Nat. Hazards Earth Syst. Sci. Discuss., 2, 3289–3349, 2014
www.nat-hazards-earth-syst-sci-discuss.net/2/3289/2014/
doi:10.5194/nhessd-2-3289-2014

Natural Hazards and Earth System Sciences
Open Access
Discussions

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

M. C. Rochoux[1,2,3], S. Ricci[1,2], D. Lucor[4], B. Cuenot[1], and A. Trouvé[5]

[1]CERFACS, 42 avenue Gaspard Coriolis, 31057 Toulouse Cedex 01, France
[2]SUC/CNRS-URA1875, 42 avenue Gaspard Coriolis, 31057 Toulouse Cedex 01, France
[3]Ecole Centrale Paris, CNRS-UPR288, Grande voie des vignes, 92295 Châtenay-Malabry, France
[4]Institut d'Alembert, Université Pierre et Marie Curie, CNRS-UMR7190, 4 place Jussieu, 75006 Paris, France
[5]Dept. of Fire Protection Engineering, University of Maryland, College Park, MD 20742, USA

Received: 19 March 2014 – Accepted: 22 March 2014 – Published: 9 May 2014

Correspondence to: M. C. Rochoux (melanie.rochoux@graduates.centraliens.net)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

This paper is the first part in a series of two articles and presents a data-driven wildfire simulator for forecasting wildfire spread scenarios, at a reduced computational cost that is consistent with operational systems. The prototype simulator features the following components: a level-set-based fire propagation solver FIREFLY that adopts a regional-scale modeling viewpoint, treats wildfires as surface propagating fronts, and uses a description of the local rate of fire spread (ROS) as a function of environmental conditions based on Rothermel's model; a series of airborne-like observations of the fire front positions; and a data assimilation algorithm based on an ensemble Kalman filter (EnKF) for parameter estimation. This stochastic algorithm partly accounts for the non-linearities between the input parameters of the semi-empirical ROS model and the fire front position, and is sequentially applied to provide a spatially-uniform correction to wind and biomass fuel parameters as observations become available. A wildfire spread simulator combined with an ensemble-based data assimilation algorithm is therefore a promising approach to reduce uncertainties in the forecast position of the fire front and to introduce a paradigm-shift in the wildfire emergency response. In order to reduce the computational cost of the EnKF algorithm, a surrogate model based on a polynomial chaos (PC) expansion is used in place of the forward model FIREFLY in the resulting hybrid PC-EnKF algorithm. The performance of EnKF and PC-EnKF is assessed on synthetically-generated simple configurations of fire spread to provide valuable information and insight on the benefits of the PC-EnKF approach as well as on a controlled grassland fire experiment. The results indicate that the proposed PC-EnKF algorithm features similar performance to the standard EnKF algorithm, but at a much reduced computational cost. In particular, the re-analysis and forecast skills of data assimilation strongly relate to the spatial and temporal variability of the errors in the ROS model parameters.

NHESSD

2, 3289–3349, 2014

1 Introduction

Real-time prediction of the direction and speed of a propagating wildfire has been identified as a valuable research objective with direct applications in both fire risk management and fire emergency response (Noonan-Wright et al., 2011). In addition, the future perspective of climate change tends to favor extreme drought events and to alter precipitations (Milly et al., 2002; Palmer et al., 2002; Boé et al., 2009); these conditions dramatically increase the risk for the development of large highly-destructive wildfires, commonly known as *megafires* (Nijhuis, 2012). In this context, accurate predictions of the resulting change in fire regime and intensity cannot only rely on the analysis of past observed wildfire events; the use of a data-driven wildfire spread simulator that takes full advantage of the recent technological advances for geo-referenced front-tracking becomes essential.

Despite our recent progress in computer-based wildfire spread modeling, our ability to accurately simulate the behavior of wildfires remains limited because the underlying dynamics feature complex multi-physics processes occurring at multiple scales (Viegas, 2011). The dynamics of wildfires are determined by interactions between pyrolysis, combustion and flow dynamics, radiation and convection heat transfer, as well as atmospheric dynamics and chemistry. These interactions occur at: vegetation scales that characterize the biomass fuel; topographical scales that characterize the terrain and vegetation boundary layer; and meteorological micro/meso-scales that characterize atmospheric conditions.

Relevant insight into wildfire dynamics has been obtained in recent years via detailed numerical simulations performed at flame scales (i.e., with a spatial resolution on the order of 1 m). For instance, FIRETEC (Linn et al., 2002) or WFDS (Mell et al., 2007) combine advanced physical modeling and classical methods of computational fluid dynamics (CFD) to accurately describe the combustion-related processes that control the fire behavior (e.g., thermal degradation of biomass fuel, buoyancy-induced flow, combustion, radiation and convection heat transfer). Note that because of the high compu-

NHESSD
2, 3289–3349, 2014

tational cost, flame-scale CFD is currently restricted to research projects (Linn et al., 2002; Mell et al., 2007; Rochoux et al., 2014) and is not compatible with operational applications. In contrast, a regional-scale viewpoint (i.e., a viewpoint that considers scales ranging from a few tens of meters up to several kilometers) is adopted in the following: the fire is described as a two-dimensional front that self-propagates normal to itself into unburnt vegetation; the local propagation speed is called the rate of spread (ROS). This viewpoint is the dominant approach used in current operational wildfire spread simulators, see for instance FARSITE (Finney, 1998), FOREFIRE (Filippi et al., 2009, 2013) and PHOENIX RapidFire (Chong et al., 2013). In particular, FARSITE uses a model due to Rothermel (1972) that treats the ROS as a semi-empirical function of biomass fuel properties associated with a pre-defined fuel category (i.e., the vertical thickness of the fuel layer, the fuel moisture content, the fuel particle surface-to-volume ratio, the fuel loading and the fuel particle mass density), topographical properties (i.e., the terrain slope) and meteorological properties (i.e., the wind velocity at mid-flame height). This approach is limited in scope because of the large uncertainties associated with the accuracy of computer models since they do not account for the interaction between the fire and the atmosphere, and since they have a limited domain of validity resulting from a calibration procedure based on experiments (Perry, 1998; Sullivan, 2007; Viegas, 2011; Cruz and Alexander, 2013; Finney et al., 2013). This approach is also limited because of the large uncertainties associated with many of the input parameters to the fire problem (Jimenez et al., 2007; Finney et al., 2011).

In order to overcome some of the current limitations of regional-scale wildfire modeling and to build predictive simulations that are compatible with operational framework, the uncertainties in the input data of the ROS semi-empirical model need to be quantified and reduced. The uncertainties inherent in wildfire spread modeling go beyond the limitations of deterministic forecast abilities of the dynamical model and thus, suggest the use of ensemble forecasts to stochastically characterize the non-linear response of the front-tracking simulator to variations in the input environmental parameters (D'Andrea et al., 2011; Finney et al., 2011). For instance, Finney et al. (2011)

NHESSD
2, 3289–3349, 2014

describes an ensemble-based forecasting capability, in which a large number of fire spread scenarios (i.e., the ensemble members) are generated based on a probabilistic uncertainty in the weather conditions and in the moisture content of biomass fuels. Model uncertainties are a combination of epistemic errors that express an imperfect knowledge of the input parameters of the ROS model (that could in theory be removed), and of aleatoric errors that result from natural and unpredictable stochastic variabilities of the physical system. These uncertainties translate inevitably into errors in the output variables of interest (e.g., time-evolving position of the front, burnt area, maximum value for the ROS). The most classical methodologies for uncertainty quantification in these output variables are random sample-based statistical methods derived from Monte-Carlo methodologies. While these methodologies are generic and robust for the simulation of stochastic models, they are however computationally expensive due to the required size of the sample (the computational cost of one realization may be already expensive itself, see Lucor et al., 2007) and each implementation typically requires ad-hoc variance reduction techniques (Boyaval, 2012). More efficient sampling methods have been developed to reach a comparable level of accuracy as Monte-Carlo-based techniques but with fewer *forward* model integrations; these sampling methods take advantage of the (possible) regularity of the model response to varying input parameters in order to increase the convergence rate compared to Monte-Carlo-based methodologies. In particular, polynomial chaos (PC) non-intrusive techniques issued from spectral-based representations and introduced by Wiener (1938) are very often efficient in terms of precision and cost (Ghanem and Spanos, 1991; Le Maître and Knio, 2010). The key idea is to build a polynomial representation of the forward model response (referred to as the *surrogate* model) to varying input parameters. Once the surrogate model is available, it is possible to benefit from a large sample of realizations (at almost no cost) in order to accurately characterize the model uncertainties. Still, the application of PC-based sampling techniques to problems of hyperbolic conservation laws remains a challenging task (Després et al., 2013).

NHESSD
2, 3289–3349, 2014

Recent progress made in airborne remote sensing provides new ways to monitor real-time fire front positions (Wooster et al., 2005, 2013; Riggan and Robert, 2009; Paugam et al., 2013). Unfortunately, these thermal-infrared measurements provide an incomplete description of the fire spread (in particular due to the opacity of the fire-induced thermal plume) and are subject to instrumental errors as well as representativeness errors (i.e., inconsistency between what the sensor can measure and what the computer model can describe). From this perspective, data assimilation (DA) offers a convenient framework for integrating fire sensor observations into a computer model in order to provide optimal estimates of poorly known model parameters and/or model state, and to improve *in fine* predictions of the fire spread behavior (Mandel et al., 2008; Cowlard et al., 2010; Lautenberger, 2013; Rochoux et al., 2014). The key idea is that, when used alone, neither measurements nor computer models can provide a reliable and complete description of the real state of the physical system. In the following, the set of model state and/or model parameters to be corrected through DA is gathered in the control vector. The DA algorithm is sequentially applied; each sequence (also referred to as the *assimilation cycle*) is decomposed into two steps: (1) a prediction step, in which the control variables are advanced in time given some uncertainty ranges; and (2) an update step based on the classical Bayes' theorem, in which new observations are considered and the probability density function (PDF) of the control variables is modified consistently with the observations in order to reduce the uncertainties in the model outputs (Gelb, 1974; Tarantola, 1987; Todling and Cohn, 1994; Ide et al., 1997; Kalnay, 2003; Reichle, 2008). The Kalman filter (KF) is the most commonly used sequential DA technique. However, the KF assumes linear dynamics between the control variables and the model outputs as well as a Gaussian statistical distribution for both modeling and observation errors. Extensions of the KF that partly overcome these limitations have been proposed, for instance the extended Kalman filter (EKF) that uses local linearization techniques (Gelb, 1974) or the ensemble Kalman filter (EnKF) that relies on a stochastic description of the model behavior (Evensen, 1994, 2009). An in-

NHESSD
2, 3289–3349, 2014

sightful comparison between EKF and EnKF is given within the framework of land DA in Reichle et al. (2002).

Although it seems appropriate to translate the inability of a fire spread model to generate accurate fire front positions into parameter uncertainty, other sources of uncertainties such as model structural errors or boundary/initial condition errors also need to be accounted for. For this purpose, DA is also used to sequentially correct the state variable of the fire spread model and thereby, provides observation-informed initial condition for model integration at future lead-times. This study, which is the first of a two-part article, concerns the estimation of ROS model parameters, while a second part is dedicated to the correction of the model state variables at the analysis time.

In this study, an ensemble-based DA methodology is considered in order to reduce the uncertainties in the ROS model parameters using measurements of the time-evolving location of the fire front. This study is an extension of our previous works presented in Rochoux et al. (2013a, b), in which a prototype data-driven wildfire simulator was developed. The initial prototype featured the following main components: a level-set-based fire propagation solver combined with a model description of the local ROS proposed by Rothermel (1972); a series of observations of the fire front position; and a cost-effective EKF-based DA algorithm. This prototype was successfully evaluated when applied for estimating the input parameters of the Rothermel-based ROS model (e.g., the fuel moisture content, the fuel particle surface-to-volume ratio, and/or the wind direction and magnitude). However, the EKF algorithm relies on the assumption that the relation between a perturbation in the ROS model parameters and the resulting changes in the fire front position (i.e., the generalized observation operator) can locally be approximated by a linear relation. While the EKF-based studies presented in Rochoux et al. (2013a, b) produced encouraging results and confirmed the value of a DA strategy for improved wildfire spread predictions, the linearity assumption is no longer valid in regional-scale fires, especially when the wind direction and magnitude vary and the vegetation properties are potentially strongly heterogeneous. To better account for non-linearities in the generalized observation operator, an extension to an EnKF ap-

NHESSD
2, 3289–3349, 2014

proach was preliminarily explored in Rochoux et al. (2012). This ensemble-based DA approach was originally developed for dynamic state estimation (Evensen, 1994) and has already been used in the field of wildfire modeling for correcting the temperature state variable (Mandel and Beezley, 2007; Mandel et al., 2008, 2011). It was also largely extended to sequential parameter estimation, for instance in the field of hydrodynamics (Durand et al., 2008, 2010; Moradkhani et al., 2005). Still, the large number of realizations required by the EnKF algorithm to obtain satisfactory results (due to the slow convergence rate of the Monte-Carlo sampling and to the accumulation of sampling errors, see Li and Xiu, 2008) may prove computationally burdensome within an operational framework. The required size of the sample significantly increases with the complexity of the physics (multi-parameter estimation) and the model non-linearities (complex physics), thus emphasizing the need for a reduced-cost EnKF. Efforts have therefore been devoted to designing more efficient EnKF schemes by reducing sampling errors (Szunyogh et al., 2008; Saad, 2007; Li and Xiu, 2008, 2009; Blanchard et al., 2010; Xiu, 2010; Rosić et al., 2013). For this purpose, and following work from Li and Xiu (2009), an EnKF strategy based on a PC approximation (PC-EnKF) is proposed in this paper; the polynomial surrogate model being used during the EnKF prediction step to generate a large number of model simulation trajectories at almost no cost and without loss of accuracy (Birolleau et al., 2014).

In this paper, we present a hybrid PC-EnKF DA algorithm that improves wildfire spread modeling by reducing uncertainty in the vegetation properties used as inputs of the Rothermel-based ROS model. The outline of the paper is as follows. Section 2 presents the available observations of the fire behavior and the wildfire spread model named FIREFLY (i.e., the forward model). The hybrid PC-EnKF algorithm developed for the wildfire application is presented in Sect. 3; in this section, the sequential implementation of the ensemble-based algorithms is also described. Section 4 illustrates how the classical EnKF and the hybrid PC-EnKF allow to properly estimate model parameters on simple test cases, in which the observations are synthetically-generated. The performance of the data-driven wildfire spread capability using the reduced-cost

NHESSD
2, 3289–3349, 2014

approach is demonstrated in a validation test corresponding to a controlled grassland fire experiment.

2 Information on wildfires at regional-scales: observations and forward model

2.1 Observations of the fire front location

2.1.1 Overview of available observations of fire spread

In practice, continental surfaces and vegetation are mainly observed within the mid- and near-infrared regions of the electromagnetic spectrum of wavelengths (0.75 to 15 µm). It is known that for high temperatures as encountered in wildfires (varying from 600 K for smoldering to 1200 K in the flaming zone), the maximum radiant intensity occurs within the mid-infrared (MIR) region. Thus, current spaceborne and airborne systems observe wildfires within a narrow waveband centered on the 3.9-micron wavelength (Butler et al., 2004; Wooster et al., 2005, 2013; Paugam et al., 2013; Rochoux et al., 2014), which is both sensitive to flaming and smoldering combustion modes. Beyond fire detection, remote sensing is regarded as a promising approach to provide a quantitative description of the fire radiation release to characterize sub-pixel fires (occupying a limited area of the sensor pixel down to 0.1 to 1 % of the pixel area) and to estimate fuel consumption as well as smoke emissions (Wooster et al., 2013). Using spaceborne or airborne platforms, the fire radiative power (FRP) emissions are detected in the burning area, while non-active areas remain blank. These information are crucial to retrieve the brightness temperature and thus, to track the time-evolving location of the fire front. For instance, Paugam et al. (2013) showed that spatio-temporal variations of the flame front ROS can be accurately retrieved using FRP analysis on a reduced-scale controlled fire experiment (the final burnt area of the reduced-scale

NHESSD
2, 3289–3349, 2014

study is about 1000 m^2); ongoing research aims at extending this FRP analysis to regional-scale wildfire spread.[12]

Currently, most spaceborne instruments, including the pioneer generation such as AVHRR (Advanced Very High Resolution Radiometer) and MODIS (MOderate resolution Imaging Spectroradiometer), neither offer a sufficiently short revisit period nor a high enough spatial resolution imagery for efficient front-tracking at regional scales. While these objectives no longer seem out of reach for the dual SPOT-Pléiades constellation,[3] airborne platforms still seem the most suitable solution for real-time geolocation of active fire contours. Typical examples are the LIVEFIRE system (Merlet, 2008; Crombette, 2010) and its US counterpart FIREMAPPER system deployed since 2004 by the US Forest Service and the US Department of Interior Bureau of Land Management (Riggan and Robert, 2009). As a complement, spaceborne data could be used for validation as well as calibration of models and DA procedures.

