Fire Safety Journal 91 (2017) 758–767

Contents lists available at ScienceDirect

Fire Safety Journal

journal homepage: www.elsevier.com/locate/firesaf

IAFSS 12th Symposium 2017

Evaluation of a data-driven wildland fire spread forecast model with spatially-distributed parameter estimation in simulations of the FireFlux I field-scale experiment

Cong Zhang[a], Mélanie Rochoux[b], Wei Tang[a], Michael Gollner[a], Jean-Baptiste Filippi[c], Arnaud Trouvé[a,*]

[a] Department of Fire Protection Engineering, University of Maryland, College Park, MD 20742, USA
[b] CECI, CERFACS – CNRS, 42 Avenue Gaspard Coriolis, 31057 Toulouse Cedex 01, France
[c] SPE, Université de Corse Pasquale Paoli – CNRS, BP 52, F-20250 Corte Cedex, France

ARTICLE INFO

Keywords:
Fire spread
Wildfires
Fire modeling
Data assimilation
Parameter estimation

ABSTRACT

The general objective of this research is to develop a prototype data-driven wildland fire spread simulator, called FIREFLY, using an ensemble-based data assimilation approach with the objective to forecast the location and speed of the fire. The specific focus of the present study is on evaluating the new features of FIREFLY at field scale in a controlled grassland fire experiment known as FireFlux I. FIREFLY features the following components: an Eulerian front-tracking solver that treats the fire as a propagating front and uses Rothermel's model for the rate of spread (ROS); a series of observations of the fire front position (based here on high-resolution fireline data previously generated by validated numerical simulations); and a data assimilation algorithm based on an ensemble Kalman filter configured in a parameter estimation mode to address model bias and uncertainties in the input data to the ROS model. In this work, FIREFLY is modified to allow for an estimation of spatially-distributed surface wind speed and direction. To generate a reliable ensemble and ensure an accurate correction, the ensemble Kalman filter requires sampling truncated probability density functions as well as localizing, *i.e.*, dynamically selecting the areas where the wind parameters are corrected. Results show that the spatialized parameter estimation approach allows for a successful reconstruction of observed fireline position and shape as well as a substantial improvement in the forecast performance compared to the standalone fire spread model. Results also show that the inferred wind parameters may not be accurate and should be viewed as effective values that incorporate multiple sources of uncertainties. Developing a better representation of fire-wind interactions is thus viewed as a key aspect to improve the FIREFLY forecast capability.

1. Introduction

Real-time forecasting of wildland fire propagation is an area that has been attracting growing interest in recent years and that can benefit both fire risk management and fire emergency response [1,2]. Current operational wildland fire spread simulators adopt a regional-scale perspective (*i.e.*, with a resolution from a few tens of square meters up to several square kilometers) and simulate a fire as a propagating front that separates burnt and unburnt vegetation. The local propagation speed of the fire front is called the rate of spread (ROS) and its value is determined by complex interactions between pyrolysis, combustion, heat transfer, near-flame flow dynamics as well as atmospheric flow dynamics and chemistry. Operational simulators use empirical or semi-empirical models to calculate the ROS. For instance, BehavePlus [3] and FARSITE [4] use a popular model due to Rothermel [5]. Rothermel's model treats the ROS as a semi-empirical function of biomass fuel properties, topographical properties and near-surface meteorological conditions. Rothermel's model, like any other ROS model, has a number of limitations. First, the model has a domain of validity that is limited to the experimental conditions used during its original development and calibration, which does not include extreme fire behavior conditions (*i.e.*, high wind conditions and/or steep slopes). Second, the model does not explicitly account for fire-atmosphere interactions. And third, the model relies on input parameters that may not be known or may only be known with limited accuracy [6].

One promising approach to overcome these limitations is data assimilation: data assimilation uses fire sensor data to correct model

* Corresponding author.

http://dx.doi.org/10.1016/j.firesaf.2017.03.057
Received 10 February 2017; Accepted 14 March 2017
Available online 10 April 2017

predictions and thereby increase the forecasting performance and reduce the range of uncertainties. Data assimilation is a well-established approach in the area of numerical weather prediction and while still at an early stage of development, this idea has also been considered over the past decade for possible applications to wildland fire behavior problems [2,7–16].

The present study is a continuation of our previous work on data-driven modeling that has led to the development of a prototype wildland fire spread simulator called FIREFLY [11–14]. FIREFLY features the following main components: a forward model using an Eulerian-based front-tracking solver and a description of the ROS based on Rothermel's formulation; a series of observations of the fire perimeter location; and an inverse model based on an ensemble Kalman filter (EnKF) that allows a stochastic representation of the forward model variability. The data assimilation algorithm adopted in the inverse model features a choice between a parameter estimation approach in which the estimation targets (*i.e.*, the control variables that are corrected by the inverse model) are the input parameters of the ROS model [13] and a state estimation approach in which the estimation targets are the positions of the fire front markers [12,14]. FIREFLY has been previously evaluated in a series of verification tests using synthetic observations (these tests are called Observing System Simulation Experiments or OSSE). The simulator has also been evaluated in a preliminary validation study corresponding to a small-scale (16 m^2) controlled grassland fire experiment in which fire perimeter observations were provided by a thermal infrared camera conveniently located at the zenith location, directly above the grassland plot [11–14].

While the initial focus on a small-scale experiment offered the benefits of controlled quasi-uniform conditions as well as proper instrumentation [17], there is an obvious need to extend the FIREFLY validation effort to the case of field-scale experiments and more realistic hazard scenarios. Note that the extension of FIREFLY to field-scale is far from trivial because this extension requires an increase in the number of unknown parameters to treat heterogeneous conditions (*i.e.*, spatial variations in fuel, topography and wind) which in turn results in new challenges in the EnKF methods as well as an increase in computational cost. The present study is a response to the need for field-scale tests in which the FIREFLY simulator is applied to the case of the FireFlux I controlled fire experiment (referred to as FireFlux in the following) [18]. The FireFlux experiment corresponds to a 63-hectare rectangular plot size, flat terrain, homogeneous tall grass vegetation, and relatively low wind conditions (2 m/s); the ROS is approximately 1 m/s in the wind direction; and the fire duration is approximately 15 min. Because the FireFlux study did not include fireline information with suitable spatial and temporal resolution, we decided to use a computational-based surrogate for experimental observations. The computational-based surrogate corresponds to fire perimeters produced by numerical simulations of the FireFlux experiment previously performed by Filippi et al. [19]: the simulations use a meso-scale atmospheric boundary layer model (Meso-NH) coupled with a wildfire spread simulator (ForeFire); the coupling in ForeFire/Meso-NH is two-way, *i.e.*, the atmospheric environment affects the fire through the wind field, while the fire affects in turn the atmospheric environment through the release of heat and water vapor. The ForeFire/Meso-NH representation of the fire front and near-surface wind are believed to be realistic and are used as reference in the present work.

In the present study, the FIREFLY simulator is configured in a parameter estimation mode and the control variables are the near-surface wind speed and direction. These variables are allowed to change temporally. In addition, because near-surface wind conditions cannot be considered as uniform at the scale of FireFlux, FIREFLY is modified to allow for spatially-varying wind parameters. This modification provides the degrees of freedom that may be required in the forward model simulations in order to match the observed fire behavior. It is also justified because it provides some representation of the expected spatial variations of wind conditions along the fireline.

The paper is organized as follows. We first present the basic FIREFLY modeling strategy (Section 2) and the FireFlux experiment (Section 3). We then proceed to a presentation of the new parameter estimation method intended to address spatial variations of parameters in large domains (Section 4). Finally, in Section 5, we present results in which the performance of FIREFLY is evaluated through its ability to forecast FireFlux "observations" at different lead times. The discussion also includes a comparison between the wind field predicted by FIREFLY and that simulated by ForeFire/Meso-NH.

