State-parameter estimation approach for data-driven wildland fire spread modeling: Application to the 2012 RxCADRE S5 field-scale experiment

Cong Zhang, Annabelle Collin, Philippe Moireau, Arnaud Trouvé, Mélanie C. Rochoux

Version of Record: https://www.sciencedirect.com/science/article/pii/S0379711218305502
Manuscript_b45c91eb4845d8326245d9e84f15c449

State-Parameter Estimation Approach for Data-Driven Wildland Fire Spread Modeling: Application to the 2012 RxCADRE S5 Field-Scale Experiment

Cong Zhang[a], Annabelle Collin[b], Philippe Moireau[c], Arnaud Trouvé[a], Mélanie C. Rochoux[d,*]

[a]*Department of Fire Protection Engineering, University of Maryland, College Park, MD 20742, USA*
[b]*Inria, Université de Bordeaux, Bordeaux INP, 351 cours de la Libération, 33405 Talence cedex, France*
[c]*Inria – LMS, Ecole Polytechnique, CNRS – Université Paris-Saclay, 1 rue Honoré d'Estienne d'Orves, Campus de l'Ecole Polytechnique, 91120 Palaiseau, France*
[d]*CECI, Université de Toulouse, CNRS, CERFACS, 42 Avenue Gaspard Coriolis, 31057 Toulouse cedex 01, France*

Abstract

Data assimilation is an emerging and powerful tool towards real-time flame front monitoring for wildland fire applications. The key idea is to regularly update the state and/or parameters of a fire spread model using observed firelines in order to improve a forecast on future fire locations. The merits of combining state estimation and parameter estimation through a hybrid state-parameter estimation algorithm are demonstrated through the 2012 RxCADRE S5 field-scale controlled burn experiment. For state estimation, we adopt a cost-effective Luenberger observer formulation to reconstruct a complete view of the burning state at a given time. For parameter estimation, we use an ensemble transform Kalman filter to solve the inverse modeling problem consisting of inferring more realistic wind conditions given observations of the actual burning state. The data-driven model relies on a front shape similarity measure derived from image segmentation theory to quantify position errors. We show that the hybrid approach provides an efficient framework to address all sources of model uncertainties and to select burning scenarios that are most likely to occur. Parameter estimation is a key component of the data-driven model by reducing model bias. Using the fire spread model in forecast mode is then an asset to accurately track the flame front dynamics at future lead times.

Keywords: Wildland fire, Front-tracking model, Data assimilation, State-parameter estimation, Inverse modeling, Front shape similarity measure, Position errors, Model forecast

*Corresponding author: Mélanie C. Rochoux, melanie.rochoux@cerfacs.fr, +33 (0) 5 61 19 30 72

1. Introduction

Data assimilation is a powerful technique to couple information coming from both measurements and simulations, taking into account that sensors or models, when used in isolation, only provide an incomplete and uncertain description of the real state of a dynamical system [1–5]. The idea is to use observations to best estimate the set of physical parameters, external forcing or initial conditions used in the model to improve its accuracy in both reanalysis and forecast modes. The application of data assimilation to combustion [6–9] raises a number of methodological questions. Combustion can be considered as a front-tracking problem with the challenges of: (1) simulating accurate flame shape and topology; and (2) comparing simulations to observation images obtained for instance from infrared camera. Classical data assimilation methods perform poorly when the simulated front deviates from observations due to the assumption that errors are of amplitude type: data assimilation methods typically compensate position errors by adjusting amplitudes, which can eventually produce non-physical fronts [10–12]. There is therefore a need to develop an original image data assimilation strategy that can handle position errors for front-tracking problems such as wildland fire spread [8, 13–19].

Wildland fire spread is represented at regional scales using a front-tracking model, where the front (or "fireline") represents the moving interface separating burnt and unburnt biomass fuel. This regional-scale modeling approach relies on an empirical parametrization of the front propagating speed (or "rate of spread") in terms of biomass, topography and near-surface meteorology conditions. But it has a number of limitations. First, rate-of-spread models rely on a series of physical simplifications; for instance, they usually do not explicitly account for fire-atmosphere interactions. Second, rate-of-spread models require a list of input parameters representing local environmental conditions but these parameters may only be known with limited accuracy. Third, rate-of-spread models have a domain of validity that is limited to the original experimental conditions used during their development and calibration, which does not include, for instance, extreme fire behavior. Still, rate-of-spread models are valuable to support decision-making by fire emergency responders [20]. And data assimilation offers an attractive framework to overcome modeling limitations by taking advantage of recent progress made in remote sensing technology. One example of typical observation data is thermal infrared imaging, which highlights areas where active flaming combustion is occurring [21] and which can be used to extract fireline position [22, 23]. However, fireline observations are likely to be available with a coarse resolution in space and time; they may also be incomplete in the sense that only a portion of the fireline may be covered, for instance due to the plume opacity. These expected limitations in the observations motivate the use of a fire spread model to reconstruct a complete fire perimeter and forecast its behavior at lead times [24].

In this context, the key idea of data-driven modeling is to design a data assimilation algorithm capable of assimilating observed fireline positions to provide an improved model representation of wildland fire behavior. The data assimilation algorithm includes the following components: (1) a model based on a rate-of-spread submodel and on a two-dimensional front-tracking numerical scheme; (2) observations of the fireline position at discrete times;

and (3) a sequential data assimilation algorithm that infers a correction on the estimation targets (or "control variables") from the differences between observed and simulated firelines and their estimated error statistics. An intuitive approach to represent the fireline evolution is to adopt a Lagrangian viewpoint and to compute the observation-simulation discrepancies in terms of a metric distance between observed and simulated markers. This approach was successfully evaluated against reduced-scale and field-scale (FIREFLUX I) experiments [17, 19]. However, this Lagrangian representation suffers limitations at regional scales where the fireline topology is highly heterogeneous and complex. To overcome these limitations, we have recently designed an object-oriented data assimilation strategy that considers the burning area as a moving object, whose shape changes in time [25], and that formulates the observation-simulation discrepancies using a non-Euclidean front shape similarity measure derived from image segmentation theory [26].

In wildland fire spread applications, the success of data assimilation strongly depends on the choice of the control variables. The control variables must be representative of the actual sources of uncertainty and the model predictions must be sensitive to variations in the control variables to make the correction physically-consistent and effective. The control variables can be of different nature. In a parameter estimation mode, corresponding to an inverse modeling problem, the control variables are the input parameters of the rate-of-spread model (which cannot be directly measured or can only be measured with great difficulty and uncertainty) [17, 19]. In a state estimation mode, the control variables are directly the spatial coordinates of the fireline [18, 24]. When using standalone parameter estimation, we found that the updated control parameters are usually over-corrected due to the presence of more sources of uncertainty than assumed in the method (e.g. initial fireline location, model formulation itself). We also found that standalone state estimation is limited to a short-term forecast due to a short temporal persistence of the correction.

The objective of the present study is to demonstrate the benefits of a hybrid state-parameter estimation approach to enhance the forecast performance. The key idea of the hybrid approach is to simultaneously control the model state and a subset of significant parameters in order to obtain more accurate physical values of the estimated parameters by attributing uncertainties not only to the model parameters, but also to the model state. Parameter estimation is carried out using an ensemble transform Kalman filter (ETKF) [27, 28]; state estimation is carried out using a Luenberger observer (LO) formulation [25]. A proof of concept of a joint state-parameter estimation approach was presented in Ref. [25] against synthetic experiments. We are now in the process of validating our front data assimilation strategy in a real-world wildland fire test corresponding to the 2012 RxCADRE S5 field-scale experiment [21]. The study in Ref. [24] was focused on state estimation applied to the RxCADRE dataset to show first, the added value of data assimilation compared to a simple extrapolation of observations for observation frequencies lower than $1/60$ s^{-1}, and second, how the front shape similarity measure can be formulated for both Eulerian and Lagrangian front-tracking models in a unified framework. In the present study, the novelty is to show the benefits of combining state estimation and parameter estimation to accurately track the fireline in analysis mode and forecast mode. The sensitivity to the observation frequency (ranging from $1/120$ s^{-1} to $1/30$ s^{-1}) is also investigated.

The outline of this paper is as follows: the fire spread model is presented in Section 2; the hybrid state-parameter estimation approach is presented in Section 3; the verification and validation tests including the RxCADRE dataset are presented in Section 4; and results are presented in Section 5.

2. Forward Modeling: the Fire Spread Model

Regional-scale wildland fire spread models feature two main components: (1) a rate-of-spread (ROS) model that gives values of ROS as a function of local vegetation and topographical properties, as well as meteorological conditions (we use the classical model due to Rothermel [29, 30]); and (2) a front-tracking algorithm that provides a solution of the front propagation equation on a two-dimensional horizontal plane $(x, y) \in \Omega$, where $\Omega \subset \mathbb{R}^2$ denotes the two-dimensional computational domain (Ω is the physical plane where the fire evolves in case of a flat terrain and is a conceptual projected plane in case of an arbitrarily complex terrain).

The fire spread model used in our work [17–19, 25] is an Eulerian front-tracking simulator based on a level-set method [31–33]. The forward model is similar to that adopted in the ELMFIRE fire simulator [34] or in the WRF-SFIRE coupled fire-atmosphere system [35]. In the Eulerian framework, a progress variable $c \equiv c(x, y, t)$ is introduced as a marker of the fireline; the fireline is then defined as the contour line $c_{\text{fr}} = 0.5$; $c < c_{\text{fr}}$ ($c > c_{\text{fr}}$) represents unburnt (burnt) vegetation. The time evolution of c is governed by

$$\frac{\partial c}{\partial t}(x, y, t) + \mathbf{V} \cdot \nabla c(x, y, t) = 0, \quad c(x, y, t_0) = c_0(x, y), \quad (x, y) \in \Omega,\ t \geq t_0, \tag{1}$$

where $\mathbf{V}$ is the displacement velocity of an iso-c surface, and where $c_0(x, y)$ is the initial condition at time t_0. In wildland fire applications, we commonly assume that the fireline self-propagates normal to itself; the normal direction to the front $\mathbf{n}_{\text{fr}}$ is defined as $\mathbf{n}_{\text{fr}} \equiv \mathbf{n}_{\text{fr}}(x, y, t) = -\nabla c(x, y, t)/|\nabla c(x, y, t)|$. Using this definition, the progress variable c is calculated as a solution of the propagation equation derived from Eq. (1), i.e.

$$\frac{\partial c}{\partial t}(x, y, t) = \mathcal{V}(x, y, t)\ |\nabla c(x, y, t)|, \quad c(x, y, t_0) = c_0(x, y), \quad (x, y) \in \Omega,\ t \geq t_0, \tag{2}$$

where $\mathcal{V} \equiv \mathbf{V} \cdot \mathbf{n}_{\text{fr}}(x, y, t)$ corresponds to the Rothermel-based ROS parameterization. Note that the near-surface wind velocity denoted by $\mathbf{U}_{\text{w}}$ is involved in the calculation of $\mathcal{V}$; it is specified using two scalar parameters, the wind flow velocity magnitude u_{w} ($u_{\text{w}} \equiv \|\mathbf{U}_{\text{w}}\|$) and the wind flow velocity direction angle d_{w} such that $\mathbf{U}_{\text{w}} = (u_{\text{w}} \sin d_{\text{w}}, u_{\text{w}} \cos d_{\text{w}})^T$. In the following, u_{w} is referred to as the "wind speed" and d_{w} as the "wind direction".

The level-set function $\phi_c(x, y, t) = (c(x, y, t) - c_{\text{fr}})$ also satisfies Eq. (2); the fireline is represented by the contour line $\phi_c(x, y, t) = 0$ referred to as the "level set". We denote the time-evolving two-dimensional fireline as $\Gamma(t) = \{(x, y) \in \Omega \,|\, c(x, y, t) = c_{\text{fr}} \Leftrightarrow \phi_c(x, y, t) = 0\}$. We also denote by $\mathcal{B}(t)$ the burnt area bounded by the fireline $\Gamma(t)$ at a given time t.

To solve Eq. (2), we follow the choices made in Ref. [36] and use a second-order Runge-Kutta scheme for time-integration and a second-order total variation diminishing (TVD) scheme with a Superbee slope limiter for spatial discretization, see Ref. [37].

3. Data-Driven Modeling: the Hybrid State-Parameter Estimation Approach

The key idea of sequential data assimilation is to find optimal estimates of a set of variables (i.e. the control variables) given uncertain observations and model predictions (i.e. the "background" or "prior" information). The model informed by the observations is referred to as the "data-driven model" by opposition to the "free run" corresponding to the model prediction without data assimilation. A general framework for sequential data assimilation is presented in Ref. [25].

