedure is represented in the algorithm in Figure 2.

The process is presented in Figure 2. Let m denote the number of chains and n denote the number of samples per chain such that the i -th element of the j -th chain is denoted as $u_{i,j}$. The distribution that is used to sample a candidate $u_{c,j}$ from element $u_{i-1,j}$ is defined by the PDF $q : \mathbf{u} \mapsto q(\mathbf{u} | \mathbf{v})$. Figure 2 presents the use of a version of MH with several chains and is motivated by the convergence diagnosis for MCMC (Markov Chain Monte Carlo) algorithms introduced by Brooks and Gelman[4]. Choosing the starting points $u_{1,1}, \dots, u_{1,m}$ quite far from each other is recommended. We can easily parallelize the loop on the m chains. The between-sequence covariance matrix B/n (of size d) and the within sequence covariance matrix W are computed as given below, based on the chains returned by the MH algorithm:

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

$$W = \frac{1}{m(n-1)} \sum_{j=1}^m \sum_{i=1}^n (u_{i,j} - \bar{u}_j) (u_{i,j} - \bar{u}_j)^T, \dots \dots \dots (21)$$

```

Define  $m, n$ , and an instrumental distribution of PDF  $q : \mathbf{u} \mapsto q(\mathbf{u} | \mathbf{v})$ 
for  $j = 1, \dots, m$  do
    Choose a starting point  $u_{1,j}$ 
    for  $i = 2, \dots, n$  do
        Sample a candidate  $u_{c,j} \sim q(\cdot | u_{i-1,j})$ 
        Compute the ratio
            
$$\tau = \frac{g_{\tilde{E}, \beta}(u_{c,j}) q(u_{i-1,j} | u_{c,j})}{g_{\tilde{E}, \beta}(u_{i-1,j}) q(u_{c,j} | u_{i-1,j})} = \frac{e^{-\beta \tilde{E}(u_{c,j})} f(u_{c,j}) q(u_{i-1,j} | u_{c,j})}{e^{-\beta \tilde{E}(u_{i-1,j})} f(u_{i-1,j}) q(u_{c,j} | u_{i-1,j})} \quad (19)$$

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

Figure 2: Metropolis-Hastings algorithm applied to $g_{\tilde{E}, \beta}$ (several chains)

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

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

where the largest eigenvalue of the symmetric, positive definite matrix $W^{-1}B/n$ is λ_1 . At convergence, \hat{R}^d tends to 1, and as per the recommendations of Gelman and Brooks[4], it can be considered that a sufficient number of MH iterations has been carried out if $\hat{R}^d < 1.1$ for the second half of the chains. This shows that the set that has the second half of all m chains makes up a representative sample of the target distribution when $\hat{R}^d < 1.1$.

3 APPLICATION TO WILDLAND FIRE SPREAD

3.1 Fire spread simulation

We use the open source fire spread solver ForeFire [7] in this study. A front-tracking technique is used to model the spread of the fire front (the interface between the burned surface and the rest of the simulation domain—unburned). Langrangian markers linked by a dynamic mesh is used to discretize the fire front. The markers are advanced according to the surface geometry and the rate of spread (ROS). As opposed to discrete time simulation methods, ForeFire is reliant on a discrete event specification. The markers are the advanced according to a given spatial increment and from its speed, we deduce the time at which the marker will reach its next position, thereby rendering the process asynchronous. Advancing a marker in time is counted as an event. With other events, we can calculate the future location and time advance of a marker, like topology checks that determine whether the markers describe a properly

burned surface and then reshapes the fire front when it is not the case.

The ROS in this study is computed according to the empirical model of Rothermel[11], widely used in wildland fire simulation, which consists of 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 this 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 n .

Some additional assumptions are also made: first, the mineral damping coefficient is assumed to be 1 and second, the fuel mineral content is negligible which means that the net initial fuel loading is equal to the fuel load σ_f . Additionally, to make up 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 W , the wind speed vector predicted by the meteorological model, so that $W_S = 0.4W \cdot n$. We also apply the revised wind speed limit function (proposed by Andrews et al.[2]) which is expressed in equation (21).