2. FIREFLY: A data-driven fire behavior model

FIREFLY is a data-driven wildland fire spread model that uses a data assimilation algorithm (the inverse model) to correct a fire propagation model (the forward model) based on estimates of the fire front position (the observations) [13,14]. The current forward model in FIREFLY is an Eulerian front-tracking simulator with two main components: a level-set-based algorithm that solves the fire propagation equation (this equation is solved on the terrain surface); and a model description of the ROS based on Rothermel's formulation [5,20]. The forward model takes the parameters of the Rothermel's model as inputs and gives the simulated time-resolved fireline position as outputs.

One difficulty in using the forward model in FIREFLY is that the Rothermel-based ROS is applied along the entire fire perimeter, not only at the head but also on the flanks and in the rear. This is inconsistent with the original intent of Rothermel's model which was developed to describe fire spread in the wind and/or upslope direction, *i.e.*, in the head fire region [5]. This is also a questionable approach since it is known that Rothermel's model, when applied to the flanks of a fire, tends to severely under-estimate the ROS [21]. We plan to address this issue in future work, for instance by implementing FARSITE [4] into FIREFLY. In the meantime, we deem this limitation acceptable because, as will be seen in the results section, our spatially-distributed parameter estimation approach will provide a way to correct model inaccuracies and match the observed fireline behavior.

The data assimilation algorithm is based on the ensemble Kalman filter (EnKF). We use FIREFLY in parameter estimation mode, for which the estimation targets $\mathbf{x}$ are the input parameters of the ROS model [12]. From an algorithmic viewpoint, the EnKF is sequentially applied over each assimilation window, each sequence corresponding to a two-step procedure (the forecast and analysis steps) triggered by new observations. Fig. 1 shows the EnKF flowchart for two successive assimilation windows [t−1, t] and [t, t+1]. During [t−1, t], starting

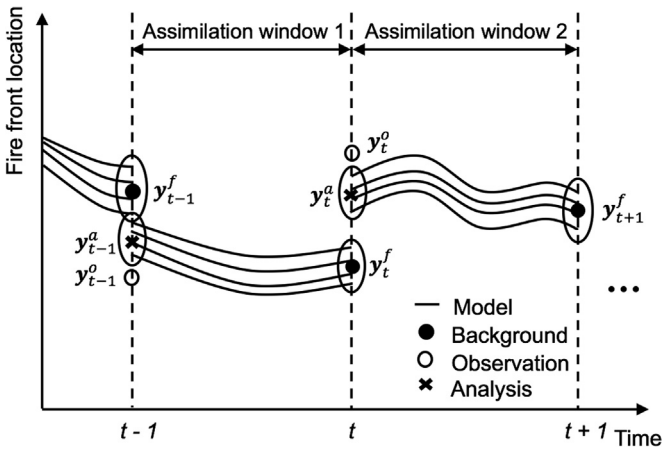

Fig. 1. Flowchart of the EnKF algorithm applied to two successive assimilation windows [t−1, t] and [t, t+1]. For clarity purpose, only 4 model trajectories are represented (instead of a large ensemble).

from an initial guess of the control parameters $\mathbf{x} = \mathbf{x}^f$ (the forecast), the forward model produces a prediction of the fire front position at time t, designated as $\mathbf{y}_t^f = G(\mathbf{x}^f)$, where G is the observation operator (G describes the mapping from the control parameters to the fire front position at the observed time t, including the fire spread model integration over $[t-1, t]$). This prediction is then compared to the observed fire front position at time t, designated as $\mathbf{y}_t^o$. Thus, the difference $(\mathbf{y}_t^o - \mathbf{y}_t^f) = (\mathbf{y}_t^o - G(\mathbf{x}^f))$ provides an estimate of the distance between observations and model predictions (the distance is formulated as an Euclidean distance between the observed markers and their closest neighbors along the simulated fire front). This distance can then be minimized through the EnKF update equation: $\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}(\mathbf{y}_t^o - G(\mathbf{x}^f))$, in which the new estimates of the control parameters $\mathbf{x} = \mathbf{x}^a$ (the analysis) are defined as a correction to the background vector $\mathbf{x}^f$ and in which the magnitude of the correction is controlled by the Kalman gain matrix $\mathbf{K}$ [12–14]. The prediction $\mathbf{y}_t^a = G(\mathbf{x}^a)$ can then be used to initialize the forecast over the next time period $[t, t+1]$.

A central idea in EnKF is to rely on a statistical ensemble of the control parameters generated assuming each of these parameters follows a Gaussian PDF and using a Monte Carlo random sampling method (see Section 4 "EnKF configuration" for the numerical settings of the Monte Carlo method used in the present study), which produces an ensemble of predictions of the fire front position *via* the observation operator. The Kalman gain matrix $\mathbf{K}$ is formulated in a statistical sense using the ensemble of input parameters as well as the ensemble of simulated fire front positions. The size of the ensemble is therefore of primary importance to have a good representation of the forward model variability and to remove noise in the Kalman gain matrix $\mathbf{K}$.

3. The FireFlux I experiment

3.1. Environmental conditions

The FireFlux experiment corresponds to a 30-hectare fire burn [18]. The main biomass fuel was tall grass, which in the modeling, is assumed to be homogeneous. During the experiment, the surface wind blew mainly from North to South (approximately 10° into the South-West direction). The fire was ignited on the North side of the lot and propagated into the southern direction (see Fig. 2 below). Details on the environmental conditions required to set up the ROS model in FIREFLY can be found in Table 1. FireFlux represents a typical case of wind-driven fire propagation over a flat terrain.

3.2. Observations

The FireFlux experiment was originally proposed as a validation experiment for coupled fire-atmosphere models and it has been used by a number of research groups as a benchmark test for Computational Fluid Dynamics (CFD) model validation. One limitation is that the instrumentation was mainly dedicated to the analysis of the smoke plume rather than the fire spread: for instance, the time evolution of the fireline was not tracked. Still, the arrival time of the fire front was recorded at two instrumented towers; this information was used by Filippi et al. [19] to establish the accuracy of ForeFire/Meso-NH simulations when configured in a two-way coupling mode. The two-way coupling mode accounts for the effects of the fire-induced micrometeorology and was shown in [19] to be an important factor in providing good agreement between numerical results and experimental data. As mentioned in the introduction, we use the ForeFire/Meso-NH fire perimeters produced by Filippi et al. [19] as a surrogate for experimental observations. In the ForeFire/Meso-NH simulations, the horizontal resolution is 10 m and the vertical resolution at ground level is 3 m. We consider observations at 100-s time intervals over a period of 800-s. Fig. 2 shows the full series of available observations; we can notice an acceleration of the fire spread between 200 and 400-s due to fire-wind interactions.

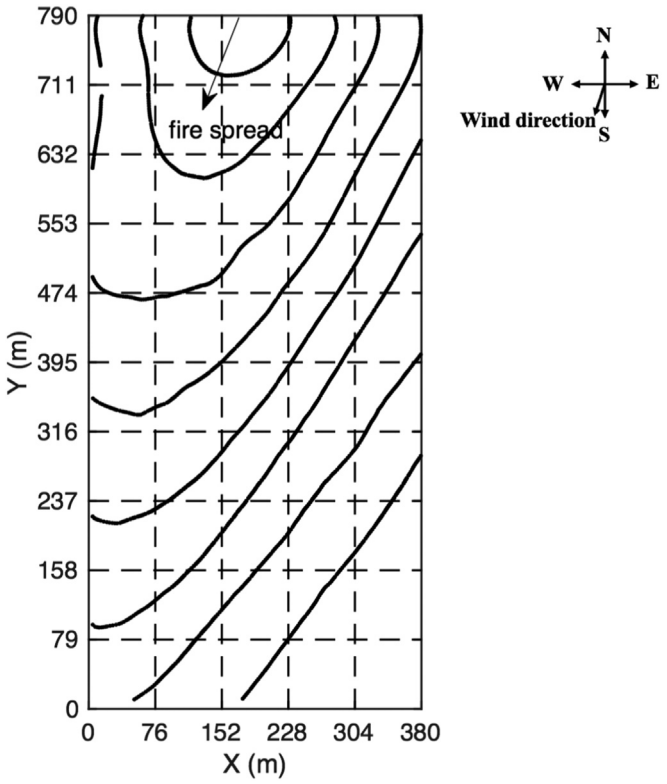

Fig. 2. Available observation data (solid lines) produced by a reference ForeFire/Meso-NH simulation at 100-s time intervals between 100 and 800 s. The fire spreads from top to bottom (North to South). The dashed lines correspond to the wind grid (see Section 4).

