Rebuttal for NHESS Discussion Manuscript

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 by M.C. Rochoux et al.

We appreciate the positive and constructive comments made by the Reviewer. Detailed answers are given below.

① **"An introductory sentence concerning the fact that this is a feasibility study could be appreciated by readers"**

▶ Consistently with the Reviewer's comment, we propose to add a sentence in the Introduction section (p. 3297): *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 rate of spread (ROS) model. The objective of this study is to show the feasibility of this approach for wildfire spread forecasting under several assumptions, i.e., a minimalist treatment of the fire front (idealized as an interface and consistent with the limited knowledge on the environmental conditions); a semi-empirical formulation of the ROS; Gaussianity of the errors on the input parameters and on the observations; the prior values for the control parameters are specified based on user-defined mean and error standard deviation.*

② **"a brief description of results obtained and discussed in the second part of the paper [would be valuable] to give a more comprehensive presentation of the research."**

▶ Consistently with the Reviewer's comment, we propose to add a sentence in the Introduction section to summarize the objective of the second part with respect to this first part (p. 3297): *For this purpose […] (Birolleau et al., 2014). In this first part, both the EnKF and PC-EnKF algorithms are limited to the estimation of spatially-uniform parameters of the rate of spread model due to computational cost constraints and a lack of high-resolution data on the environmental conditions. In contrast, in the second part of this series of two articles (reference), a state estimation strategy is designed to address anisotropy uncertainties in wildfire spread and to correct the shape of the fire front for forecast initialization. Thus, parameter estimation and state estimation are complementary approaches that are valuable for wildfire behavior forecasting; it is therefore important to discuss their benefits and drawbacks for tests with increasing complexity.*

③ **"the explanation of some chosen values of parameters can also be given."**

▶ We agree with the Reviewer that it is important to clarify how the setting of a data assimilation experiment is done. That is why we propose to add a generic comment in

Section 3.1.2 (p. 3310) about the modeling of the error statistics of the input parameters: *As shown in Fig. 4, the forecast control parameters are stochastically represented at time t […] with k varying between 1 and N_e. These realizations are randomly-generated based on mean and error standard deviations according to user-defined confidence interval for each control parameter over the first assimilation cycle and to previous analysis results for next assimilation cycles.*

④ "a longer explanation of figure 9 and also of chosen values x^t = 0.4 1/s and x^f = 0.2 1/s"

▶ We agree with the Reviewer that it is important to clarify the meaning of x^t and x^f in a synthetic data assimilation experiment such as those reported in Section 4.1, but we feel that figure 9 is fully described by the paragraph in Section 4.1.3, p. 3321-3322. With regards to the comment on x^t and x^f, we propose to add the following sentence at the beginning of Section 4.1 (p. 3318-3319): *The ensemble of prior values is drawn from a Gaussian distribution centered in x^f = 0.2 1/s with an error STD σ^f = 0.05 1/s (assumed constant along the assimilation cycles). Note that the true value of the control parameter x^t is at the tail of the Gaussian PDF associated with the forecast estimates. This case is chosen on purpose, in order to evaluate the capability of the parameter estimation approaches to retrieve accurate values of the control parameter, even though the prior value is far from the true control parameter and its uncertainty (with respect to the observation uncertainty) is high.*

⑤ I wonder if there is the possibility of an automatic best selection of some parameters while the code runs.

▶ From one assimilation cycle to the next, there is indeed the possibility to change the parameters that are included in the control vector according to the level of information available (the uncertainties in the surface wind or in the biomass fuel properties are not time-invariant). Automatic sensibility tests could be performed, *a priori* or along the assimilation cycles, to modify the control vector if required. We propose to add a comment in the Conclusion section on this point.

⑥ "Authors could briefly discuss analogies and differences with this approach [the approach proposed by Pagnini et al.]"

▶ We agree with the Reviewer that the proposed approach only addresses uncertainties in the input parameters of the forward model. In addition, Kalman filters are designed to specifically address Gaussian error statistics since the Kalman update equation (see Eq. 20) is obtained by assuming Gaussian error statistics for the control variables. One advantage of the ensemble Kalman filter over the classical or extended Kalman filter is that it is able to account for non-linearities in the forward model via the ensemble of forward model integrations to generate the members. Still, it relies on the assumption of Gaussian error statistics. In order to evaluate the impact of this assumption on the filter results, a recent study by da Silva et al. (paper under revision) was performed with

particle filters applied to FIREFLY, showing very similar results between the ensemble Kalman filter and particle filters. Thus, the assumption of Gaussian error statistics seems valid in the present study.