2.1.2 Choice of observations for data assimilation

In the present study, we assume that observations of the fire front position are available and that these observations can be made at different relevant times with a low measurement error (typically, 0–30 m for the LIVEFIRE system). In the following, the observed fire front is represented as a segmented line using a pre-defined number of equally-spaced markers (i.e., the N_{fr}^{o} observation markers); the observation vector noted $\boldsymbol{y}_t^o$ contains the two-dimensional coordinates (x_i^o, y_i^o) of the fire front markers (the subscript i is the index of a particular marker in the observation vector, with $i = 1, \cdots, N_{fr}^{o}$) observed at the analysis time t. The coordinates of the fire front markers are assumed to have independent Gaussian-like random errors $\boldsymbol{\epsilon}^o$ with zero mean and with standard deviation (STD) σ^o. The size of the observation vector $\boldsymbol{y}_t^o$ is $2N_{fr}^{o}$.

[1] http://wildfire.geog.kcl.ac.uk/

[2] http://gofc-fire.umd.edu/

[3] https://directory.eoportal.org/web/eoportal/satellite-missions/p/pleiades

NHESSD
2, 3289–3349, 2014

Two types of experiments are presented in the following: observation system simulation experiments (OSSE), in which observations are synthetically-generated using a reference solution of the FIREFLY fire spread model (called the true evolution) that is modified by random observation errors $\boldsymbol{\epsilon}^o$; and a controlled grassland fire experiment, in which the observations are reconstructed from measured temperature maps and using a definition of the fire front as the 600 K iso-temperature contour.

2.2 The fire spread model (the forward model)

The front-tracking solver, called FIREFLY and formally noted $\mathcal{M}$ in the following, simulates the propagation of surface wildfires within the biomass fuel bed and at regional scales, as illustrated in Fig. 1. Note that the present study is limited to flat terrains and problems with complex topography are outside its scope. FIREFLY tracks the time-evolving location of the fire front using the following three components: (1) a sub-model for the ROS noted Γ; (2) a level-set-based solver for the fire front propagation equation; and (3) an iso-contour algorithm for the reconstruction of the fire front.

2.2.1 The Rothermel-based rate of spread sub-model

(a) Original one-dimensional formulation

The ROS sub-model is based on the widely-used semi-empirical model due to Rothermel (1972) that describes Γ as a function of the local environmental conditions (e.g., vegetation and weather properties). The ROS is derived from the one-dimensional formulation of the energy balance equation per unit volume of the unburnt biomass fuel located ahead of the flame; the physical quantities involved in this energy balance are then parameterized using wind-tunnel experiments. In this formulation, Γ [$\mathrm{m\,s^{-1}}$] is expressed as the ratio between the heat flux received by the unburnt

NHESSD
2, 3289–3349, 2014

vegetation I_p [J m^{-2} s^{-1}] and the energy required to ignite the fuel H_{ig} [J m^{-3}]. Γ reads:

$$\Gamma = \frac{I_p}{H_{ig}} = \frac{\xi I_r}{\rho_b \varepsilon Q_{ig}} \left(1 + \Phi_w\right), \quad (1)$$

where I_p is a function of the energy release rate of the combustion I_r, of the dimensionless propagating flux ratio ξ (that describes the proportion of energy that is released by the flame and transferred to the vegetation in the non-flaming zone) as well as of the wind correction coefficient Φ_w that was determined for a one-dimensional case corresponding to a head fire configuration and that non-linearly depends on the wind velocity magnitude at mid-flame height u_w such that:

$$\Phi_w \equiv \Phi_w(u_w) = C_w u_w^{B_w} \left(\frac{\beta_v}{\beta_{v,\,opt}}\right)^{-E_w}, \quad (2)$$

with C_w, B_w and E_w calibrated parameters depending on the biomass fuel surface-to-volume ratio Σ_v [m^{-1}], with β_v the biomass fuel packing ratio and $\beta_{v,\,opt} \equiv \beta_{v,\,opt}(\Sigma_v)$ its optimum value. The ignition energy H_{ig} is formulated as $H_{ig} = \rho_b \varepsilon Q_{ig}$, with Q_{ig} [J kg^{-1}] the heat of ignition, ε the dimensionless effective heating number (i.e., amount of fuel effectively involved in the ignition process) and ρ_b [kg m^{-3}] the biomass fuel bulk mass density that satisfies $\rho_b = \beta_v \rho_p$ for a porous medium, ρ_p [kg m^{-3}] being the biomass fuel particle mass density.

The expression for the local ROS Γ due to Rothermel may be written in the following compact form that is equivalent to Eq. (1):

$$\Gamma \equiv \Gamma\left(\delta_v, M_v, M_{v,\,ext}, \Sigma_v, m''_v, \rho_p, \Delta h_c, u_w\right), \quad (3)$$

where the nomenclature for the input parameters are summarized in Table 1. Note that the fuel loading m''_v [kg m^{-2}] satisfies $m''_v = \rho_b \delta_v/(1 + s_t) = (\beta_v \rho_p)\delta_v/(1 + s_t)$, s_t being the fuel particle total mineral content.

NHESSD
2, 3289–3349, 2014

(b) Extension to two-dimensional surface wildfire spread

The original Rothermel's one-dimensional model is extended to two-dimensional configurations, in order to account for the wind effects on the shape of the fireline, while still maintaining a simple parameterization of the ROS with respect to local environmental conditions. Accounting for wind-induced wildfire spread in FIREFLY is such that when the wind blows in the direction of the fire spread (i.e., a head fire configuration), the wind contribution to the ROS is maximum. On the contrary, the wind contribution to the ROS is zero when the wind blows in the direction opposite to the direction of the fire spread (i.e., a rear fire configuration), meaning that the fire propagates at the value of no-wind ROS on this section of the fire front (i.e., $\Phi_w = 0$). On the flanks, the fire front advances faster than in the absence of wind (i.e., $\Phi_w > 0$). This implies that the ROS can drastically change along the fireline at a given time. For this purpose, characteristic angles in the horizontal plane (x, y) are defined to represent the direction angle of the wind noted α_w^* and the direction angle of the fire propagation noted α_{fr} (the index fr referring to the front); α_{fr} indicates the outward-pointing normal direction to the fire front noted $\boldsymbol{n}_{fr}$ (see Fig. 1). These angles are defined from the North direction, namely from the positive y-coordinates and increasing in the clockwise direction. Since the propagation of wildfires is anisotropic, the normal vector $\boldsymbol{n}_{fr}$ is not uniform along the fireline and is modified over time, with:

$$\boldsymbol{n}_{fr} \equiv \boldsymbol{n}_{fr}(x, y, t) = \begin{pmatrix} \sin \alpha_{fr}(x, y, t) \\ \cos \alpha_{fr}(x, y, t) \end{pmatrix}. \quad (4)$$

Thus, the wind velocity magnitude at mid-flame height u_w (see Table 1) corresponds to the projection of the wind velocity vector $\boldsymbol{u}_w^*$ along the normal direction to the front $\boldsymbol{n}_{fr}$:

$$u_w \equiv u_w(x, y, t) = \boldsymbol{u}_w^* \cdot \boldsymbol{n}_{fr}(x, y, t), \quad (5)$$

NHESSD
2, 3289–3349, 2014

with $\boldsymbol{u}^*_w$ defined by its magnitude, u^*_w [m s^{-1}], and direction angle, α^*_w [°]:

$$\boldsymbol{u}^*_w = \begin{pmatrix} u^*_w \sin \alpha^*_w \\ u^*_w \cos \alpha^*_w \end{pmatrix}. \quad (6)$$

In the following, u^*_w and α^*_w are treated as spatially-uniform and time-independent. The projected wind velocity at mid-flame height $u_w = u_w(x,y,t)$ is a time-dependent and spatially-varying quantity along the propagating fireline. It is worth noting that the wind contribution Φ_w is forced to a zero-value in FIREFLY when the scalar product $\boldsymbol{u}^*_w \cdot \boldsymbol{n}_{fr}(x,y,t)$ is negative (see Eq. 5) to ensure that the ROS Γ remains positive. This is consistent with the common assumption in the field of fire spread modeling that the fire propagates at least at the no-wind ROS. As for biomass fuel properties, the fuel depth $\delta_v = \delta_v(x,y)$ is treated as a time-independent, spatially-varying quantity; all other ROS model parameters are treated as constant and uniform.

(c) Sensitivity study of the rate of spread

The identification of which parameters are important to include in the control vector (denoted by $\boldsymbol{x}$) is an essential step towards the application of DA to the front-tracking simulator FIREFLY. The key idea when dealing with parameter estimation is to focus the correction on a reduced set of parameters that have significant uncertainties and to which FIREFLY is the most sensitive.

In order to identify to which input parameters the ROS Γ is the most sensitive among biomass fuel properties and weather conditions, a sensitivity study is carried out with the classical one-dimensional Rothermel's model for short grass. Nominal environmental conditions are as follows: the head fire propagates in presence of a moderate wind $u^*_w = 1\,\mathrm{m\,s^{-1}}$, and the vegetation is characterized by the moisture content $M_v = 20\,\%$, the particle surface-to-volume ratio $\Sigma_v = 11\,485\,\mathrm{m^{-1}}$, the layer thickness $\delta_v = 0.5\,\mathrm{m}$, and the layer packing ratio $\beta_v = 0.106\,\%$. These 4 parameters are perturbed around these nominal conditions. Note that the moisture at extinction $M_{v,\,ext} = 30\,\%$, the fuel

NHESSD
2, 3289–3349, 2014

particle mass density $\rho_p = 512.6$ kg m^{-3}, the effective fuel mineral content $s_e = 1\,\%$ ($s_t = 5.55\,\%$), and the heat of combustion $\Delta h_c = 1.861 \times 10^7$ J kg^{-1} remain constant and correspond to the standard values of the Rothermel's fuel database (Rothermel, 1972).

Figure 2 compares the variability in the ROS Γ when uncertainties are assumed in 4 parameters, u_w^*, M_v, Σ_v and β_v. It is found that the ROS values are the most sensitive to u_w^*. As for biomass fuel properties, they feature a wide scatter for M_v and Σ_v, while Γ is less sensitive to β_v, indicating that a lack of information in u_w^*, M_v and Σ_v results in a significant uncertainty range in the ROS model predictions. It is also shown that the ROS Γ depends non-linearly on the pair of parameters M_v and Σ_v; in particular, there is a ROS acceleration when the biomass fuel becomes drier or when the biomass fuel particles become thinner. Note that these non-linearities will be more important when the wind magnitude fluctuates over time or when the fire active area is covered by heterogeneous biomass fuels. This highlights the importance of applying a DA methodology able to handle multiple sources of non-linearities in the fire spread model.

2.2.2 The level-set-based solver

A level-set-based solver is used to propagate the fire front at the Rothermel-based ROS. FIREFLY adopts a classical approach taken from the premixed combustion literature (Poinsot and Veynante, 2005), in which a reaction progress variable noted $c = c(x,y,t)$ is introduced as a flame marker: $c = 0$ in the unburnt vegetation, $c = 1$ in the burnt vegetation, and the flame is the region where c takes values between 0 and 1 (the flame front is identified as the iso-contour $c_{fr} = 0.5$ as illustrated in Fig. 1). Within the level-set framework, the progress variable c is calculated as a solution of the following propagation equation:

$$\frac{\partial c}{\partial t} = -\boldsymbol{\gamma} \cdot \nabla c = \Gamma \, |\nabla c|, \tag{7}$$

NHESSD
2, 3289–3349, 2014

with $\Gamma = \boldsymbol{\gamma} \cdot \boldsymbol{n}_{\mathrm{fr}}$ the projected ROS given by Eq. (3) and defined along the normal direction to the fire front that satisfies $\boldsymbol{n}_{\mathrm{fr}} = -\nabla c/|\nabla c|$. Equation (7) is solved using a second-order Runge–Kutta scheme for time-integration and an advection algorithm for spatial discretization based on a second-order total variation diminishing (TVD) scheme combined with a Superbee slope limiter (Rehm and McDermott, 2009; Mallet et al., 2009). The validation of the FIREFLY level-set-based solver was presented in a previous work (Rochoux et al., 2013a), which showed a consistent average fire front speed with the ROS sub-model Γ and a relatively constant flame front thickness over time (that demonstrates the non-diffusive behavior of the numerical scheme). Further details are provided in Rochoux et al. (2014).

2.2.3 Reconstruction of the simulated fire front and comparison with the observed fire front

Once the spatio-temporal variations of the progress reaction c are known, the position of the fire front is extracted using a simple iso-contour algorithm such that, formally, the outputs of the FIREFLY model are:

$$\left[(x_i,\ y_i),\ 1 \leq i \leq N_{\mathrm{fr}}\right] = \mathcal{M}_{[t-1,t]}(c_{t-1}, \lambda), \tag{8}$$

where (x_i, y_i) represents the two-dimensional coordinates of the N_{fr} fire front markers obtained at time t (the index i indicating the marker), where c_{t-1} designates the initial condition (i.e., the spatial distribution of the progress variable c at time $(t-1)$), and where λ designates the list of input parameters of the ROS model presented in Table 1, $\lambda = (\delta_{\mathrm{v}}, M_{\mathrm{v}}, M_{\mathrm{v,\,ext}}, \Sigma_{\mathrm{v}}, m''_{\mathrm{v}}, \rho_{\mathrm{p}}, \Delta h_{\mathrm{c}}, u_{\mathrm{w}})$.

The correction provided by the DA algorithm relies on the comparison between the FIREFLY-simulated fire front described by the N_{fr} markers (corresponding to a fine-grained discretization of the simulated fire front) and the observed fire front at time t. Since observations of the fire front position are likely to be provided with a much coarser resolution and since they may cover only a fraction of the fire front perimeter,

NHESSD
2, 3289–3349, 2014

the observed fire front is discretized with a set of N_{fr}^{o} markers such that the observation vector $\boldsymbol{y}_t^o$ reads:

$$\boldsymbol{y}_t^o = \left[(x_1^o, y_1^o), (x_2^o, y_2^o), \ldots, (x_{N_{fr}^o}^o, y_{N_{fr}^o}^o)\right], \tag{9}$$

with N_{fr}^{o} much lower than N_{fr}. In the following, we assume for simplicity that $N_{fr}^{o} = (N_{fr}/r)$, where r is an integer taking values much larger than 1. Preliminary tests have indeed shown that this simple treatment (taking 1 out of every r points) provides reasonable results (Rochoux et al., 2014); the generalization of this treatment to complex fire front topology will be re-visited in future work. In order to compare the simulated fire front (SFF) with the observed fire front (OFF), a selection operator $\mathcal{H}$ is introduced. This operator pairs a subset of N_{fr}^{o} markers along SFF with the N_{fr}^{o} markers along OFF, associating each marker of OFF with its closest neighbor along SFF (see Fig. 3).

3 Data assimilation algorithm: the polynomial chaos-based ensemble Kalman filter

3.1 The standard ensemble Kalman filter

We present here the ensemble Kalman filter (EnKF) algorithm applied, in the context of parameter estimation, for one assimilation cycle between time $(t-1)$ and time t. The vector $\boldsymbol{x}_t \in \mathbb{R}^n$ corresponds to the control vector that includes the n uncertain parameters to be estimated over the assimilation cycle $[t-1, t]$.