Table 1
Input parameters of Rothermel's ROS model corresponding to the FireFlux conditions. These parameters are treated as invariant except for U_w and α_w that are the estimation targets.

Parameter name (unit)	Symbol	Value
Fuel depth (m)	δ_v	1.5
Fuel loading (kg/m^2)	m_v''	1.08
Fuel moisture content (dead fuel)	M_v	0.09
Fuel heat of combustion (J/kg)	Δh_c	1.543e+7
Fuel particle surface-to-volume ratio (1/m)	Σ_v	5000
Fuel particle mass density (kg/m^3)	ρ_P	400
Fuel moisture content at extinction	$M_{v,ext}$	0.3
Wind speed (at mid-flame height) (m/s)	U_w	2.0
Wind direction (at mid-flame height) (°)	α_w	10

3.3. Numerical configuration for the forward model

The front-tracking solver uses a computational domain that covers the size of the grass area burnt during the experiment, and is 380-m long in the x-direction (the West-East direction) and 790-m long in the y-direction (the South-North direction); the total size of the computational domain is therefore 380×790 m^2 or 30 ha. The spatial resolution is 1-m. Initial conditions for the fireline are treated in an approximate way using the observed fireline shortly after ignition time (14-m wide in the x-direction and 4-m deep in the y-direction) and without accounting for the fact that in the experiment, fire ignition was set by two walking fire crews [19]. The total simulation time is 800-s. The temporal resolution of the fire propagation solver is 0.02-s.

4. Spatially-distributed parameter estimation

4.1. Modifications of the control vector

In previous work with FIREFLY configured in parameter estimation mode, the input parameters of the ROS model were allowed to change in time but were treated as uniform in space. The assumption of uniform parameters will obviously break down at field scale in which vegetation, topography and wind are likely to change spatially. In the application of FIREFLY to FireFlux, vegetation and topography parameters are still treated as uniform, but wind parameters (speed and direction) are now treated as functions of space. These wind spatial variations are described on a uniform rectangular Cartesian grid that consists of a certain number of nodes and that is typically much coarser than the computational grid used in the front-tracking solver. In the following, we will present results obtained with a forward model grid featuring 1-m resolution and a wind grid featuring 76-m resolution in the x-direction and 79-m resolution in the y-direction (the dashed lines in Fig. 2 show the lines of the wind grid). Both wind speed and direction are linearly interpolated between the wind grid nodes. The maximum size of the control vector is thus equal to the total number of wind grid nodes (66 nodes) multiplied by two. The control vector is assumed constant in time over an assimilation window ($[t-1, t]$ in Fig. 1) and is only modified when a new observation is available and the algorithm proceeds to the next assimilation cycle ($[t, t+1]$ in Fig. 1). Note that the wind information manipulated in FIREFLY corresponds to near-surface wind defined at mid-flame height (*i.e.*, here at 2-m elevation). Note also that numerical tests with wind grids featuring from 1 to 231 nodes have shown that while the FIREFLY solution is sensitive to the wind grid at resolutions coarser than 100 m, it becomes grid-converged (*i.e.*, independent of the wind grid) at resolutions smaller than 100 m. Our baseline configuration with 66 nodes is grid-converged.

4.2. Modifications of the EnKF ensemble generation

As previously noted in Ref. [13], the EnKF algorithm becomes ill-posed when perturbations in the control parameters do not result in observable changes of the system state. A solution to this problem consists in introducing bounds in the search space for the wind parameters. The EnKF algorithm was therefore modified to use truncated Gaussian probability density functions (PDF) when generating ensemble members. The bounds introduced in the wind direction are due to a lack of sensitivity of the ROS model to this parameter when the wind blows against the direction of fire propagation (see Fig. 3(a) where wind-opposed spread corresponds to wind angles smaller than −90° or greater than +90°). In the present study, the search space for wind direction was restricted to the South-West/South and South/South-East quadrants. Similarly, the bounds introduced in the wind speed are due to a lack of sensitivity of the ROS model to wind speed and direction when ROS takes low values. In the FireFlux experiment, the ROS of the head fire was close to 1 m/s, which corresponds to a wind speed close to 2 m/s (see Fig. 3(b)); the search space for wind speed was therefore restricted to values above 1 m/s.

An ill-posed problem is also found when the EnKF algorithm tries to optimize the values of wind speed and direction at wind grid nodes that are remote from the fire region. To avoid this problem, we adopt a simple localization methodology in which only the wind grid nodes that are sufficiently close to the fireline (*i.e.*, within a user-specified threshold distance) are included in the control vector. Thus, the size of the control vector is not constant and changes dynamically over successive assimilation cycles. In the following, we use a threshold distance approximately equal to twice the wind grid spacing (140 m). With this choice, the total number of wind grid nodes grows from a value of 13 at 100-s to a peak value of 32 at 600-s. Thus, in addition to filtering out noise in the Kalman gain estimation, localization allows a significant reduction in the problem size (starting from a wind grid with 66 nodes, the number of nodes included in the control vector remains below 32).

4.3. Special treatment of the first assimilation cycle

At ignition time, the initial guess of the wind parameters (the first forecast) corresponds to a uniform wind field. This initial guess leads to predictions of the fire front position at the first analysis time (100 s) that are far from the observations and that cannot be totally corrected by the analysis step. To improve the performance of FIREFLY, the EnKF two-step procedure is applied a second time using the analysis of the first iteration as initial guess for the second iteration. This iterative procedure is applied to the first assimilation cycle only and is implemented following the simplified quasi-outer-loop (QOL) algorithm proposed by Yang et al. [22].

4.4. EnKF configuration

With observation data available at 100-s time intervals, each data assimilation cycle is 100-s long. In a typical data assimilation cycle, starting at time t, we first integrate the forward model to the time of the next observation, at time (t+100 s); the integration is performed many times, for all members of the EnKF statistical ensemble, thereby producing an ensemble forecast of the fireline. We then consider the new observation at time (t+100 s) and perform a Kalman filter analysis; this analysis is used to update the wind parameters and we integrate again the forward model from t to (t+100 s), and thereby produce optimized estimates of the fireline. We then repeat the data assimilation cycle and produce new firelines corresponding to a forecast at time (t+200 s). In the results section, we evaluate the EnKF forecast performance at say time (t+200 s) by comparing predictions and observations taken at that same time; in the baseline case, predictions are based on an analysis performed at time (t+100 s) but we also consider predictions based on an analysis performed at an earlier time (*e.g.*, at time t, (t−100 s) or earlier) as well as predictions based on forward model simulations without data assimilation.

The EnKF ensemble has 200 members. The values of the control parameters are selected using (truncated) Gaussian distributions with an assumed standard deviation of 1 m/s for wind speed and 30° for wind direction. These values are selected based on rough estimates of uncertainties as well as with the intent to avoid the classical problem of ensemble collapse [23]. The observation error is assumed small and the quality of the EnKF performance is evaluated by measuring distance from the observed fireline position.

Each simulation is run using 20 processors on a large-scale Linux cluster with typical runs requiring 470 h of CPU time. Note that no attempt has been made to optimize the algorithms and minimize the computational cost. Note also that much of the cost is associated with running the simulations for a relatively large EnKF ensemble of 200 members. For comparison purposes, a run with uniform wind parameters requires 340 h and the overhead associated with solving for spatially-distributed wind parameters is approximately 40%.

5. Results

5.1. Verification test

We first present results from a representative verification test in which synthetic observations are generated using the fire propagation solver in FIREFLY and user-prescribed values of the control parameters (an OSSE test in which the "true" vector $\mathbf{x}^t$ exists and is known). The performance of FIREFLY is in that case simply evaluated by examining its ability to re-construct the "true" fire behavior. In this test, synthetic observations are generated using a spatially-varying (and time-independent) wind field created on a (76 m×79 m) wind grid.