3.1. Sources of Uncertainty and Control Variables

The choice of the control variables is a critical step in the design of a data assimilation procedure. This choice must reflect the main sources of uncertainty in the forward model, and the model predictions must be sensitive to variations of the selected uncertain variables. An incorrect identification of the sources of uncertainty would indeed lead to an artificial adjustment of the model through data assimilation and thus to a poor-quality forecast.

In the present study, the uncertain variables are the initial location of the fire (i.e. the initial condition c_0) and some parameters θ informing the ROS model $\mathcal{V}(x, y, t)$ in Eq. (2). The uncertainties in the front position due to uncertainties in the initial fire location c_0 are addressed through state estimation, meaning that the model state c is updated over time to locally correct the shape of the fireline and restart the forward model from a fire location that is informed by observations. This is a classical approach in numerical weather prediction [3]. Furthermore, uncertainties in the front position due to uncertainties in the ROS parameters θ are addressed through parameter estimation, meaning that the parameters included in θ are explicitly updated when observations become available [38]. Time-averaged estimates of the control parameters are retrieved over the time window between two successive observations; we thereby assume that the errors on the ROS model parameters vary slowly in time [15, 17, 18]. The control parameters are here the near-surface wind speed u_{w} and direction d_{w}.

3.2. Front Shape Similarity Measure

The estimation algorithms used in this work are not classical since we are dealing with a front-tracking problem: we need to have a specific measure (referred to as the "front shape similarity measure") to represent the discrepancies between the observations $\mathrm{y^o}$ and the simulated firelines y (or "observable quantities") in terms of position errors [10, 25, 26]. Here the observation $\mathrm{y^o}$ is defined as the instantaneous two-dimensional binary field such that field values are equal to 0 in the unburnt region and 1 in the burnt region.

The front shape similarity measure is derived from the Chan-Vese contour fitting functional introduced for object detection in image processing theory [32, 39, 40]. In a level-set formalism, this measure can be written as

$$\mathcal{J}(\mathrm{y^o}, \phi_c) = \int_\Omega H_{\mathrm{v}}(\phi_c)\, \left[\mathrm{y^o} - C_{\max}(\mathrm{y^o}, \phi_c)\right]^2 + (1 - H_{\mathrm{v}}(\phi_c))\, \left[\mathrm{y^o} - C_{\min}(\mathrm{y^o}, \phi_c)\right]^2 \, dx\, dy, \quad (3)$$

where ϕ_c is the level-set function ($\phi_c = c - c_{\mathrm{fr}}$), H_{v} the Heaviside function ($H_{\mathrm{v}}(\phi_c) = 0$ if $\phi_c < 0$; $H_{\mathrm{v}}(\phi_c) = 1$ if $\phi_c > 0$), and where $C_{\min}$ and $C_{\max}$ are scalar coefficients defined by

$C_{\min} = \min(C_0, C_1)$ and $C_{\max} = \max(C_0, C_1)$ with

$$\begin{cases} C_0(\mathrm{y^o}, \phi_c) = \dfrac{\displaystyle\int_\Omega (1 - H_\mathrm{v}(\phi_c))\, \mathrm{y^o}\, dx\, dy}{\displaystyle\int_\Omega (1 - H_\mathrm{v}(\phi_c))\, dx\, dy}, \; 0 \leq C_0 \leq 1, \\ C_1(\mathrm{y^o}, \phi_c) = \dfrac{\displaystyle\int_\Omega H_\mathrm{v}(\phi_c)\, \mathrm{y^o}\, dx\, dy}{\displaystyle\int_\Omega H_\mathrm{v}(\phi_c)\, dx\, dy}, \; 0 \leq C_1 \leq 1. \end{cases} \tag{4}$$

C_1 measures the level of agreement between the observed and simulated burnt areas; C_0 measures the level of disagreement between the observed and simulated unburnt areas.

The objective in image segmentation is to minimize the functional $\mathcal{J}$; this is achieved when the observed and simulated fronts coincide, i.e. $C_1 = 1$ and $C_0 = 0$. In data assimilation applications, the objective is to retrieve an optimal estimate of the control variables that is compatible with both the observations and the background information. We need to introduce estimators that are compatible with the front shape similarity measure (Eqs. 3–4) and that also account for observation and background uncertainties.

3.3. State Estimation (SE)

To present the state estimation (SE) procedure, we assume in this section that the control parameters θ are perfectly known, i.e. the control variables correspond to the model state c and uncertainties are only due to the initial condition c_0. We correct the model state using a Luenberger observer (LO), i.e. by directly modifying the prognostic Eq. (2) through the introduction of a relaxation term that steers the model state towards the observations $\mathrm{y^o}$ in a continuous way. This methodology corresponds to a continuous-model continuous-data estimation problem.

The key idea of a LO is to use the physical constraints of the model equation to implicitly propagate the correction over the computational domain Ω [41]. In the present front-tracking problem, the LO implicitly propagates the correction from one point of the fireline to the rest of the fireline and thus preserves a coherent structure for the simulated front [26]. Note that this is different from the SE algorithm used in Ref. [18] based on an ensemble Kalman filter (EnKF). Using a LO avoids the formulation of high-dimensional matrices required within the EnKF framework to compute the Kalman gain matrix. This indeed becomes a computational issue when the wildland fire extends over a large spatial domain, while the fireline still needs to be well-resolved in space (the RxCADRE experiment already extends over a few hectares, while the fireline requires meter-scale resolution).

We use the same LO formulation as in Refs. [24, 25] so that the estimated progress

variable $\widehat{c}$ is governed by

$$\begin{cases} \dfrac{\partial \widehat{c}}{\partial t}(x,y,t) = \mathcal{V}\ |\nabla \widehat{c}(x,y,t)| \\ \qquad -\lambda(x,y)\,\delta(\phi_c)\left\{[\mathrm{y^o}(x,y,t) - C_{\max}(\mathrm{y^o},\phi_c)]^2 - [\mathrm{y^o}(x,y,t) - C_{\min}(\mathrm{y^o},\phi_c)]^2\right\},\ t \geq t_0 \\ \widehat{c}(x,y,t_0) = \widehat{c}_0(x,y), \end{cases} \tag{5}$$

where $\widehat{c}_0$ is the (presumably incorrect) initial condition at time t_0 and where $\delta \equiv \delta(\phi_c)$ is a Dirac delta-function that ensures that the correction is only active along the simulated fireline [26]. The rate at which the simulated burnt area y converges towards the observation $\mathrm{y^o}$ is controlled by the gain factor λ. Note that we can also add a topological feedback term in Eq. (5) to track multiple fires simultaneously [25]; however this is not required in the RxCADRE application since there is no fire spotting. Note also that Eq. (5) has an equivalent formulation for a Lagrangian front-tracking solver [24].

In practice, Eq. (5) is discretized in space and time following the choices made for the standalone model integration (TVD with Superbee slope limiter and Runge-Kutta scheme, see Section 2).

3.4. Parameter Estimation (PE)

To present the parameter estimation (PE) procedure, we assume in this section that the initial condition c_0 is perfectly known, i.e. the control variables correspond to the uncertain parameters θ and uncertainties are only due to the parameters θ. We correct the control parameters using a Kalman-based filtering framework, which is usually presented as a discrete-model discrete-observation estimation problem. There is a mathematical correspondence between the continuous and discrete estimation problems; this framework is summarized in Ref. [25]. For simplicity purpose, we only describe the Kalman correction after space and time discretization of the forward model and observations in this paper. From now on, we therefore use bold letters to refer to the quantities associated with the discretized solutions; t refers to the time index and $[t-1,t]$ refers to a given assimilation time window (i.e. the time between two successive observations).

In the Kalman-based filtering framework, the control parameters are noted $\boldsymbol{\theta}$. The variables in $\boldsymbol{\theta}$ are considered as independent random variables; each random variable is characterized by a Gaussian marginal probability density function (PDF) defined by its mean value and its standard deviation (STD). Kalman filtering is a two-step procedure over each window $[t-1,t]$ made of: (1) a forecast step to quantify uncertainties in the background control parameters and measure the sensitivity of the simulated observable quantities with respect to variations in the control parameters over the time window $[t-1,t]$; and (2) an analysis step to provide new estimates of the control parameters at time t, given available observations and the following update equation:

$$\boldsymbol{\theta}_t^{\mathrm{a}} = \boldsymbol{\theta}_t^{\mathrm{f}} + \mathbf{K}_t\left[\boldsymbol{D}\left(\mathbf{y}_t^{\mathrm{o}}, \boldsymbol{G}(\boldsymbol{\theta}_t^{\mathrm{f}})\right)\right], \tag{6}$$

where $\boldsymbol{\theta}_t^{\mathrm{f}}$ ($\boldsymbol{\theta}_t^{\mathrm{a}}$) is the forecast (analysis) estimate of the control vector, where $\boldsymbol{G}$ is the observation operator mapping the control vector (here the parameter space) onto the obser-

vation space such that $\mathbf{y}_t^{\mathrm{f}} = \boldsymbol{G}(\boldsymbol{\theta}_t^{\mathrm{f}})$, and where $\boldsymbol{D}$ is the discrepancy operator of least-square type associated with the front shape similarity measure in Eq. (3) and of the form $\boldsymbol{D}\left(\mathbf{y}_t^{\mathrm{o}}, \boldsymbol{G}(\boldsymbol{\theta}_t^{\mathrm{f}})\right) = (\boldsymbol{D}^{\epsilon,+}, \boldsymbol{D}^{\epsilon,-})^T$ [25] with $\boldsymbol{D}^{\epsilon,+}$ and $\boldsymbol{D}^{\epsilon,-}$ discretizing the quantities

$$\begin{cases} \boldsymbol{D}^{\epsilon,+} = \left[1 + \frac{2}{\pi}\arctan\left(\frac{\phi_c}{\epsilon}\right)\right]\left[\mathbf{y}_t^{\mathrm{o}} - C_{\max}(\mathbf{y}_t^{\mathrm{o}}, \boldsymbol{\phi}_c)\right], \\ \boldsymbol{D}^{\epsilon,-} = \left[1 - \frac{2}{\pi}\arctan\left(\frac{\phi_c}{\epsilon}\right)\right]\left[\mathbf{y}_t^{\mathrm{o}} - C_{\min}(\mathbf{y}_t^{\mathrm{o}}, \boldsymbol{\phi}_c)\right], \end{cases} \tag{7}$$

where ϵ is a small parameter defined with respect to the contour sharpness [42]. $\mathbf{K}_t$ is the Kalman gain matrix that gives a weight to the observation-simulation discrepancy according to the uncertainties in the observations and in the background estimate. These uncertainties are described using error covariance matrices denoted by $\mathbf{P}^{\mathrm{f}}$ ($\mathbf{P}^{\mathrm{a}}$) for the background (analysis) and $\mathbf{R}$ for the observations. This two-step filtering procedure is applied sequentially when new observations become available.

Due to the presence of nonlinearities in the wildland fire spread model, we rely on an EnKF algorithm [4] to perform sequential PE. There are several EnKF algorithms in the literature. In Refs. [17, 19], we use a stochastic version of the EnKF algorithm in the sense that observations are randomly perturbed to obtain an ensemble of observations to be compared to the ensemble of simulated firelines; this approach is known to introduce additional sampling error and to underestimate the error in the new analysis estimate [43]. Therefore, in the present work, we rely on a deterministic version of the EnKF called ensemble transform Kalman filter (ETKF), which avoids observation randomization [27, 28].

Ensemble-based Kalman filters rely on the use of an ensemble of simulations to estimate the sensitivities between model inputs and outputs. We note N_{e} the size of the ensemble. Each particle or "member" of the ensemble can be decomposed as $\boldsymbol{\theta}^{(k)} = \overline{\boldsymbol{\theta}} + \widehat{\boldsymbol{\theta}}^{(k)}$, where k refers to the index of the member, $\overline{\boldsymbol{\theta}}$ is the mean (ensemble-averaged) value defined as

$$\overline{\boldsymbol{\theta}} = \frac{1}{N_{\mathrm{e}}} \sum_{k=1}^{N_e} \boldsymbol{\theta}^{(k)}, \tag{8}$$

and $\hat{\boldsymbol{\theta}}^{(k)}$ is the statistical spread characterized by the error covariance matrix $\mathbf{P}$,

$$\mathbf{P} = \left(\frac{1}{N_{\mathrm{e}} - 1}\right) \boldsymbol{\Theta}\, \boldsymbol{\Theta}^T, \tag{9}$$

with $\boldsymbol{\Theta} = [\boldsymbol{\theta}^{(1)} - \overline{\boldsymbol{\theta}}, \cdots, \boldsymbol{\theta}^{(N_{\mathrm{e}})} - \overline{\boldsymbol{\theta}}] = [\widehat{\boldsymbol{\theta}}^{(1)}, \cdots, \widehat{\boldsymbol{\theta}}^{(N_{\mathrm{e}})}]$ the ensemble perturbation matrix associated with the control parameters.