The process used to advance the markers of the fire front is based on a first-order approximation. Assuming a marker that at time t_i is located at x_i , with its normal to the front denoted as n_i (oriented towards the unburned area), the next location is determined by:

$$x_{i+1} = x_i + \delta l n_i, \dots \dots \dots (23)$$

and the advance in time is dependent on ROS_i , which is the ROS calculated with the values of the environmental inputs at location x_i at time t_i as defined here:

$$t_{i+1} = t_i + \frac{\delta l}{ROS_i} \dots \dots \dots (24)$$

3.2 Prior uncertainty in input data

| Input | Unit | Perturbation | Distribution | Notes |
|-------------------------------|---------------------|----------------------------|---|---|
| Wind direction | ° | Additive | $\mathcal{N}(0, 60^2)$ | Truncated to [-180, 180] |
| Wind speed norm | 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 | MJ kg ⁻¹ | Additive | $\mathcal{U}(-5, 5)$ | |
| Particle density | kg m ⁻³ | Additive | $\mathcal{U}(-300, 300)$ | |
| Fuel height | m | Multiplicative, individual | $\mathcal{U}(0.4, 1.6)$ | |
| Fuel load | kg m ⁻² | Multiplicative, individual | $\mathcal{U}(0.4, 1.6)$ | |
| Surface-volume ratio | m ⁻¹ | Multiplicative, individual | $\mathcal{U}(0.4, 1.6)$ | |
| Direction from ignition point | ° | 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 | $\mathcal{U}(-1, 1) \times \Delta_{\max}$ | $\Delta_{\max} \in \{10, 60, 120, 180, 240\}$ |

Figure 3: Prior probability distributions of the perturbations in simulation inputs

Figure 3 shows the marginals of the prior distribution for the perturbation variables used in this study. We assume the inputs to be independent. The perturbation variables are described in a previous study [1], where the distributions were mostly truncated normal (table 2 in [1]). Expert knowledge and the support of the

marginals followed values found in the scientific literature were used to choose the value of the standard deviations. In this study, we choose a wider support for some of the distributions, which can be substituted for uniform distributions. This choice is a result of initial efforts at calibration where evaluation was performed on ensembles of simulations that came from a few specific input distributions. This gave way to obtain better probabilistic scores for most of the fires studied even though it can be considered as somewhat arbitrary.

In these simulations, up to 13 burnable fuel types may be involved and are linked to the Corine Land Cover classification [6]. In Figure 3, when the perturbation is "individual", it signifies that one perturbation coefficient is sampled for each fuel type.

For one simulation, the number of perturbation coefficients generated are $d = 48$. The distributions are truncated for wind direction and wind speed norm, and the "Distribution" column corresponds to the distribution before truncation is applied. The main reason we choose these distributions with finite support is to avoid sampling extreme values that may turn out to be unrealistic or can lead to non-physical parameter values.

The perturbation parameters "direction from ignition point" and "distance to ignition point" specify the uncertainty in the location of the ignition point. In order to sample a perturbed ignition point, we could 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 Δ_{\max} depends on the fire case: the range for sampling the perturbed ignition point may therefore be within a radius ranging from 100 m to 1 km around the reference. Similarly, the maximum perturbation Δ_{\max} for the time of start of the fire and time of end of the fire depends on the fire case. The varying uncertainty is because of the information available about each fire. Δ_{\max} is specific to each fire case for the last three variables. But for the calibrations, one "reduced" variable exists for each of these three inputs, whose support is either [0, 1] or [-1, 1]. We obtain the actual perturbation used to run the fire spread simulations for a given fire case from the reduced variable after multiplication by Δ_{\max} .

3.3 Application to seven Corsican wildfires

The procedure for emulation and calibration is applied to $K = 7$ fires that took place in Corsica in the period 2017-2018 and are presented in [1]. From here on, we refer to "reference ensembles" as the ensembles obtained in this previous study.

We build an emulator with training sample of size 4000 and is evaluated with a test sample of size 2000. Following equation (5), for computation of the Wasserstein distance, we can consider an orthogonal uniform grid that covers the burned surface which is used to approximate the PDF by a sum of Dirac delta distributions at these points. The size of the surfaces influences the spatial resolution of the grid because of the drastic increase in computational cost with the number of points. The resolution is approximately 20 m for small burned surfaces and about 80 m for the largest ones. We carry out these computations using the package *ot* from the Python toolbox POT. [8]

The MH algorithm is then applied to several distributions with varying values of β ranging in $\{\frac{1}{10}, \frac{1}{7}, \frac{1}{4}, \frac{1}{2}, 1, 2\}$. $n = 150000$ iterations are carried out for each value of β for $m = 8$ chains. $q(u | v)$, which describes the instrumental distribution is a product of independent univariate truncated normal distributions. The k -th normal distribution is focused on the k -th component of v before truncation the standard deviation is equal to a twentieth of the perturbation range's width. The logarithm of the perturbation follows a truncated normal distribution for the perturbation of wind speed norm.