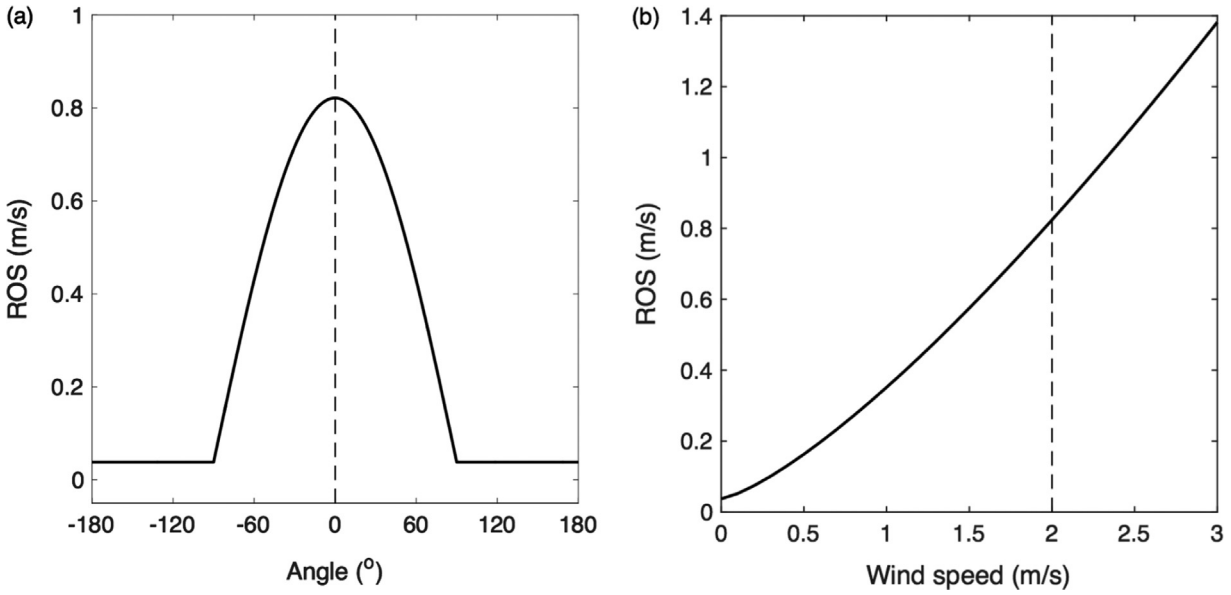

Fig. 3. Sensitivity of FIREFLY's Rothermel-based ROS model to the wind angle (angle between the wind direction and the fire spread direction) (a) and wind speed (b). (a) The wind angle is 0°. (b) The wind speed is 2 m/s. In wind-assisted directions, FIREFLY uses a correction to Rothermel's ROS expression based on the cosine of the wind angle; in wind-opposed directions, FIREFLY uses Rothermel's ROS value without wind (see Ref. [13]).

Note that at initial time, the data-driven FIREFLY simulator assumes a uniform wind field. This test is used to check the general performance of the data assimilation algorithm as well as the performance of the two-iteration procedure applied during the first assimilation cycle.

Fig. 4 presents a comparison, at time t=100 s, between the true fireline position (the observation) and different predictions based on FIREFLY: the mean (*i.e.*, averaged over the EnKF ensemble) fireline position produced without any analysis, called the free forecast; the mean fireline position based on an analysis but produced without the two-iteration procedure, called Analysis-1; and the mean fireline position based on an analysis with the two-iteration procedure, called Analysis-2. The free forecast does not correctly capture the direction of fire spread. In contrast, the Analysis-1 and Analysis-2 predictions correctly capture the South-East direction of the spread. The Analysis-2 prediction is located very close to the true fireline position, thereby demonstrating the benefits of the proposed two-iteration procedure.

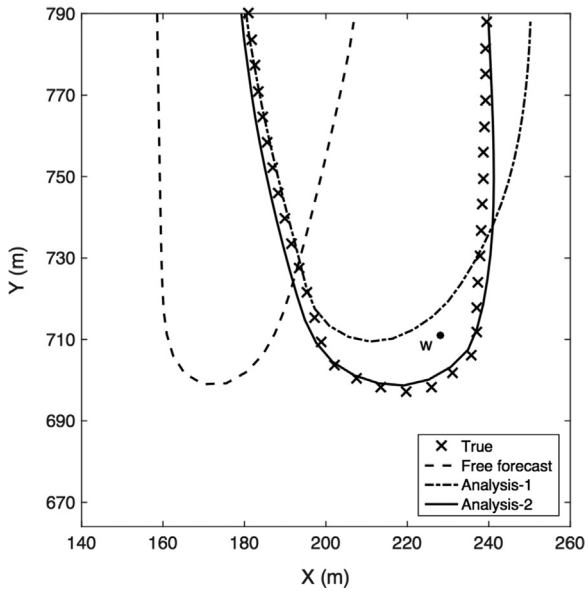

Fig. 4. Verification test with synthetic observations and a prescribed wind field. Comparison of fire front positions at time t=100 s: true fireline (cross symbols); mean free forecast (dashed line); mean analysis with one iteration, Analysis-1 (dash-dot line); mean analysis with two iterations, Analysis-2 (solid line). The location of Point W used for analysis in Fig. 5 is indicated.

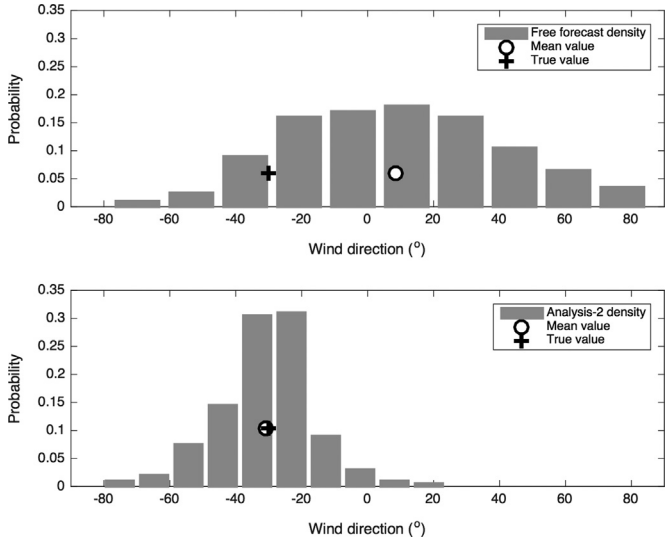

Fig. 5. PDF of the wind direction at Point W over the time window [0; 100 s], associated with the forecast (*i.e.*, without data assimilation - top plot) and with the analysis (*i.e.*, with data assimilation - bottom plot). 0° corresponds to the South direction; positive (negative) values correspond to the South-West (South-East) direction. Histograms are reconstructed from the EnKF ensemble; circle symbols correspond to the ensemble mean and cross symbols correspond to the true value.

The differences in fire perimeters observed in Fig. 4 are due to differences in the predicted wind field. We examine these differences in Fig. 5 in terms of estimates of the wind direction at one particular point located near the head of the fire, (x, y) =(228 m, 711 m) (marked as Point W in Fig. 4). Fig. 5 shows the probability density function (PDF) of angular values of the wind direction at Point W and at time 100 s, before and after the EnKF analysis. In both plots, the mean estimate of the wind direction is indicated and compared to the true value (−30°) used to generate the synthetic observations. The PDF before analysis features a large scatter (standard deviation of 32°) and an incorrect ensemble mean (8.5°); in contrast, the PDF after analysis features a reduced scatter (standard deviation of 14.5°) and an ensemble mean (−31°) close to the true value.