In the ETKF framework, the analysis mean estimate is obtained using the Kalman update equation (Eq. 6) and the analysis error covariance matrix $\mathbf{P}^{\mathrm{a}}$ is obtained as a transformation of the background error covariance matrix $\mathbf{P}^{\mathrm{f}}$ as

$$\mathbf{P}^{\mathrm{a}} = \frac{1}{N_{\mathrm{e}} - 1}\, \boldsymbol{\Theta}^{\mathrm{a}} \boldsymbol{\Theta}^{\mathrm{a}\,T} = \frac{1}{N_{\mathrm{e}} - 1}\, \boldsymbol{\Theta}^{\mathrm{f}}\, \mathbf{W}\, \mathbf{W}^T\, \boldsymbol{\Theta}^{\mathrm{f}\,T}, \tag{10}$$

where $\mathbf{W}$ is the transform matrix satisfying

$$\mathbf{W} = \sqrt{N_\mathrm{e} - 1}\, \left[(N_\mathrm{e} - 1)\, \mathbf{I} + \mathbf{Y}^{\mathrm{f}T}\, \mathbf{R}^{-1}\, \mathbf{Y}^{\mathrm{f}} \right], \tag{11}$$

such that $\mathbf{Z}^{\mathrm{a}} = \mathbf{Z}^{\mathrm{f}}\, \mathbf{W}$ with $\mathbf{Z}^{\mathrm{f}} = 1/\sqrt{N_\mathrm{e} - 1}\, \mathbf{\Theta}^{\mathrm{f}}$, and where $\mathbf{Y}^{\mathrm{f}} = [\widehat{\boldsymbol{y}}^{\mathrm{f},(1)}, \cdots, \widehat{\boldsymbol{y}}^{(N_\mathrm{e})}]$ is the forecast ensemble perturbation matrix associated with the model outputs. The analysis ensemble can then be easily constructed from the analysis mean and the analysis spread. Details of the ETKF algorithms can be found in Ref. [44].

To account for unrepresented error sources when generating the ensemble and to avoid underestimating the ensemble variance, we adopt the relaxation-to-prior perturbations (RTPP) scheme to artificially inflate the analysis ensemble spread [45]:

$$\widehat{\boldsymbol{\theta}}^{\mathrm{f}}_{t+1} = (1 - \alpha)\, \widehat{\boldsymbol{\theta}}^{\mathrm{a}}_{t} + \alpha\, \widehat{\boldsymbol{\theta}}^{\mathrm{f}}_{t}, \tag{12}$$

where α ranges from 0 to 1 featuring a small-to-large inflation effect.

Note that for the RxCADRE experiment, the wind control parameters u_w and d_w are assumed to be uniform over the computational domain Ω since environmental conditions are quasi-uniform (an extension to spatially-distributed PE is possible for heterogeneous conditions [19]).

3.5. Hybrid State-Parameter Estimation (HE)

In this study, our data assimilation algorithm combines a SE approach based on the LO presented in Eq. (5) and a PE approach based on the ETKF algorithm presented in Eqs. (8)–(10). Since we use two distinct methods for SE and PE, we refer to this algorithm as "hybrid".

The hybrid data assimilation algorithm (HE) is sequentially applied over an assimilation window $[t-1, t]$. Each assimilation cycle is decomposed into two successive forecast-analysis steps for each ensemble member. During the forecast step, the state of the kth ensemble member ($k = 1, \cdots, N_\mathrm{e}$) is forwarded from time $(t-1)$ to time t using Eq. (5) with a given gain λ, instead of the original Eq. (2) of the forward model. The kth forecast model state at time t (after spatial and temporal discretization using TVD and Runge-Kutta schemes, see Section 2) is noted $\mathbf{c}_t^{(k)}$. The corresponding forecast observable quantity at time t, $\mathbf{y}_t^{\mathrm{f},(k)}$, is different for each member k due to changes in the control parameters $\boldsymbol{\theta}_t^{\mathrm{f},(k)}$ and in subsequent changes in the LO feedback term. The idea is to perform a LO-based SE for each ensemble member during the ETKF forecast step; we only update the control parameters $\boldsymbol{\theta}_t^{\mathrm{f},(k)}$ during the ETKF analysis step following Eqs. (6)–(11). The new state $\mathbf{c}_t^{\mathrm{a}}$ and control parameters $\boldsymbol{\theta}_t^{\mathrm{a},(k)}$ serve as the prior-known information, $\mathbf{c}_{t+1}^{\mathrm{f}}$ and $\boldsymbol{\theta}_{t+1}^{\mathrm{f},(k)}$, for the next assimilation window $[t, t+1]$ (with RTPP-inflation for the control parameters, see Eq. 12).

Note that the hybrid method is based on a dual state-parameter estimation, meaning that LO and ETKF are applied successively over a given assimilation time window [46, 47]. This could be extended to a joint estimation as in Refs. [25, 48, 49], where both LO and ETKF are applied simultaneously; however, this is beyond the scope of the present study, which is more oriented towards wildland fire applications.

To show the benefits of the HE approach, we compare in Section 5, the results obtained with different choices of the control variables: standalone PE using ETKF, standalone SE using LO (corresponding to the first step of HE applied to each ensemble member, also presented in detail in Ref. [24]), and HE (corresponding to the successive application of LO and ETKF over each assimilation time window). The same value of λ is used for standalone SE and HE. λ is specified so as to reduce uncertainties induced by wrong initial condition and imperfect model predictions, not those induced by wrong input parameters to the fire spread model (this is the role of the forecast error covariance matrix $\mathbf{P}^{\mathrm{f}}$). We therefore consider a moderate level of nudging $\lambda = 0.2$ so as to balance the correction in-between the two SE/PE steps of the HE approach.

4. Verification and Validation Procedure

4.1. Observing System Simulation Experiment (OSSE) Verification Procedure

In data assimilation, verification tests are typically done through the observing system simulation experiment (OSSE) framework, in which the reference (or "true") state of the system is assumed to be known and is constructed from a solution of the forward model. Figure 1 presents a typical OSSE framework, in which the true values of the control variables (the initial condition $\mathbf{c}_0^{\mathrm{t}}$ and/or the input parameters $\boldsymbol{\theta}^{\mathrm{t}}$ denoted by $\mathbf{x}^{\mathrm{t}}$ for genericity) exist and are known. Thus, the observation and background errors $\boldsymbol{\epsilon}^{\mathrm{o}}$ and $\boldsymbol{\epsilon}^{\mathrm{f}}$ are known.

The background information $\mathbf{x}^{\mathrm{f}}$ (prior to data assimilation) is obtained by perturbing the true parameters $\mathbf{x}^{\mathrm{t}}$; the perturbation $\boldsymbol{\epsilon}^{\mathrm{f}}$ is of Gaussian-type for the ETKF algorithm and is characterized by the STD $\boldsymbol{\sigma}^{\mathrm{f}}$. Note that the initial condition $\mathbf{c}_0$ is formulated in a parametric form with respect to the "center of mass" of the initial burning area, whose position is denoted by $(x_{\mathrm{ign}}, y_{\mathrm{ign}})$. Thus the parameters that are perturbed to generate a forecast are the initial conditions $(x_{\mathrm{ign}}, y_{\mathrm{ign}})$ and the ROS parameters $(u_{\mathrm{w}}, d_{\mathrm{w}})$. The N_{e}-sets of parameters are obtained using Monte Carlo random sampling.

The observations $\mathbf{y}^{\mathrm{o}}$ are synthetically generated by integrating the fire spread model (Eq. 2) using the true values of the uncertain variables and by adding noise to the observed fireline. In the OSSE test case (Section 5.1), the observation error $\boldsymbol{\epsilon}^{\mathrm{o}}$ (characterized by the STD $\boldsymbol{\sigma}^{\mathrm{o}}$) is assumed to be small ($\boldsymbol{\sigma}^{\mathrm{o}} = 0.1$); the observations are thus considered as the "targets" of the data assimilation algorithm.

The performance of the HE approach is evaluated by examining its ability to retrieve the true structure of the fire state $\mathbf{c}^{\mathrm{t}}$ and the true values of the control parameters $\boldsymbol{\theta}^{\mathrm{t}}$ in a situation in which observations are considered accurate. This configuration aims at pushing the data-driven model to its limit in a situation in which the error in the prior is large and the estimation involves an important correction to match the observations. Diagnostics comparing the background $\mathbf{x}^{\mathrm{f}}$, the analysis $\mathbf{x}^{\mathrm{a}}$ and the truth $\mathbf{x}^{\mathrm{t}}$ are used as verification tools of the proposed data assimilation algorithm.

4.2. RxCADRE Experimental Data

In this section, we briefly review the experimental data used for validation in Section 5.2. RxCADRE corresponds to a series of field experiments aimed at model development and

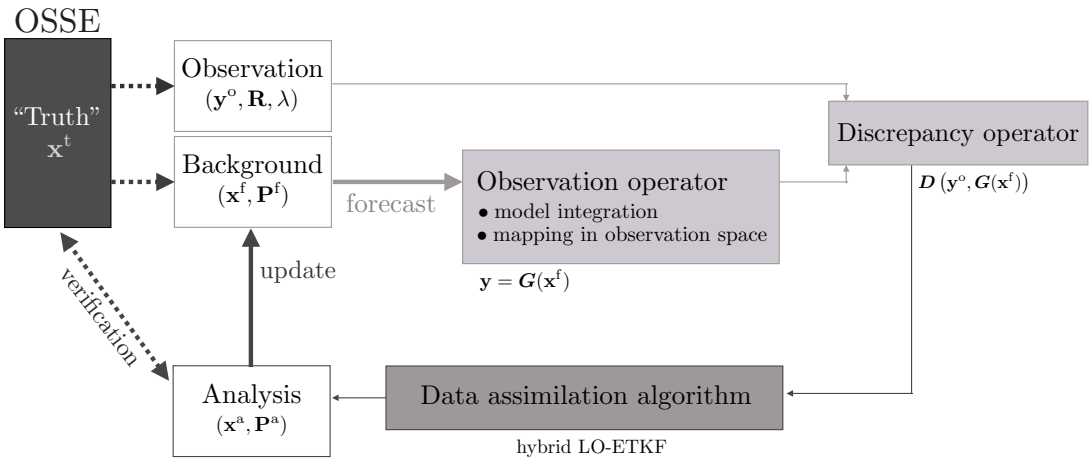

Figure 1: Schematic of the OSSE framework, in which observation data are generated by the forward model using the true values of the control variables $\mathbf{x}^{\mathrm{t}}$. A correction on the background estimates of the control variables $\mathbf{x}^{\mathrm{f}}$ is derived; the analysis estimate $\mathbf{x}^{\mathrm{a}}$ is then used to run the forward model in forecast mode. The two-step forecast-analysis procedure is applied sequentially over time.

validation [50]. In the present work, we use the 2012 RxCADRE S5 experimental dataset [21] already adopted in Ref. [24]. The S5 experiment was a 15-minute-long prescribed fire conducted on a flat terrain characterized by a surface area of approximately 180 m $\times$ 180 m and by a mixed grass and shrub vegetation. The fire was ignited on the North side of the lot and propagated into the southern direction. Fire propagation was recorded through a series of temperature maps starting at time $t = 34$ s after ignition and recorded at 1-Hz frequency using a long-wave thermal infrared imaging system [21].

Since the initial fire only covered a very small area, we use the observed fire at time 60 s as initial condition for fire spread simulations. Hence in the following, time $t_0 = 0$ s corresponds to time 60 s in the RxCADRE dataset. Figure 2a presents the map of flame arrival times, from $t_0 = 0$ to 780 s (corresponding to the time interval [60 s; 840 s] in the RxCADRE dataset), showing the time at which the fireline arrives at a given pixel of the S5 burn lot. To demonstrate the benefits of the HE approach, we assume that observations are available at 60-s time intervals (Ref. [24] shows that when the assimilation time period is shorter than 60 s, the forecast performance of data assimilation is no better than a simple method based on extrapolation of observation data). In the following, we also test the sensitivity of the data assimilation results with respect to changes in the assimilation time period, considering observations at 30-s and 120-s time intervals.

The present object-oriented data assimilation procedure requires two-dimensional instantaneous binary fields as observation $\mathbf{y}^{\mathrm{o}}$. We generate these binary fields at the observation times using the corresponding map of flame arrival times after some filtering to remove small-scale holes and outliers. Figure 2b shows the contour of these binary fields at 60-s time intervals. The shape of these contours (or "firelines") present some irregularities, especially on the flanks, which may not be captured by the fire spread model.