3.1.1 Generalized observation operator

The generalized observation operator $\mathcal{G}_t$ maps the control space of $\boldsymbol{x}_t$ onto the observation space of $\boldsymbol{y}_t^o$. Within the framework of parameter estimation, $\mathcal{G}_t$ is a composition of the fire spread model $\mathcal{M}_{[t-1,t]}$ (providing the N_{fr} front marker locations associated

NHESSD
2, 3289–3349, 2014

with a realization of the control vector $\boldsymbol{x}_t$) with the selection operator $\mathcal{H}_t$ (taking 1 out of every $r = N_{\mathrm{fr}}/N_{\mathrm{fr}}^{\mathrm{o}}$ markers along SFF at time t). Formally, $\mathcal{G}_t$ reads:

$$\boldsymbol{y}_t = \mathcal{G}_t(\boldsymbol{x}_t) = \mathcal{H}_t \circ \mathcal{M}_{[t-1,t]}(\boldsymbol{x}_t), \tag{10}$$

with $\boldsymbol{y}_t$ the location of the $N_{\mathrm{fr}}^{\mathrm{o}}$ fire front markers associated with a set of control parameters $\boldsymbol{x}_t$ at time t (corresponding to the model counterparts of the observed quantities). In the following, both $\boldsymbol{x}_t$ and $\boldsymbol{y}_t^{\mathrm{o}}$ are considered as random variables.

3.1.2 Sequential estimation

The EnKF algorithm is sequentially applied over an assimilation window $[t-1, t]$; each assimilation cycle decomposes into two successive steps for each member of the ensemble indexed by the exponent k as illustrated in Fig. 4:

1. **a prediction step (forecast)**, in which the system is evolved from time $(t-1)$ to time t (t being the next observation time) through an integration of FIREFLY to forecast the fire front position $\boldsymbol{y}_t$ given some uncertainty ranges in the control vector $\boldsymbol{x}_t$. We note $\mathrm{p}^{\mathrm{f}}(\boldsymbol{x}_t)$ this PDF of the control vector (also called the *forecast* PDF) at time t. We also note $\mathcal{F}_{[t-1,t]}$ the operator describing the temporal evolution of the control parameters from time $(t-1)$ to time t, with $\boldsymbol{x}_t = \mathcal{F}_{[t-1,t]}(\boldsymbol{x}_{t-1})$. A temporal evolution of the control vector is introduced here to fit with the classical description of the EnKF algorithm: since there is no dynamic model available to describe the evolution of the control parameters, the parameter evolution model $\mathcal{F}_{[t-1,t]}$ is artificially set up with a random walk model (West, 1993; Moradkhani et al., 2005) and will be described below.

2. **an update step (analysis)**, in which new observations are considered at the analysis time t and the forecast PDF of the control parameters is modified consistently with the observations $\boldsymbol{y}_t^{\mathrm{o}}$, in order to reduce the uncertainties in the computer model outputs $\boldsymbol{y}_t$. The new PDF, called the *analysis* and noted $\mathrm{p}^{\mathrm{a}}(\boldsymbol{x}_t)$, is given by

NHESSD
2, 3289–3349, 2014

the classical Bayes' theorem:

$$p^a(\boldsymbol{x}_t) \propto p(\boldsymbol{y}_t^o|\boldsymbol{x}_t)\, p^f(\boldsymbol{x}_t), \tag{11}$$

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

Based on Bayesian theory, the EnKF algorithm assumes that the errors on the control parameters $\boldsymbol{x}_t$ and the errors on the observations $\boldsymbol{y}_t^o$ are random variables defined by Gaussian PDFs with a zero mean value and an error covariance model. Under these assumptions, the forecast PDF may be written as:

$$p^f(\boldsymbol{x}_t) \propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{x}_t - \boldsymbol{x}_t^f\right)^T (\mathbf{P}_t^f)^{-1}\left(\boldsymbol{x}_t - \boldsymbol{x}_t^f\right)\right\}, \tag{12}$$

where $\boldsymbol{x}_t^f$ is the forecast estimate of the control vector, and where $\mathbf{P}_t^f \in \mathbb{R}^{n\times n}$ is the forecast error covariance matrix representing errors in the ROS model parameters. The data likelihood may be similarly expressed as:

$$p(\boldsymbol{y}_t^o|\boldsymbol{x}_t) \propto \exp\left\{-\frac{1}{2}\,\mathbf{d}_t^T\,\mathbf{R}^{-1}\,\mathbf{d}_t\right\}, \tag{13}$$

with $\mathbf{R} \in \mathbb{R}^{2N_{fr}^o \times 2N_{fr}^o}$ the observation error covariance matrix representing observation errors (assumed constant over time in this study), and with $\mathbf{d}_t$ the innovation vector of size $2N_{fr}^o$ corresponding to the differences between simulated and observed fire fronts:

$$\mathbf{d}_t = \boldsymbol{y}_t^o - \boldsymbol{y}_t^f = \boldsymbol{y}_t^o - \mathcal{G}_t(\boldsymbol{x}_t^f). \tag{14}$$

Using the selection procedure (see Fig. 3), $\mathbf{d}_t$ is simply defined as the vector formed by the directed distances between the paired SFF-OFF markers. Note that the statistical

NHESSD
2, 3289–3349, 2014

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

Within this framework, the analysis PDF from Eq. (11) is also Gaussian and is written as:

$$\mathrm{p}^{\mathrm{a}}(\boldsymbol{x}_t) \propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{x}_t - \boldsymbol{x}_t^{\mathrm{f}}\right)^{\mathrm{T}} (\mathbf{P}_t^{\mathrm{f}})^{-1}\left(\boldsymbol{x}_t - \boldsymbol{x}_t^{\mathrm{f}}\right) - \frac{1}{2}\mathbf{d}_t^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{d}_t\right\}, \tag{15}$$

$$\propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{x}_t - \boldsymbol{x}_t^{\mathrm{a}}\right)^{\mathrm{T}} (\mathbf{P}_t^{\mathrm{a}})^{-1}\left(\boldsymbol{x}_t - \boldsymbol{x}_t^{\mathrm{a}}\right)\right\}, \tag{16}$$

where $\boldsymbol{x}_t^{\mathrm{a}}$ is the analysis estimate of the control vector, and where $\mathbf{P}_t^{\mathrm{a}} \in \mathbb{R}^{n\times n}$ is the analysis error covariance matrix. Conditional mode estimation searches for the mode of the PDF $\mathrm{p}^{\mathrm{a}}(\boldsymbol{x}_t)$, i.e., the value of the control vector $\boldsymbol{x}_t$ that maximizes the probability to estimate its true value $\boldsymbol{x}_t^{\mathrm{t}}$. Under Gaussian assumption, this maximum likelihood estimation is equivalent to a minimization problem:

$$\max_{\boldsymbol{x}_t\in\mathbb{R}^n} \mathrm{p}_{\mathrm{a}}(\boldsymbol{x}_t) \iff \min_{\boldsymbol{x}_t\in\mathbb{R}^n} \{-\ln[\mathrm{p}_{\mathrm{a}}(\boldsymbol{x}_t)]\} = \min_{\boldsymbol{x}_t\in\mathbb{R}^n} \mathcal{J}(\boldsymbol{x}_t), \tag{17}$$

with $\mathcal{J}$ the cost function of the estimation problem defined as follows:

$$\mathcal{J}(\boldsymbol{x}_t) = \frac{1}{2}\left(\boldsymbol{x}_t - \boldsymbol{x}_t^{\mathrm{f}}\right)^{\mathrm{T}} (\mathbf{P}_t^{\mathrm{f}})^{-1}\left(\boldsymbol{x}_t - \boldsymbol{x}_t^{\mathrm{f}}\right) + \frac{1}{2}\mathbf{d}_t^{\mathrm{T}}(\mathbf{R})^{-1}\mathbf{d}_t. \tag{18}$$

The direct minimization of $\mathcal{J}$ leads to the classical KF equations when the generalized observation operator $\mathcal{G}_t$ is linear (denoted by $\mathbf{G}_t$). In the present case, this implies that the fire spread model $\mathcal{M}_{[t-1,t]}$ is linear and that the parameter evolution model $\mathcal{F}_{[t-1,t]}$ is linear (denoted by $\mathbf{F}_{[t-1,t]}$). Using these assumptions, it could be shown that the forecast in the prediction step is obtained via the integration of the following equations:

$$\boldsymbol{x}_t^{\mathrm{f}} = \mathbf{F}_{[t-1,t]}\boldsymbol{x}_{t-1}^{\mathrm{a}}, \quad \mathbf{P}_t^{\mathrm{f}} = \mathbf{F}_{[t-1,t]}\mathbf{P}_{t-1}^{\mathrm{f}}\mathbf{F}_{[t-1,t]}^{\mathrm{T}}, \tag{19}$$

NHESSD
2, 3289–3349, 2014

assuming there is no error in the formulation of the parameter evolution model. In this context, the analysis update in Eq. (16) leads to the following equations:

$$\boldsymbol{x}_t^{\mathrm{a}} = \boldsymbol{x}_t^{\mathrm{f}} + \mathbf{K}_t \left(\boldsymbol{y}_t^{\mathrm{o}} - \mathbf{G}_t \, \boldsymbol{x}_t^{\mathrm{f}} \right), \tag{20}$$

$$\mathbf{K}_t = \mathbf{P}_t^{\mathrm{f}} \, \mathbf{G}_t^{\mathrm{T}} \left(\mathbf{G}_t \mathbf{P}_t^{\mathrm{f}} \, \mathbf{G}_t^{\mathrm{T}} + \mathbf{R} \right)^{-1}, \tag{21}$$

$$\mathbf{P}_t^{\mathrm{a}} = \left(\mathbf{I}_n - \mathbf{K}_t \, \mathbf{G}_t \right) \mathbf{P}_t^{\mathrm{f}}, \tag{22}$$

where $\mathbf{K}_t$ is called the gain matrix. Starting from a prior value of the control parameters (i.e., the forecast $\boldsymbol{x}_t^{\mathrm{f}}$) and using the observations $\boldsymbol{y}_t^{\mathrm{o}}$ available at time t, the analysis estimate $\boldsymbol{x}_t^{\mathrm{a}}$ is a feedback information for the fire spread model; $\boldsymbol{x}_t^{\mathrm{a}}$ is optimal when the variance of its distance to the true value $\boldsymbol{x}_t^{\mathrm{t}}$ gets to a minimum, meaning, for Gaussian cases, that its PDF is dense around its mean. The expressions in Eqs. (20–22) are the basis of the EKF algorithm used in Rochoux et al. (2013a, b), $\mathbf{F}_{[t-1,t]}$ and $\mathbf{G}_t$ being the tangent linear operators (Jacobian) of $\mathcal{F}_{[t-1,t]}$ and $\mathcal{M}_{[t-1,t]}$ in the vicinity of the control vector $\boldsymbol{x}_t$, respectively. Thus, in the EKF, a linearized and approximate equation is used for the prediction of errors statistics as well as for the relation between the control space and the observation space.

In contrast, the EnKF algorithm does not require the explicit use of the linear operators $\mathbf{F}_{[t-1,t]}$ and $\mathbf{G}_t$ in the prediction step. As shown in Fig. 4, the forecast control parameters $\boldsymbol{x}_t^{\mathrm{f}}$ are stochastically represented at time t based on N_{e} realizations called the ensemble members

$$\left[\boldsymbol{x}_t^{\mathrm{f},(1)}, \cdots, \boldsymbol{x}_t^{\mathrm{f},(k)}, \cdots, \boldsymbol{x}_t^{\mathrm{f},(N_{\mathrm{e}})} \right],$$

with k varying between 1 and N_{e}. The temporal evolution of the control parameters is artificially set up using a random walk model so that the kth ensemble member reads:

$$\boldsymbol{x}_t^{\mathrm{f},(k)} = \mathcal{F}_{[t-1,t]} \left(\boldsymbol{x}_{t-1}^{\mathrm{a},(k)} \right) = \overline{\boldsymbol{x}_{t-1}^{\mathrm{a}}} + \mathbf{e}_{t-1}^{(k)}, \tag{23}$$

NHESSD
2, 3289–3349, 2014

where $\overline{\boldsymbol{x}^{\mathrm{a}}_{t-1}}$ is the mean of the posterior estimates obtained at the previous analysis time ($t-1$), and where $\mathbf{e}^{(k)}_{t-1}$ is a randomly-generated white noise following a Gaussian distribution of zero mean and given STD σ^x (taken equal to the forecast error STD σ^{f} in the following). A series of N_{e} independent forward model integrations up to the analysis time t based on these N_{e} realizations of the control parameters is performed; this forecast step provides N_{e} fire front positions at time t corresponding to the model counterparts of the observed quantities and designated as $[\boldsymbol{y}^{\mathrm{f},(1)}_t, \cdots, \boldsymbol{y}^{\mathrm{f},(k)}_t, \cdots, \boldsymbol{y}^{\mathrm{f},(N_{\mathrm{e}})}_t]$, with $\boldsymbol{y}^{\mathrm{f},(k)}_t = \mathcal{G}_t(\boldsymbol{x}^{\mathrm{f},(k)}_t)$ for the kth ensemble member. We note $\mathbf{C}_{\boldsymbol{xy}} \in \mathbb{R}^{n \times 2N^{\mathrm{o}}_{\mathrm{fr}}}$ the matrix that represents the stochastically-based relation between the control space (of size n) and the observation space (of size $2N^{\mathrm{o}}_{\mathrm{fr}}$); $\mathbf{C}_{\boldsymbol{xy}}$ is expressed as:

$$\mathbf{C}_{\boldsymbol{xy}} = \mathbf{P}^{\mathrm{f}}_t \mathbf{G}^{\mathsf{T}}_t = \sum_{k=1}^{N_{\mathrm{e}}} \frac{\left(\boldsymbol{x}^{\mathrm{f},(k)}_t - \overline{\boldsymbol{x}^{\mathrm{f}}_t}\right)\left(\mathcal{G}_t(\boldsymbol{x}^{\mathrm{f},(k)}_t) - \overline{\mathcal{G}_t(\boldsymbol{x}^{\mathrm{f}}_t)}\right)^{\mathsf{T}}}{N_{\mathrm{e}} - 1}, \tag{24}$$

where the overline denotes the mean value over the ensemble. Similarly, the symmetric error covariance matrix on the predicted measurements denoted by $\mathbf{C}_{\boldsymbol{yy}} \in \mathbb{R}^{2N^{\mathrm{o}}_{\mathrm{fr}} \times 2N^{\mathrm{o}}_{\mathrm{fr}}}$ is stochastically formulated as:

$$\mathbf{C}_{\boldsymbol{yy}} = \mathbf{G}_t \mathbf{P}^{\mathrm{f}}_t \mathbf{G}^{\mathsf{T}}_t = \sum_{k=1}^{N_{\mathrm{e}}} \frac{\left(\mathcal{G}_t(\boldsymbol{x}^{\mathrm{f},(k)}_t) - \overline{\mathcal{G}_t(\boldsymbol{x}^{\mathrm{f}}_t)}\right)\left(\mathcal{G}_t(\boldsymbol{x}^{\mathrm{f},(k)}_t) - \overline{\mathcal{G}_t(\boldsymbol{x}^{\mathrm{f}}_t)}\right)^{\mathsf{T}}}{N_{\mathrm{e}} - 1}. \tag{25}$$

This means that the EnKF algorithm approximates the mean and the covariance of the forecast by the mean and the covariance of an ensemble, while still making the assumption that all PDFs are Gaussian. Additionally, the distance between the kth prediction $\boldsymbol{y}^{\mathrm{f},(k)}_t = \mathcal{G}_t(\boldsymbol{x}^{\mathrm{f},(k)}_t)$ and the observation vector $\boldsymbol{y}^{\mathrm{o}}_t$ is computed according to Burgers et al. (1998), meaning that an additional noise $\boldsymbol{\xi}^{\mathrm{o},(k)}$ is added to the observation vector

NHESSD
2, 3289–3349, 2014

to avoid ensemble collapse. Thus, for the kth member, the innovation vector $\mathbf{d}_t^{(k)}$ reads:

$$\mathbf{d}_t^{(k)} = \boldsymbol{y}_t^{\mathrm{o}} + \boldsymbol{\xi}^{\mathrm{o},(k)} - \boldsymbol{y}_t^{\mathrm{f},(k)}. \tag{26}$$

During the analysis, each ensemble member is updated based on the classical KF formulation presented in Eqs. (20–22), with the difference than the generalized observation operator $\mathcal{G}_t$ is non-linear and that the gain matrix $\mathbf{K}_t^{\mathrm{e}}$ is now stochastically calculated using Eqs. (24–25). The kth member analysis satisfies:

$$\boldsymbol{x}_t^{\mathrm{a},(k)} = \boldsymbol{x}_t^{\mathrm{f},(k)} + \mathbf{K}_t^{\mathrm{e}}\left(\boldsymbol{y}_t^{\mathrm{o}} + \boldsymbol{\xi}^{\mathrm{o},(k)} - \mathcal{G}_t\left(\boldsymbol{x}_t^{\mathrm{f},(k)}\right)\right), \tag{27}$$

$$\mathbf{K}_t^{\mathrm{e}} = \mathbf{C}_{xy}(\mathbf{C}_{yy} + \mathbf{R})^{-1}. \tag{28}$$

One of the advantages of the EnKF formulation in Eqs. (27)–(28) is that the explicit estimation of the tangent-linear of the observation operator $\mathbf{G}_t$ (including the tangent-linear of the fire spread model for parameter estimation) is avoided. This ensemble-based method allows the non-linearity in the observation operator $\mathcal{G}_t$ to be better taken into account than a local estimation $\mathbf{G}_t$ achieved for instance through a finite difference scheme as in the EKF (Ros et al., 1997; Rochoux et al., 2013a, b). The use of Eqs. (27)–(28) provides an ensemble of posterior estimates at time t

$$\left[\boldsymbol{x}_t^{\mathrm{a},(1)},\cdots,\boldsymbol{x}_t^{\mathrm{a},(k)},\cdots,\boldsymbol{x}_t^{\mathrm{a},(N_{\mathrm{e}})}\right],$$

easily used to simulate over the time window $[t-1,t]$ an ensemble of retrospective posterior estimates of the fire front positions $[\boldsymbol{y}_t^{\mathrm{a},(1)},\cdots,\boldsymbol{y}_t^{\mathrm{a},(k)},\cdots,\boldsymbol{y}_t^{\mathrm{a},(N_{\mathrm{e}})}]$ as well as an ensemble of forecasts of the fire spread beyond time t.