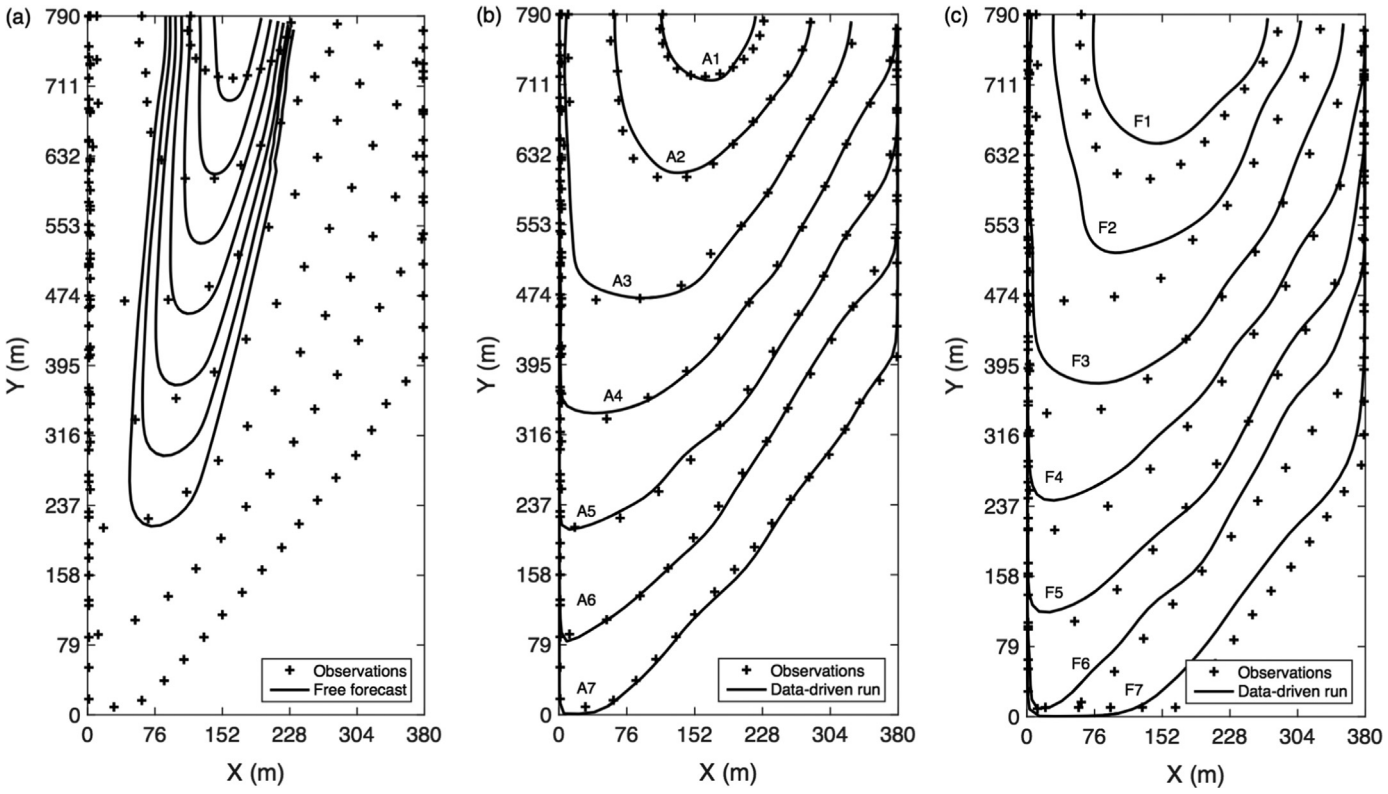

Fig. 6. Comparison of predicted (solid lines) and observed (cross symbols) fire front positions for FireFlux. (a) The predictions correspond to the mean free forecast (100≤ t≤700 s). (b) The predictions correspond to the mean analysis performed at events A1-A7 (100≤ t≤700 s). (c) The predictions correspond to the mean forecast performed at events F1-F7 (200≤ t≤800 s). The fire spreads from top to bottom (North to South). The time interval between plotted perimeters is 100 s.

Overall, these results are satisfactory and demonstrate the ability of FIREFLY to match observed fireline position at analysis time.

5.2. Validation test: Predicted time evolution of the fire perimeter

We now turn to the application of FIREFLY to the FireFlux experiment with observations produced by a reference ForeFire/Meso-NH simulation at 100-s time intervals. The 800-s simulation time is divided into eight 100-s-long data assimilation cycles: the cycles are characterized by seven analysis events, called A1-A7 with Ai designating an analysis performed at time (i×100 s) based on an observation made at the same time, and by seven forecast events, called F1-F7 with Fi corresponding to a forecast performed at time ((i+1) ×100 s) based on the analysis Ai. The quality of the analysis is evaluated through comparisons between predicted and observed fireline positions at time (i×100 s); the quality of the forecast is evaluated through similar comparisons at time ((i+1) ×100 s).

Fig. 6 presents a comparison between model predictions and observations at analysis and forecast times and in terms of the mean fireline position. Fig. 6(a) presents the mean free forecast estimate, *i.e.*, without data assimilation; Fig. 6(b) and (c) present the mean analysis and forecast estimates, respectively, based on data assimilation. The results in Fig. 6(a) illustrate the poor accuracy of the forward model in FIREFLY, primarily due to an incorrect description of the fire spread on the flanks. Also, while the observations indicate that the fire accelerates between 200 and 300-s, the model incorrectly predicts a constant rate of spread in the wind direction.

The results are much improved with data assimilation. The results in Fig. 6(b) evaluate the quality of the Kalman filter: they demonstrate the ability of FIREFLY to match observed fireline positions at analysis time and thereby confirm the general ability of data assimilation to steer an inaccurate model towards observations. The results in Fig. 6(c) evaluate the quality of the forecast (performed 100 s after the EnKF analysis). Due to both model inaccuracies and input parameter uncertainties, the predicted fire fronts are found to deviate from the observations and to under-estimate the observed ROS at the head of the fire. The magnitude of these deviations, however, remains moderate and is much smaller than that observed without data assimilation (Fig. 6(a)). We will show below that the magnitude of the deviations increases in time, and remains acceptable in Fig. 6 because of the availability of observations at relatively short time intervals (100 s). In other words, the quality of the forecast depends on the frequency of the observations and of the corresponding updates of the model.

As mentioned above, the data-driven predictions (forecast and analysis) are based on an EnKF statistical ensemble of 200 members and while we choose in Fig. 6 to demonstrate the quality of these predictions by plotting a single fire perimeter (the mean perimeter), it is also worth considering the entire distribution of 200 perimeters and the associated deviations from the mean. Fig. 7 presents a representative sample of 10 firelines plotted with the observed fireline at analysis time A1 (t=100 s), corresponding to the free forecast in Fig. 7(a) and to the analysis in Fig. 7(b). In Fig. 7(a), the predicted fire perimeters deviate significantly from the observations and feature a large scatter. In contrast, in Fig. 7(b), the analyzed fire perimeters are close to the observations and feature a much reduced scatter. Because of this reduced scatter, we adopt in the following the viewpoint of Fig. 6 and use the mean fire perimeter to quantify the performance of FIREFLY.