Note that the RxCADRE validation test is challenging for the data-driven model: the

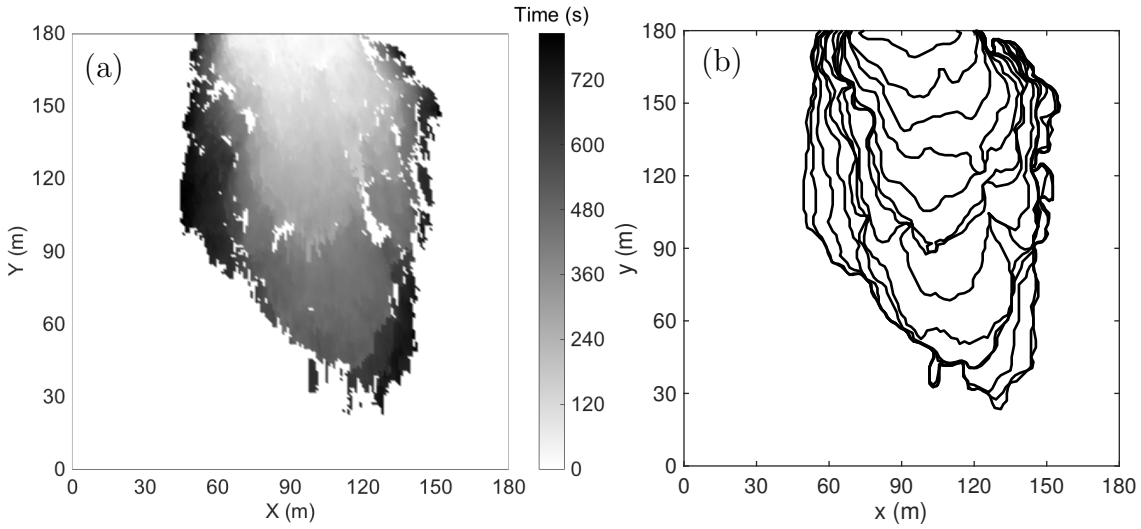

Figure 2: Dataset for the 2012 RxCADRE S5 fire. (a) Map of flame arrival times (0-780 s). (b) Series of firelines over the time period [0; 780 s] at 60-s time intervals.

values of the control variables for which the model predictions agree with the observations may not exist.

5. Results and Discussion

5.1. Verification Test

In this section, we present results from an OSSE verification test based on synthetic observations (see Section 4.1). The computational domain Ω is 180 m $\times$ 180 m (with a computational grid cell size $\Delta x = \Delta y = 1$ m). The duration is 50 s (with a constant time step $\Delta t = 0.1$ s); a single assimilation cycle is considered so that the analysis is performed at time $t = 50$ s. We consider a flat terrain and assume a bimodal biomass moisture distribution: $M_{\rm v}^{(1)} = 5$ % or $M_{\rm v}^{(2)} = 20$ % (see the colormap in Fig. 4). Except for the moisture content, the input parameters of the ROS model are assumed uniform: the biomass fuel is characterized by $\delta_{\rm v} = 0.2$ m (vegetation layer thickness), $m_{\rm v}^{''} = 0.28$ kg m^{-2} (fuel surface loading), $\Sigma_{\rm v} =$ 9,000 m^{-1} (fuel particle surface-to-volume ratio), $\rho_{\rm p} = 512.6$ kg m^{-3} (fuel particle density), and $\Delta h_{\rm v} = 18.6 \times 10^6$ J kg^{-1} (heat of combustion).

5.1.1. Convergence of the Parameter Estimation (PE) Approach

We first consider the standalone ETKF-based PE approach (see Section 3.4). We assume that uncertainties only come from the uniform wind speed and direction so that the control vector is simply defined as $\boldsymbol{\theta} = (u_{\rm w}, d_{\rm w})^T$. The "true" fire is initialized at $(x_{\rm ign}, y_{\rm ign}) = (90\ {\rm m}, 90\ {\rm m})$ as a circular front with a radius $r_{\rm ign} = 10$ m. The true wind speed is $u_{\rm w}^{\rm t} = 3$ m s^{-1}; the true wind direction is $d_{\rm w}^{\rm t} = 180°$ (the wind is blowing into the north direction). The background simulations are initialized at the "true" position $(x_{\rm ign}, y_{\rm ign}) = (90\ {\rm m}, 90\ {\rm m})$ but with values of wind parameters that are randomly-selected

and correspond to the following mean values: $u_{\mathrm{w}}^{\mathrm{f}} = 4\ \mathrm{m\,s^{-1}}$ (25 % error) and $d_{\mathrm{w}}^{\mathrm{f}} = 225°$ (20 % error, the mean background wind is blowing into the North-East direction). The corresponding STD $\boldsymbol{\sigma}^{\mathrm{f}}$ is 0.5 $\mathrm{m\,s^{-1}}$ for the wind speed and 25° for the wind direction. Changes in the assumed wind conditions result in significantly different front positions between the truth and the free run. In addition, heterogeneities in the biomass fuel moisture content result in different front shapes and differences keep increasing in time.

In this test, we analyze the sensitivity of the results to the ETKF ensemble size N_{e}, with N_{e} varying between 10 and 200. Figure 3 presents the mean and STD of the analysis wind speed and direction as a function of N_{e}. The mean wind speed and direction produced by the analysis converge towards the true values provided that the ETKF ensemble size is large enough, $N_{\mathrm{e}} > 30$. In contrast, the STD is insensitive to variations in N_{e}. In the following, we consider $N_{\mathrm{e}} = 40$. Note that in Ref. [19], a higher number of members ($N_{\mathrm{e}} = 200$) was required for spatially-distributed PE due to the larger dimension of the control vector that included wind speed and direction at multiple nodes in the vicinity of the fireline.

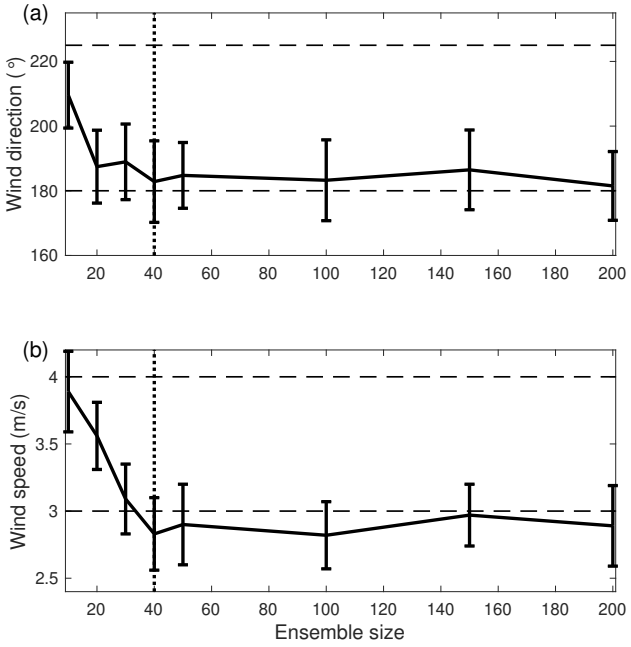

Figure 3: Convergence test of the ETKF-based PE approach for a variable ensemble size N_{e} for (a) the wind direction d_{w} [°], and (b) the wind speed u_{w} [$\mathrm{m\,s^{-1}}$]; OSSE test with $\boldsymbol{\theta} = (u_{\mathrm{w}}, d_{\mathrm{w}})^T$. The plots compare mean values produced by the analysis step (black solid lines) with the mean values produced by the free run (upper horizontal dashes lines) and with the true values (lower horizontal dashes lines). Error bars represent the STD values; the vertical dotted lines represent the ensemble size ($N_{\mathrm{e}} = 40$) adopted in the present study.

Figure 4 presents a representative sample from the ensemble of simulated fire perimeters at the analysis time $t = 50$ s, without and with PE. In the absence of data assimilation, there is a wide scatter in the estimate of the fireline location due to the uncertainties in wind speed and direction. As expected, with the application of data assimilation, all the predictions from the ensemble are steered towards the observed fire location that is representative of the truth. In addition, the ensemble features a much reduced scatter compared to the free run

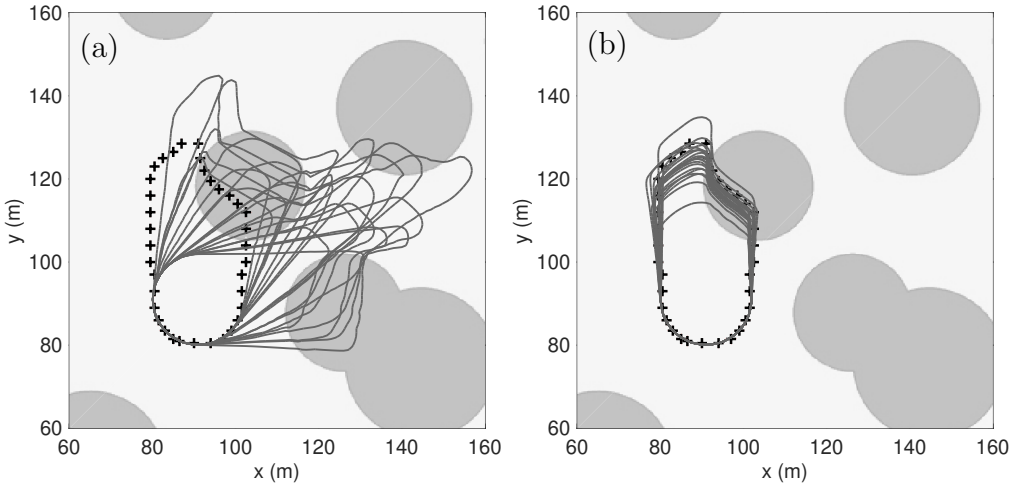

Figure 4: Comparison of fireline locations at time $t = 50$ s; OSSE test with $\boldsymbol{\theta} = (u_{\mathrm{w}}, d_{\mathrm{w}})^T$ and $N_{\mathrm{e}} = 40$ – Representative sample from the ensemble of simulated fire perimeters obtained using (a) the free run (without data assimilation); and (b) the data-driven run (with PE-based data assimilation). Cross symbols correspond to the observed fireline. The colormap visualizes the heterogeneous distribution of the biomass moisture content: light gray areas correspond to $M_{\mathrm{v}}^{(1)} = 5$ %; dark gray areas correspond to $M_{\mathrm{v}}^{(2)} = 20$ %.

ensemble. Note that the ETKF algorithm uses a RTPP inflation scheme with a moderate value, $\alpha = 0.5$ in Eq. (12), providing enough statistical variations to avoid filter divergence issues and accounting for the small variability observed in the headfire predictions.

5.1.2. Merits of Hybrid State-Parameter Estimation (HE)

We now assume that uncertainties are due to both wind parameters and ignition location so that the control variables include $\boldsymbol{\theta} = (u_{\mathrm{w}}, d_{\mathrm{w}})^T$ and the model state $\mathbf{c}$ at a given time. The true fire is initialized as before at $(x_{\mathrm{ign}}, y_{\mathrm{ign}})^{\mathrm{t}} = (90\ \mathrm{m}, 90\ \mathrm{m})$ with $u_{\mathrm{w}}^{\mathrm{t}} = 3\ \mathrm{m\,s^{-1}}$ and $d_{\mathrm{w}}^{\mathrm{t}} = 180°$; the background simulations are initialized at an incorrect location $(x_{\mathrm{ign}}, y_{\mathrm{ign}})^{\mathrm{f}} = (85\ \mathrm{m}, 85\ \mathrm{m})$, with randomly-selected wind parameters characterized by mean values equal to $u_{\mathrm{w}}^{\mathrm{f}} = 4\ \mathrm{m\,s^{-1}}$ and $d_{\mathrm{w}}^{\mathrm{f}} = 225°$ (the wind parameters are the same as in Section 5.1.1).