Taking the second half of the chains obtained using the MH algorithm, we get samples of size $m \times n/2 = 600000$ for every value of β . What we will refer to as "calibrated ensembles" from here on are the ensembles of wildland fire simulations that are generated for the seven fire cases based on these empirical distributions. A calibrated ensemble's size ranges between 2000 and 10000. The evaluation domain follows that of the reference ensembles but in contrast to the previous study, no computational time limit is applied to the simulations. "Prior ensembles" are obtained when the ensemble generation procedure is carried out based on the prior distributions.

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

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

This score lies between 0 and 1 and is negatively oriented. From the ensembles that forecast a constant probability, the probability $p_c = \frac{1}{N} \sum_{i=1}^N o_i$ is used to obtain the one with the lowest Brier score and the Brier score of this ensemble is $BS_c = p_c (1 - p_c)$. The Brier skill score (BSS) is defined as;

$$BSS = 1 - \frac{BS}{BS_c} \dots\dots\dots(26)$$

which is positively oriented. In case of several fires, the Brier scores are summarized by their mean. The corresponding value of BS_c is obtained with the mean of the p_c . Using these two global Brier scores, we can define the Brier skill score as can be seen in equation (26). In practice though, a Monte Carlo method is used to estimate the probabilities p_i , which means we only have an estimate of the true values of BS and BSS. The standard deviation of the estimator may be estimated with bootstrap [5], where we re-estimate the probabilities p_i by sampling with replacement among the burned surfaces. Given a significantly large set of bootstrap samples, we get σ_{BSS}^b , an estimator of the standard deviation of BSS. We can show that the estimators of both BS and BSS are asymptotically normal and that the bootstrap estimation is consistent by the regularity of BS. Now it becomes possible to establish an approximate confidence interval at level $1 - \alpha$ for BSS based on σ_{BSS}^b and the quantile of the standard normal distribution for probability $1 - \alpha/2$. To provide 95% confidence intervals of BSS, bootstrap re-sampling was carried out.

In order to create a summary of the information given by the other tools used for evaluation on several fires (rank histogram,

reliability and sharpness diagrams), we weigh the contribution of each fire case by the size of the evaluation domain before adding the contributions of the seven fires. The varying ensemble size causes the values of the rank to be normalized for the rank histogram.

4 RESULT

4.1 Emulation

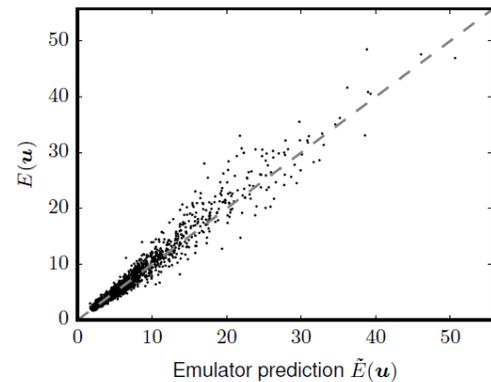


Figure 4: Error of emulator function on test data

The result of the predicted emulated test data is shown in Figure 4. The grey line indicates a perfect line fit. As shown in Figure 1, test sets are deviating from this grey dotted line which indicates some amount of error in the emulation result. The author used two error metrics to describe the test results which are MAE and Q2. Values for their approximation are found to be 0.73 and 95.3% respectively. As mentioned by the author use of logarithm favors a good approximation in comparison with non-logarithmic computation (MAE = 0.97 and Q2 = 93.2%). A decrease in error results MAE value being closer to 0 and Q2 tends to move towards 1. It took them like 0.6 seconds to create results for one energy function. This means to perform 150000 iterations of the MH algorithm, took them more than a day.

4.2 Ensemble Evaluation

The author presented the values for Brier Skill Score for all the seven cases of fire as shown in Table 1. The size of the ensemble is considered for all the ensembles except the one with reference ensembles that are of size 500. It can be seen that prior ensembles are showing better results than the reference ensemble. The scores are also improved when calibrated ensembles are compared with the prior ensemble but it is not true for all fire cases.