Fig. 8(a) shows the time variations of the FIREFLY error defined as the average distance between the predicted and observed fireline positions. This average distance is evaluated through an approximate treatment in which at any given time, the predicted and observed fronts are first discretized by the same number of markers, called P- and O-markers, the P- and O-markers are then paired together, and the FIREFLY error is finally calculated as the average distance between paired P- and O-markers. Fig. 8(a) compares the evolution of the distance error in the free forecast and data-driven modes. The

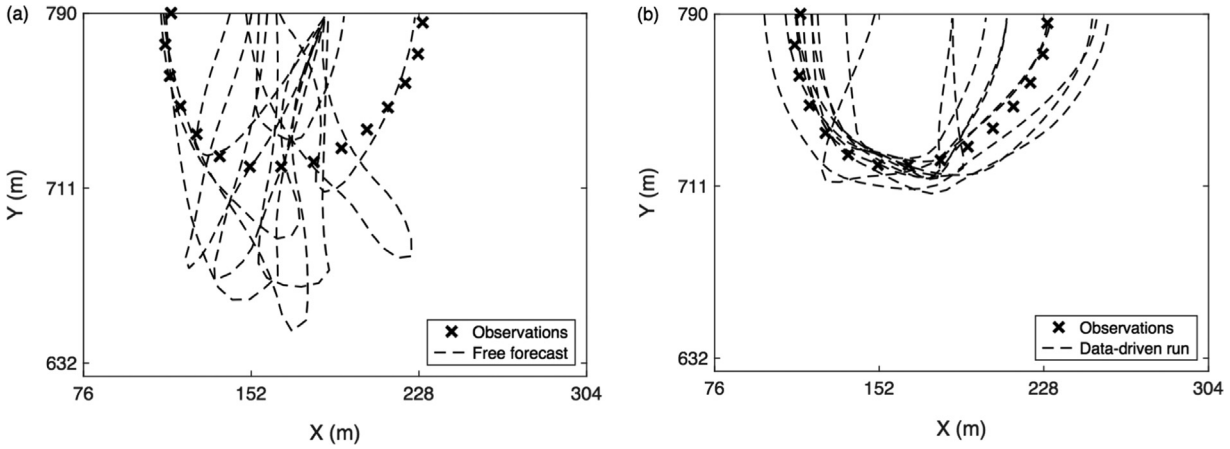

Fig. 7. Comparison of predicted (dashed lines) and observed (symbols) fire front positions at t=100 s. The predictions correspond to a representative sample of 10 EnKF ensemble members (corresponding to different possible values of the wind parameters) for: (a) a free forecast; and (b) an analysis.

evolution of the error in the free forecast mode is a smooth, continuously increasing function of time. Consistent with the observed fire propagation acceleration between 200 and 300-s (Fig. 6), the error grows at a faster rate (twice faster) during that period. This error takes very large values (more than 150 m at time t=600 s) primarily due to the incorrect description of the fire spread on the flanks. In contrast, the evolution of the error in the data-driven mode is a discontinuous function: deviations of model predictions from observations are periodically reduced (to less than 10 m) during the analysis events A1-A7. After each analysis event, the error increases but remains bounded and takes small-to-moderate values: the error features a peak value of 40 m at time t=300 s (due to changes in the fire dynamics) and decreases to less than 10 m (after 600 s, the fire can be considered as a flank fire and is therefore easier to track).

Fig. 8(b) presents a slightly different perspective that illustrates the effect of the assimilation frequency by comparing the evolution of the distance error between the following variations: the free forecast; a first (second; third) data-driven curve, called DD1 (DD2; DD3), with analysis at time t=100 s (t=100 and 200-s; t=100, 200 and 300-s) and a free forecast thereafter; and a fourth data-driven curve, called DDB, corresponding to our baseline case with model updates at all analysis events A1-A4. The growth of the forecast error is slower for DD3 and DDB than for DD1 and DD2, confirming that the fire intensity decreases after its peak around 300-s and that the forecast is becoming more reliable at late times. At time t=500 s, the distance error of the free forecast is approximately 140 m; the error decreases to 115 m, 75 m, 30 m and 10 m, in cases DD1, DD2, DD3 and DDB, respectively.

The information in Fig. 8(b) is particularly valuable for the design of a data-driven model. Let us assume for instance that the design objective is to predict the fire location with a 50-m accuracy: the results in Fig. 8(b) suggest that the model requires data assimilation and that the assimilation frequency should correspond to observations made at least every 200-s. Lower assimilation frequencies will result in less accurate predictions of the fireline location. We conclude here that a 50-m accuracy on the fireline location requires relatively frequent observations made (approximately) every 3 min. Note that this conclusion is specific to the FireFlux experiment: it will be further examined in the future for other burns. Note also that in addition to

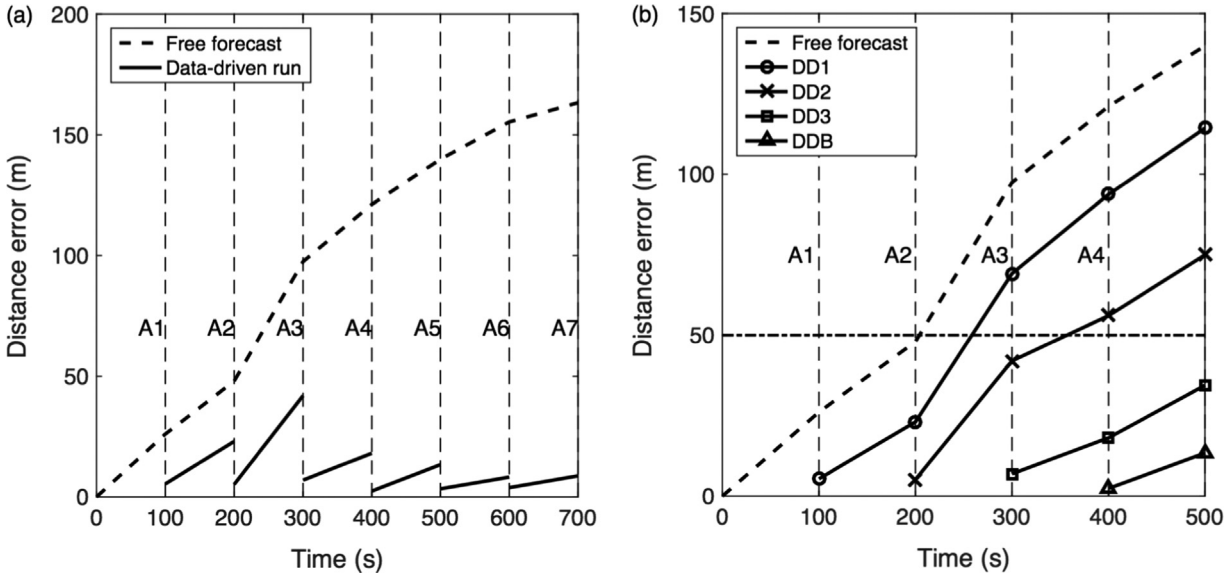

Fig. 8. Time variations of the mean distance between the predicted and observed fire front positions for FireFlux. (a) Comparison between the free forecast (dashed line) and baseline data-driven run (solid line), 0≤ t≤700 s (b) Comparison between the free forecast (dashed line) and different variations of the data-driven run, 0≤ t≤500 s: DD1 (circles); DD2 (crosses); DD3 (squares); DDB (triangles). The dash-dotted horizontal line at 50 m indicates a possible target value for the distance error.

