Towards predictive simulation of wildfire spread at regional scale using ensemble-based data assimilation to correct the fire front position

MÉLANIE C. ROCHOUX[1,2,3], CHARLOTTE EMERY[1,2,4], SOPHIE RICCI[1,2], BÉNÉDICTE CUENOT[1], and ARNAUD TROUVÉ[4]
[1] CERFACS, 42 avenue G. Coriolis, 31057 Toulouse cedex 01, France
[2] SUC-URA1875, CNRS, 42 avenue G. Coriolis, 31057 Toulouse cedex 01, France
[3] Ecole Centrale Paris, CNRS-UPR288, Grande voie des vignes, 92295 Châtenay-Malabry, France
[4] Department Fire Protection Engineering, University of Maryland, College Park, MD 20742, USA

ABSTRACT

The objective of this study is to develop a prototype data-driven wildfire simulation capability capable of forecasting the fire spread dynamics. The prototype simulation capability features the following main components: a level-set-based fire propagation solver that adopts a regional scale viewpoint, treats wildfires as propagating fronts, and uses a description of the local rate of spread (ROS) of the fire as a function of vegetation properties and wind conditions based on Rothermel's model; a series of observations of the fire front position; and a data assimilation algorithm based on an Ensemble Kalman Filter (EnKF). Members of the EnKF ensemble are generated through variations in estimates of the fire ignition location and/or variations in the ROS model parameters; the data assimilation algorithm also features a state estimation approach in which the estimation targets (the control variables) are the two-dimensional coordinates of the discretized fire front. The prototype simulation capability is first evaluated in a series of verification tests using synthetically-generated observations; the tests include representative cases with spatially-varying vegetation properties and temporally-varying wind conditions. The prototype simulation capability is then evaluated in a validation test corresponding to a controlled grassland fire experiment. The results indicate that data-driven simulations are capable of correcting inaccurate predictions of the fire front position and of subsequently providing an optimized forecast of the wildfire behavior.

KEYWORDS: wildfire, fire modeling, front tracking, level set, data assimilation, Ensemble Kalman Filter

NOMENCLATURE LISTING

c	progress variable
$\mathbf{d}$	innovation vector
H	observation operator
$\mathbf{K}$	gain matrix
M_v	fuel moisture content (%)
m_v''	fuel loading (kg/m^2)
N_e	number of members in the EnKF ensemble
N_{fr}	number of points in discretized simulated fire front
N_{fr}^o	number of points in discretized observed fire front
$\mathbf{n}_{fr}$	fire front normal vector
p	probability density function (PDF)
$\mathbf{P}$	forecast or analysis error covariance matrix
r	(integer) ratio between N_{fr} and N_{fr}^o
$\mathbf{R}$	observation error covariance matrix
$\mathbf{u}_w$	wind velocity (m/s)
x,y	two-dimensional coordinates (m)
$\mathbf{x}$	control vector
$\mathbf{y}^o$	observation vector

Greek

Γ	rate of spread of the fire (ROS, in m/s)
δ_v	fuel depth (m)
Δh_c	heat of combustion (J/kg)
ε	error treated as a random variable
ρ_p	fuel particle mass density (kg/m^3)
σ	standard deviation of the error
Σ_v	fuel surface-to-volume ratio (1/m)

Superscript

a	analysis
f	forecast
o	observation
t	true

Subscript

e	ensemble
i	front marker index
ign	ignition
t	time
v	vegetation
w	wind

1. INTRODUCTION

Computer-based wildfire spread modeling has emerged during the past two decades as a powerful tool for applications in both fire risk management and fire emergency response. However, because the underlying dynamics feature complex multi-physics occurring at multiple scales, our ability to accurately simulate the behavior of wildfires remains limited [1]. The dynamics of wildfires are determined by interactions between pyrolysis, combustion, flow dynamics as well as atmospheric dynamics. These interactions occur at: vegetation scales that characterize the biomass fuel; topographical scales that characterize the terrain and vegetation boundary layer; and meteorological micro/meso-scales that characterize atmospheric conditions. The mathematical models proposed to simulate wildfire spread are limited because of their inability to cover the entire range of relevant scales, because also of knowledge gaps and/or inaccuracies in the description of the physics as well as knowledge gaps and/or inaccuracies in the description of the controlling input parameters (i.e., the vegetation, topographical and meteorological properties).

Relevant insight into wildfire dynamics has been obtained in recent years via detailed numerical simulations performed at flame scales (i.e., with a spatial resolution on the order of 1 m). For instance, FIRETEC [2] or WFDS [3] combine advanced physical modeling and classical methods of Computational Fluid Dynamics (CFD) to accurately describe the combustion-related processes that control the fire spread. Note that because of their high computational cost, flame-scale CFD is currently restricted to research projects and is not compatible with operational applications. In contrast, we adopt in the following a regional-scale viewpoint (i.e., a viewpoint that considers scales ranging from a few tens of meters up to several kilometers), in which the fire is described as a two-dimensional flame front that self-propagates normal to itself into unburnt vegetation; the local propagation speed is called the rate of spread (ROS). This viewpoint is the dominant approach used in current operational wildfire spread models, see for instance BehavePlus [4] and FARSITE [5] in the United States, the MacArthur Fire Danger Meters [6] in Australia and the Forest Fire Behavior Prediction System (FBPS) [7] in Canada. BehavePlus and FARSITE use a model due to Rothermel [8] that treats the ROS as a function of vegetation (fuel) properties associated with a pre-defined fuel category (i.e., the vertical thickness of the fuel layer, the fuel moisture content, the fuel particle surface-to-volume ratio, the fuel loading and the fuel particle mass density), topographical properties (i.e., the terrain slope) and meteorological properties (i.e., the wind velocity at mid-flame height). A first limitation in this approach is that the ROS model does not explicitly account for the interaction of the fire and the atmosphere and has a domain of validity that is limited to the conditions of the calibration experiments used during its original development. A second limitation is that the input parameters that determine the ROS are often unknown or are only known with limited accuracy.

The uncertainties inherent in wildfire spread modeling suggest the use of ensemble forecasts: ensemble forecasts stochastically characterize the non-linear response of models to variations in the input data. For instance, Finney *et al*. [9] describes an ensemble-based forecasting capability based on possible weather scenarios and moisture content evolution. Furthermore, recent progress made in airborne remote sensing provides new ways to monitor real-time fire front positions [10,11]. Ensemble-based data assimilation (DA) techniques that integrate these fire sensor observations into a computer modeling tool provide an attractive framework to correct and optimize the model outputs and to thereby produce improved forecasting capabilities. In this framework, the DA algorithm is sequentially applied; each sequence (also called the analysis cycle) is decomposed into two steps: 1) a prediction step in which the control variables (i.e., model state and/or parameters) are advanced in time given some uncertainty ranges in the model input data; 2) an update step based on Bayes' theorem in which new observations are considered and the probability density function (PDF) of the control variables is modified consistently with the observations in order to reduce the uncertainties in the model outputs [12–15]. The Kalman filter (KF) is the most commonly used sequential DA technique. However, KF assumes linear dynamics and a Gaussian statistical distribution for both modeling and observation errors. Extensions of KF that overcome in part these limitations have been proposed, for instance the Extended Kalman Filter (EKF) [12] that uses local linearization techniques and the Ensemble Kalman filter (EnKF) [16,17] that uses a stochastic description of the model behavior.