For the fire of Sant'Andrea di Cotone, Olmeta di Tuda and Ghisoni, there is an increase in BSS value with β . For the fire of Chiatria, there is a decrease in BSS value with β . For the fire of Calenzana and Ville di Paraso, influence of change in BSS is not so significant with β but for some optimum value of β , BSS is maximum. This raises the question that which β value is best to consider. For three cases of fire, the value of $\beta=2$ gives us the best result but for the three of the remaining four results value of $\beta=2$ gives us the lowest score. In general, the value of $\beta=2$ is considered best since the decrease

Table 1: Brier Skill Score of the reference prior and calibrated ensembles for different values of β in all seven fire cases. The best result for all the seven fire cases is shown in bold.

| 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 |
| Olmata 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 2: Ranking of Brier Skill Score

| 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 |
| Olmata 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 |

in BSS value is relatively low as compared to the increase in BSS value when β increases.

To find the best value of β , the authors have presented another table indicating the rank of the Brier Skill Score for all the ensembles as shown in Table 2. Rank score from 1-8 is given to all the results from the different ensembles and the overall ranking is calculated by summing the rank. According to the method, it is found that $\beta = 1/2$ gives the best distribution for the ensembles. 95% confidence interval is also estimated using 10000 bootstrap samples for each of the ensembles and reported in Table 3 Appendix B.

Figure 5 shows the reliability and sharpness diagram for all the ensembles used. It can be seen that in the case of calibrated ensembles lines tends to move near the dotted line which indicates a perfect reliable prediction for a probability of 0.4 or less. In general, prediction is more reliable for calibrated ensembles as compared to a reference or prior ensembles. Sharpness is also plotted for the reference, prior, $\beta = 1/4$ and $\beta = 2$ distributions.

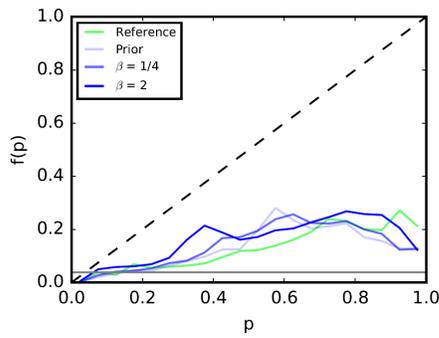
In Figure 6, global rank histogram for various ensembles distribution is shown. The dotted horizontal line indicates the ideal line for consistency. Higher bars on the left side of the histogram describe that there is some tendency to over-predict the probabilities. This over-prediction can be limited by calibration which results in an ideal flat histogram. Burn Probability Map for reference, prior and calibrated ($\beta = 2$) ensembles is shown in Figure 7a in Appendix A. As we move from reference to a prior ensemble, the non-zero probability region reaches further location due to high uncertainty

in the input parameters. This extension is limited in the calibrated ensemble because calibrated distribution favors lower ROS(Rate of Spread). This limitation results in better accuracy of the prediction.

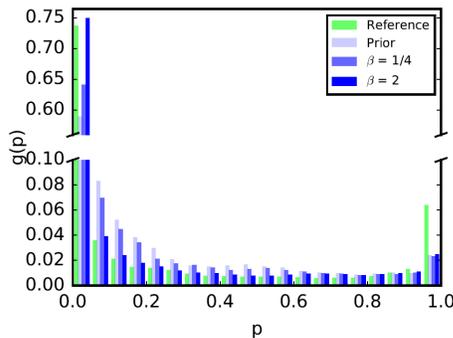
5 CONCLUSION

The study purposed a novel method to generate calibrated ensembles using input parameters whose distribution depends on the posteriori pseudo-likelihood function. A gaussian emulator is built to obtain various calibrated ensembles for all the seven fires which involve Wasserstein distance between the simulated and observed burned surface. Emulation of the energy function resulted in a good result ($Q_2 > 95\%$) so that the fire spread prediction has better overall accuracy. It is safe to note that increasing the value of β leads to a lower marginal distribution for the heat of combustion(ROS) and also lower uncertainty in wind direction. The best overall BSS score is found for $\beta=1/2$, although it is not good globally but rather best for most fire cases. In all the seven fire cases, prior ensembles tend to over-predict the burn probability. The calibrated ensembles limit this more significantly and ultimately lead to great overall accuracy.