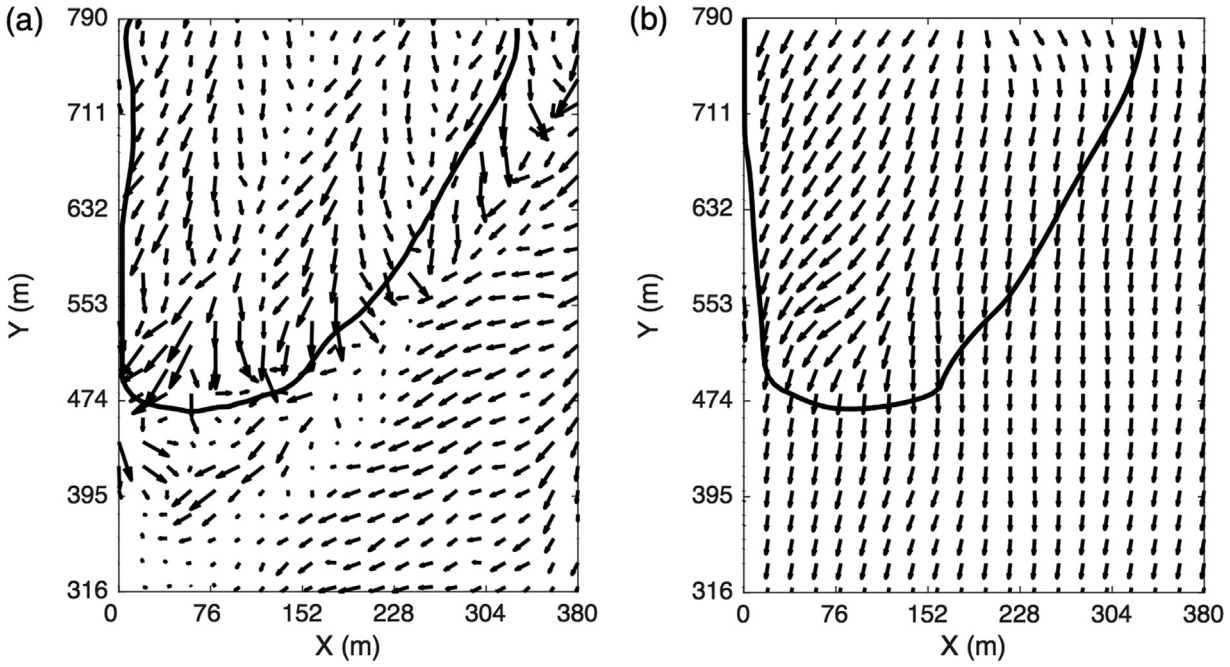

Fig. 9. Instantaneous spatial variations of the horizontal wind velocity vector at 2-m elevation and at time t=300 s. Comparison between: (a) "observations" (reference ForeFire/Meso-NH simulation); (b) predictions produced by the baseline data-driven FIREFLY simulation (ensemble-averaged field). The solid line indicates the fire front location.

being decreased by increasing the assimilation frequency, the forecast error can also be decreased by providing a more accurate fire spread model.

5.3. Validation test: Analysis of the wind field

While the previous results demonstrate that the FIREFLY manipulation of the wind field allows for a successful reconstruction of the observed fireline positions, there is an open question about the exact meaning of the inferred wind parameters and whether these parameters can be used to provide information on the near-flame wind dynamics. To answer this question, we present in Fig. 9(a) comparison between the "true" values of the wind field produced by the reference ForeFire/Meso-NH simulation and the predicted values produced by the baseline data-driven FIREFLY simulation. The comparison is made through instantaneous snapshots of the spatial variations of horizontal wind velocity, at 2-m elevation and at time t=300-s. In Fig. 9, the near-surface ForeFire/Meso-NH data, initially estimated at 3-m elevation, are linearly interpolated to 2-m elevation; in addition, the FIREFLY data, initially estimated at 1-m horizontal resolution, are plotted on the 10-m resolution ForeFire/Meso-NH grid. The "true" and predicted wind fields are found to be quite different, especially near the fire front: the wind field in the ForeFire/Meso-NH simulation features turbulent-like small-scale variations (Fig. 9(a)) that are absent in the FIREFLY simulation (Fig. 9(b)). The presence of small-scale variations in Fig. 9(a) may be explained by the fire-atmosphere interactions that are captured in ForeFire/Meso-NH when configured in two-way coupling mode. FIREFLY cannot represent those small-scale variations, partly due to the coarse resolution of the wind grid and partly due to the limited modeling components (there is no description of the atmospheric boundary layer flow and of the fire plume in FIREFLY).

The discrepancies between ForeFire/Meso-NH and FIREFLY results are further characterized in Fig. 10 through scatter plots of the wind magnitude and direction: a perfect correlation between ForeFire/Meso-NH and FIREFLY results would correspond in these plots to a cluster of points near the 45° line; by the same token, large deviations from this line indicate a lack of correlation between the "true" and predicted values of the wind parameters. Note that the points near the 2 m/s vertical line in Fig. 10(a) and near the 10° vertical line in Fig. 10(b) correspond to data at inactive wind nodes (*i.e.*, points located at a distance greater than the threshold distance from the fireline). The results in Fig. 10 suggest that the inferred wind parameters in FIREFLY are not accurate and should be interpreted as effective values that incorporate multiple sources of uncertainties that are not identified here. More realistic values of the wind parameters may be obtained if the control vector **x** is extended to include additional parameters (*e.g.*, vegetation parameters) and/or if the accuracy of the ROS model is increased (in particular to better describe the fire spread on the flanks).

6. Conclusions

This study presents an application of the FIREFLY data-driven wildland fire spread simulator to the FireFlux I field-scale controlled grassland fire experiment. The FIREFLY simulator features a front-tracking solver based on a ROS formulation (a forward model) and a data assimilation algorithm based on an ensemble Kalman filter (an inverse model). Observations correspond to fire perimeters. Because the FireFlux I experiment did not include fireline information, we use CFD-based fire perimeters produced by Filippi et al. [19] as a surrogate for experimental observations. Compared to our previous work [11–14], the present study features an extension of FIREFLY to field-scale problems through a new spatially-distributed parameter estimation method. In the present application of FIREFLY to FireFlux I (a 30-hectare problem), the control parameters are the surface wind parameters (speed and direction) of the ROS model; they are treated as functions of space (and time) on a wind grid that is much coarser (100-m resolution) than the forward model computational grid (1-m resolution).

The results demonstrate the ability of FIREFLY to steer an inaccurate fire spread model towards observed fireline positions and to provide an improved forecast of the fire behavior compared to the standalone fire spread model. The quality of the forecast depends on the frequency of the observations. In the present study, the performance of the FIREFLY forecast is quantified for the first time: it is found that a 50-m accuracy on the fireline location requires a frequency of observation of 3 min or better. While this figure of merit is not general and is specific to the FireFlux experiment, it provides a valuable

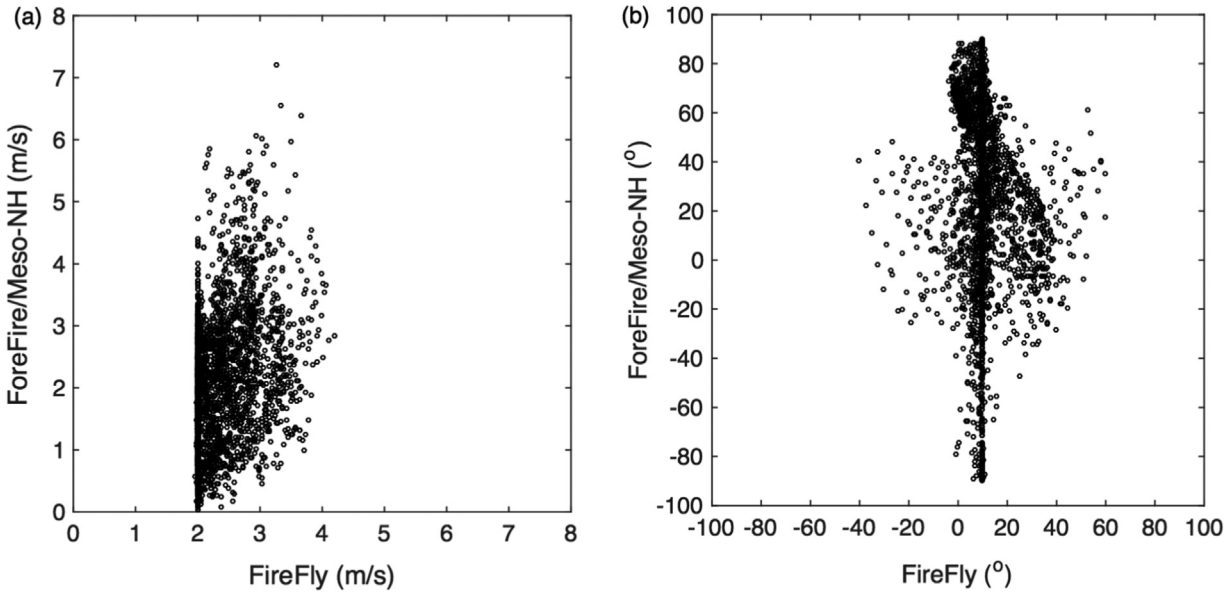

Fig. 10. Scatter plots comparing FIREFLY (*x*-axis) and ForeFire/Meso-NH (*y*-axis) results for: (a) wind magnitude; (b) wind direction. The points correspond to the wind data presented in Fig. 9.

benchmark for future studies. We note in particular that the forecast error can be decreased by providing a more accurate fire spread model: the present Rothermel-based ROS model is inaccurate on the flanks and in the rear of the fire and therefore negatively impacts the FIREFLY performance.