The present study is an extension of our previous work in Refs. [18, 19] in which a prototype data-driven wildfire simulation capability was developed. The initial prototype featured the following main components: a level-set-based fire propagation solver combined with a model description of the local ROS proposed by Rothermel [8]; a series of observations of the fire front position; and an EKF-based data assimilation algo-

rithm. The DA prototype was successfully evaluated when applied for estimation of the parameters used in the Rothermel-based ROS model (e.g., the vertical thickness of the fuel layer, the fuel moisture content, the fuel particle surface-to-volume ratio, and/or the wind direction and magnitude); the evaluation was performed in the context of a controlled grassland fire experiment. While the studies presented in Refs. [18,19] produced encouraging results and confirmed the value of a DA strategy for improved wildfire spread predictions, some of the design choices made during the initial development of the prototype simulator were proposed on a preliminary and temporary basis with the understanding that they would have to be re-visited in subsequent work. The most questionable choices were the selection of the EKF algorithm and that of the stand-alone parameter estimation approach. The choice of the EKF algorithm is considered questionable because it assumes a linear relationship between the control variable space and the observation space (i.e. a linear relationship between changes in the parameters of the Rothermel-based ROS model and the resulting changes in the fire front position); this linear assumption is believed to be of limited value in general wildfire problems. The modification to an EnKF approach was explored in Ref. [20]. In this work, some of the non-linearities in the relationship between control variable space and observation space were taken into account as these are stochastically characterized over an ensemble of simulations. The choice of a parameter estimation approach is also considered questionable: while well-suited for statistically spatially-homogeneous problems (in which corrections to the parameters of the ROS model can be applied uniformly), this choice is no longer adapted to more general wildfire problems in which the vegetation, topographical and possibly meteorological properties exhibit arbitrary spatial variations. The extension to the estimation of spatialized vegetation/wind ROS parameters would be indeed computationally prohibitive in the context of real-time forecast of the fire behavior. The objective of the present study is to remove some of the main limitations in the initial design of our prototype data-driven wildfire simulator.

This study is a continuation of our previous studies presented in Refs. [18–20] and an extension to the case of spatially-varying vegetation properties and temporally-varying wind conditions. This extension is based on a change from a parameter estimation approach to a state estimation approach. This change was inspired in part by previous studies by Mandel *et al*. [21–23], in which the control variable is the temperature field and is characterized by a bimodal PDF in the fire region (burning state or not-burning state). In order to satisfy the Gaussian assumption in EnKF, the idea of morphing from image processing was introduced [21]; however, this choice led to technical difficulties in the EnKF implementation. In the present study, the control variable is the fire front position and is characterized by an approximate Gaussian PDF, which allows for a straightforward application of EnKF. The paper is organized as follows. The fire front observations and the wildfire spread model (called FIREFLY) are presented in Section 2. The EnKF algorithm is presented in Section 3. The performance of the resulting data-driven wildfire simulation capability is evaluated in Section 4 using first, academic tests in which observations are synthetically-generated and second, a validation test in which observations are taken from a controlled grassland fire experiment.

2. FIRE FRONT OBSERVATIONS AND SIMULATIONS

Observations of the fire front position

We assume in the present study that observations of the fire front position are available and that these observations can be made at different relevant times. Note that there is a growing body of literature on recent technological developments for geo-referenced wildfire front tracking, see for instance [10,11]. So far, spaceborne and airborne systems observe fires in the Middle InfraRed (MIR) region but only airborne platforms provide spatial and temporal resolution suitable for real-time geo-location of active fire contours. For instance, Paugam *et al*. [10] show that spatio-temporal variations of the rate of spread of the fire can be accurately retrieved using a Fire Radiative Power (FRP) analysis on a thousand-meter-square controlled fire experiment.

In the following, the observed fire front is represented as a segmented line using a pre-defined number of equally-spaced markers (the observation points); the observation vector, noted $\mathbf{y}_t^o$, contains the two-dimensional coordinates (x_i^o, y_i^o) of the fire front markers (the subscript i is the index of a particular marker in the observation vector) observed at time t. The fire front coordinates are assumed to have independent Gaussian-like random errors ε^o with zero mean and with standard deviation noted σ^o. Two types of tests are presented in Section 4: Observation System Simulation Experiments (OSSE) in which observations are synthetically

generated using a reference solution of the fire spread model (called the true evolution) modified by random observation errors ε^o; and a controlled grassland fire experiment in which the observations are reconstructed from measured temperature maps and using a definition of the fire front as the 600 K iso-temperature contour.

The fire spread model (the forward model)

The front-tracking solver, called FIREFLY, simulates the propagation of surface wildland fires at regional scales as illustrated in Fig. 1. The solver is currently limited to flat terrains and problems with complex topography are outside the scope of the present study. FIREFLY tracks the time evolution of the fire front location using the following three components:

- *A sub-model for the rate of spread*. The expression for the local ROS, noted Γ, proposed by Rothermel [8] may be written as:

$$\Gamma = \Gamma(x,y,t) = \Gamma\Big(\delta_v(x,y), M_v, M_{v,ext}, \Sigma_v, m_v'', \rho_p, \Delta h_c, \mathbf{u}_w(t)\Big) \tag{1}$$

  where input parameters are summarized in Table 1. In the following, the wind velocity $\mathbf{u}_w$ is treated as a time-dependent, spatially-uniform quantity; the fuel depth δ_v is treated as a time-independent, spatially-varying quantity; all other parameters are treated as constant and uniform.

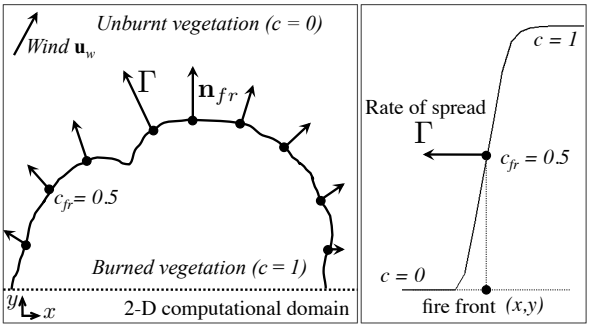


Fig. 1. Level-set-based fire spread simulator. Left: the fire front is the isocontour $c_{fr} = 0.5$; Γ measures the local rate-of-spread of the fire along the normal direction $\mathbf{n}_{fr}$. Right: profile of the spatial variations of the progress variable c across the fire front.

Table 1. Input parameters in the Rothermel-based ROS model.