The calibrated ensembles generally result in a lower heat value and limit the overprediction of burn probability. If the prior ensemble is underpredicted, increasing β will make underprediction more significant. In the simulation, firefighting actions are not modeled, so underprediction is less preferred than overprediction. This raises a question about the choice of β for the pseudo-likelihood function. It is found by ranking the BSS score for the seven fire cases, the



(a) Reliability plot



(b) Sharpness plot

Figure 5: Reliability and Sharpness plot for the results of various ensemble. (a) This shows a dotted black line indicating the perfect reliable prediction.

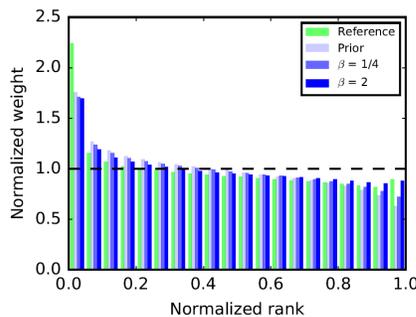


Figure 6: Rank Histogram for the results of various ensemble.

value of $\beta = 1/2$ leads to better overall accuracy. Another question that is to be considered is the choice of the fire cases. In this paper, all large fire for one season and region is considered. So, there is no guarantee that the results will be the same for all the other fire cases as well.

Overall, a novel method was purposed for calibrating the burn probabilistic predictions of wildland fire spread. Prediction accuracy is very crucial and needs to be improved since it adversely affects endangered species, human lives, infrastructures, and ecosystems. Further research on this topic can be considered to combine all

those calibrated ensembles with models for probability of ignition and value at stake to access the next day’s wildlife risk.

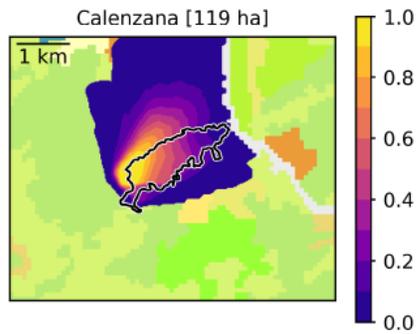
REFERENCES

- [1] Frédéric Allaire, Jean-Baptiste Filippi, and Vivien Mallet. 2020. Generation and evaluation of an ensemble of wildland fire simulations. *International Journal of Wildland Fire* 29, 2 (2020), 160–173. <https://doi.org/10.1071/WF19073>
- [2] Patricia Andrews, Miguel Cruz, and Richard Rothermel. 2013. Examination of the wind speed limit function in the Rothermel surface fire spread model. *International Journal of Wildland Fire* 22 (01 2013), 959. <https://doi.org/10.1071/WF12122>
- [3] Nicolas Bonneel, Michiel Panne, Sylvain Paris, and Wolfgang Heidrich. 2011. Displacement Interpolation Using Lagrangian Mass Transport. *ACM Trans. Graph.* 30 (12 2011), 158. <https://doi.org/10.1145/2024156.2024192>
- [4] Stephen P. Brooks and Andrew Gelman. 1998. General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics* 7, 4 (1998), 434–455. <http://www.jstor.org/stable/1390675>
- [5] B. Efron and R. Tibshirani. 1986. Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statist. Sci.* 1, 1 (1986), 54 – 75. <https://doi.org/10.1214/ss/1177013815>
- [6] J. Feranec, T. Soukup, G. Hazeu, and G. Jaffrain. 2016. *European Landscape Dynamics: CORINE Land Cover Data*. CRC Press. <https://books.google.de/books?id=EsHBDAAAQBAJ>
- [7] Jean-Baptiste Filippi, Jacques Balbi, and Frederic Morandini. 2010. Discrete Event Front-tracking Simulation of a Physical Fire-spread Model. *Simulation* 86 (10 2010). <https://doi.org/10.1177/0037549709343117>
- [8] Rémi Flamary, Nicolas Courty, Alexandre Gramfort, Mokhtar Z. Alaya, Aurélie Boisbunon, Stanislas Chambon, Laetitia Chapel, Adrien Corenflos, Kilian Fatras, Nemo Fournier, Léo Gautheron, Nathalie T.H. Gayraud, Hicham Janati, Alain Rakotomamonjy, Ievgen Redko, Antoine Rolet, Antony Schutz, Vivien Seguy, Danica J. Sutherland, Romain Tavenard, Alexander Tong, and Titouan Vayer. 2021. POT: Python Optimal Transport. *Journal of Machine Learning Research* 22, 78 (2021), 1–8. <http://jmlr.org/papers/v22/20-451.html>
- [9] Marc C. Kennedy and Anthony O’Hagan. 2001. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63 (03 2001), 425–464. <https://doi.org/10.1111/1467-9868.00294>
- [10] Jeong-Soo Park and Jangsun Baek. 2001. Efficient computation of maximum likelihood estimators in a spatial linear model with power exponential covariogram. *Computers Geosciences* 27 (02 2001), 1–7. [https://doi.org/10.1016/S0098-3004\(00\)00016-9](https://doi.org/10.1016/S0098-3004(00)00016-9)
- [11] R.C. Rothermel. 1972. *A Mathematical Model for Predicting Fire Spread in Wildland Fuels*. Intermountain Forest & Range Experiment Station, Forest Service, U.S. Department of Agriculture. <https://books.google.de/books?id=AfyMv5NBSjoC>
- [12] Olivier Roustant, David Ginsbourger, and Yves Deville. 2012. DiceKriging, DiceOptim: Two R Packages for the Analysis of Computer Experiments by Kriging-Based Metamodeling and Optimization. *Journal of Statistical Software* 51 (2012), 1–55.