The results also show that the inferred wind parameters should be viewed as effective values that incorporate multiple sources of uncertainties that are not identified here. While the objective of the parameter estimation method is not to provide a detailed representation of fire-wind interactions, these interactions play a role in the FIREFLY forecast capability and therefore should be captured with some level of accuracy. We plan to address this issue in future work by both extending the control vector to include additional parameters (*e.g.*, vegetation parameters) and by providing a more accurate fire spread model.

In the present work, FIREFLY relies on the formulation of an Euclidean distance between the observations and the simulated fire fronts, which is possible by discretizing the fronts with a set of markers that are paired two-by-two. While efficient in all test cases carried out until now, we plan to evaluate the robustness of this distance metrics when applying the EnKF methodology to large-scale wildfires, where the active burning areas may feature complex front topology.

Acknowledgements

This research was funded by the National Science Foundation (Award number 1331615) through the WIFIRE project (http://wifire.ucsd.edu). The authors would like to gratefully acknowledge Dr. Craig Clements from San José State University (USA) for sharing FireFlux I experiment data, as well as Thierry Morel, Anthony Thévenin and Dr. Florent Duchaine from CERFACS (France) for support on OpenPALM.

References

[1] E.K. Noonan-Wright, T.S. Opperman, M.A. Finney, G.T. Zimmerman, R.C. Seli, L.M. Elenz, D.E. Calkin, J.R. Fiedler, Developing the US Wildland Fire Decision Support System, J. Combust. 2011 (2011) 1–14. http://dx.doi.org/10.1155/2011/168473.

[2] T. Artés, A. Cencerrado, A. Cortés, T. Margalef, D. Rodríguez-Aseretto, T. Petroliagkis, J. San-Miguel-Ayanz, Towards a dynamic data driven wildfire behavior prediction system at European level, Procedia Comput. Sci. 29 (2014) 1216–1226. http://dx.doi.org/10.1016/j.procs.2014.05.109.

[3] P.L. Andrews, BEHAVE: Fire Behavior Prediction and Fuel Modeling System-burn Subsystem, Part 1, USDA Forest Service, Intermountain Forest and Range Experiment Station, General Technical Report INT-194, Ogden, UT, 1986, p. 130.

[4] M.A. Finney, FARSITE: Fire Area Development and Evaluation, USDA Forest Service, Rocky Mountain Research Station, Research Paper RMRS-RP-4, Ogden, UT, 2004, p. 47.

[5] R.C. Rothermel, A Mathematical Model for Predicting Fire Spread in Wildland Fuels, USDA Forest Service, Intermountain Forest and Range Experiment Station, Research Paper INT-115, Ogden, UT, 1972, p. 40.

[6] M.A. Finney, J.D. Cohen, S.S. McAllister, W.M. Jolly, On the need for a theory of wildland fire spread, Int. J. Wildland Fire 22 (2013) 25–36. http://dx.doi.org/10.1071/WF11117.

[7] J. Mandel, L.S. Bennethum, J.D. Beezley, J.L. Coen, C.C. Douglas, M. Kim, A. Vodacek, A wildland fire model with data assimilation, Math. Comput. Simul. 79 (2008) 584–606. http://dx.doi.org/10.1016/j.matcom.2008.03.015.

[8] J. Mandel, J.D. Beezley, A.K. Kochanski, Coupled atmosphere-wildland fire modeling with WRF 3.3 and SFIRE 2011, Geosci. Model Dev. 4 (2011) 591–610. http://dx.doi.org/10.5194/gmd-4-591-2011.

[9] M. Denham, K. Wendt, G. Bianchini, A. Cortés, T. Margalef, Dynamic data-driven genetic algorithm for forest fire spread prediction, J. Comput. Sci. 3 (2012) 398–404. http://dx.doi.org/10.1016/j.jocs.2012.06.002.

[10] T. Artés, A. Cencerrado, A. Cortés, T. Margalef, Real-time genetic spatial optimization to improve forest fire spread forecasting in high-performance computing environments, Int. J. Geogr. Inf. Sci. 30 (2016) 594–611. http://dx.doi.org/10.1080/13658816.2015.1085052.

[11] 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, Proc. Combust. Inst. 34 (2013) 2641–2647. http://dx.doi.org/10.1016/j.proci.2012.06.090.

[12] M.C. Rochoux, C. Emery, S. Ricci, B. Cuenot, A. Trouvé, Towards predictive simulation of wildfire spread at regional-scale using ensemble-based data assimilation to correct the fire front position, Fire Saf. Sci. 11 (2014) 1443–1456. http://dx.doi.org/10.3801/iafss.fss.11-1443.

[13] 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, Nat. Hazards Earth Syst. Sci. 14 (2014) 2951–2973. http://dx.doi.org/10.5194/nhess-14-2951-2014.

[14] 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, Nat. Hazards Earth Syst. Sci. 15 (2015) 1721–1739. http://dx.doi.org/10.5194/nhess-15-1721-2015.

[15] O. Rios, E. Pastor, M.M. Valero, E. Planas, Short-term fire front spread prediction using inverse modelling and airborne infrared images, Int. J. Wildland Fire 25 (2016) 1033–1047. http://dx.doi.org/10.1071/WF16031.

[16] M.L. Tardivo, P. Caymes-Scutari, G. Bianchini, M. Méndez-Garabetti, A. Cortés, Three evolutionary statistical parallel methods for uncertainty reduction in wildland fire prediction, in: Proceedings of the International Conference on High Performance Computing & Simulation (HPCS), Innsbruck, 2016, pp. 721–728. doi: 10.1109/HPCSim.2016.7568406.

[17] R. Paugam, M.J. Wooster, G. Roberts, Use of handheld thermal imager data for airborne mapping of fire radiative power and energy and flame front rate of spread, IEEE Trans. Geosci. Remote Sens. 51 (2013) 3385–3399. http://dx.doi.org/10.1109/TGRS.2012.2220368.

[18] C.B. Clements, S. Zhong, S. Goodrick, J. Li, B.E. Potter, X. Bian, W.E. Heilman,

J.J. Charney, R. Perna, M. Jang, D. Lee, M. Patel, S. Street, G. Aumann, Observing the dynamics of wildland grass fires: Fireflux – a field validation experiment, Bull. Am. Meteorol. Soc. 88 (2007) 1369–1382. http://dx.doi.org/10.1175/BAMS-88-9-1369.

[19] J.B. Filippi, X. Pialat, C.B. Clements, Assessment of ForeFire/Meso-NH for wildland fire/atmosphere coupled simulation of the FireFlux experiment, Proc. Combust. Inst. 34 (2013) 2633–2640. http://dx.doi.org/10.1016/j.proci.2012.07.022.

[20] P.L. Andrews, M.G. Cruz, R.C. Rothermel, Examination of the wind speed limit function in the Rothermel surface fire spread model, Int. J. Wildland Fire 22 (2013) 959–969. http://dx.doi.org/10.1071/WF12122.

[21] P. Cunningham, R.R. Linn, Numerical simulations of grass fires using a coupled atmosphere-fire model: dynamics of fire spread, J. Geophys. Res.: Atmos. 112 (2007) 1–19. http://dx.doi.org/10.1029/2006JD007638.

[22] S.C. Yang, E. Kalnay, B. Hunt, Handling nonlinearity in an ensemble Kalman filter: experiments with the three-variable Lorenz model, Mon. Weather Rev. 140 (2012) 2628–2646. http://dx.doi.org/10.1175/MWR-D-11-00313.1.

[23] Z. Meng, F. Zhang, Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part III: comparison with 3DVAR in a real-data case study, Mon. Weather Rev. 136 (2008) 522–540. http://dx.doi.org/10.1175/2007MWR2106.1.