Note that in the present study, we assume that observation errors are uncorrelated, i.e., the observation error covariance matrix $\mathbf{R}$ is treated as a diagonal matrix, in which each diagonal term is the error variance $(\sigma^{\mathrm{o}})^2$ associated with the error in the x- or y-coordinate of the markers along the observed fire front.

NHESSD
2, 3289–3349, 2014

3.2 Polynomial chaos-based ensemble Kalman filter

In the classical EnKF algorithm, a Monte-Carlo sampling is used to generate the forecast members $\boldsymbol{x}_t^{\mathrm{f}}$ and their associated fire front trajectory $\boldsymbol{y}_t^{\mathrm{f}}$. While this provides accurate access to the full statistics of the modeling uncertainties (provided the ensemble N_{e} is sufficiently large), it involves a large number of forward model integrations (as illustrated in Fig. 4) that becomes time-consuming for regional-scale fire spread problems. To maintain the computational cost of the EnKF algorithm compatible with the objective of wildfire spread forecasting, a numerical strategy based on a polynomial chaos (PC) expansion is introduced; this PC expansion is used in the prediction step of the EnKF algorithm as outlined in Fig. 5. This hybrid EnKF algorithm is denoted by PC-EnKF in the following.

3.2.1 General formulation of the surrogate model

The PC-based surrogate model approximates the generalized observation operator $\mathcal{G}_t$ at time t and is therefore denoted by $\mathcal{G}_{\mathrm{pc},t}$. It is parameterized with respect to the multi-dimensional control vector $\boldsymbol{x}_t^{\mathrm{f}} \in \mathbb{R}^n$ following the forecast PDF $\mathrm{p}^{\mathrm{f}}(\boldsymbol{x}_t)$. This random vector may be regarded as a set of second-order random variables (i.e., with finite variance) expressed in terms of a random event ω such that $\boldsymbol{x}_t^{\mathrm{f}} = \boldsymbol{x}_t^{\mathrm{f}}(\omega)$. It can be projected onto a stochastic space spanned by orthogonal PC functions of independent Gaussian random variables $\boldsymbol{\zeta}(\omega)$ as follows:

$$\boldsymbol{x}_t^{\mathrm{f}}(\omega) = \left[\mathrm{x}_{1,t}^{\mathrm{f}}, \mathrm{x}_{2,t}^{\mathrm{f}}, \cdots, \mathrm{x}_{\mathrm{n},t}^{\mathrm{f}}\right] = \sum_{q=0}^{\infty} \hat{\boldsymbol{x}}_q \, \varphi_q\left(\boldsymbol{\zeta}(\omega)\right). \tag{29}$$

The simulated positions of the fire front $\boldsymbol{y}_t^{\mathrm{f}} = \mathcal{G}_t(\boldsymbol{x}_t^{\mathrm{f}}(\boldsymbol{\zeta}))$ can also be viewed as a random variable and therefore, they can be projected onto a stochastic space spanned by

NHESSD
2, 3289–3349, 2014

orthogonal PC functions as follows:

$$\boldsymbol{y}_t^{\mathrm{f}} = \mathcal{G}_{\mathrm{pc},t}\left(\boldsymbol{x}_t^{\mathrm{f}}(\boldsymbol{\zeta})\right) = \sum_{q=0}^{\infty} \hat{\boldsymbol{y}}_q(t)\,\varphi_q(\boldsymbol{\zeta}), \tag{30}$$

where $\hat{\boldsymbol{y}}_q \equiv \hat{\boldsymbol{y}}_q(t)$ are time-dependent coefficients, and where $(\varphi_q)_{q=0,\cdots,\infty}$ designate the multi-dimensional approximating polynomial functions forming an orthogonal basis with respect to the joint PDF $\mathrm{p}^{\mathrm{f}}(\boldsymbol{x}_t) = \mathrm{p}^{\mathrm{f}}(\mathrm{x}_{1,t}, \mathrm{x}_{2,t}, \cdots, \mathrm{x}_{\mathrm{n},t})$. The choice for the basis functions may depend on the type of random variable functions (Xiu and Karniadakis, 2002). Since the control vector $\boldsymbol{x}_t^{\mathrm{f}}$ is assumed to follow a Gaussian PDF $\mathrm{p}^{\mathrm{f}}(\boldsymbol{x}_t)$ within the framework of the EnKF, the surrogate model of the observation operator $\mathcal{G}_{\mathrm{pc,\,t}}$ is built upon the basis of the Hermite polynomials (Ghanem and Spanos, 1991). Stated differently, the Hermite polynomials form the optimal basis for random variables following multi-variate Gaussian PDF. Note that the model outputs $\boldsymbol{y}_t^{\mathrm{f}}$ are represented in terms of the same random event ω as the model inputs $\boldsymbol{x}_t^{\mathrm{f}}$, since the uncertainty in the model outputs is assumed to be mainly due to the uncertainty in the ROS model parameters within the framework of parameter estimation.

In practice, a truncated version of Eq. (30) is used; there are several ways of constructing the approximation space. The most common choice is to constrain the number of terms N_{pc} in the PC expansion by the number of control parameters n and by the maximum order of the polynomial basis Q_{po} such that:

$$N_{\mathrm{pc}} = \frac{(n + Q_{\mathrm{po}})!}{(n!\; Q_{\mathrm{po}}!)}. \tag{31}$$

This choice of N_{pc} ensures that the PC approximation is of highest order Q_{po}. Note that Q_{po} is a user-defined quantity that must be chosen carefully, according to the model non-linearity, in order to obtain an accurate representation of the model outputs $\boldsymbol{y}_t^{\mathrm{f}}$ with a high-order convergence rate. Theoretically, $Q_{\mathrm{po}} = 1$ (i.e., only two terms for $n = 1$

NHESSD
2, 3289–3349, 2014

corresponding to the mean and STD of the control variable) is enough to approximate exactly a Gaussian random variable. Note also that N_{pc} rapidly grows with n and Q_{po}, implying that a balance between accuracy and computational cost must be found. For instance, if $n = 2$ and $Q_{po} = 2$, there are $N_{pc} = 6$ terms retained in the PC expansion. Using this formalism, the surrogate model $\mathcal{G}_{pc,t}$ can be formulated as follows:

$$\boldsymbol{y}_t^f \cong \mathcal{G}_{pc,t}\left(\boldsymbol{x}_t^f(\boldsymbol{\zeta})\right) = \sum_{q=0}^{N_{pc}} \hat{\boldsymbol{y}}_q(t)\,\varphi_q(\boldsymbol{\zeta}), \tag{32}$$

where the unknowns are the following time-dependent vectors:

$$\hat{\boldsymbol{y}}_q \equiv \hat{\boldsymbol{y}}_q(t) = \left[(\hat{x}_1, \hat{y}_1)_q, \ldots, (\hat{x}_{N_{fr}^o}, \hat{y}_{N_{fr}^o})_q\right]_t, \tag{33}$$

q varying between 1 and N_{pc}, with N_{fr}^o the number of markers along OFF at time t. Note that the size of the qth vector $\hat{\boldsymbol{y}}_q$ is $2N_{fr}^o$ (each marker location being represented with both the x- and y-coordinate on the horizontal plane) and thereby, the computation of $(2N_{fr}^o N_{pc})$ coefficients (also referred to as the *PC modes*) is necessary to build the surrogate model $\mathcal{G}_{pc,t}$.

3.2.2 Calculation of the polynomial chaos modes

Due to the orthogonality of the PC basis, it can be shown that the qth PC coefficients $\hat{\boldsymbol{y}}_q$ are given by:

$$\hat{\boldsymbol{y}}_q = \frac{\mathbb{E}\left[\mathcal{G}_{pc,t}(\boldsymbol{x}_t^f)\,\varphi_q(\boldsymbol{\zeta})\right]}{\mathbb{E}\left[\varphi_q(\boldsymbol{\zeta})^2\right]}, \tag{34}$$

where:

NHESSD
2, 3289–3349, 2014

– $\mathbb{E}[\cdot]$ refers to the expectation operator satisfying $\mathbb{E}[\varphi_q(\boldsymbol{\zeta})\varphi_l(\boldsymbol{\zeta})] = 0$ if $q \neq l$, with the following definition for the inner product:

$$\mathbb{E}\left[\varphi_q(\boldsymbol{\zeta})\varphi_l(\boldsymbol{\zeta})\right] = \int_{\mathbb{R}^n} \varphi_q(\boldsymbol{\zeta})\varphi_l(\boldsymbol{\zeta})\mathrm{p}(\boldsymbol{\zeta})\mathrm{d}\boldsymbol{\zeta} = \delta_{\mathrm{ql}}\left[\varphi_q^2\right], \tag{35}$$

with δ_{ql} the Kronecker delta-function;

– $\mathbb{E}[\varphi_q(\boldsymbol{\zeta})^2]$ is a normalization factor equal to 1 if the basis is constructed orthonormal;

– $\mathbb{E}[\mathcal{G}_{\mathrm{pc},t}(\boldsymbol{x}_t^{\mathrm{f}})\varphi_q(\boldsymbol{x}_t^{\mathrm{f}})]$ is computed using a Gauss-Hermite quadrature rule, with $[\boldsymbol{x}_t^{\mathrm{f},(1)},\cdots,\boldsymbol{x}_t^{\mathrm{f},(j)},\cdots,\boldsymbol{x}_t^{\mathrm{f},((N_{\mathrm{quad}})^n)}]$ the quadrature roots vector of size $(N_{\mathrm{quad}})^n$ constrained by the maximum order of the polynomial basis Q_{po} such that $2Q_{\mathrm{po}} \leq 2(N_{\mathrm{quad}} - 1)$. Thus, this term is computed as follows:

$$\mathbb{E}\left[\mathcal{G}_{\mathrm{pc},t}(\boldsymbol{x}_t^{\mathrm{f}})\varphi_q(\boldsymbol{\zeta})\right] = \int_{\mathbb{R}^n} \mathcal{G}_t(\boldsymbol{x}_t^{\mathrm{f}})\varphi_q(\boldsymbol{\zeta})\mathrm{dp}(\boldsymbol{\zeta}) \cong \sum_{j=1}^{(N_{\mathrm{quad}})^n} \mathcal{G}_t(\boldsymbol{x}_t^{\mathrm{f},(j)})\varphi_q(\boldsymbol{\zeta}^{(j)})w^{(j)}, \tag{36}$$

where $\boldsymbol{y}_t^{\mathrm{f},(j)} = \mathcal{G}(\boldsymbol{x}_t^{\mathrm{f},(j)})$ corresponds to the FIREFLY forward model integration evaluated at the jth quadrature root $\boldsymbol{x}_t^{\mathrm{f},(j)}$ with its associated weight $w^{(j)}$, and where φ_q is the qth multi-dimensional basis function formulated as tensor products of one-dimensional polynomial functions:

$$\varphi_q \equiv \varphi_q(\boldsymbol{\zeta}) = \prod_{l=1}^{n} \varphi_{i(l)}^{1D}\left(\zeta_l\right), \tag{37}$$

with $\varphi_{i(l)}^{1D}$ the one-dimensional polynomial basis and its multi-index $i(l)$ varying between 0 and Q_{po} to determine the proper term in the multi-variable space.

NHESSD
2, 3289–3349, 2014

Based on this formulation, the construction of the surrogate model $\mathcal{G}_{\mathrm{pc},t}$ over the assimilation window $[t-1,t]$ requires a limited number of $(N_{\mathrm{quad}})^n$ forward model integrations (see the first step in Fig. 5). The polynomial approximation $\mathcal{G}_{\mathrm{pc},t}$ calculated in Eq. (32) is then used in the prediction step of the EnKF algorithm (instead of the observation operator $\mathcal{G}_t$) to compute the predictions of the time-evolving fire front locations $\boldsymbol{y}^{\mathrm{f}}_{\mathrm{pc},t}$ for a large number of members N_{e} (see the second step in Fig. 5). This ensemble of forecasts is used to accurately estimate the covariance matrices $\mathbf{C}_{xy}$ and $\mathbf{C}_{yy}$ that are required in the formulation of the Kalman gain matrix. Thus, the EnKF update can be performed with reliable covariance matrices at a reduced computational cost compared to the standard EnKF algorithm based on a Monte-Carlo sampling. This approach leads to analysis estimates of the control parameters $\boldsymbol{x}^{\mathrm{a},(k)}_t$ and to accurate PDF of the fire front locations $\boldsymbol{y}^{\mathrm{a},(k)}_t$ $(k=1,\cdots,N_{\mathrm{e}})$ using the same surrogate model as for the forecast estimates.

In order to reduce the computational cost of the EnKF algorithm, a surrogate model based on a PC expansion is used in place of the forward model (i.e., the FIREFLY regional-scale wildfire spread model) in the DA procedure. The performance of the resulting PC-EnKF algorithm is assessed on synthetically-generated fire spread cases based on preliminary work presented in Rochoux et al. (2012) as well as on the controlled grassland fire experiment.

3.3 Numerical implementation

In practice, the EnKF and PC-EnKF ensemble-based DA algorithms were implemented with the fire spread simulator FIREFLY using the Open-PALM dynamic coupling software (Lagarde et al., 2001), co-developed at CERFACS and ONERA.[4] Open-PALM allows for the coupling of independent code components with a high-level of modularity in the data exchanges and treatment, while providing a straightforward parallelization environment (Fouilloux et al., 1999; Buis et al., 2006). In this study, it is used as a task-

[4] http://www.cerfacs.fr/globc/PALM_WEB/

NHESSD
2, 3289–3349, 2014

parallelism manager to handle communications and data exchanges between FIREFLY and the mathematical units required to sequentially apply the EnKF and PC-EnKF algorithms. The PALM-PARASOL functionality in Open-PALM was used to efficiently and independently run the FIREFLY time-integrations in parallel, on the available processors. The master processor of PALM-PARASOL spawns multiple copies of the same computer program (i.e., the slaves), each on one or several processors with a different set of input parameters of the ROS model. Each slave integrates FIREFLY using one realization of the control vector $\boldsymbol{x}_t$ to provide the associated fire front position $\boldsymbol{y}_t$, subsequently used for the computation of the covariance matrices $\mathbf{C}_{xy}$ and $\mathbf{C}_{yy}$. As illustrated in Figs. 4 and 5, this integration is performed for the N_e ensemble members using the forward model FIREFLY for the classical EnKF. In contrast, for the PC-EnKF algorithm, a limited number of FIREFLY model integrations $(N_{quad})^n$ is used to build the surrogate model and subsequently, a large number of evaluations of the surrogate model (N_e) are computed using the PALM-PARASOL functionality.

4 Data assimilation experiments

4.1 Convergence of the ensemble-based algorithms

The EnKF and PC-EnKF algorithms are compared on an OSSE experiment, in which the Rothermel-based ROS model of Eq. (3) is reformulated as $\Gamma(x,y) = P\,\delta_v(x,y)$ with P [s^{-1}] a proportionality coefficient and $\delta_v = \delta_v(x,y)$ a spatially-varying function that is assumed to be perfectly known. Note that this formulation takes advantage of the proportionality between the ROS Γ and the fuel layer thickness δ_v in the Rothermel's formulation. Thus, the control vector is limited to a single parameter, $\boldsymbol{x} = P$, which encompasses different uncertainties that are not distinguished here.