We compare the HE approach with the standalone PE approach in terms of PDFs of the wind parameters. Figure 5a presents the PDFs of the wind direction obtained after the analysis step in the HE approach; Fig. 5b presents the counterpart for the wind speed. These PDFs are compared to those obtained without data assimilation (corresponding to the background estimate) or with the standalone PE approach; the true wind speed and direction are also indicated as reference. As expected, the background prediction is inaccurate (the mean error in wind direction is about 40°; the mean error in wind speed is 1.25 $\mathrm{m\,s^{-1}}$). Results show that the PE approach provides improved results but is also subject to a bias (the mean error in wind direction is about 25°; the mean error in wind speed is 0.5 $\mathrm{m\,s^{-1}}$). The PE approach is indeed unable to predict the true fireline position because it incorrectly interprets errors in the initial fire location as errors in the wind parameters. The bias in the predicted fireline position due to an incorrect initial fire location at time $t = 0$ s is much reduced in the HE approach. The HE approach accurately retrieves the true parameter

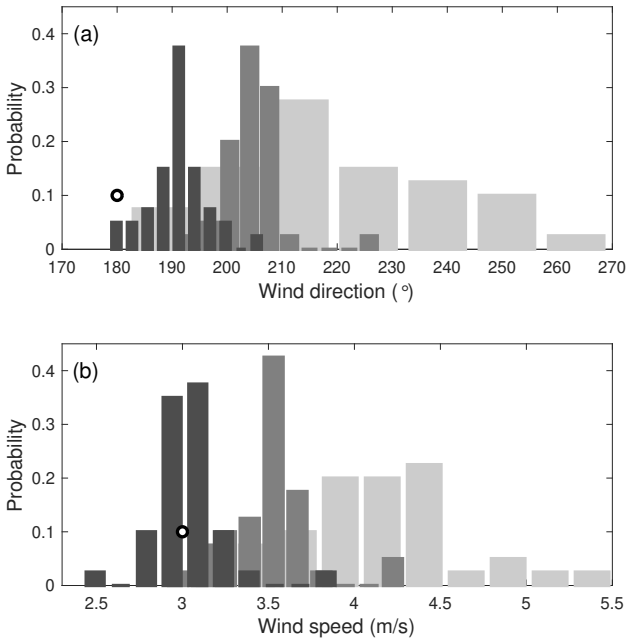

Figure 5: Comparisons of PDFs of the wind parameters; OSSE test with $N_e = 40$. (a) Wind direction d_w [°]. (b) Wind speed u_w [$m\,s^{-1}$]. Circle symbols represent the true (reference) values of the wind parameters. The plots compare PDFs produced without data assimilation corresponding to the background estimate (light gray histogram), by the analysis step of the PE approach (medium gray histogram) and by the analysis step of the HE approach (dark gray histogram).

values (the mean error in wind direction is about 10°; the mean error in wind speed is close to 0). The scatter of the ensemble is also much reduced. Thus, the HE approach is able to overcome uncertainties in both ROS model parameters and initial fire location.

5.2. Validation Test: Application to the RxCADRE Experiment

We now apply the HE approach to the RxCADRE experimental dataset (see Section 4.2). The computational domain Ω is 180 m × 180 m (with a computational grid cell size $\Delta x = \Delta y = 1$ m). The time window is 780 s (with a constant time step $\Delta t = 0.05$ s). During the experiment, the fire was ignited near the northern boundary and propagated from North to South. In the simulations, we initialize the fire at time $t = 0$ s as a semi-circular front located at $(x_{ign}, y_{ign})^f = (90\text{ m}, 180\text{ m})$ with a radius $r_{ign} = 15$ m. We assume uniformly-distributed biomass fuel as well as uniform and constant wind. The input biomass fuel parameters required by Rothermel's model are based on experimental measurements: $\delta_v = 0.2$ m (vegetation layer thickness), $m''_v = 0.28\ kg\,m^{-2}$ (fuel surface loading), $\Sigma_v = 9{,}000\ m^{-1}$ (fuel particle surface-to-volume ratio), $M_v = 10$ % (fuel moisture content). The values of the fuel particle density and heat of combustion are standard values, $\rho_p = 512.6\ kg\,m^{-3}$ and $\Delta h_v = 18.6 \times 10^6\ J\,kg^{-1}$, respectively. For the wind parameters, we start with the following mean background estimates for speed and direction: $u_w^f = 2\ m\,s^{-1}$ and $d_w^f = 360°$ (the wind is blowing into the south direction). The corresponding background STD $\boldsymbol{\sigma}^f$ is $0.5\ m\,s^{-1}$ for the wind speed and 25° for the wind direction. The size of the ETKF-ensemble is $N_e = 40$.

The LO-gain value is $\lambda = 0.2$.

5.2.1. Analysis and Forecast Performance

We first evaluate the analysis performance by comparing observed and predicted firelines at the analysis time $t = 60$ s of the first assimilation cycle. Figure 6(a) presents free run predictions. Figures 6(b)–(c)–(d) present data-driven predictions obtained with data assimilation. Figure 6(b) corresponds to the PE approach; the corresponding mean values for the wind parameters are $u_{\mathrm{w}}^{\mathrm{a}} = 2.4\ \mathrm{m\,s^{-1}}$ and $d_{\mathrm{w}}^{\mathrm{a}} = 335°$; the associated STD values are $0.30\ \mathrm{m\,s^{-1}}$ and 11.6°. Figure 6(c) presents the simulated firelines after the first step of the HE approach corresponding to a SE step. Figure 6(d) presents the simulated firelines after the second step of the HE approach corresponding to a PE step; the corresponding mean values for the wind parameters are $u_{\mathrm{w}}^{\mathrm{a}} = 2.1\ \mathrm{m\,s^{-1}}$ and $d_{\mathrm{w}}^{\mathrm{a}} = 325°$; the associated STD values are $0.30\ \mathrm{m\,s^{-1}}$ and 13.3°. These results show that compared to the free run, the data-driven runs require a significant change in the wind parameters (in particular in the wind direction) in order to match the observed fireline shape and position. The scatter in the wind parameters obtained in the PE and HE approaches is similar and is reduced by a factor of two compared to the free run. The HE approach is seen to provide a better match of the observed fireline than the standalone PE approach at the analysis time.

We now evaluate the forecast performance at future lead times. Figure 7 presents a comparison between observed and predicted firelines in terms of ensemble mean prediction (instead of a sample of individual members as in Fig. 4). Also, the comparison between predictions and observations is not only made at the analysis time, $t = 60$ s, i.e. the time at which observation data are assimilated, but also at forecast times $t = 120$ and 180 s, i.e. at times that benefit from the assimilation performed at $t = 60$ s but also rely on the quality of the forward model and the calibration of wind parameters to maintain a predictive capability. At times $t = 120$ and 180 s, the observations are used for diagnostic purposes but are not assimilated (i.e. the forward model runs free after $t = 60$ s for all approaches). Figure 7 compares results obtained with a PE approach (a), a SE approach (b) and a HE approach (c). In addition, Fig. 7 includes a comparison to a simple extrapolation method based on observations made at times $t = 0$ and 60 s. This observation extrapolation method does not require a physics-based fire spread model: the fireline position is projected at times $t = 120$ and 180 s assuming that the magnitude and direction of the ROS does not change and following a simple methodology introduced in Ref. [24]. Results show that the SE approach fails to capture the direction of the fire spread (between South and South-East). Results also show that both the PE and HE approaches are capable of forecasting the fire spread dynamics. These results suggest that the background wind direction is incorrect and needs to be updated, as done in the PE and HE approaches. The HE approach provides a slightly better forecast of the fireline shape on the flanks. Finally, the observation extrapolation method significantly overestimates the ROS over the entire fire perimeter. These results indicate the importance of using a physics-based forward model to provide realistic constraints to the fire propagation predictions.

Next, in order to quantify the performance of the data assimilation algorithms, we consider the Chan-Vese data fitting functional $\mathcal{J}$ (Eq. 3) to evaluate a posteriori the discrepan-

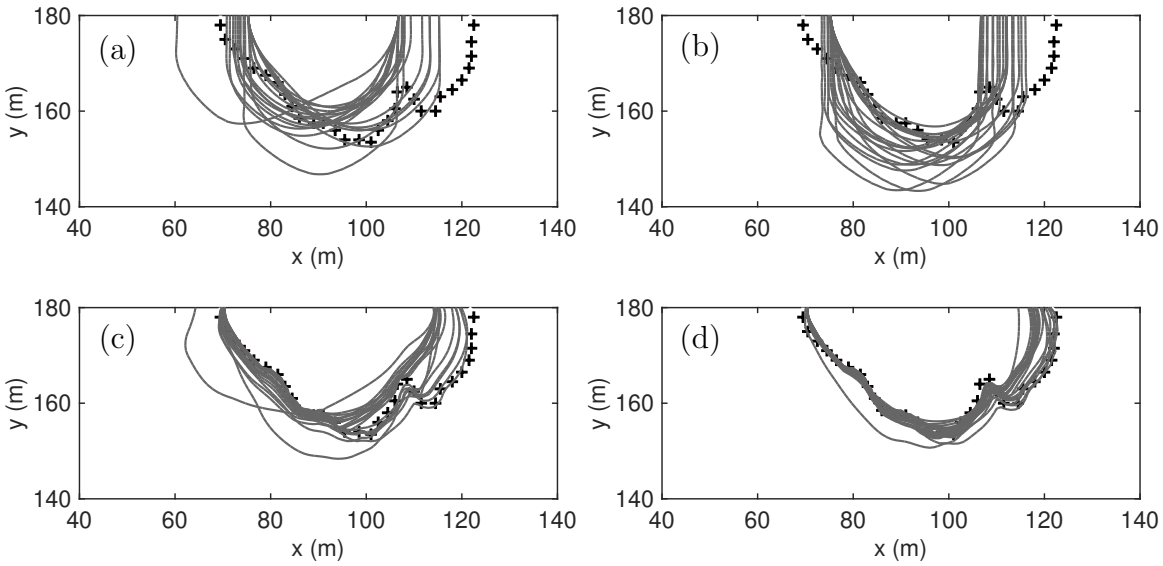

Figure 6: Comparison of fireline locations at time $t = 60$ s; RxCADRE S5 validation test with $N_e = 40$ and $\lambda = 0.2$. Representative sample from the ensemble of simulated fire perimeters obtained using (a) the free run (without data assimilation); (b) the standalone PE approach; (c) after the first (forecast) step in the HE approach corresponding to a SE step; and (d) after the second (analysis) step in the HE approach corresponding to a PE step. Cross symbols correspond to the observed fireline.

cies between observations and model predictions in analysis and forecast modes. Figure 8 presents the temporal variations of the Chan-Vese functional computed with the data presented in Fig. 7. The minimum distance error between observed and predicted firelines is achieved when the Chan-Vese functional is close to 0. Results show that the performance of the free run is worst among all approaches and gets worse over time. All data assimilation approaches provide an improved performance at the analysis time, $t = 60$ s. In the case of the SE approach, the positive impact of data assimilation at analysis time is rapidly lost during the forecast time and results suggest that the SE approach behaves like the free run, but with a time delay. In contrast, the positive impact of data assimilation at analysis time is more persistent in the PE and HE approaches. The HE approach provides the best performance among all approaches at the analysis time $t = 60$ s and at the lead time $t = 120$ s; it provides similar performance as the PE approach at the lead time $t = 180$ s. This suggests that the PE and HE approaches are capable of re-calibrating and finding physical values of unknown or uncertain parameters. Moreover, the HE approach provides a better forecast of the fireline geometry on the flanks and of its irregularities than the PE approach. Note that diagnostics in Fig. 8 are consistent with diagnostics from the literature such as Jaccard's and Sorensen's indices [15, 51] (not shown here).

5.2.2. Sequential Application of the Hybrid State-Parameter Estimation (HE) Approach

We now consider the entire duration of the RxCADRE experiment, from $t = 0$ to $t = 780$ s; we consider that observations are available at 60-s time intervals. The 780-s time is divided into thirteen 60-s-long data assimilation cycles: the cycles are characterized by

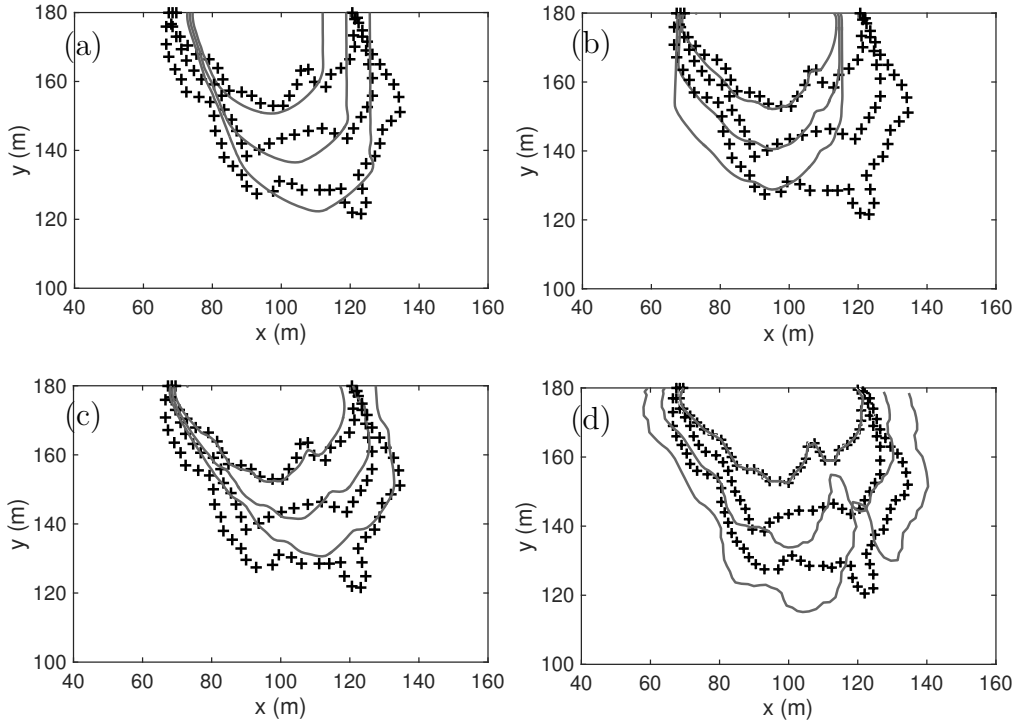

Figure 7: Comparison of fireline locations at analysis time $t = 60$ s and at forecast times $t = 120$ and 180 s; RxCADRE S5 validation test with $N_e = 40$. (a) Standalone PE approach. (b) First SE step of the HE approach. (c) Second PE step of the HE approach. (d) Simple extrapolation method based on observations made at times $t = 0$ and 60 s. Cross symbols correspond to the observed firelines; solid gray lines correspond to ensemble-mean predictions.