Name	Symbol	Unit
Fuel depth (vertical thickness of the vegetation layer)	δ_v	m
Fuel moisture (mass of water divided by mass of dry vegetation)	M_v	-
Fuel moisture at extinction	$M_{v,ext}$	-
Fuel particle surface-to-volume ratio	Σ_v	1/m
Fuel loading	m_v''	kg/m^2
Fuel particle mass density	ρ_p	kg/m^3
Fuel heat of combustion	Δh_c	J/kg
Wind velocity vector at mid-flame height (projected into horizontal plane)	$\mathbf{u}_w$	m/s

- *A level-set-based solver for the fire front propagation equation*. A progress variable noted $c = c(x,y,t)$ is introduced as a flame marker: $c = 0$ in the unburnt vegetation, $c = 1$ in the burnt vegetation, and the flame is the region where c takes values between 0 and 1 (the flame front is identified as the isocontour

$c = 0.5$). The progress variable is calculated as a solution of the following propagation equation:

$$\frac{\partial c}{\partial t} = \Gamma \, |\nabla c| \,, \tag{2}$$

where Γ is modeled using Eq. (1). Equation (2) is solved using a second-order Runge-Kutta scheme for time-integration and an advection algorithm for spatial discretization based on a second-order total variation diminishing (TVD) scheme combined with a Superbee slope limiter [19,24].

- *An isocontour algorithm for the reconstruction of the fire front*. Once the spatio-temporal variations of the progress reaction c are known, the position of the fire front is extracted using a simple isocontour algorithm such that, formally, the outputs of the FIREFLY model are $[(x_i, y_i), 1 \leq i \leq N_{fr}] = M_{[t-1,t]}(c_{t-1}, \lambda)$, where (x_i, y_i) represents the two-dimensional coordinates of the N_{fr} fire front markers obtained at time t, where c_{t-1} designates the initial condition (i.e., the spatial distribution of the progress variable c at time $(t-1)$) and where λ designates the list of input parameters of the ROS model, $\lambda = (\delta_v, M_v, M_{v,ext}, \Sigma_v, m_v'', \rho_p, \Delta h_c, \mathbf{u}_w)$ (see Table 1).

3. THE ENSEMBLE-BASED DATA ASSIMILATION ALGORITHM

Comparison between simulations and observations

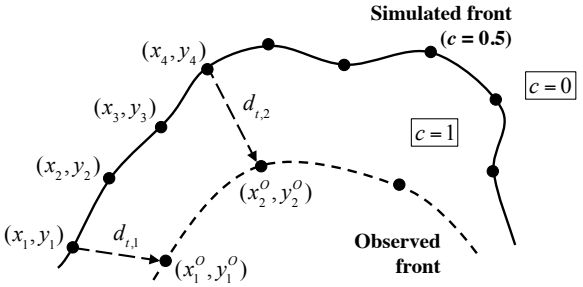


Fig. 2. Construction of the innovation vector $\mathbf{d}_t$ introduced to quantify the differences between simulated and observed fire fronts. In this illustration, $r = 4$.

The data assimilation algorithm uses a discretization of both the simulated and observed fire fronts, called SFF and OFF, respectively. The discretization of SFF is a set of N_{fr} markers; the control vector $\mathbf{x}_t$, also called the state vector, contains the two-dimensional coordinates of the markers at time t, $\mathbf{x}_t = [(x_1, y_1), \ldots, (x_{N_{fr}}, y_{N_{fr}})]$. Similarly, the discretization of OFF is a set of N_{fr}^o markers; the observation vector $\mathbf{y}_t^o$ is defined as $\mathbf{y}_t^o = [(x_1^o, y_1^o), \ldots, (x_{N_{fr}}^o, y_{N_{fr}}^o)]$. Note that the FIREFLY solver uses a high-resolution computational grid that allows for a detailed representation of the local conditions (the spatial resolution is on the order of 1 m). In contrast, observations of the fire front position are likely to be provided with a much coarser resolution; in addition, observations may be incomplete and cover only a fraction of the fire front perimeter. Thus we may expect N_{fr}^o to be much lower than N_{fr}; in the following, we assume for simplicity that $N_{fr}^o = (N_{fr}/r)$ where r is an integer taking values (much) larger than 1. In order to map the state variable space (SFF) onto the observation space (OFF), an observation operator H is introduced that selects a subset of N_{fr}^o markers among the fine-grained discretization of SFF and pairs each one of those markers with one of those used in the coarse-grained discretization of OFF (see Fig. 2). The observation operator H may be defined in several ways (for instance using a projection scheme) but preliminary tests have shown that a simple treatment (taking 1 out of every r points) provided reasonable results; this issue will be re-visited in future work. The innovation vector, noted $\mathbf{d}_t$, is now simply defined as the vector formed by the directed distances between the paired SFF-OFF markers:

$$\mathbf{d}_t = \mathbf{y}_t^o - H(\mathbf{x}_t). \tag{3}$$

The statistical moments of $\mathbf{d}_t$ (e.g., mean and standard deviation) provide a convenient measure of the deviations of model predictions from observations.

The Ensemble Kalman Filter algorithm

The objective of data assimilation is to estimate the state vector $\mathbf{x}_t$ using the observations $\mathbf{y}_t^o$ made at time t (see Fig. 3). In the prediction step, the PDF of the state vector is evolved from time $(t-1)$ until time t. We note $p_f(\mathbf{x}_t)$ this PDF (also called the forecast PDF) at time t. At the analysis time t, also called the update step, this forecast PDF is corrected in order to be more consistent with the observations $\mathbf{y}_t^o$. The new PDF, called the analysis and noted $p_a(\mathbf{x}_t)$, is given by Bayes' theorem:

$$p_a(\mathbf{x}_t) \propto p(\mathbf{y}_t^o|\mathbf{x}_t)\, p_f(\mathbf{x}_t), \tag{4}$$

where the symbol $\propto$ means "proportional to" and where $p(\mathbf{y}_t^o|\mathbf{x}_t)$ represents the data likelihood, i.e. the conditional PDF of having the observations $\mathbf{y}_t^o$ given the state $\mathbf{x}_t$.