- [13] Filippo Santambrogio. 2015. Optimal Transport for Applied Mathematicians. (2015). <https://doi.org/doi:10.1007/978-3-319-20828-2>
- [14] Andrew L. Sullivan. 2009. Wildland surface fire spread modelling, 1990–2007. 1: Physical and quasi-physical models. *International Journal of Wildland Fire* 18 (07 2009), 349–368. <https://doi.org/10.1071/WF06144>
- [15] Andrew L. Sullivan. 2009. Wildland surface fire spread modelling, 1990–2007. 3: Simulation and mathematical analogue models. *International Journal of Wildland Fire* 18 (07 2009), 387–403. <https://doi.org/10.1071/WF06144>
- [16] Xu Wu, Tomasz Kozłowski, Hadi Meidani, and Koroush Shirvan. 2018. Inverse Uncertainty Quantification using the Modular Bayesian Approach based on Gaussian Process, Part 1: Theory Nuclear Engineering and Design. 335 (01 2018), 339–355. <https://doi.org/10.1016/j.nucengdes.2018.06.004>

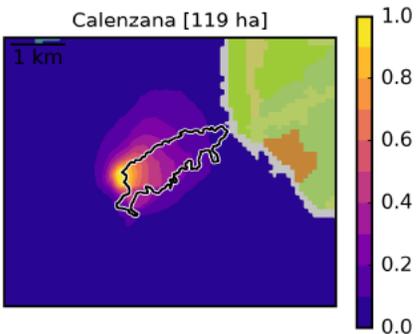
APPENDICES

A ADDITIONAL FIGURES

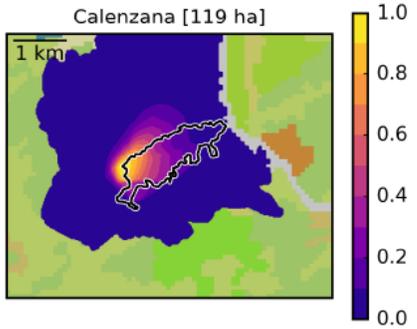
B ADDITIONAL TABLES



(a) Reference ensemble



(b) Prior ensemble



(c) Calibrated ensemble, $\beta = 2$

Figure 7: Burn probability map for the results of various ensembles. Colored scale on the right represent the burn probability. Black and white line in the center is the observed burn surface and the background coloured region is the Corine Land Cover.

Table 3: Bootstrap 95% confidence interval of the Brier Skill Score from Table 1

| Fire name | 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] |
| Olmata 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) | [-5.45,-5.18] | [-3.16,-3.01] | [-10.01,-9.64] | [-3.21,-3.05] | [-3.20,-3.04] | [-3.12,-2.96] | [-3.13,-2.97] | [-8.47,-8.19] |
| Ghisoni (2000) | [-10.11,-9.85] | [-10.47,-10.06] | [-10.01,-9.64] | [-10.03,-9.66] | [-9.51,-9.15] | [-9.18,-8.85] | [-8.78,-8.49] | [-8.47,-8.19] |
| Global | [-1.63, -1.58] | [-1.35, -1.30] | [-1.29, -1.24] | [-1.29, -1.24] | [-1.21, -1.16] | [-1.15, -1.11] | [-1.10, -1.06] | [-1.05, -1.01] |