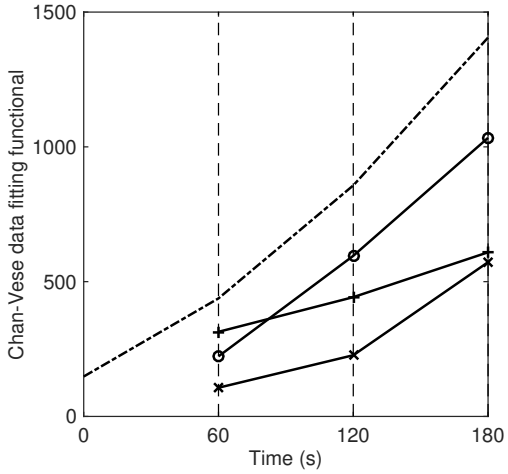

Figure 8: Temporal variations of the Chan-Vese contour fitting functional to quantify the discrepancies between observed and predicted firelines; RxCADRE S5 validation test. The plots compare results obtained with a free run (dashed lines), a SE approach using $\lambda = 0.2$ (solid lines with circle symbols), a PE approach using $N_e = 40$ (solid lines with plus symbols) and a HE approach using $N_e = 40$ and $\lambda = 0.2$ (solid lines with cross symbols). In the data-driven models, data are assimilated at $t = 60$ s; predictions at $t = 120$ and 180 s correspond to a forecast.

twelve analysis events, called A_1–A_{12} with A_i designating an analysis performed at time ($i \times 60$ s) based on an observation made at the same time, and by twelve forecast events, called F_1–F_{12} with F_i corresponding to a forecast performed at time ($(i+1) \times 60$ s) based on the analysis A_i. The lead time of the forecast is therefore fixed and equal to 60 s. The quality of the analysis is evaluated through comparisons between predicted and observed fireline positions at time ($i \times 60$ s) using the Chan-Vese functional as in Fig. 8; the quality of the forecast is evaluated through similar comparisons at time ($(i+1) \times 60$ s).

In Section 5.2.1, we discuss results obtained during the first assimilation cycle A_1/F_1. Here we also consider subsequent assimilation cycles A_2–A_{12}/F_2–F_{12} but limit our discussion to the HE approach. The wind parameters are sequentially updated every 60 s to track temporal changes of the wind. Figure 9 presents the temporal variations of the wind speed and direction produced by the HE approach. The variations are characterized in terms of mean and STD values. In addition, Figures 10–11 present a comparison between observed and predicted firelines in terms of ensemble mean prediction. Comparisons are made at the analysis times A_1–A_{12} (Fig. 10) and at the forecast times F_1–F_{12} (Fig. 11).

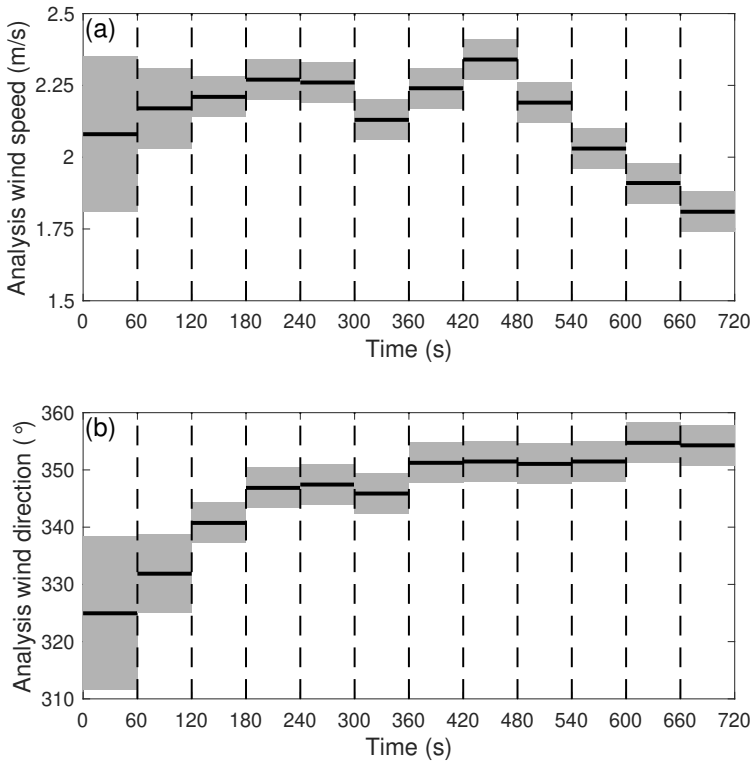

Figure 9: Temporal variations of the corrected wind parameters of the ROS model for all (60-s-long) assimilation cycles A_1–A_{12} using the HE approach; RxCADRE S5 validation test. (a) Wind speed [$\mathrm{m\,s^{-1}}$]. (b) Wind direction [°]. The plots show the mean values (solid lines) and the STD (the vertical thickness of the gray areas is equal to twice the STD).

Results suggest that the mean wind direction significantly changes over the first 180 s of the RxCADRE test and then stabilizes: in Fig. 9, the wind shifts slightly from a northwest

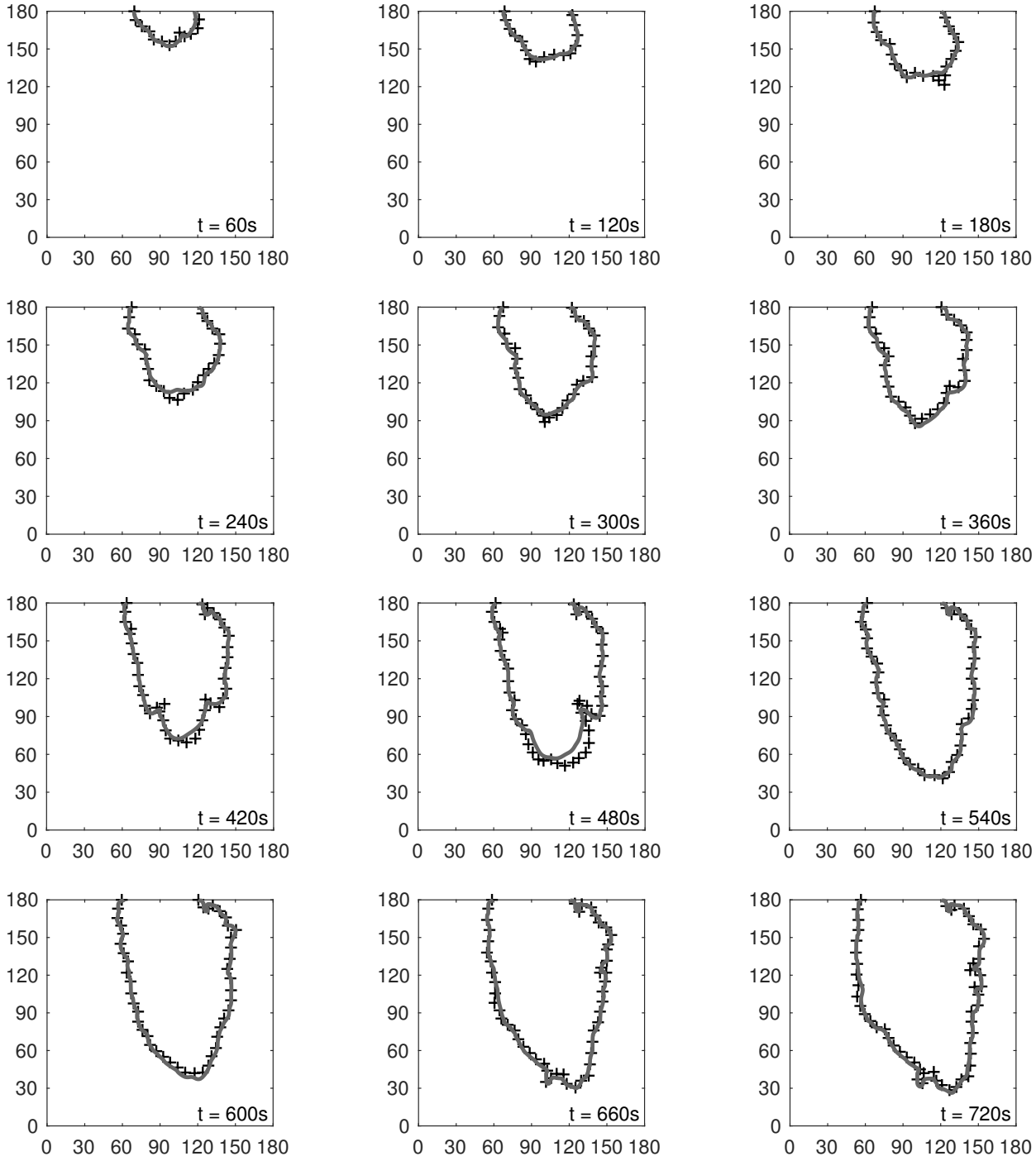

Figure 10: Comparison of predicted (solid lines) and observed (cross symbols) fireline locations at the analysis times A_1-A_{12} ($60 \leq t \leq 720$ s); RxCADRE S5 validation test. Predictions correspond to the ensemble mean obtained with the HE approach using $N_e = 40$ and $\lambda = 0.2$.

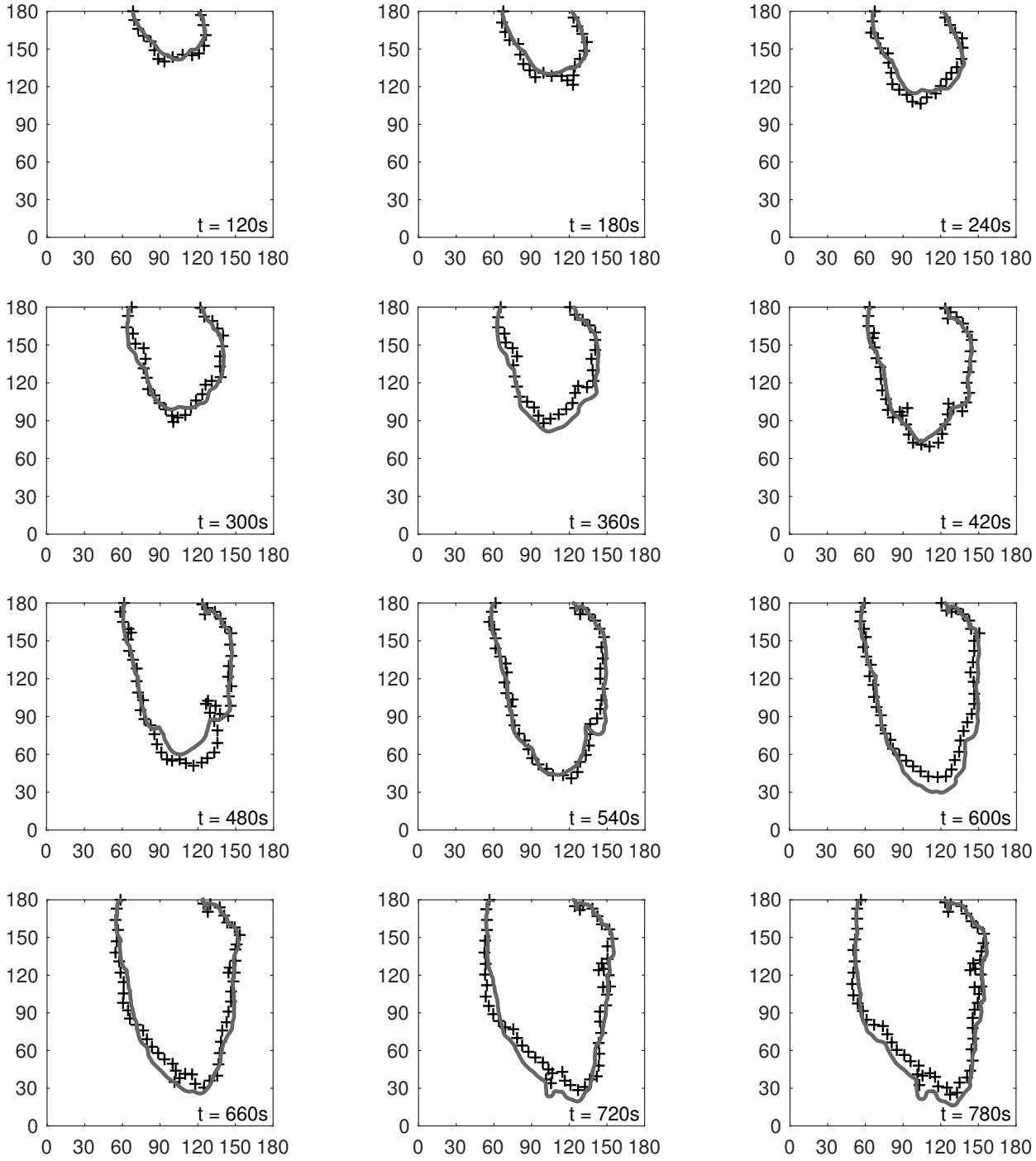

Figure 11: Comparison of predicted (solid lines) and observed (cross symbols) fireline locations at the forecast times F_1-F_{12} ($120 \leq t \leq 780$ s); RxCADRE S5 validation test. Predictions correspond to the ensemble mean obtained with the HE approach using $N_e = 40$ and $\lambda = 0.2$.