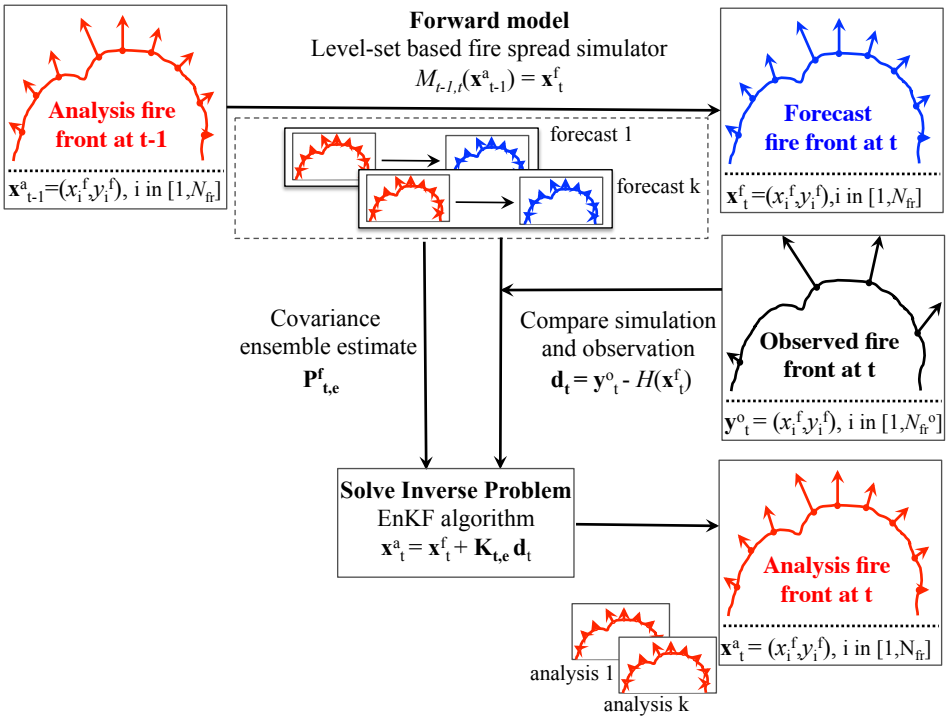


Fig. 3. Flow chart of the EnKF algorithm during the $[t-1, t]$ analysis cycle.

The EnKF algorithm assumes that both the model state $\mathbf{x}_t$ and the observations $\mathbf{y}_t^o$ are random variables defined by Gaussian PDF with a zero mean value and an error covariance model. Under these assumptions, the forecast PDF may be written as:

$$p_f(\mathbf{x}_t) \propto \exp\left\{-\frac{1}{2}\left(\mathbf{x}_t - \mathbf{x}_t^f\right)^T \left(\mathbf{P}^f\right)^{-1}\left(\mathbf{x}_t - \mathbf{x}_t^f\right)\right\}, \tag{5}$$

where $\mathbf{x}_t^f$ is the forecast estimate of the true state vector and where $\mathbf{P}^f$ is the forecast error covariance matrix representing modeling errors. The data likelihood may be similarly expressed as:

$$p(\mathbf{y}_t^o|\mathbf{x}_t) \propto \exp\left\{-\frac{1}{2}\left(\mathbf{y}_t^o - H(\mathbf{x}_t)\right)^T \mathbf{R}^{-1}\left(\mathbf{y}_t^o - H(\mathbf{x}_t)\right)\right\}, \tag{6}$$

with $\mathbf{R}$ the observation error covariance matrix representing observation errors. Within this framework, the analysis PDF is also Gaussian and is written as:

$$p_a(\mathbf{x}_t) \propto \exp\left\{-\frac{1}{2}\left(\mathbf{x}_t - \mathbf{x}_t^a\right)^T (\mathbf{P}^a)^{-1}\left(\mathbf{x}_t - \mathbf{x}_t^a\right)\right\}, \tag{7}$$

where $\mathbf{x}_t^a$ is the analysis estimate of the true state vector and where $\mathbf{P}^a$ is the analysis error covariance matrix.

The classical Kalman filter algorithm assumes that the observation operator H is linear (denoted by $\mathbf{H}$); in that case, it may be shown that the analysis update in Eq. (7) leads to the following equations:

$$\mathbf{x}_t^a = \mathbf{x}_t^f + \mathbf{K}(\mathbf{y}_t^o - H(\mathbf{x}_t^f)), \quad \mathbf{K} = \mathbf{P}^f\mathbf{H}^T(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R})^{-1}, \quad \mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H}), \tag{8}$$

where $\mathbf{K}$ is called the gain matrix. The expressions in Eq. (8) are the basis of the EKF algorithm used in Refs. [18,19].

In contrast, the Ensemble Kalman Filter (EnKF) algorithm approximates the forecast PDF of the state vector $p_f(\mathbf{x}_t)$ by performing a series (an ensemble) of N_e independent forward model integrations up to the analysis time t, thereby providing N_e forecast estimates of the state vector, called the ensemble members $\mathbf{x}_{t,e}^f = [\mathbf{x}_t^{f,1}, \mathbf{x}_t^{f,2}, ..., \mathbf{x}_t^{f,N_e}]$. The EnKF algorithm approximates the mean and the covariance of the forecast by the mean and the covariance of the ensemble, while still making the assumption that all PDF are Gaussian. During the analysis, each ensemble member is updated using the classical Kalman filter formulation in Eq. (8), with the difference that the gain matrix $\mathbf{K}$ is now calculated from an estimate of the forecast error covariance matrix, noted $\mathbf{P}_e^f$, and using the ensemble-based stochastic representation of the relationship between state space and observation space. Note that in the present study, we use the EnKF version proposed by Burgers *et al*. [25], and assume that the observation errors are uncorrelated, i.e. the observation error covariance matrix $\mathbf{R}$ is treated as a diagonal matrix in which each diagonal term is the error variance $(\sigma^o)^2$ associated with the x- or y-coordinate of the OFF markers.

4. EVALUATION OF THE DATA-DRIVEN SIMULATION CAPABILITY

We first present results from OSSE experiments in which synthetic observations are generated from a reference FIREFLY solution using particular values of the ROS model input parameters and a particular ignition location (the true state is thus known and can be represented by the model). The OSSE tests are representative of field-scale fires where the rate of spread takes values on the order of 10 cm/s. Because in real-world applications, the measurements will be sparse and imperfect, we study the sensitivity of the data-driven solution to the standard deviation of the observation error, σ^o, to the number of observation points along the fire front, N_{fr}^o, and to the level of completeness of the observations. We then present results from a validation study in which observations are taken from an experimental database corresponding to a controlled grassland fire (a case for which it may not be possible to represent the true state with the FIREFLY model). This validation study corresponds to a real reduced-scale fire in which the rate of spread takes values on the order of 1 cm/s.

Observation System Simulation Experiments (OSSE)

In the first series of OSSE tests, the numerical configuration corresponds to a 200 m $\times$ 200 m domain with uniform vegetation properties and no wind; the ROS is constant and uniform and is taken equal to 0.2 m/s. The computational grid resolution is $\Delta x = \Delta y = 1$ m and the temporal resolution is $\Delta t = 0.5$ s. The true fire front is initialized as a circular front centered at (x_{ign},y_{ign}) = (100 m, 100 m) and with a radius of 5 m. The FIREFLY model is first integrated in time in order to produce at the analysis time (chosen to be $t = 200$ s) the true fire front position. An ensemble of $N_e = 25$ forecasts is then produced based on spatial variations of the ignition location (x_{ign},y_{ign}) around a mean value (97 m, 103 m) and with a standard deviation of 10 m, see Fig. 4(a). In this first configuration, uncertainties in the forecast ensemble are only due to the initial location of the fire. The simulated fire fronts are discretized using $N_{fr} = 100$ points.