The fire is ignited at $(x_{ign}, y_{ign}) = (100\,m, 100\,m)$ as a circular front with a radius of 5 m; it spreads upon a random fuel distribution $\delta_v(x,y)$ over a 200 m × 200 m domain. Observations (represented using $N_{fr}^o = 20$ front markers) are synthetically-generated at

NHESSD
2, 3289–3349, 2014

50 s intervals with FIREFLY and a chosen true value $\boldsymbol{x}^{\mathrm{t}} = P^{\mathrm{t}} = 0.4\,\mathrm{s}^{-1}$. An observation error characterized by the error STD σ^{o} is also introduced. The ensemble of prior values is drawn from a Gaussian distribution centered in $\boldsymbol{x}^{\mathrm{f}} = 0.2\,\mathrm{s}^{-1}$ with an error STD $\sigma^{\mathrm{f}} = 0.05\,\mathrm{s}^{-1}$ (assumed constant along the assimilation cycles).

A PC approximation (with a polynomial order $Q_{\mathrm{po}} = 4$ and subsequently a quadrature order $N_{\mathrm{quad}} = 5$, see Sect. 3.2) is used to build the model response surface $\mathcal{G}_{\mathrm{pc}}$ to the control parameter $\boldsymbol{x} = P$ corresponding to the forecast $(\boldsymbol{x}^{\mathrm{f}},\ \sigma^{\mathrm{f}})$.

4.1.1 Sensitivity to sampling errors

Convergence properties of the EnKF-based analysis estimates are studied in Fig. 6 with respect to the number of ensemble members N_{e} for a fixed observation error STD $\sigma^{\mathrm{o}} = 2\,\mathrm{m}$ and for one assimilation cycle. Since there is no analytical solution of the problem, the convergence of the EnKF is assumed to be achieved if the mean value of the control parameter and its STD remain constant when increasing N_{e}. The performance of the PC-EnKF algorithm is compared to that of the standard EnKF algorithm (black squares) for different PC polynomial orders, $Q_{\mathrm{po}} = 2$ (orange triangled-dashed line) and $Q_{\mathrm{po}} = 4$ (red circled-dashed line).

Figure 6 shows that in the present configuration, the EnKF algorithm converges for a minimum of $N_{\mathrm{e}} = 48$ members (meaning that FIREFLY is integrated 48 times to produce 48 fire front trajectories associated with each realization of the control parameter). It is shown that the PC-EnKF algorithm provides a comparable result as the EnKF (in terms of mean and STD) above $N_{\mathrm{e}} = 40$ members for a polynomial order $Q_{\mathrm{po}} = 4$. However, the results achieved with PC-EnKF are obtained for a lower number of FIREFLY time-integrations (i.e., 5 FIREFLY model integrations only since $N_{\mathrm{quad}} = 5$ quadrature points are used to build the model surface response $\mathcal{G}_{\mathrm{pc}}$) than the standard EnKF, while considering the same number of members N_{e} to generate the forecast/analysis estimates. Thus, the PC-EnKF algorithm provides a solution that reproduces the converged solution of the EnKF for a computational cost that is reduced by a factor of

NHESSD
2, 3289–3349, 2014

at least 8. This implies that for more complex fire spread cases where more members are required to track spatial variations in wind and vegetation conditions, the PC-EnKF algorithm appears as a promising alternative to obtain accurate simulations of fire spread at a reasonable computational cost. Additionally, the PC-EnKF algorithm provides a mean estimate that is less fluctuating than the EnKF algorithm, with a slightly reduced scatter for low values of N_e, indicating that the PC-EnKF strategy requires less ensemble members N_e to reach convergence.

Figure 6 also illustrates the sensitivity of the PC-EnKF-based analysis to the choice of the PC polynomial order Q_{po} for a varying number of ensemble members N_e. While $Q_{po} = 2$ (i.e., $N_{quad} = 3$) provides a reasonable approximation of the mean analysis estimate when considering the standard EnKF as reference, $Q_{po} = 4$ (i.e., $N_{quad} = 5$) leads to a more accurate estimate without loss of accuracy. Even though the fire front marker locations exhibit approximate Gaussian PDF and in theory $n = 1$ is sufficient to characterize their distributions, a high polynomial order is required in this case since the true value ($P^t = 0.4\,s^{-1}$) is not in the zone of high probability occurrence of the forecast estimates ($P^f = 0.2\,s^{-1}$ with $\sigma^f = 0.05\,s^{-1}$). Indeed, the true fire front locations are at the tail of the forecast PDF, which makes the estimation of the fire front locations more difficult. This difficulty shows the ability of the PC-EnKF procedure to retrieve accurate estimates of the fire spread at a low computational cost and without loss of accuracy, even though the prior information is very uncertain.

4.1.2 Example of polynomial chaos-based surface response

Figure 7 provides a comparison in the observation space between the observed fire front and the forecast/analysis estimates obtained through the PC-EnKF algorithm for an observation error STD $\sigma^o = 2\,m$, a PC polynomial order $Q_{po} = 4$ and a number of ensemble members $N_e = 1000$ (a single assimilation cycle is considered). As expected, the analysis estimates (red squared-solid line) provide a more accurate approximation of the observed fire front location (black crosses) than the forecast estimates (blue circled-dashed line).

NHESSD
2, 3289–3349, 2014

To offer insight into the main ideas underlying the PC-EnKF algorithm, Fig. 8 illustrates the mapping between the control space and the observation space for one marker of the fireline (its position on the forecast/analysis fireline is indicated in Fig. 7 by the character m). The variations in the x- and y-coordinates of this marker are represented with respect to variations in the control parameter P: black crosses indicate the simulated marker positions associated with the $N_{quad} = 5$ quadrature roots (i.e., FIREFLY model integrations) corresponding to the first step of the PC-EnKF algorithm; and blue circles indicate the forecast estimates obtained through the surrogate model evaluation combined with a Monte-Carlo sampling ($N_e = 1000$) corresponding to the second step of the PC-EnKF algorithm (see Fig. 5). These fire front estimates are associated with the forecast control parameter $P^f = 0.20\,s^{-1}$ and its error STD $\sigma^f = 0.05\,s^{-1}$. In contrast, red squares are produced by the EnKF update applied for any of the 1000 ensemble members, they correspond to the analysis estimates related to $P^a = 0.38\,s^{-1}$ and $\sigma^a = 0.01\,s^{-1}$. The scatter of the ensemble is significantly reduced in the analysis, around the true value $P^t = 0.40\,s^{-1}$, highlighting the uncertainty reduction achieved through the ensemble-based DA.

4.1.3 Sensitivity to observation errors

For verification purposes on the behavior of the PC-EnKF algorithm, Fig. 9 examines the influence of the observation error on the performance of the EnKF and PC-EnKF algorithms (the EnKF algorithm is used as reference). Statistics (in terms of mean value and STD) of the analysis obtained for $N_e = 48$ members over one assimilation cycle at time $t = 50\,s$ are presented as a function of the magnitude of the observation errors measured by σ^o (up to $\sigma^o = 30\,m$); vertical bars give a graphical representation of the magnitude of the STD within the analysis ensemble. The results show the consistency of the PC-EnKF algorithm with the EnKF in retrieving realistic values for the control parameter, even though the observation error is significant. When the observation error STD σ^o is small, the PC-EnKF algorithm successfully drives the analysis ensemble towards the true value of the parameter $P^t = 0.4\,s^{-1}$; the resulting analysis exhibits

NHESSD
2, 3289–3349, 2014

a much reduced scatter by at least a factor 4 in comparison to the forecast STD $\sigma^f = 0.05\,s^{-1}$. In contrast, when σ^o is large, the PC-EnKF algorithm has reduced effects and the analysis ensemble remains close to the forecast ensemble (the analysis STD is similar to the forecast STD $\sigma^f = 0.05\,s^{-1}$). For intermediate values of σ^o, the PC-EnKF algorithm produces optimized analyses lying between forecast and observation; as expected, the more accurate the observations, the more certain the analysis for a given forecast error.

4.2 Temporal variability of the parameter error

Sequential application of the EnKF allows for a temporal correction of the parameter P for a case in which the time-varying profile of the true parameter P^t was artificially set up between 0.3 and $0.6\,s^{-1}$ over 7 assimilation cycles (the true profile is shown in Fig. 10a in black solid line). While the mean value of the forecast estimates is set to $0.2\,s^{-1}$ for the first assimilation cycle, it is set to the mean analysis estimate from the previous assimilation cycle for all further assimilation cycles. Since there is no explicit dynamic model for the control parameter P, it is of primary importance to track the temporal variability of the error in the parameter through a sequential estimation combined with a random walk model in the EnKF prediction step (see Eq. 23). Note that for this experiment, $N_e = 48$ members are considered in the ensemble and a constant observation error STD $\sigma^o = 5\,m$ is assumed.

Figure 10a shows the temporal variations of the EnKF estimates along the assimilation cycles. The EnKF solution (red squared-solid curve) provides an optimal mean value of the control parameter, resulting in an ensemble of fire fronts that is coherent with the observation error statistics (see Fig. 10b). In contrast, the model without DA (green triangled-dashed-dotted curve) significantly underestimates the ROS. While being not as accurate as the analysis at the assimilation time, the forecast (blue circled-dashed curve) provides a significant improvement in the prediction of wildfire spread at future lead-times compared to the model without DA (i.e., free run). Note that there is

NHESSD
2, 3289–3349, 2014

a temporal shift between the forecast and analysis estimates in Fig. 10a. The analysis estimates are obtained at the current observation time and thereby, provide the most recently-updated information. In contrast, the forecast estimates only contain information up to the previous analysis time:

- if the error in the control parameter does not change, the correction obtained at the previous analysis time is still valid and adapted to track the actual fire front position: the forecast estimates provide reliable scenarios of wildfire spread at future lead-times;
- if this error significantly varies in-between two successive analysis times (this is the case of the present DA experiment), the correction is no longer suitable to predict wildfire spread at long lead-times. For instance, over the assimilation cycle indexed by 5, the analysis estimates provide a good approximation of the actual fire front location at time t_4; however, the forecast obtained at time t_5 when starting from the analysis at t_4 over-estimates the true ROS, meaning that a new observation is required to gain information on the wildfire behavior.

In the present case, Fig. 10b shows that the error in the forecast is systematically higher than that of the error in the analysis in the observation space. This means that the assimilation needs to be renewed according to the temporal variability of the error in the control parameter to ensure a high-level performance of the data-driven simulated forecast.

4.3 Application to a controlled grassland fire

The EnKF and PC-EnKF algorithms are applied to a real-world case study, corresponding to a reduced-scale controlled grassland fire (4 m × 4 m), propagating over a flat terrain and occurring under moderate wind conditions (Paugam et al., 2013). These wind conditions are assumed to be uniform and constant, $u_w^* = 1.0\,\mathrm{m\,s^{-1}}$, blowing into a western direction, $\alpha_w^* = 307°$. The grass is assumed to exhibit a uniform layer thickness, $\delta_v = 0.08\,\mathrm{m}$, and controlled properties, for instance the moisture content and the

NHESSD
2, 3289–3349, 2014

grass surface-to-volume ratio are (approximately) known, $M_v = 22\,\%$ (field measure) and $\Sigma_v = 11\,485\,m^{-1}$ (values related to short grass taken from the Rothermel's fuel database). The observed fire front locations are extracted from thermal-infrared imaging at 28 s intervals; they are discretized with $N_{fr}^{o} = 40$ front markers with a measurement error estimated to $\sigma^{o} = 0.05\,m$ (based on the spatial resolution of the camera).

The performance of both algorithms is studied over one assimilation cycle [50, 78 s] with an EnKF update at time $t_1 = 78\,s$ (analysis mode) as well as over a forecast time period [78, 106 s] with an EnKF forecast at time $t_2 = 106\,s$ (forecast mode); the initial condition at $t_0 = 50\,s$ is taken as the observed fire front at t_0. The control parameters are the fuel moisture M_v and the fuel particle surface-to-volume ratio Σ_v such that $\boldsymbol{x} = [M_v, \Sigma_v]$ with $n = 2$. A PC approximation (with a polynomial order $Q_{po} = 4$ and a quadrature order $N_{quad} = 5$) is used to build the model response surface to the two control parameters M_v and Σ_v; this response surface is shown for one particular simulated front marker in Fig. 11a (black crosses). A forecast ensemble of $N_e = 1000$ members (blue circles) is generated at no-cost using the PC-based surrogate model assuming uncertainties in M_v and Σ_v; the forecast estimates of these control parameters are described in Table 2 along with the associated STD. Note that the blue circles are contained within the surface response described by the black crosses that represent the $(N_{quad})^2 = 25$ FIREFLY forward model integrations performed to build the PC-based surface response. Thus, the PC decomposition properly approximates the observation operator.

4.3.1 Analysis mode

In the analysis mode, the forecast ensemble is corrected by assimilating the fire front at time $t_1 = 78\,s$. The comparison between the observations (black crosses), the forecast estimates (blue circled-dashed line) and the PC-EnKF-based analysis estimates (red squared-solid line) obtained at time $t_1 = 78\,s$ are presented in Fig. 12a. The forecast trajectory represents the ensemble mean of the surrogate model simulations obtained

NHESSD
2, 3289–3349, 2014

without DA (i.e., using the forecast estimates of the control parameters), while the analysis trajectory derives from an EnKF update at t_1 using the analysis estimates in the surrogate model integrations.

It is found that the PC-EnKF strategy allows to significantly decrease the distance between the observations and the simulated fronts with a comparable level of accuracy as the standard EnKF algorithm (the PC-EnKF algorithm provides similar analysis mean and STD, see Table 2). As illustrated in Fig. 11b, the uncertainty in the fire front positions is significantly reduced in comparison to the forecast since the STD related to the analysis estimates is much smaller than that of the forecast estimates. This indicates that the PC-EnKF algorithm allows reliable statistical information to be retrieved for only 25 FIREFLY model integrations (in contrast, the standard EnKF algorithm requires 1000 members to correct $n = 2$ control parameters and thereby, correct $2N^{o}_{fr} = 80$ fire front marker coordinates).

Consistently, Fig. 13 shows that the support of the analysis PDF (see Fig. 13b) is significantly reduced compared to the forecast PDF (see Fig. 13a) for the x- and y-coordinates of the $N^{o}_{fr} = 40$ observed front markers. The topology of the PDF along the observed fire front is found to be overall preserved through the EnKF update, implying that the assumption of Gaussian error statistics for the modeling error statistics seems not to deteriorate the performance of the ensemble-based DA algorithms. Some regions of the PDF related to the x-coordinates of the front marker locations (nearby $x = 2\,\mathrm{m}$) are not sensitive to variations in M_v and Σ_v. These regions correspond to the flank of the fire, meaning that the x-coordinates of the surrounding front markers do not vary and the growth of the burning area only induces variations in the y-coordinates.

As discussed for the OSSE test cases, the non-linear response of the observation operator to the control parameters induces a slightly non-Gaussian PDF for the forecast estimates: it is indeed found that the mode of the PDF does not exactly coincide with the mean value. Note that the PDF exhibits a relatively flat tail for decreasing x- and increasing y-coordinates of the observed fire front markers: this is due to a sharp

NHESSD
2, 3289–3349, 2014

ROS acceleration when decreasing the fuel moisture content M_v or alternatively, when increasing the fuel particle surface-to-volume ratio Σ_v (see Fig. 2b and c).

4.3.2 Forecast mode

In the forecast mode, Fig. 12b compares the fire front position at $t_2 = 106\,s$ obtained using the forecast estimates (without DA) and the analysis estimates derived from a DA update at $t_1 = 78\,s$. The PC-EnKF algorithm appears to properly represent the forecast trajectory at $t_2 = 106\,s$ in comparison to the standard EnKF. This result illustrates that a PDF sampling based on PC (instead of Monte-Carlo in the standard EnKF) can significantly reduce the computational cost of the EnKF prediction/update steps (in terms of number of FIREFLY model integrations that constitute the most time-consuming task in the PC-EnKF algorithm) and thereby, provide accurate error statistics on the inputs and outputs of the wildfire spread model. For instance, 1000 FIREFLY model integrations were used in the EnKF algorithm to accurately represent the error statistics; in contrast, only 25 FIREFLY model integrations were performed in the PC-EnKF algorithm. Thus, the number of FIREFLY model integrations is here divided by a factor of 40.