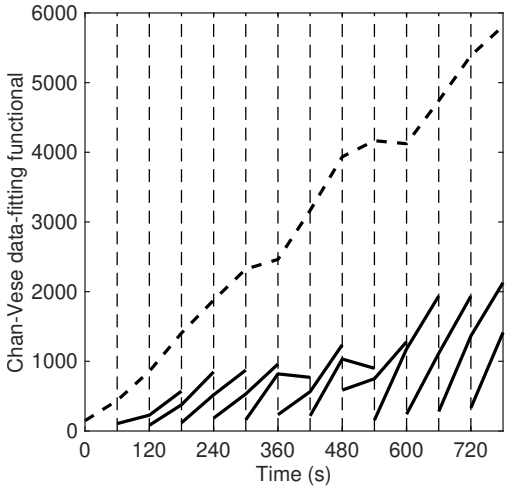

Figure 12: Temporal variations of the Chan-Vese contour fitting functional used to measure the discrepancies between observed and predicted firelines; RxCADRE S5 validation test. The plot compares results obtained with a free run (dashed line) and a HE-based data-driven run (piecewise-linear solid line), where data are assimilated at times A_1-A_{12} and predictions are made at times F_1-F_{12} with a 60-s-long lead time.

wind (the wind is blowing into the South-East direction), $d_{\rm w}^{\rm a} \approx 325°$, to a northerly wind (the wind is blowing into the South direction), $d_{\rm w}^{\rm a} \approx 350°$. This predicted shift is consistent with the observed location of the headfire presented in Figs. 10–11. Results also suggest that the mean wind speed is stable over the first 300 s of the RxCADRE test and then decreases: in particular, in Fig. 9, the wind speed suddenly decreases by 10 % at time $t = 300$ s, due to possible wind stagnancy; the wind speed subsequently relaxes back to a higher value apparently reached at time $t = 420$ s before decreasing again. These predicted variations are consistent with the slightly slower or faster propagation of the predicted and observed fire perimeters presented in Figs. 10–11. For instance, the predicted fireline at time $t = 360$ s in Fig. 11 is seen to propagate too fast in the headfire region compared to observations: this explains the drop in wind speed observed in Fig. 9 during the assimilation cycle $[300; 360$ s$]$; the HE approach responds to the discrepancy detected at time $t = 360$ s by reducing the value of the wind speed and therefore the value of the ROS. Similarly, the predicted fireline at time $t = 420$ s in Fig. 11 is seen to propagate too slowly in the headfire region compared to observations: this explains the increase in wind speed observed in Fig. 9 during the assimilation cycle $[420; 480$ s$]$; the HE approach responds to the discrepancy detected at time $t = 480$ s by increasing the value of the wind speed and therefore the value of the ROS. Thus, the corrections in the wind parameters of the ROS model are designed to provide the best possible match between predicted and observed fireline positions.

Figure 12 presents a quantification of the performance of the HE approach by plotting the temporal variations of the Chan-Vese contour fitting functional. The evolution of the distance error in the free run mode is a smooth, continuously increasing function of time. In contrast, the evolution of the distance error in the data-driven model is a discontinuous piecewise-linear function: deviations of model predictions from observations are periodically reduced during the analysis events A_1-A_{12} due to the benefit of assimilating observations.

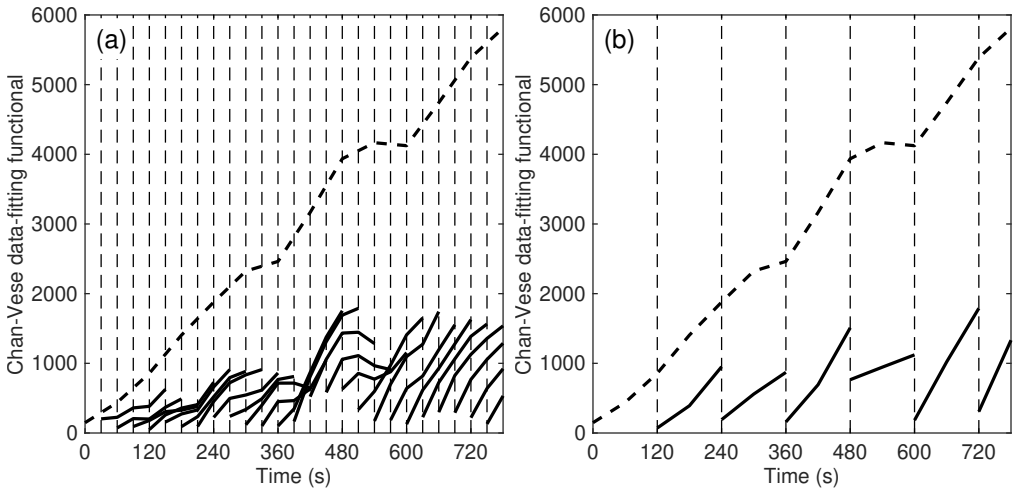

Figure 13: Temporal variations of the Chan-Vese contour fitting functional – RxCADRE S5 validation test. See caption of Fig. 12. In the HE approach, data are assimilated with (a) a 30-s-long lead time; or (b) a 120-s-long lead time.

After each analysis event, when the data-driven model is used in forecast mode, the error increases with time but remains bounded. It is clear from this figure that data assimilation shall be carried out at regular time intervals to avoid an increase in the simulation-observation discrepancies and control the accuracy of the data-driven run. Note that the performance of the HE approach is not sensitive to the assimilation time period (not shown here); the question of the influence of the observation frequency will be better addressed in future work when considering large-scale and long-duration wildfire events.

5.2.3. Sensitivity to the Data Assimilation Frequency

So far, we have assumed that observations are made at 60-s time intervals. We now consider variations in the duration of the assimilation time period, considering observations at 30-s and 120-s time intervals. Note that because of the limited duration of the RxCADRE experiment, it is not possible to study lower observation frequencies. Figure 13 presents the temporal variations of the Chan-Vese contour fitting functional. The results are similar to those already presented in Fig. 12: in particular the values of the distance error do not seem to increase when the observation frequency is decreased, which suggests that for the present test, the performance of the HE approach is not sensitive to the assimilation time period. Figure 14 presents the temporal variations of the wind speed and direction produced by the HE approach. The results are similar to those already presented in Fig. 9: in particular, the progressive shift from a North-West wind to a northerly wind, as well as the wind speed decrease around $t = 300$ s are also observed, which again suggests that for the present test, the performance of the HE approach is not sensitive to the assimilation time period. The question of the influence of the observation frequency will be better addressed in future work, i.e. in applications of the present fire spread modeling capability to large-scale and long-duration wildfire events.

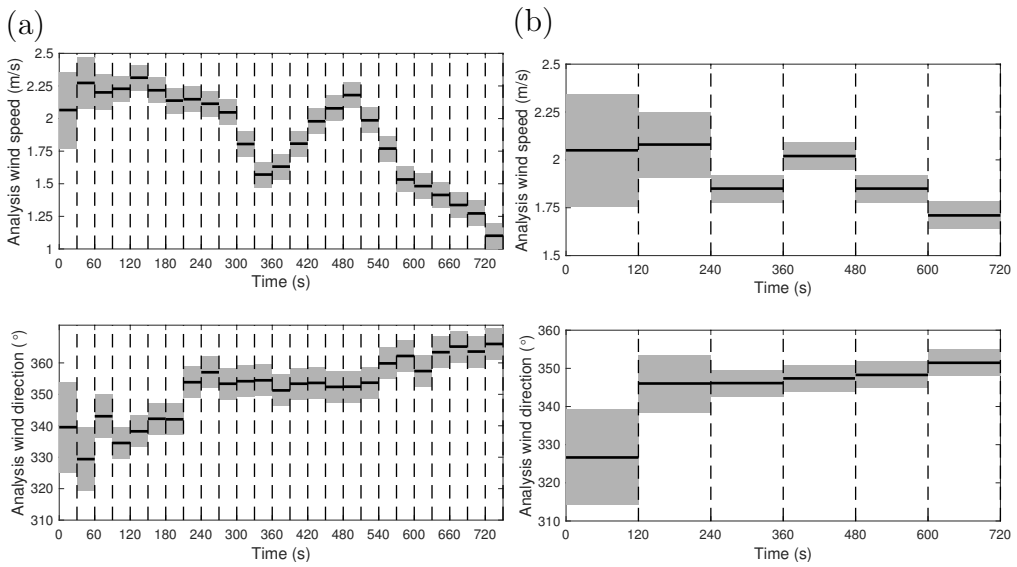

Figure 14: Temporal variations of the corrected wind parameters of the ROS model using (a) a 30-s-long lead time; or (b) a 120-s-long lead time – RxCADRE S5 validation test. See caption of Fig. 9.

6. Conclusions

Data-driven wildland fire spread prediction is still at an early stage of development compared to numerical weather prediction. While the application of classical data assimilation methods to wildland fire spread modeling clearly benefits from past developments made for weather forecasting applications, it is important to recognize that there are yet a number of application-specific technical barriers that need to be overcome. The application of data assimilation to wildland fire problems first requires the formulation of an adequate measure of the discrepancies between model predictions and observations. Regional-scale wildland fires are commonly described as propagating fronts and are cast as a front-tracking problem. Errors in wildland fire spread models therefore correspond to position errors that are addressed here using a front shape similarity measure derived from image segmentation theory. The similarity measure provides a useful diagnostic to evaluate the analysis and forecast performance of the data-driven system. The application of data assimilation to wildland fire problems also requires a suitable choice of the control variables; the control variables are here the near-surface wind parameters and the initial position of the fire front.

Our objective in the present study is to demonstrate the performance of data assimilation in general, and of a hybrid state-parameter estimation approach in particular towards an improved forecast of wildland fire spread in the context of an Eulerian front-tracking solver (similar results could be obtained using a Lagrangian front-tracking solver [24]). The hybrid state-parameter estimation approach relies on an ensemble transform Kalman filter (ETKF) for parameter estimation and a fire propagation equation modified by a Luenberger observer (LO) for state estimation. Parameter estimation is used to update the near-surface wind parameters of the rate-of-spread (ROS) model used in the fire propagation equation; in this sense, it can be regarded as an inverse modeling procedure derived from a Bayesian

estimation problem. State estimation is used to update the fireline position. The updates are driven by observations of the fire perimeter that are assimilated at discrete times. The performance of the hybrid state-parameter estimation approach is evaluated in an observing system simulation experiment (OSSE) verification test and in a validation test corresponding to the 2012 RxCADRE S5 prescribed fire experiment [21]. Results show that data assimilation provides an interesting framework to produce accurate forecast of the fire dynamics, in which parameter estimation is an essential component to reduce bias. A simple method based on extrapolation of observation data fails to provide accurate forecast. With the benefit of data assimilation, a physics-based wildland fire spread model is capable of correcting for uncertainties in both the ROS model parameters and the initial fire location and thereby of providing an accurate forecast. The hybrid state-parameter estimation model provides a general and flexible framework to account for all sources of model uncertainties.