Figure 4(a) presents a comparison between the true and the forecasted fire front positions at time $t = 200$ s and shows that due to uncertainties in the ignition location of the fire, the predicted front positions are scattered over a large area. Since in this test, uncertainties in the distribution of the vegetation properties are not taken into account in the generation of the ensemble, the propagation of the front is isotropic (simulated fire fronts

remain circular). Thus, within the ensemble, the errors in the position of the N_{fr} simulated fire front markers are highly correlated. As a result, the DA algorithm translates the information observed at one point into a uniform correction along the fire front. Figure 4(b) presents the comparison between the true and the analyzed fire front positions, i.e., the updated front positions that are produced by the EnKF algorithm at the end of the analysis cycle (at time $t = 200$ s), when only one observation is available (black cross). As expected, the analyzed front positions feature a much reduced scatter, they are located close to the true front position and the EnKF correction is isotropic. It should be noted that when several observations are used, the analysis produces a circular-shaped fire fronts located at an optimal distance of the observations.

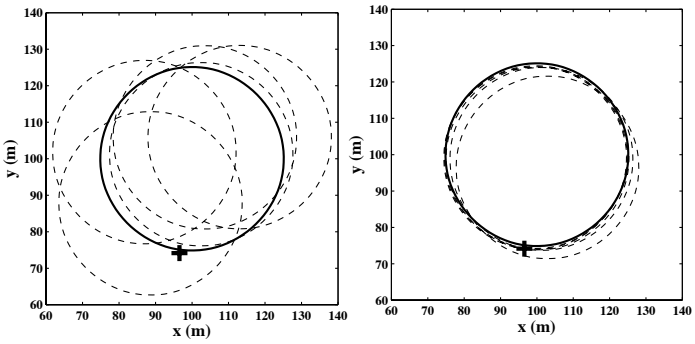
The results in Fig. 4(b) were produced with a low value of the observation error standard deviation, $\sigma^o = 1$ m, and Fig. 4(c) examines the influence of this error on the EnKF performance. Figure 4(c) presents the root mean square (RMS) distance between the true and the forecasted fire front positions and between the true and the analyzed fire front positions as a function of the magnitude of the observation errors measured by σ^o; the vertical bars in Fig. 4(c) give a graphical representation of the magnitude of the standard deviations in the forecast and analysis ensembles. The figure shows that when the observation error is small, the EnKF algorithm successfully drives the analysis ensemble towards the true state; in contrast, when the observation error is large, the EnKF algorithm has reduced effects and the analysis ensemble remains close to the forecast ensemble; for intermediate values of the observation error, the EnKF algorithm produces optimized predictions lying between forecast and observation. These different regimes illustrate how data assimilation combines information from both models and observations, and produces better results that those that would be obtained if models or observations were considered separately. In the following tests, the observation errors are assumed to be small and the performance of EnKF will be evaluated by its ability to track the observed fire front location.

Next we present results from a second series of OSSE tests, the numerical configuration corresponds to a 700 m $\times$ 700 m domain with spatially-varying vegetation properties and with wind; the ROS is estimated using Rothermel's model; the fuel depth is assumed to be spatially-varying taking different values in the 4 quadrants of the square-shaped computational domain. The FIREFLY model is first integrated in time (starting from a small circular front) in order to produce at the analysis time (chosen to be $t = 400$ s) the true fire front position. An ensemble of $N_e = 20$ forecasts is then produced based on assumed uncertainties in the ROS input model parameters, specifically the fuel depth, moisture content and particle surface-to-volume-ratio, (δ_v, M_v, Σ_v), the wind velocity vector $\mathbf{u}_w$ (i.e., magnitude and direction), as well as uncertainties in the ignition location (x_{ign},y_{ign}). Thus, in this second configuration, uncertainties in the forecast ensemble are now due to spatial variations in as many as 10 input parameters (4 values of δ_v for the 4 quadrants, plus values of M_v, Σ_v, $\mathbf{u}_w$ and (x_{ign},y_{ign})); corresponding values for the means and standard deviations are presented in Table 2 (the wind parameters are presented in terms of a magnitude and a direction angle). The simulated fire fronts are discretized using $N_{fr} = 100$ points. The observation error standard deviation is small, $\sigma^o = 1$ m.

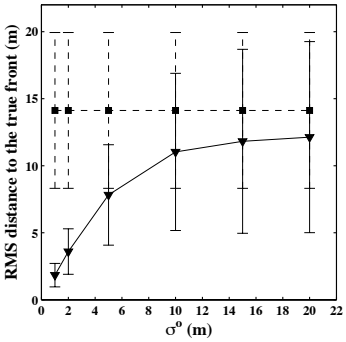
Table 2. Properties of the ensemble forecast in the spatially-varying OSSE test.

Input parameter	True value	Mean ensemble value	Standard deviation
$\delta_{v,1}$ (m)	0.25	0.25	0.10
$\delta_{v,2}$ (m)	1.25	1.25	0.10
$\delta_{v,3}$ (m)	0.75	0.75	0.10
$\delta_{v,4}$ (m)	1.75	1.75	0.10
M_v (%)	20	20	10
Σ_v (1/m)	11500	11500	4000
$\mathbf{u}_w$ (m/s, deg)	(1.0, 315)	(0.75, 315)	(0.15, 45)
x_{ign} (m)	350	350	20
y_{ign} (m)	350	350	20

Figure 5(a) presents a comparison between the true and the forecasted fire front positions at time $t = 400$ s and shows that due to uncertainties in the fire spread model parameters, the predicted fire fronts deviate significantly from the true state. In contrast to previous results presented in Fig. 4(a), the propagation is now



(a) Forecast. (b) Analysis with $\sigma^o = 1$ m.



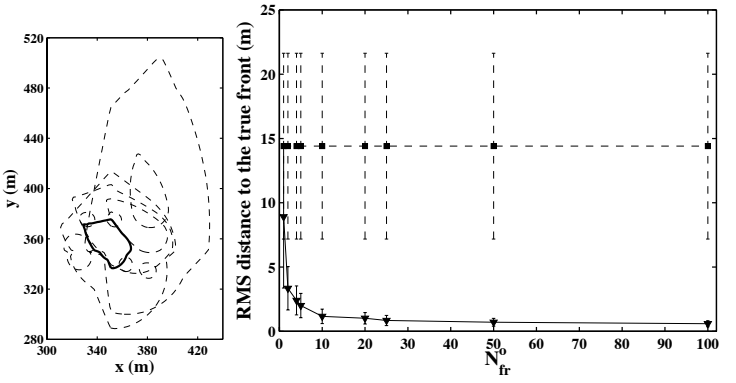
(c) EnKF performance versus σ^o.

Fig. 4. Spatially-uniform OSSE test with constant ROS but uncertain ignition location; single analysis cycle; all figures correspond to time $t = 200$ s. (a) Comparison between true (solid line) and forecasted (dashed lines) fire front positions; the cross symbol is the observation. (b) Similar comparison between true (solid line) and analyzed (dashed lines) fire front positions. (c) Averaged distance between the true and forecasted fire front positions (squares) and between the true and analyzed fire front positions (triangles) as a function of the observation error standard deviation σ^o.