Additionally, Fig. 12b shows that the errors in the control parameters do not significantly change in-between the two observation times (i.e., at $t_1 = 78\,s$ and $t_2 = 106\,s$), meaning that an observation time period of 28 s seems appropriate for applying DA (relatively to the temporal variability of the errors in the control vector $\boldsymbol{x}$).

While the improved accuracy of EnKF-based data-driven simulations is obtained at the expense of heavy computational cost (in the context of multi-parameter estimation for instance), the PC-EnKF strategy appears as a promising strategy for solving Bayesian filtering problems at a low computational cost that is a requirement of operational frameworks.

NHESSD
2, 3289–3349, 2014

5 Conclusions

A data assimilation strategy based on the ensemble Kalman filter (EnKF) with parameter estimation is demonstrated to account for both experimental and modeling uncertainties in wildfire spread modeling and thereby, to provide optimized forecast of wildfire behavior. The sequential parameter estimation is efficient at reducing uncertainty in the numerical predictions of fire spread for synthetic measurements as well as for a controlled grassland fire experiment. It was found that the non-linear relation between the environmental parameters and the fire front positions induced by the non-linearities of the fire spread can be stochastically described over the ensemble members. It is also highlighted that the duration of the assimilation cycle is of primary importance in the success of the proposed data assimilation approaches: the assimilation must be renewed frequently, according to the temporal variability of the parameter errors, in order to track the real fire behavior. In order to reduce computational cost and balance sampling errors due to frequent assimilation cycles, a probabilistic sampling based on polynomial chaos techniques (PC-EnKF) was shown to significantly reduce the computational cost of the EnKF-based parameter estimation approach (by a factor of at least 10 in the present configurations), and thereby to provide access to accurate error statistics on both model inputs and outputs for the formulation of the Kalman gain matrix.

Future plans include the extension of the proposed PC-EnKF approach to cases with spatially-varying biomass fuel conditions and time-varying wind conditions to further validate and optimize this strategy in terms of computational cost, in order to meet operational requirements.

Even though both EnKF and PC-EnKF data assimilation algorithms used in this study are effective in correcting the input parameters of the rate of spread model, the treatment of highly anisotropic uncertainties remains an important challenge for wildfire spread forecasting. Due to the operational constraints on the computational cost, the parameter estimation approach could be extended to the case of weak spatial varia-

NHESSD
2, 3289–3349, 2014

tions of the ROS model parameters. Assuming that the errors on the parameters vary slowly in time, the correction provided by data assimilation could then reasonably be used for forecast at future lead-times, thus allowing for mid- to long-term forecast. Still, there is a need to design a strategy to address spatially-distributed error correlations along the fireline in order to be able to correct the shape of the time-evolving fire front.

Acknowledgements. The financial support provided by the Agence Nationale de la Recherche under the IDEA[5] project grant ANR-09-COSI-006 (2010–2013), by the LEFE-MANU grant (INSU-CNRS program, 2011–2013) and by the 2012 CTR Summer Program[6] (Uncertainty Quantification group) is greatly appreciated. The authors also acknowledge Ronan Paugam and Martin Wooster (King's College London) for sharing the experimental data, Florent Duchaine and Thierry Morel (CERFACS) for support on Open-PALM, Albert Simeoni (University of Edinburgh) for clarifications on Rothermel's model as well as Gianluca Iaccarino and Paul Constantine (CTR) for helpful discussions on uncertainty quantification.

References

Birolleau, A., Poëtte, G., and Lucor, D.: Adaptive Bayesian inference for discontinuous inverse problems, application to hyperbolic conservation laws, Commun. Comput. Phys., 16, 1-34, 2014. 3297

Blanchard, E. D., Sandu, A., and Sandu, C.: A polynomial chaos-based Kalman filter approach for parameter estimation of mechanical systems, J. Dyn. Syst.-T. ASME, 132, 061404, 2010. 3297

Boé, J., Terray, L., Martin, E., and Habets, F.: Projected changes in components of the hydrological cycle in French river basins during the 21st century, Water Resour. Res., 45, W08426, doi:10.1029/2008WR007437, 2009. 3292

Bouttier, F. and Courtier, P.: Data Assimilation Concepts and Methods, ECMWF, Meteorological Training Course Lecture Series, March 1999.

[5] http://anridea.univ-corse.fr

[6] http://www.stanford.edu/group/ctr/SummerProgram/

Boyaval, S.: A fast Monte-Carlo method with a reduced basis of control variates applied to uncertainty propagation and bayesian estimation, Comput. Methods Appl. Mech. Engrg., 241–244, 190–205, 2012. 3294
Buis, S., Piacentini, A., and Declat, D.: PALM: a computational framework for assembling high performance computing applications, Concurrency Computat. Pract. Exper., 18, 247–262, 2006. 3317
Burgers, G., van Leeuwen, P., and Evensen, G: Analysis scheme in the ensemble kalman filter, Mon. Weather Rev., 126, 1719–1724, 1998. 3311, 3340
Butler, B., Cohen, J., Latham, D., Schuette, R., Sopko, P., Shannon, K., Jimenez, D., and Bradshaw, L.: Measurements of radiant emissive power and temperatures in crown fires, Can. J. Forest Res., 34, 1577–1587, 2004. 3298
Chong, D., Tolhurst, K. G., Duff, T. J., and Cirulis, B.: Sensitivity Analysis of PHOENIX RapidFire, Bushfire CRC, University of Melbourne, 2013. 3293
Cowlard, A., Jahn, W., Abecassis-Empis, C., Rein, G., and Torero, J.: Sensor-assisted fire fighting, Fire Technol., 46, 719–741, 2010. 3295
Crombette, P.: Optimisation et poursuite des développements du système LIVEFIRE de géolocalisation automatisée et temps réel de prises de vue aéroportées, Master thesis, Université de Toulouse (France), 2010. 3299
Cruz, M. G. and Alexander, M. E.: Uncertainty with model predictions of surface and crown fire rates of spread, Environ. Modell. Softw., 47, 16–28, 2013. 3293
Daley, R.: Atmospheric data analysis, Cambridge atmospheric and space science series, Cambridge University Press, 1991.
D'Andrea, M., Fiorucci, P., and Holmes, T. P.: A stochastic Forest Fire Model for future land cover scenarios assessment, Nat. Hazards Earth Syst. Sci., 10, 2161–2167, doi:10.5194/nhess-10-2161-2010, 2010. 3293
Després, B., Poëtte, G., and Lucor, D.: Review of robust uncertainty propagation in systems of conservation laws with the entropy closure method, in: Uncertainty Quantification in Computational Fluid Dynamics, edited by: Bijl, H., Lucor, D., Mishra, S., and Schwab, C., vol. 92 of Lecture Notes in Computational Science and Engineering, Springer-Verlag, 105–149, 2013. 3294
Durand, M., Andreadis, K., Alsdorf, D, Lettenmaier, D., and Moller, D.: Estimation of bathymetric depth and slope from data assimilation of swath altimetry into a hydrodynamic model, Geophys. Res. Lett., 35, L20401, 2008. 3297

NHESSD
2, 3289–3349, 2014

Durand, M., Lee-Lang, F., Lettenmaier, D. P., Alsdorf, D. E., Rodriguez, E., and Esteban-Fernandez, D.: The surface water and ocean topography mission: observing terrestrial surface water and oceanic submesoscale eddies, IEEE, 98, 766–779, 2010. 3297

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, J. Geophys. Res., 99, 143–162, 1994. 3295, 3297

Evensen, G.: Data Assimilation – The Ensemble Kalman Filter, Springer, 2009. 3295

Filippi, J.-B., Bosseur, F., Mari, C., Lac, C., Le Moigne, P., Cuenot, B., Veynante, D., Cariolle, D., and Balbi, J.-H.: Coupled atmosphere-wildland fire modelling, Journal of Advances in Modeling Earth Systems, 1, 210–226, 2009. 3293

Filippi, J.-B., Pialat, X., and Clements, C. B.: Assessment of ForeFire/Meso-NH for wildland fire/atmosphere coupled simulation of the FireFlux experiment, Proc. Combust. Inst., 34, 2633–2640, 2013. 3293

Finney, M. A.: FARSITE: Fire Area Simulator – model development and evaluation, Forest Service, US Dept. of Agriculture, Research Paper RMRS-RP-4, 1998. 3293

Finney, M. A., Grenfell, I. C., McHugh, C. W., Seli, R. C., Trethewey, D., Stratton, R. D., and Brittain, S.: A method for ensemble wildland fire simulation, Environmental Modeling and Assessment, 16, 153–167, 2011. 3293

Finney, M. A., Cohen, J. D., McAllister, S. S., and Jolly, W. M.: On the need for a theory of wildland fire spread, Int. J. Wildland Fire, 22, 25–36, doi:10.1071/WF11117, 2013. 3293

Fouilloux, A. and Piacentini, A.: The PALM Project: MPMD paradigm for an oceanic data assimilation software, Lect. Notes Comput. Sc., 1685, 1423–1430, 1999. 3317

Gelb, A.: Applied Optimal Estimation, Cambridge Massachusetts MIT Press, 1974. 3295

Ghanem, R. and Spanos, P.: Stochastic Finite Elements, A Spectral Approach, Dover, 1991. 3294, 3314

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

Ide, K., Courtier, P., Ghil, M., and Lorenc, A. C.: Unified notation for data assimilation: operational, sequential and variational, J. Meteorol. Soc. Jpn., 75, 181–189, 1997. 3295

Jimenez, E., Hussaini, M. Y., Goodrick, S. L.: Uncertainty quantification in Rothermel's model using an efficient sampling method, in: The fire environment–innovations, management, and policy; conference proceedings, edited by: Butler, Bret, W.; Cook, Wayne, comps, 26–30 March 2007; Destin, FL, 2007. 3293

NHESSD

2, 3289–3349, 2014

Kalnay, E.: Atmospheric Modeling, Data Assimilation and Predictability, Cambridge University Press, 2003. 3295
Lagarde, T., Piacentini, A., and Thual, O.: A new representation of data assimilation methods: the PALM flow charting approach, Q. J. Roy. Meteor. Soc., 127, 189–207, 2001. 3317
Lautenberger, C.: Wildland fire modeling with an Eulerian level-set method and automated calibration, Fire Safety J., 62, 289–298, 2013. 3295
Le Maître, O. P. and Knio, O. M.: Spectral Methods for Uncertainty Quantification with Applications to Computational Fluid Dynamics, Scientific Computation, Springer, 2010. 3294
Li, J. and Xiu, D.: On numerical properties of the ensemble Kalman filter for data assimilation, Comput. Meth. Appl. Math. Eng., 197, 3574–3583, 2008. 3297
Li, J. and Xiu, D.: A generalized polynomial chaos based ensemble Kalman filter with high accuracy, J. Comput. Phys., 228, 5454–5469, 2009. 3297
Linn, R., Reisner, J., Colman, J. J., and Winterkamp, J.: Studying wildfire behavior using FIRETEC, Int. J. Wildland Fire, 11, 233–246, 2002. 3292, 3293
Lucor, D., Meyers, J., and Sagaut, P.: Sensitivity analysis of LES to subgrid-scale-model parametric uncertainty using Polynomial Chaos, J. Fluid Mech., 585, 255–279, 2007. 3294
Mallet, V., Keyes, D. E., and Fendell, F. E.: Modeling wildland fire propagation with level-set methods, Computers and Mathematics with Applications, 57, 1089–1101, 2009. 3305
Mandel, J. and Beezley, J. D.: Morphing Ensemble Kalman filter, Tellus A, 60, 2007. 3297
Mandel, J., Bennethum, L. S., Beezley, J. D., Coen, J. L., Douglas, C. C., Minjeong, K., and Vodacek, A.: A wildland fire model with data assimilation, Math. Comput. Simulat., 79, 584–606, 2008. 3295, 3297
Mandel, J., Beezley, J. D., and Kochanski, A. K.: Coupled atmosphere-wildland fire modeling with WRF 3.3 and SFIRE 2011, Geosci. Model Dev., 4, 591–610, doi:10.5194/gmd-4-591-2011, 2011. 3297
Mell, W., Jenkins, M. A., Gould, J., and Cheney, P.: A physics-based approach to modeling grassland fires, Int. J. Wildland Fire, 16, 1–22, 2007. 3292, 3293
Merlet, N.: Evaluation des possibilités de géolocalisation automatisée et temps réel de prises de vue aéroportées dans le cadre de la lutte contre les feux de forêt, Master thesis, Université de Toulouse (France), 2008. 3299
Milly, P. C., Wetherald, R. T., Dunne, K. A., and Delwort, T. L.: Increasing risk of great floods in a changing climate, Nature, 415, 514–517, 2002. 3292

NHESSD
2, 3289–3349, 2014

Moradkhani, H., Sorooshian, S, Gupta, H., and Houser, P.: Dual state-parameter estimation of hydrological models using ensemble Kalman filter, Adv. Water Resour., 28, 135–147, 2005. 3297, 3307

Nijhuis, M.: Forest fires: burn out, Nature, 489, 352–354, doi:10.1038/489352a, 2012. 3292

Noble, I. R., Bary, G. A. V., and Gill, A. M.: McArthur's fire danger meters expressed as equations, Aust. J. Ecol., 5, 201–203, 1980.

Noonan-Wright, E. K., Opperman, T. S., Finney, M. A., Zimmerman, G. T., Seli, R. C., Elenz, L. M., Calkin, D. E., and Fiedler, J. R.: Developing the US Wildland Fire Decision Support System, J. Combust., ID 168473, 2011. 3292

Palmer, T. N., and Räisänen, J.: Quantifying the risk of extreme seasonal precipitation events in a changing climate, Nature, 415, 512–514, 2002. 3292

Paugam, R., Wooster, M. J., and Roberts, G.: Use of handheld thermal imager data for airborne mapping of fire radiative power and energy and flame front rate of spread, Geosci. Remote Sens., 51, 3385–3399, 2013. 3295, 3298, 3323

Perry, G. L. W.: Current approaches to modelling the spread of a wildland fire: a review, Prog. Phys. Geogr., 22, 222–245, 1998. 3293

Poinsot, T. and Veynante, D.: Theoretical and Numerical Combustion, 2nd edn., R. T. Edwards, 2005. 3304

Rehm, R. G., McDermott, R. J.: Fire front propagation using the level-set method, NIST, Technical Report 1611, 2009. 3305

Reichle, R. H.: Data assimilation methods in the Earth sciences, Adv. Water Resour., 31, 1411–1418, 2008. 3295

Reichle, R. H., Walker, J. P., Koster, R. D., and Houser, P. R.: Extended versus Ensemble Kalman filtering for land data assimilation, J. Hydrometeorol., 3, 728–740, 2002. 3296

Riggan, P. and Robert, G.: Airborne remote sensing of wildland fires, in: Wildland Fires and Air Pollution, Developments in Environmental Science 8, edited by: Bytnerowicz, A., Arbaugh, M., Andersen, C., and Riebau, A., Elsevier, 139–168. 3295, 3299

Rochoux, M. C., Ricci, S., Lucor, D., Cuenot, B., Trouvé, A., and Bart, J.-M.: Towards predictive simulations of wildfire spread using a reduced-cost Ensemble Kalman filter based on Polynomial Chaos approximations, Proceedings of the Summer Program, Center for Turbulence Research, July 2012, NASA AMES, Stanford University, USA, 2012. 3297, 3317

NHESSD

2, 3289–3349, 2014

Rochoux, M. C., Cuenot, B., Ricci, S., Trouvé, A., Delmotte, B., Massart, S., Paoli, R., and Paugam, R.: Data assimilation applied to combustion, C. R. Mecanique, 341, 266–276, doi:10.1016/j.crme.2012.10.011, 2013a. 3296, 3305, 3310, 3312

Rochoux, M. C., Delmotte, B., Cuenot, B., Ricci, S., and Trouvé, A.: Regional-scale simulations of wildland fire spread informed by real-time flame front observations, Proc. Combust. Inst., 34, 2641–2647, doi:10.1016/j.proci.2012.06.090, 2013b. 3296, 3310, 3312

Rochoux, M. C.: Vers une meilleure prévision de la propagation d'incendies de forêt: evaluation de modèles et assimilation de données, Ph.D. Thesis (written in English), Ecole Centrale Paris (France), 2014. 3293, 3295, 3298, 3305, 3306

Ros, D. D. and Borga, M.: Adaptive use of a conceptual model for real time flood forecasting, Nord. Hydrol., 28, 169–188, 1997. 3312

Rosić, B. V., Kucerová, A., Sýkora, J., Pajonk, O., Litvinenko, A., and Matthies, H. G.: Parameter identification in a probabilistic setting, Eng. Struct., 50, 179–196, 2013. 3297

Rothermel, R. C.: A mathematical model for predicting fire spread in wildland fuels, USDA Forest Service, Research Paper INT-115, Intermountain Forest and Range Experiment, Ogden, UT:40, 1972. 3293, 3296, 3300, 3304

Saad, G. A.: Stochastic data assimilation with application to multi-phase flow and health monitoring problems, Ph.D. Thesis, Faculty of the Graduate School, University of Southern California, 2007. 3297

Sullivan, A. L.: Wildland surface fire spread modeling, 1990–2007, 2. Empirical and quasi-empirical models, Int. J. Wildland Fire, 18, 369–386, 2009. 3293

Szunyogh, I., Kostelich, E., Gyarmati, G., Kalnay, E., Hunt, B., Ott, E., Satterfield, E., and Yorke, J.: A local ensemble transform Kalman filter data assimilation system for the NCEP global model, Tellus A, 60, 113–130, 2008. 3297

Talagrand, O.: Assimilation of observations, an introduction, J. Meteorol. Soc. Jpn., 75(1B), 191–209, 1997.