Future work includes extending the application of our data-driven system to the case of large-scale long-duration wildfire events, in which both models and observations are subject to significant error and uncertainty, in which observation data will be available at medium-to-low spatial and temporal resolution, and in which the number of control variables is large (model parameters are no longer treated as spatially uniform and are treated instead as functions of spatial coordinates) and the computational cost of data assimilation is much increased. In particular, the sensitivity of the data assimilation results to the observation frequency, the observation resolution and the choice of the control parameters will be studied to provide guidelines to apply data assimilation to wildfire hazards. From a more theoretical viewpoint, future work also includes extending the object-oriented data assimilation algorithm to a more consistent treatment of the observation error, in particular of the spatial correlations in the observation error in the thermal-infrared imaging data. In practice, this will be useful to reconstruct a complete fire perimeter through data assimilation, even if for instance only a portion of the fire perimeter is actually observed due to the opacity of the thermal plume and/or due to monitoring limitations.

Acknowledgments

CZ and AT gratefully acknowledge support from NSF (Award number 1331615, WIFIRE). The project is also supported by supercomputing resources made available by the University of Maryland (http://hpcc.umd.edu). AC, PM and MR gratefully acknowledge support from SMAI, LEFE/INSU, LabEx AMIES and ANR (ANR-16-CE04-0006, FIRECASTER). The authors also acknowledge the flowchart provided by VerdandInMatlab, a MATLAB version of the Verdandi library developed at INRIA and used in the present study.

References

[1] A. Gelb, Applied optimal estimation, Cambridge MIT Press, Massachusetts, USA, 1974.
[2] A. Tarantola, Inverse problem theory and methods for model parameter estimation, Society for Industrial and Applied Mathematics, 1987.
[3] R. Daley, Atmospheric data analysis, Cambridge University Press, 1991.
[4] G. Evensen, Data assimilation: the ensemble Kalman filter, Springer, Berlin, 2007.

[5] A. Carrassi, M. Bocquet, L. Bertino, G. Evensen, Data assimilation in the geosciences: An overview of methods, issues, and perspectives, Wiley Interdisciplinary Reviews: Climate Change 9 (5) (2018) e535. doi:10.1002/wcc.535.
[6] A. Cowlard, W. Jahn, C. Abecassis-Empis, G. Rein, J. Torero, Sensor assisted fire fighting, Fire Technology 46 (2010) 719–741. doi:10.1007/s10694-008-0069-1.
[7] W. Jahn, G. Rein, J. Torero, Forecasting fire dynamics using inverse computational fluid dynamics and tangent linearisation, Advances in Engineering Software 47 (1) (2012) 114–126.
[8] M. C. Rochoux, B. Delmotte, B. Cuenot, S. Ricci, A. Trouvé, Regional-scale simulations of wildland fire spread informed by real-time flame front observations, Proceedings of the Combustion Institute 34 (2013) 2641–2647. doi:10.1016/j.proci.2012.06.090.
[9] J. Gray, M. Lemke, J. Reiss, C. Paschereit, J. Sesterhenn, J. Moeck, A compact shock-focusing geometry for detonation initiation: Experiments and adjoint-based variational data assimilation, Combustion and Flame 183 (2017) 144–156. doi:10.1016/j.combustflame.2017.03.014.
[10] Y. Chen, C. Snyder, Assimilating vortex position with an ensemble Kalman filter, Monthly Weather Review 135 (2007) 1828–1845. doi:10.1175/MWR3351.1.
[11] J. Beezley, High-dimensional data assimilation and morphing ensemble Kalman filters with applications in wildfire modeling, Ph.D. thesis, University of Colorado, USA (2009).
[12] T. Nehrkorn, B. K. Woods, R. N. Hoffman, T. Auligné, Correcting for position errors in variational data assimilation, Monthly Weather Review 143 (4) (2015) 1368–1381. doi:10.1175/MWR-D-14-00127.1.
[13] M. Denham, K. Wendt, G. Bianchini, A. Cortés, T. Margalef, Dynamic data-driven genetic algorithm for forest fire spread prediction, Journal of Computational Science 3 (2012) 398–404.
[14] J. Mandel, J. D. Beezley, A. K. Kochanski, V. Y. Kondratenko, M. Kim, Assimilation of perimeter data and coupling with fuel moisture in a wildland fire-atmosphere dddas, Procedia Computer Science 9 (2012) 1100–1109. doi:10.1016/j.procs.2012.04.119.
[15] O. Rios, E. Pastor, M. Valero, E. Planas, Short-term fire front spread prediction using inverse modelling and airborne infrared images, International Journal of Wildland Fire 25 (2016) 1033–1047. doi:10.1071/WF16031.
[16] T. Artes, A. Cencerrado, A. Cortes, T. Margalef, D. Rodriguez-Aseretto, T. Petroliagkis, J. San-Miguel-Ayanz, Towards a dynamic data driven wildfire behavior prediction system at european level, Procedia Computer Science 29 (2014) 1216–1226. doi:10.1016/j.procs.2014.05.109.
[17] M. C. Rochoux, S. Ricci, D. Lucor, B. Cuenot, A. Trouvé, Towards predictive data-driven simulations of wildfire spread – part i: Reduced-cost ensemble kalman filter based on a polynomial chaos surrogate model for parameter estimation, Natural Hazards and Earth System Science 14 (11) (2014) 2951–2973. doi:10.5194/nhess-14-2951-2014.
[18] M. C. Rochoux, C. Emery, S. Ricci, B. Cuenot, A. Trouvé, Towards predictive data-driven simulations of wildfire spread – part ii: Ensemble kalman filter for the state estimation of a front-tracking simulator of wildfire spread, Natural Hazards and Earth System Science 15 (8) (2015) 1721–1739. doi:10.5194/nhess-15-1721-2015.
[19] C. Zhang, M. C. Rochoux, W. Tang, M. Gollner, J. B. Filippi, A. Trouvé, Evaluation of a data-driven wildland fire spread forecast model with spatially-distributed parameter estimation in simulations of the fireflux i field-scale experiment, Fire Safety Journal 91 (2017) 758–767. doi:10.1016/j.firesaf.2017.03.057.
[20] M. G. Cruz, M. E. Alexander, A. L. Sullivan, J. S. Gould, M. Kilinc, Assessing improvements in models used to operationally predict wildland fire rate of spread, Environmental Modelling & Software 105 (2018) 54 – 63. doi:10.1016/j.envsoft.2018.03.027.
[21] J. O'Brien, E. Loudermilk, B. Hornsby, A. Hudak, B. Bright, M. Dickinson, J. Hiers, R. Ottmar, High-resolution infrared thermography for capturing wildland fire behaviour: RxCADRE 2012, International Journal of Wildland Fire 25 (1) (2016) 62–75.
[22] R. Paugam, M. Wooster, G. Roberts, 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 (2013) 3385–3399.
[23] M. Valero, O. Rios, E. Pastor, E. Planas, Automated location of active fire perimeters in aerial infrared

imaging using unsupervised edge detectors, International Journal of Wildland Fire 27 (4) (2018) 241–256.
[24] C. Zhang, A. Collin, P. Moireau, A. Trouvé, M. Rochoux, Front shape similarity measure for data-driven simulations of wildland fire spread based on state estimation: Application to the rx-cadre field-scale experiment, Proceedings of the Combustion Institute 37 (3) (2019) 4201–4209. doi:10.1016/j.proci.2018.07.112.
[25] M. Rochoux, A. Collin, C. Zhang, A. Trouvé, D. Lucor, P. Moireau, Front shape similarity measure for shape-oriented sensitivity analysis and data assimilation for Eikonal equation, ESAIM: Proceedings and Surveys 63 (2018) 215–236.
[26] A. Collin, D. Chapelle, P. Moireau, A Luenberger observer for reaction–diffusion models with front position data, Journal of Computational Physics 300 (C) (2015) 288–307.
[27] B. R. Hunt, E. J. Kostelich, I. Szunyogh, Efficient data assimilation for spatiotemporal chaos: A local ensemble transform kalman filter, Physica D: Nonlinear Phenomena 230 (1) (2007) 112–126. doi:10.1016/j.physd.2006.11.008.
[28] B. Bonan, M. Nodet, C. Ritz, V. Peyaud, An etkf approach for initial state and parameter estimation in ice sheet modelling, Nonlinear Processes in Geophysics 21 (2) (2014) 569–582. doi:10.5194/npg-21-569-2014.
[29] R. Rothermel, A mathematical model for predicting fire spread in wildland fuels, Tech. rep., U.S. Department of Agriculture, Forest Service, Intermountain forest and range experiment station, Ogden, Utah, USA (1972).
[30] P. L. Andrews, M. G. Cruz, R. C. Rothermel, Examination of the wind speed limit function in the rothermel surface fire spread model, International Journal of Wildland Fire 22 (2013) 959–969. doi:10.1071/WF12122.
[31] J. Sethian, Level set methods and fast marching methods, Cambridge University Press, 1999.
[32] S. Osher, R. Fedkiw, Level set methods and dynamic implicit surfaces, Vol. 153, Applied Mathematical Sciences - Springer, 2003.
[33] V. Mallet, D. Keyes, F. Fendell, Modeling wildland fire propagation with level set methods, Computers and Mathematics with Applications 57 (7) (2009) 1089–1101. doi:10.1016/J.CAMWA.2008.10.089.
[34] C. Lautenberger, Wildland fire modeling with an eulerian level set method and automated calibration, Fire Safety Journal 62 (2013) 289–298. doi:10.1016/j.firesaf.2013.08.014.
[35] J. Mandel, J. Beezley, A. Kochanski, Coupled atmosphere-wildland fire modeling with WRF 3.3 and SFIRE 2011, Geoscientific Model Development 4 (2011) 591–610. doi:10.5194/gmd-4-591-2011.
[36] R. Rehm, R. McDermott, Fire front propagation using the level set method, Tech. Rep. March, National Institute of Standards and Technology (2009).
[37] M. C. Rochoux, Vers une meilleure prévision de la propagation d'incendies de forêt : évaluation de modèles et assimilation de données (written in english: Towards a more comprehensive monitoring of wildfire spread – contributions of model evaluation and data assimilation strategies), Ph.D. thesis, Ecole Centrale Paris, France (2014).
[38] J. J. Ruiz, M. Pulido, T. Miyoshi, Estimating model parameters with ensemble-based data assimilation: A review, Journal of the Meteorological Society of Japan. Ser. II 91 (2) (2013) 79–99. doi:10.2151/jmsj.2013-201.
[39] D. Mumford, J. Shah, Optimal approximations by piecewise smooth functions and associated variational problems, Comm. Pure Appl. Math. 42 (5) (1989) 577–685.
[40] T. F. Chan, L. A. Vese, A level set algorithm for minimizing the mumford-shah functional in image processing, in: Proceedings of the IEEE Workshop on Variational and Level Set Methods (VLSM'01), VLSM '01, IEEE Computer Society, Washington, DC, USA, 2001, pp. 161–168.
[41] D. Pazo, A. Carrassi, J. M. Lopez, Data assimilation by delay-coordinate nudging, Quarterly Journal of the Royal Meteorological Society 142 (696) (2016) 1290–1299. doi:10.1002/qj.2732.
[42] B. Engquist, A.-K. Tornberg, R. Tsai, Discretization of Dirac delta functions in level set methods, Journal of Computational Physics 207 (1) (2005) 28–51.
[43] J. S. Whitaker, T. M. Hamill, Ensemble data assimilation without perturbed observations, Monthly

Weather Review 130 (7) (2002) 1913–1924.
[44] C. Zhang, Data-driven simulations of wildfire spread at regional scales, Ph.D. thesis, University of Maryland, College Park (2018).
[45] F. Zhang, C. Snyder, J. Sun, Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble kalman filter, Monthly Weather Review 132 (5) (2004) 1238–1253.
[46] E. A. Wan, A. T. Nelson, Dual kalman filtering methods for nonlinear prediction, smoothing, and estimation, in: In Advances in Neural Information Processing Systems 9, 1997.
[47] H. Moradkhani, S. Sorooshian, H. Gupta, P. Houser, Dual state–parameter estimation of hydrological models using ensemble Kalman filter, Advances in Water Resources 28 (2005) 135–147. doi:10.1016/j.advwatres.2004.09.002.
[48] P. Moireau, D. Chapelle, P. L. Tallec, Joint state and parameter estimation for distributed mechanical systems, Computer Methods in Applied Mechanics and Engineering 197 (6–8) (2008) 659–677. doi:10.1016/j.cma.2007.08.021.
[49] P. Moireau, D. Chapelle, P. L. Tallec, Filtering for distributed mechanical systems using position measurements: perspectives in medical imaging, Inverse Problems 25 (3) (2009) 035010.
[50] G. Wells, Capturing fire: Rxcadre takes fire measurements to whole new level, Fire Science Digest.
[51] J. B. Filippi, V. Mallet, B. Nader, Evaluation of forest fire models on a large observation database, Natural Hazards and Earth System Sciences 14 (11) (2014) 3077–3091. doi:10.5194/nhess-14-3077-2014.