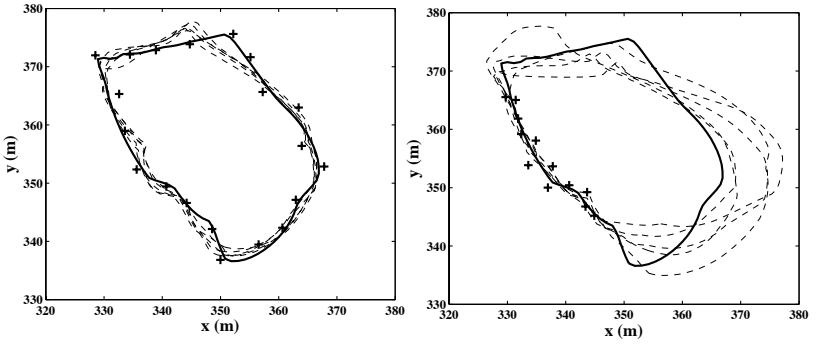
anisotropic and the present fire fronts are characterized by stronger irregularities and more complex shapes; these features are the consequence of both the presence of wind and the spatial variations in fuel depth. Figure 5(c) presents a similar comparison between the true and analyzed fire front positions. As expected, the analyzed front positions feature a much reduced scatter in terms of fire front shapes and are located close to the true front position. Since the anisotropy of the propagation is now represented by a wide range of uncertainties in the ensemble, the error in the location of one observed point is only correlated with the error in the location of the other adjacent points along the fire front. The DA algorithm translates the information coming from one observation point into a correction restricted to the closets neighbors only. As a consequence, when several observations are available, a non-uniform correction is obtained and the DA algorithm is then able to change the shape of the fire front and more easily match the observations.

Figure 5(b) examines the influence on the EnKF performance of the number of uniformly-distributed observation points along the front. This figure presents the RMS distance between the true and the forecasted fire front positions and between the true and the analyzed fire front positions as a function of N_{fr}^o. The figure shows that when N_{fr}^o is large (see Fig. 5(c) where $N_{fr}^o = 20$), the EnKF algorithm successfully drives the analysis ensemble towards the true state; in contrast, when N_{fr}^o is small, the EnKF algorithm has reduced effects and the analysis remains close to the forecast. In other words, the performance of the DA algorithm and its ability to capture the high-resolution features of the fire front depend strongly on the density of the observation network.

While Figs. 5(b)-(c) show that the direct observation of the fire front position can overcome various uncertainties in the wildfire spread model parameters, Fig. 5(d) illustrates that the spatial distribution of the observations along the fire contour has a significant impact on the analysis. This figure considers a practically relevant situation in which the observations are limited to a certain section of the fire front (the informed section) and therefore provide an incomplete picture. In this situation, while the EnKF algorithm produces an analysis that is close to the true state in the informed section, the benefits of data assimilation are much reduced in the non-informed sections. It is worth pointing out, however, that despite a reduced level of performance, EnKF informed by incomplete observations remains capable of improved performance compared to a stand-alone forecast.



(a) Forecast. (b) EnKF performance versus N^o_{fr}.



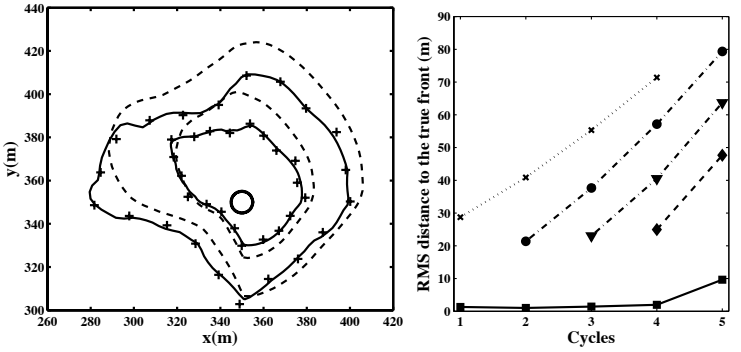
(c) Analysis with $N^o_{fr} = 20$. (d) Analysis with $N^o_{fr} = 12$ for an incomplete set of observations.

Fig. 5. Spatially-varying OSSE test with uncertain ROS model parameters and uncertain ignition location; single analysis cycle; all figures correspond to time $t = 400$ s. (a) Comparison between true (solid line) and forecasted (dashed lines) fire front positions. (b) Averaged distance between the true and forecasted fire front positions (squares) and between the true and analyzed fire front positions (triangles) as a function of the number of observation points along the fire front N^o_{fr}. (c) Comparison between true (solid line) and analyzed (dashed lines) fire front positions; the cross symbols are the observations. (d) See caption of Fig.(c); case of an incomplete set of observations.

We now consider multiple analysis cycles for the spatially-varying OSSE test and examine the behavior of the forecast between successive observations or after the last observation is made. The true fire front spread is simulated for time-varying wind conditions, whereas the forecast ensemble is simulated for constant wind conditions, using the parameters presented in Table 2. The performance of the forecast is expected to deteriorate in time for two reasons. First, because the impact of the fire front correction at a previous time decreases as the forecast lead time increases. Second, because the present implementation of the EnKF does not provide any correction for ROS modeling errors or any correction for uncertainties in the ROS model

input parameters (including the incorrect assumption of a constant wind). Such corrections may be introduced through a parameter estimation approach, see Refs. [18–20]. In this test, we apply the EnKF algorithm over four successive analysis cycles: the EnKF update is performed at times $t_1 = 150$ s, $t_2 = 300$ s, $t_3 = 450$ s and $t_4 = 600$ s.

Figure 6(a) presents a comparison between the averaged forecasted fire front positions, the observations (considered to be close to the true state) and the averaged analyzed fire front positions at time $t_2 = 300$ s and $t_4 = 600$ s. Similar to the results obtained in previous OSSE tests, the RMS distance to the true front is significantly reduced by EnKF, from 20 m for the forecast to less than 1 m for the analysis (see also the curve with square symbols in Fig. 6(b)). Figure 6(b) shows the typical cyclic evolution of the deviations of model predictions from observations: during the update step of the analysis cycle n, the analysis provides a correction to the front position and the distance between the true state and the forecast is greatly reduced; the analysis ensemble at the end of cycle n provides the initial conditions for the next cycle $(n+1)$; during the prediction step of cycle $(n+1)$, the wildfire spread model simulates the fire evolution (i.e., produces a forecast) but the distance between the true state and the forecast increases significantly; during the update step of the analysis cycle $(n+1)$, the distance between the true state and the forecast is again reduced and the cycle may be repeated. For instance, in Fig. 6(b), the curve with circles correspond to an averaged forecast that has been updated at time $t_1 = 150$ s, thereby leading to a close approximation of the true state; without additional observations, this forecast is seen to deviate from the true state (the RMS distance between the true state and the forecast is approximately 80 m at time $t_5 = 750$ s). In summary, these results show that in a state estimation approach, the EnKF updates have to be performed at regular time intervals in order to allow for an accurate tracking of the true fire front position.