Tarantola, A.: Inverse Problem Theory, Methods for Data Fitting and Parameter Estimation, Elsevier, 1987. 3295

Thirel, G., Martin, E., Mahfouf, J.-F., Massart, S., Ricci, S., and Habets, F.: A past discharges assimilation system for ensemble streamflow forecasts over France – Part 1: Description and validation of the assimilation system, Hydrol. Earth Syst. Sci., 14, 1623–1637, doi:10.5194/hess-14-1623-2010, 2010.

NHESSD

2, 3289–3349, 2014

Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M., and Whitaker, J. S.: Ensemble Square Root Filters, Mon. Weather Rev., 131, 1485–1490, 2003.
Todling, R. and Cohn, S. E.: Suboptimal schemes for atmospheric data assimilation based on the Kalman Filter, Mon. Weather Rev., 122, 2530–2557, 1994. 3295
Viegas, D. X.: Overview of forest fire propagation research, Proceedings of the International Association of Fire Safety Science, 10, 95–108, 2011. 3292, 3293
West, M.: Approximating posterior distributions by mixture, J. R. Stat. Soc., 55, 409–422, 1993. 3307
Wiener, N.: The homogeneous chaos, Am. J. Math., 60, 897–936, 1938. 3294
Wooster, M., Robert, G., Smith, A., Johnston, J., Freeborn, P., Amici, S., and Hudak, A.: Thermal remote sensing of active vegetation fires and biomass burning events, Volume Remote Sensing and Digital Image Processing 17, Chapter Thermal infrared remote sensing, Springer, 347–390, 2013. 3295, 3298
Wooster, M. J., Roberts, G., Perry, G., and Kaufman, Y. J.: Retrieval of biomass combustion rates and totals from fire radiative power observations: FRP derivation and calibration relationships between biomass consumption and fire radiative energy release, J. Geophys. Res., 110, D24311, 2005. 3295, 3298
Xiu, D.: Numerical methods for stochastic computations – a spectral method approach, Princeton University Press, Princeton, NJ, 2010. 3297
Xiu, D., and Karniadakis, G.: The Wiener–Askey polynomial chaos for stochastic differential equations, SIAM Journal on Scientific Computing, 24, 619–644, 2002. 3314

NHESSD
2, 3289–3349, 2014

Table 1. Input parameters of the Rothermel-based ROS sub-model.

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

NHESSD
2, 3289–3349, 2014

Table 2. PC-EnKF-based experiment for the controlled grassland fire experiment: error statistics (in the parameter space) of the forecast and analysis ensemble estimates for $\boldsymbol{x} = [M_v, \Sigma_v]$ ($n = 2$). The number of FIREFLY integrations is also presented as indicator of the computational cost.

FORECAST	Cost	Ens. mean	Ens. STD
PC-EnKF	25	15.0 % 11 500 m^{-1}	4.0 % 3000 m^{-1}
EnKF	1000	15.0 % 11 500 m^{-1}	4.0 % 3000 m^{-1}
ANALYSIS	Cost	Ens. mean	Ens. STD
PC-EnKF	25	13.8 % 22 583 m^{-1}	1.4 % 1157 m^{-1}
EnKF	1000	13.5 % 22 345 m^{-1}	1.4 % 1170 m^{-1}

NHESSD
2, 3289–3349, 2014

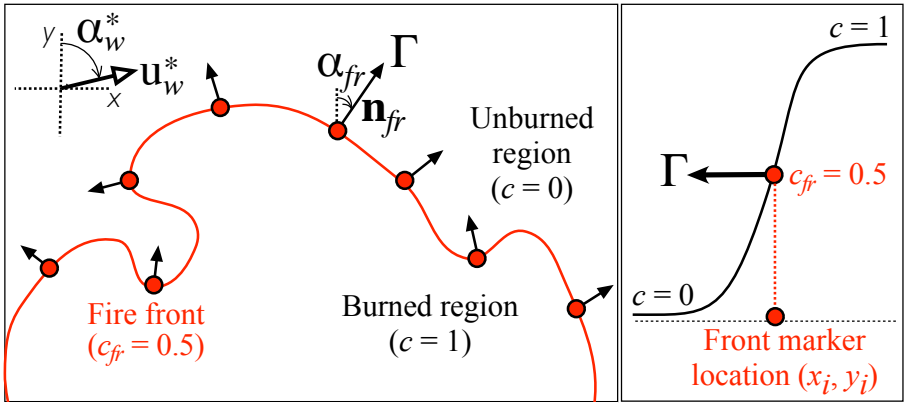
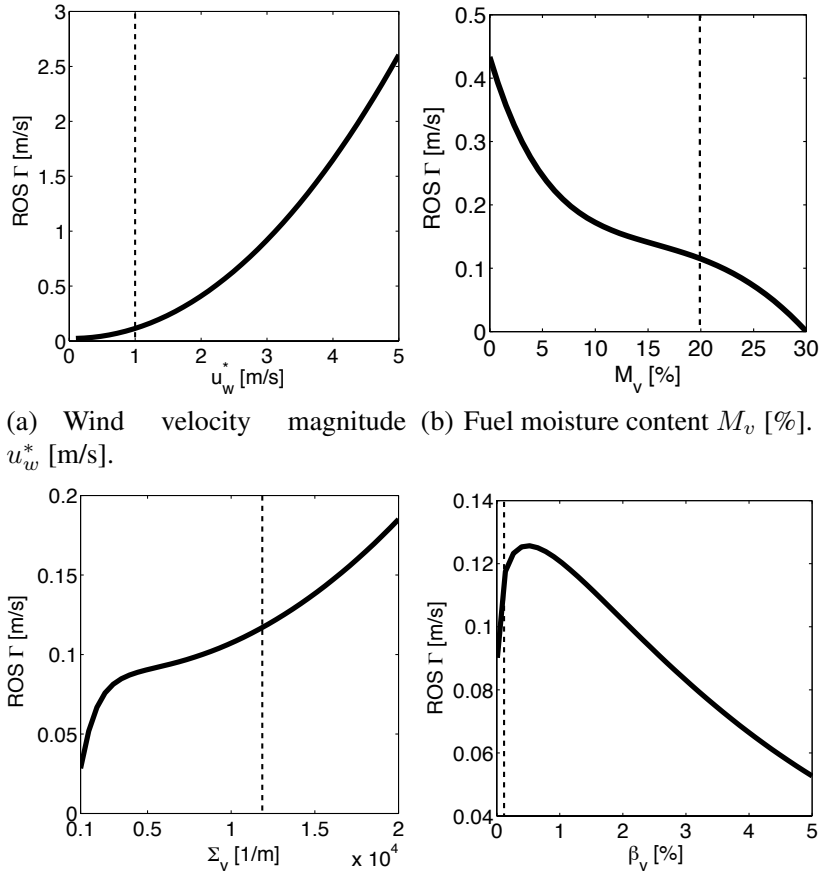


Fig. 1. Level-set-based front-tracking simulator FIREFLY. Left: The fire front is the isoline $c_{\mathrm{fr}} = 0.5$; Γ measures the local ROS of the fire along the normal direction to the front $\boldsymbol{n}_{\mathrm{fr}}$ (defined by the direction angle of fire propagation α_{fr}) given the wind velocity vector $(u^*_{\mathrm{w}}, \alpha^*_{\mathrm{w}})$. Right: Profile of the spatial variations of the progress variable c across the fire front, (x_i, y_i) representing the location of the ith fire front marker.

NHESSD
2, 3289–3349, 2014



(a) Wind velocity magnitude u_w^* [m/s].

(b) Fuel moisture content M_v [%].

(c) Fuel particle surface-to-volume ratio Σ_v [1/m].

(d) Fuel packing ratio β_v [%].

Fig. 2. Sensitivity of the Rothermel-based ROS Γ to environmental parameters; nominal conditions are indicated by vertical lines.

NHESSD
2, 3289–3349, 2014

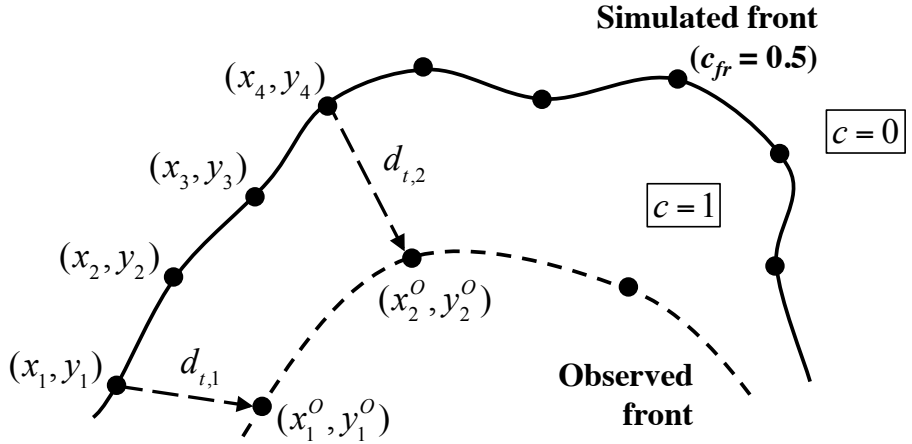


Fig. 3. Construction of the differences between simulated fire front (SFF) and observed fire front (OFF) noted $\mathbf{d}_t = [\mathrm{d}_{t,1}, \cdots, \mathrm{d}_{t,N_{\mathrm{fr}}^{\mathrm{o}}}]$. In this illustration, $r = N_{\mathrm{fr}}/N_{\mathrm{fr}}^{\mathrm{o}} = 4$.

NHESSD
2, 3289–3349, 2014

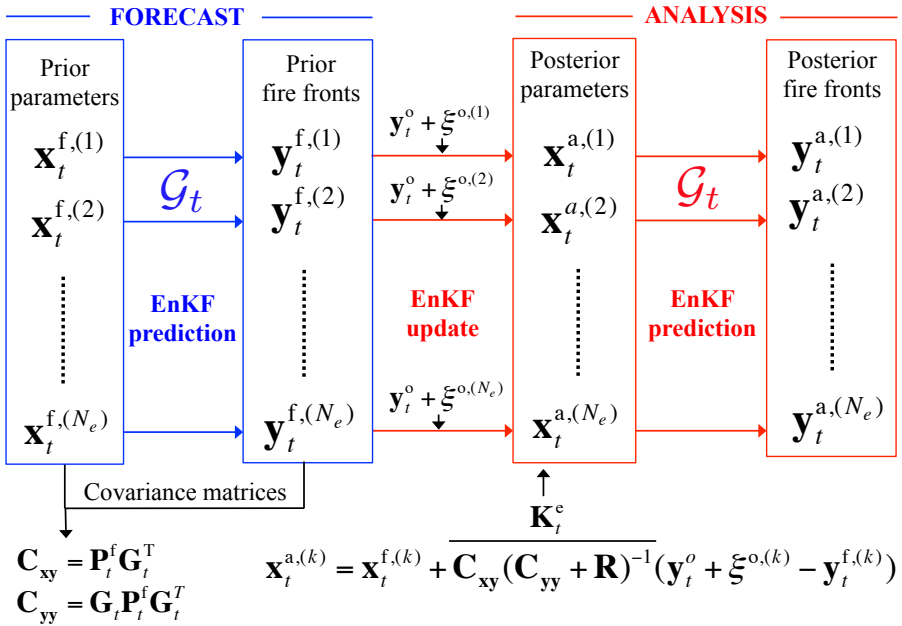


Fig. 4. Flowchart of the EnKF algorithm during the $[t-1, t]$ assimilation cycle for a parameter estimation approach. Data randomization (Burgers et al., 1998) is used in the EnKF with $\xi^{o,(k)}$ following observation error statistics for each member $k = 1, \cdots, N_e$.

NHESSD
2, 3289–3349, 2014

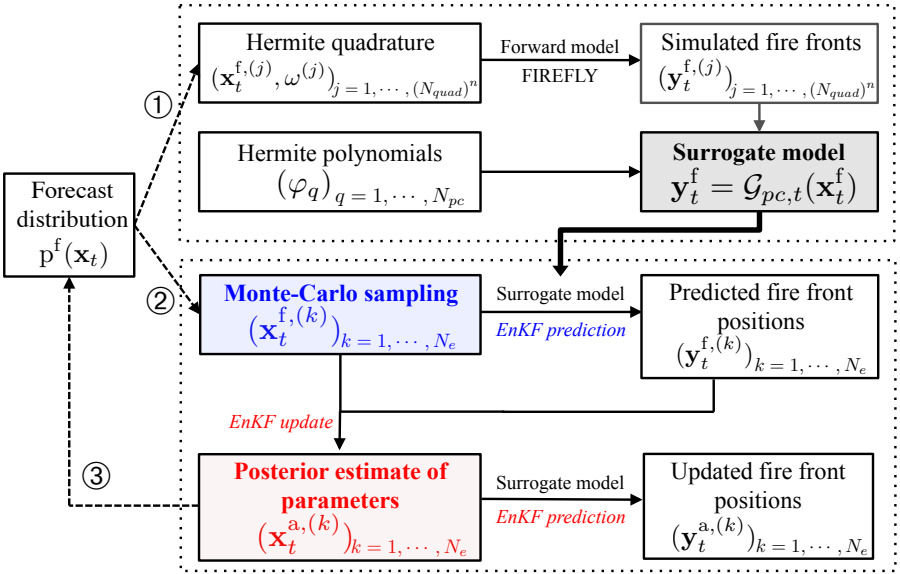


Fig. 5. Flowchart of the PC-EnKF algorithm during the assimilation cycle $[t-1, t]$ decomposed into 3 steps: (1) construction of the PC expansion of the generalized observation operator; (2) EnKF prediction and update for the assimilation cycle $[t-1, t]$; and (3) parameter evolution to the next assimilation cycle $[t, t+1]$.