It is correct that the ensemble Kalman filter and the stochastic approach proposed by Pagnini and Mentrelli (2013)* are complementary and could be combined to address both epistemic and aleatoric errors. We propose to mention this study in the introduction to complete the literature review: *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 (that can be addressed by stochastic models, see for instance Pagnini and Mentrelli, 2013, that relies on a stochastic component to represent the transport of firebrands).*

While very interesting, the comparison between the two approaches is not further discussed since the authors believe that the paper is already rich in complex concepts with the development of the polynomial chaos strategy and with the objective to detail different techniques to address uncertainties in environmental conditions. Furthermore, the second part of this series of two papers addresses a state estimation approach that it is able to account for all the possible sources of uncertainty in FIREFLY, i.e., in the input parameters of the rate of spread model as well as in the parameterization of the rate of spread. To summarize the key aspects of all this discussion, we propose to add the following comment at the end of the Conclusion section: *There is also a need to address all possible sources of uncertainty in the fire spread model, not only in the input parameters of the rate of spread model but also in the parameterization of the rate of spread that is limited in scope due to a lack of physical modeling (e.g., steady-state assumption, transport of firebrands). It is worth mentioning that the second part of this series of two papers is dedicated to the evaluation of a state estimation approach that is able to account for both anisotropic uncertainties and modeling uncertainties. While out of the scope of this series of two papers, a proper representation of the model errors could be performed by introducing a model error covariance matrix in the ensemble Kalman filter (Trémolet, 2007*); a stochastic model such as introduced by Pagnini and Mentrelli (2013)* could be useful to describe this model error covariance matrix.*

*Pagnini G., Mentrelli A., Modelling wildland fire propagation by tracking random fronts. Nat. Hazards Earth Syst. Sci. Discuss. 1, 6521–6557 (2013).

*Trémolet, Y. (2007b). Model-error estimation in 4D-Var. Quarterly Journal of the Royal Meteorological Society 133, 1267–1280.

⑦ "I think that a better explanation of [the parameter r] should be given"

▶ The role of the parameter r is explained in Section 2.2.3 entitled *Reconstruction of the simulated fire front and comparison with the observed fire front* (p. 3305-3306). Let me first summarize the main ideas behind this parameter r.

Within the framework of data assimilation, there is a need to compute the distance between the simulated fire front and the observed fire front. In the present study, we assume that the observed fire front is discretized with a finite number of segments. Furthermore, as mentioned in the paper, it is expected that the observed fire front is provided with a much coarser resolution than the simulated fire front. Multiple synthetic cases reported in the PhD thesis of the first author (Rochoux, 2014) showed that a simple selection algorithm provides a reasonable answer to this problem by comparison to a projection algorithm. Thus, the computation of the distance between simulated and observed fire fronts relies on the following steps:

1) to discretize the simulated contour line $c_{fr} = 0.5$ with a finite number of markers according to the model resolution (N_{fr}), see Eq. 8
2) to discretize the observed fire front with a finite number of markers according to the observation resolution (N_{fr}^o), see Eq. 9
3) to identify to which simulated markers the observed fire front markers can be paired through the selection procedure: 1 out of $r = N_{fr}/ N_{fr}^o$ markers is taken along the simulated fire front. The resulting N_{fr}^o markers selected along the simulated fire front correspond to the closest neighbors of the observed front markers along their trajectory over time.

The authors propose to summarize the main steps of this algorithm at the end of Section 2.2.3 (p. 3305-3306) to clarify this selection procedure and to mention the assumptions/limitations underlying the current selection procedure:

- *[In the following, we assume for simplicity that …, where r is an integer taking values much larger than 1]. In this context, the global parameter r represents the difference in resolution between the simulated fire front and the observed fire front that is crucial to pair the simulated and observed markers.*

- *One of the advantages of this representation of the simulated and observed fire fronts is that it provides a local information on the discrepancies between simulated and observed fire fronts and not only a global information such as the difference in the burnt area or in the fireline perimeter. This local information is efficient at representing the anisotropy in wildfire spread. Still, the topology of the fire front can be complex in real-world wildfire spread cases, and/or only a section of the fire front can be observed due to the opacity of the fire-induced thermal plume or due to a limited monitoring. Thus, the pairing between simulated markers and observed markers becomes more challenging for complex fire front topologies. The generalization of this treatment to complex fire front topology will be revisited in future work. Projection schemes reported in Rochoux (2014) are expected to provide a valuable answer to this issue and*

could be integrated to the proposed data assimilation algorithms. However, this issue is out of the scope of this study that aims at showing the feasibility of data assimilation for wildfire spread forecasting.

To conclude, the minor comments pointed out by the Reviewer will be taken into account in the new version of the manuscript except for the comment no. 2: the notations for the simulated fire front (Eq. 8) and the notations for the observed fire front (Eq. 9) are different on purpose. The objective is to distinguish the model outputs (the front with N_{fr} markers) to the model counterparts of the observations (the front with N_{fr}^o markers) to avoid confusion in the description of the filter.