(a) Forecast and analysis. (b) EnKF performance versus time.

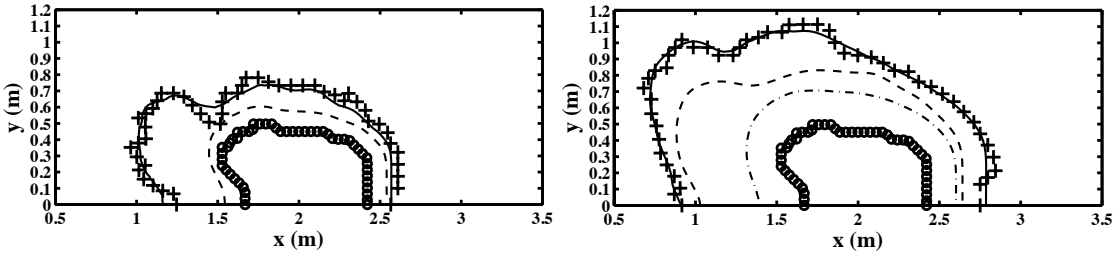
Fig. 6. Spatially-varying OSSE test with uncertain ROS model parameters and uncertain ignition location; multiple analysis cycles. (a) Comparison between mean forecast (dashed lines), observations (crosses) and mean analysis (solid lines) at times $t_2 = 300$ *s* and $t_4 = 600$ *s*; the small circle in the center of the figure corresponds to the initial conditions. (b) Averaged distance between the true and forecasted/analyzed fire front positions as a function of the analysis cycle number; crosses correspond to a forecast with no update; circles, triangles, diamonds correspond to a forecast with an update at $t_1 = 150$ s, $t_2 = 300$ s, $t_3 = 450$ s, respectively; square symbols correspond to an analysis performed at times t_1, t_2, t_3 and t_4.

Validation study: Application to a controlled grassland fire experiment

We now evaluate EnKF-FIREFLY in a validation study corresponding to a controlled grassland fire experiment. The experimental configuration corresponds to a small (4 m x 4 m), flat and and horizontal, open-field grassland lot burning under moderate wind conditions. The properties of the grass are (approximately) known: $\delta_v = 8$ cm (field measurement), $M_v = 22$ % (field measurement) and $\Sigma_v = 11500$ m^{-1} (Rothermel's fuel model for short grass); the wind conditions are also approximately known: the magnitude and direction angle of the wind are constant and equal to 1 m/s and 307 deg. The fire spread is recorded during 350 s using a thermal-infrared camera; the thermal maps are post-processed (the fire front is defined at the 600 K iso-temperature

contour) and thereby provide full fire contours at 14 s intervals; based on the spatial resolution of the camera, the estimated standard deviation of the measurement error is $\sigma^o = 0.05$ m. An ensemble of $N_e = 50$ forecasts is produced based on assumed uncertainties in the ROS input model parameters, (δ_v,M_v,Σ_v,$\mathbf{u}_w$) as well as uncertainties in the initial conditions taken at time $t_0 = 50$ *s*. When generating the forecast ensemble, the mean values of the ROS input model parameters are the known properties of the grass and and known conditions of the wind; the uncertainties are characterized by relatively large levels, $\sigma(M_v) = 6$ %, $\sigma(\Sigma_v) = 4000$ m^{-1} and $\sigma(\mathbf{u}_w) = (0.4$ m/s$, 45$ deg$)$. In addition the grassland lot is divided into 4 equally-sized rectangular-shaped sections and the fuel depth is treated as different in each zone, δ_v = 6, 10, 8 and 12 cm, and $\sigma(\delta_v) = 4$ cm; following the conclusions from the OSSE experiements this treatment is adopted in order to generate a rich forecast ensemble with a wide range of simulated fire front shapes and locations. The uncertainty in the initial position of the fire at time t_0 is also accounted for in the generation of the ensemble, the mean is the observed fire front position taken from the experiment; the stochastic perturbation corresponds to a shift of the front along the *x*-axis; the magnitude of the shift has a standard deviation of 65 cm. Finally, in these simulations, $N_{fr} = 100$ and $N^o_{fr} = 50$.

Figure 7 presents a comparison between the averaged forecasted fire front positions, the observations and the averaged analyzed fire front positions, at time $t_1 = 64$ s and $t_2 = 78$ s. Note that in Fig. 7(b), the mean forecast (dashed line) is initialized at time t_1 by the analysis produced by the first DA cycle (solid line in Fig. 7(a)) while the free forecast (dash-dot line) is initialized at time t_0 and does not use any analysis. In Fig. 7(a) and Fig. 7(b), it is seen that the mean forecast (that illustrates the performance of FIREFLY without data assimilation) significantly underestimates the rate of spread of the fire; in contrast, the mean analysis (that illustrates the performance of FIREFLY coupled with EnKF-based assimilation of the experimental fronts) provides accurate estimates of the fire front position. The agreement between the analyzed and observed fire front positions is remarkable and significantly better than that previously obtained in Refs. [18, 20] using a spatially-uniform parameter estimation approach for the same test case. In particular, the analyzed fire fronts feature a topology that is very close to that of the observed front, a result that requires an accurate and non-uniform correction of the locations of the fire front markers. However, in spite of the quality of the correction provided by EnKF, the performance of the forecast remains limited: for instance, Fig. 7(b) shows that the mean forecast (initialized by the analysis at time t_1 and integrated until time t_2), while still significantly more accurate than the free forecast (initialized by the initial conditions at time t_0 and integrated until time t_2), is not in agreement with the observation. These results are similar to those obtained in Fig. 6 and suggest that while a state estimation approach provides excellent forecasting performance at short lead times, this level of performance is not persistent and needs to be renewed by frequent observations.



(a) First analysis cycle (b) Second analysis cycle

Fig. 7. Validation test performed using data from a controlled grassland fire experiment and EnKF-FIREFLY results assuming uncertain ROS model parameters and uncertain initial conditions; multiple analysis cycles. Comparison between initial conditions at time $t_0 = 50$ s (open circles), mean forecast (dashed line), observations (crosses) and mean analysis (solid line). (a) time $t_1 = 64$ s; (b) time $t_2 = 78$ s; the dash-dot line is the free forecast obtained without any assimilation.

CONCLUSIONS

This study presents a prototype data-driven wildfire simulator capable of forecasting the fire spread dynamics. The prototype simulator features a regional-scale wildfire spread model coupled with a data assimilation algorithm based on an Ensemble Kalman Filter (EnKF) and a state estimation approach. The study assumes that observations of the fire front position are available at frequent times but possibly provide an inaccurate and incomplete description of the fire front. The prototype simulator is first evaluated in a series of verification tests using synthetically-generated observations and including configurations with spatially-varying vegetation properties and temporally-varying wind conditions. It is subsequently evaluated in a validation test corresponding to a controlled grassland fire experiment. The results indicate that data-driven simulations are capable of correcting inaccurate predictions of the fire front position and of subsequently providing an optimized forecast of the wildfire behavior. It was demonstrated that in order to allow for a spatially-varying correction of the front position, the generation of the EnKF ensemble should represent the anisotropy in fire propagation that results from spatial variations in vegetation properties and from the presence of wind. This anisotropy was implicitly introduced in EnKF by selecting spatially-dependent vegetation properties and different wind conditions between the members. Finally, the results also indicate that the forecasting performance of a state estimation approach is limited to near-term predictions (i.e., short lead times).