NHESSD
2, 3289–3349, 2014

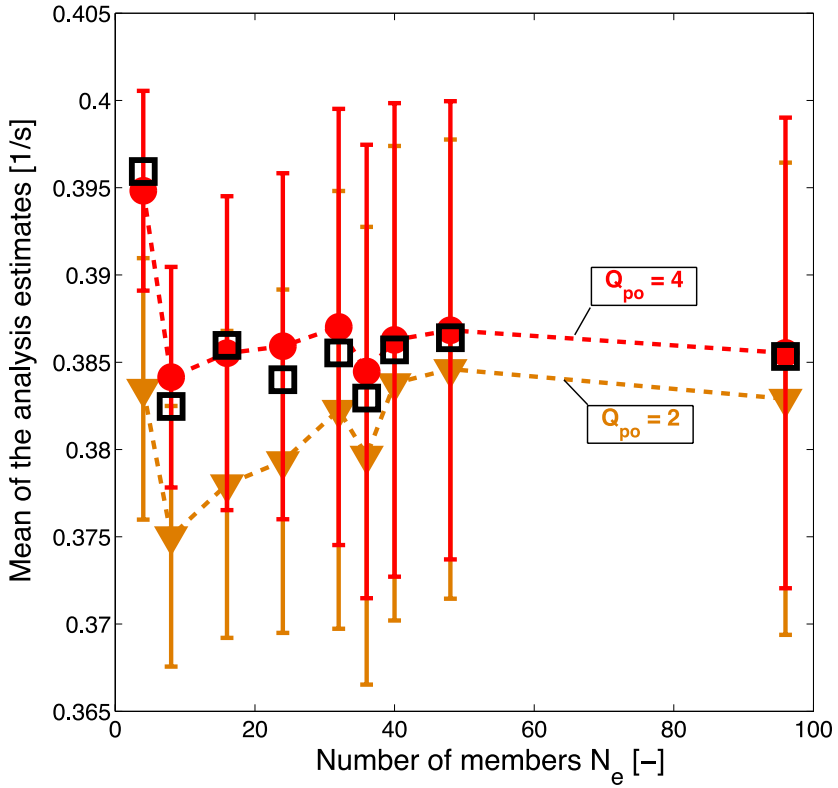


Fig. 6. Convergence of the mean analysis estimates of the proportionality coefficient P [s^{-1}] with respect to the number of ensemble members N_e for a fixed observation error STD $\sigma^o = 2$ m and a single assimilation cycle: comparison of the performance between the EnKF and PC-EnKF algorithms. The orange triangled-dashed line corresponds to $Q_{po} = 2$; and the red circled-dashed line corresponds to $Q_{po} = 4$ for the PC-EnKF algorithm. Black squares correspond to the analysis estimates obtained using the standard EnKF. Vertical error bars correspond to the associated error STD.

NHESSD
2, 3289–3349, 2014

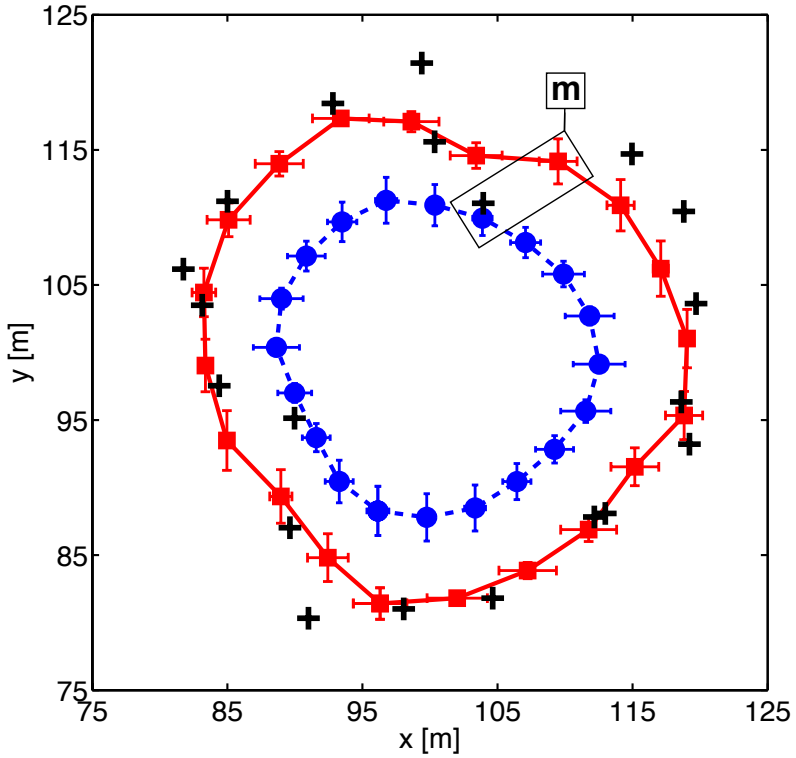
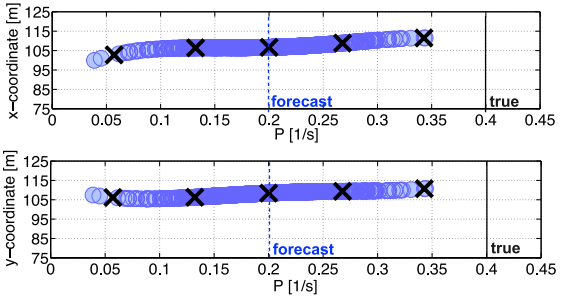
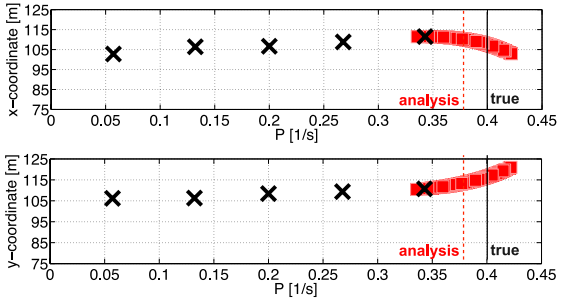


Fig. 7. Comparison of fire front locations using the PC-EnKF approach with an EnKF update at 50 s for $\sigma^{o} = 2$ m, $Q_{po} = 4$ and $N_e = 1000$; estimation of the proportionality coefficient P [s^{-1}]; all fronts correspond to time 50 s. Black crosses correspond to observations; the blue circled-dashed line corresponds to the mean forecast estimate of the fire front and the red squared-solid line corresponds to the analysis counterpart. Horizontal and vertical error bars correspond to the associated error STD along the x- and y-directions, respectively. The location of the fire front marker indexed by the character m is indicated.

NHESSD
2, 3289–3349, 2014



(a) Forecast estimates of the proportionality coefficient P [1/s].



(b) Analysis estimates of the proportionality coefficient P [1/s].

Fig. 8. Model surface response (or surrogate model) of the x- and y-coordinates of the front marker indexed by m on the fireline (see Fig. 7), with respect to the control parameter P [s^{-1}]. Black crosses correspond to quadrature roots (i.e., FIREFLY forward model integrations); blue circles correspond to **(a)** forecast estimates and red squares to **(b)** analysis estimates obtained through the PC-EnKF algorithm at time 50 s. The vertical solid line indicates the true value $P^t = 0.4\,s^{-1}$; the vertical dashed lines indicate the mean forecast and analysis estimates of the proportionality coefficient P [s^{-1}].

NHESSD
2, 3289–3349, 2014

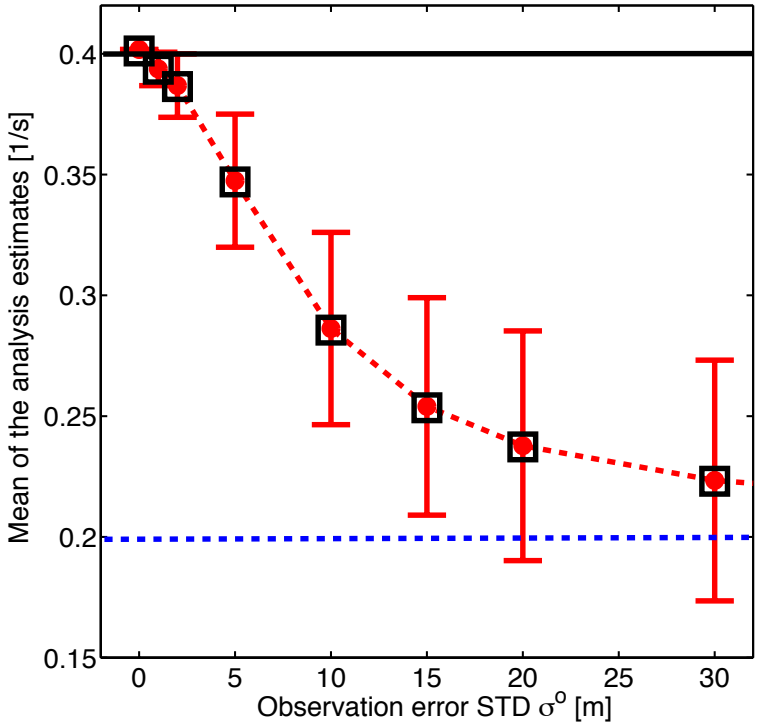
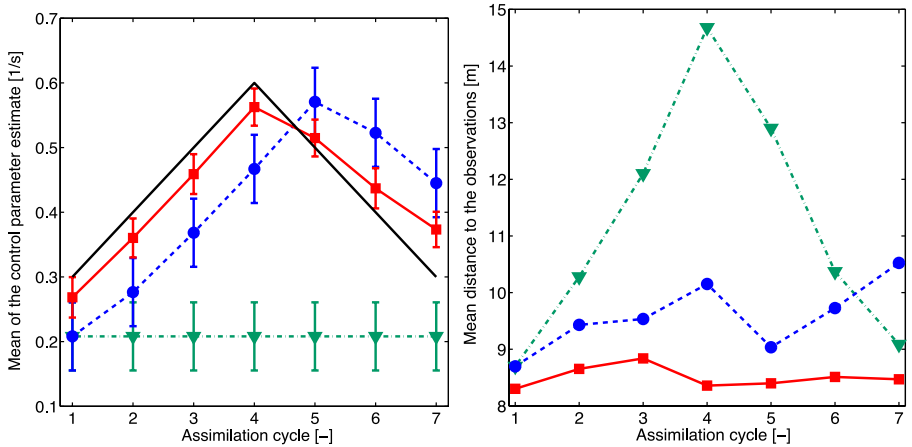


Fig. 9. Mean and STD of the analysis estimates of the proportionality coefficient P [s^{-1}] as a function of the observation error STD σ^o for a fixed number of members N_e = 48 and for one assimilation cycle (with an EnKF update at 50 s): comparison between EnKF and PC-EnKF. The black solid line corresponds to the true value 0.4 s^{-1}; the blue dashed line corresponds to the mean value of the forecast 0.2 s^{-1}; and the red circled-dashed line corresponds to the mean analysis estimate obtained using the PC-EnKF algorithm. Black squares correspond to the mean analysis estimates obtained by the standard EnKF. Vertical error bars correspond to the associated error STD.

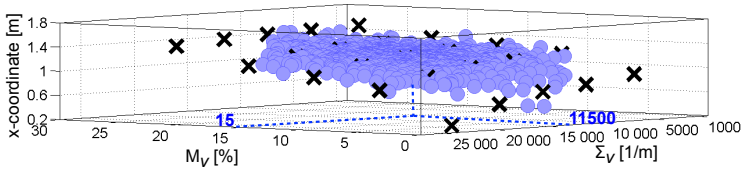
NHESSD
2, 3289–3349, 2014



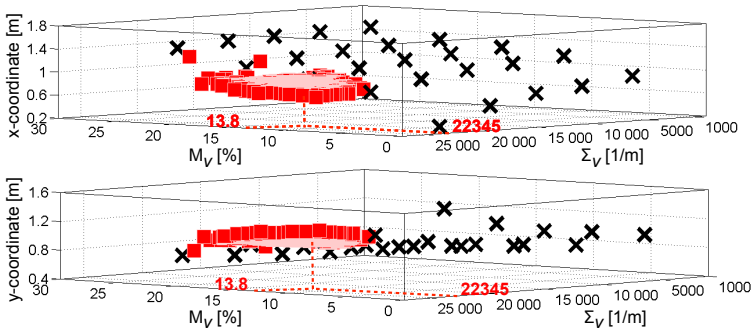
(a) Control parameter space. (b) Observation space.

Fig. 10. Sequential EnKF estimation of the coefficient P [s^{-1}] over 7 assimilation cycles with $N_e = 48$ members and $\sigma^o = 5$ m; time-varying true control parameter. The green triangled-dashed-dotted curve corresponds to the free run (without DA); the blue circled-dashed curve corresponds to the mean forecast estimate; the red squared-solid curve corresponds to the mean analysis estimate; and the black solid line corresponds to the true control parameter. **(a)** Parameter estimates (vertical error bars correspond to the associated error STD). **(b)** Mean distance to the observed fire front.

NHESSD
2, 3289–3349, 2014



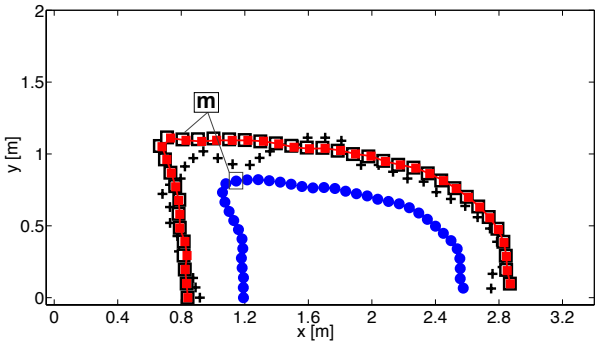
(a) Forecast estimates.



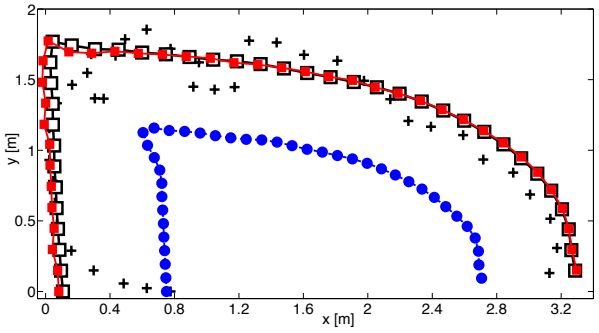
(b) Analysis estimates.

Fig. 11. Model surface response (or surrogate model) of the x- and y-coordinates of the front marker indexed by m on the fireline (see Fig. 12) with respect to the control vector $\boldsymbol{x} = [M_v, \Sigma_v]$ ($n = 2$). Black crosses correspond to quadrature roots (FIREFLY integrations). **(a)** Forecast estimates (blue circles) and **(b)** analysis estimates (red squares) of the x- (top) and y-coordinates (bottom) of the fire front positions are mapped onto the PC-based model surface response.

NHESSD
2, 3289–3349, 2014



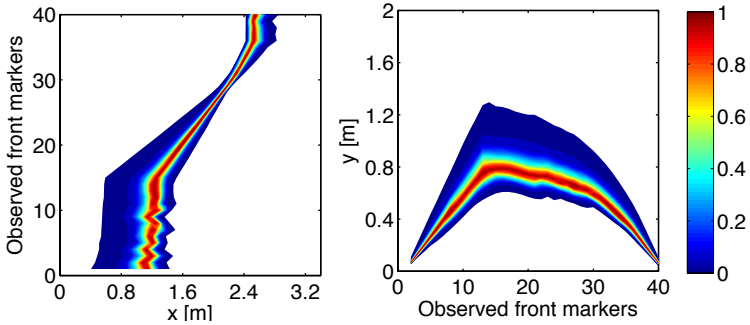
(a) Analysis time, $t_1 = 78$ s.



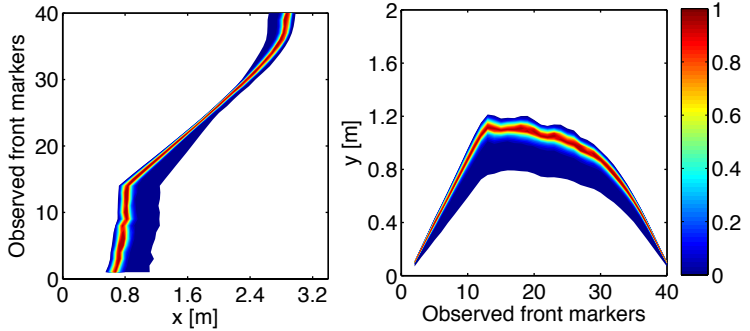
(b) Forecast time, $t_2 = 106$ s.

Fig. 12. Comparison between simulated and measured fire front positions for the controlled grassland fire experiment: black crosses correspond to observations, the blue circled-dashed line corresponds to the mean forecast estimate constructed through the PC-based surrogate model; the red squared-solid line corresponds to the mean analysis estimate obtained by the PC-EnKF procedure applied at time t_1 = 78 s. Black squares correspond to the standard EnKF used as reference. The location of the fire front marker indexed by m is indicated.

NHESSD
2, 3289–3349, 2014



(a) Forecast PDF with respect to the x- (left) and y- (right) coordinates of the observed fire front markers.



(b) Analysis PDF with respect to the x- (left) and y- (right) coordinates of the observed fire front markers.

Fig. 13. Colormap of the PDF of the fire front marker locations (in terms of *x* and *y*- coordinates) for the controlled grassland fire experiment at the analysis time t_1 = 78 s. **(a)** PDF related to the ensemble of forecast estimates. **(b)** PDF related to the ensemble of analysis estimates.

NHESSD
2, 3289–3349, 2014