Future work will be aimed at developing a dual state estimation/parameter estimation approach that would overcome the limitations illustrated in the present and past studies. The parameter estimation approach could be extended to the case of weak spatial variations of the ROS model parameters. Assuming that the errors on the parameters vary slowly in time, the correction provided by data assimilation can reasonably be used for forecast, thus allowing for mid- to long-term forecast. In addition, the state estimation approach could be used for short-term forecast in order to locally correct the shape of the fire front. Furthermore, future plans also include the extension of the wildfire spread model FIREFLY to treat configurations with complex topography (the model is currently limited to the case of flat horizontal terrains). And finally, future plans include the integration of the DA algorithm into a CFD atmospheric model [26] in order to describe the interactions between the fire and the atmosphere. The ultimate goal of this research is to provide real-time fire forecasts using thermal-infrared imaging data including a description of both wildfire dynamics and fire plume emissions.

ACKNOWLEDGMENT

The authors would like to gratefully acknowledge useful interactions with Dr. Ronan Paugam (Department of Geography, King's College London, UK) for sharing the data of the grassland fire experiment as well as with Drs. Florent Duchaine and Thierry Morel (CERFACS) for support on Open-PALM.

REFERENCES

[1] Viegas, D.X. Overview of forest fire propagation research. In *Proceedings of the International Association of Fire Safety Science*, 10:95-108, 2011.

[2] Linn, R., Reisner, J., Colman, J.J. and Winterkamp, J. Studying wildfire behavior using FIRETEC. *International Journal of Wildland Fire*, 11:233-246, 2002.

[3] Mell, W., Jenkins, M.A., Gould, J. and Cheney, P. A physics-based approach to modeling grassland fires. *International Journal of Wildland Fire*, 16:1-22, 2007.

[4] Andrews, P.L. BEHAVE: Fire behavior prediction and modeling system, BURN Subsystem Part 1. Forest Service, U.S. Dept. of Agriculture, Technical Report INT-194, 1986.

[5] Finney, M.A. FARSITE: Fire area simulatorModel development and evaluation. Forest Service, U.S. Dept. of Agriculture, Research Paper RMRS-RP-4, 1998.

[6] Noble, I.R., Bary, G.A.V. and Gill, A.M. McArthur's fire danger meters expressed as equations. *Australian Journal of Ecology*, 5:201-203, 1980.

[7] Hirsch, K.G. Canadian forest Fire Behavior Prediction (FBP) system: User's guide. Northern Forest Centre, Special Report No. 7, 1996.

[8] Rothermel, R.C. A Mathematical Model for Predicting Fire Spread in Wildland Fuels. Forest Service, U.S. Dept. of Agriculture, Research Paper INT-115, 1972.

[9] Finney, M.A., Grenfell, I.C., McHugh, C.W., Seli, R.C., Trethewey, D., Stratton, R.D. and Brittain, S. A method for ensemble wildland fire simulation. *Environmental Modeling and Assessment*, 16:153-167, 2011.

[10] Paugam, R., Wooster, M.J. and Roberts, G. Use of handheld thermal imager data for airborne mapping of fire radiative power and energy and flame front rate of spread. *Geoscience and Remote Sensing*, 51:3385-3399, 2013.

[11] Wooster, M.J., Roberts, G., Perry, G. and Kaufman, Y.J. Retrieval of biomass combustion rates and totals from fire radiative power observations: FRP derivation and calibration relationships between biomass consumption and fire radiative energy release. *Journal of Geophysical Research*, 110, D24311, 2005.

[12] Gelb, A. *Applied Optimal Estimation*. Cambridge Massachusetts MIT Press, 1974.

[13] Kalnay, E. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 2003.

[14] Tarantola, A. *Inverse Problem Theory, Methods for Data Fiiting and Parameter Estimation*. Elsevier, 1987.

[15] Todling, R. and Cohn, S.E. Suboptimal schemes for atmospheric data assimilation based on the Kalman filter. *Monthly Weather Review*, 122:2530-2557, 1994.

[16] Evensen, G. Sequential data assimilation with non linear quasi-geotrophic model using Monte-Carlo methods to forecast error statistic. *Journal of Geophysical Research*, 99:143-162, 1994.

[17] Evensen, G. *Data assimilation - The Ensemble Kalman Filter*. Springer, 2009.

[18] Rochoux, M.C., Delmotte, B., Cuenot, B., Ricci, S. and Trouvé, A. Regional-scale simulations of wildland fire spread informed by real-time flame front observations. *Proceedings of the Combustion Institute*, 34:2641-2647, 2013, http://dx.doi.org/10.1016/j.proci.2012.06.090.

[19] Rochoux, M.C., Cuenot, B., Ricci, S., Trouvé, A., Delmotte, B., Massart, S., Paoli, R. and Paugam, R. Data assimilation applied to combustion. *Compte-Rendus Mécanique*, 241:266-276, 2013, http://dx.doi.org/10.1016/j.crme.2012.10.011.

[20] Rochoux, M.C., Ricci, S., Lucor, D., Cuenot, B., Trouvé, A., and Bart, J.-M. Towards predictive simulations of wildfire spread using a reduced-cost Ensemble Kalman filter based on Polynomial Chaos approximations. In *Proceedings of the Summer Program, Center for Turbulence Research*, July 2012, NASA AMES, Stanford University, USA.

[21] Mandel, J., Beezley, J.D. Morphing Ensemble Kalman filter. *Tellus 60A*, 2007.

[22] Mandel, J., Bennethum, L.S., Beezley, J.D., Coen, J.L., Douglas, C.C., Minjeong, K. and Vodacek, A. A wildland fire model with data assimilation. *Mathematics and Computers in Simulation*, 79:584-606, 2008.

[23] Mandel, J., Beezley, J.D., Kochansky, A.K. Coupled atmosphere-wildland fire modeling with WRF 3.3 and SFIRE 2011. *Geosciences Model Development*, 4:591-610, 2011.

[24] Rehm, R.G. and McDermott, R.J. Fire front propagation using the level-set method. National Institute of Standards and Technology, Technical Note 1611, 2009.

[25] Burgers, G. and Van Leeuwen, P. and Evensen, G. Analysis scheme in the Ensemble Kalman filter. *Monthly Weather Review*, 126:1719-1724, 1998.

[26] Filippi, J.-B., Pialat, X. and Clements, C.B. Assessment of ForeFire/Meso-NH for wildland fire/atmosphere coupled simulation of the FireFlux experiment. In *Proceedings of the Combustion Institute*, 34:2633-2640, 2013.