
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) “need for a reduced-cost EnKF. Why? There are many examples where EnKF is applied to the very complex models and does not require large ensembles (30 to 50 members)”

► Tests have shown that in the context of wildfire spread, a large number of members are required in the ensemble to retrieve accurate error statistics on the control parameters (these tests can be found in the first author’s recently published PhD Thesis, which is in the list of references, see Rochoux, 2014). This behavior of the ensemble Kalman filter in the context of parameter estimation is due to three main reasons: (1) the non-linear interrelation between the control space and the observation space; (2) the complexity of retrieving the specific signature of each control parameter on the resulting distribution of the simulated fire front (the required number of members in the ensemble increases with the number of parameters in the control vector); (3) the accumulation of sampling errors along assimilation cycles that can only be addressed by increasing the size of the ensemble as pointed out in reference by Li and Xiu (2009) already cited in our paper.

We believe that the Reviewer mainly alludes to state estimation problems in the comment. In these problems, the control vector is composed of the spatially-distributed values of the variables of interest. These values of the control variables are spatially-correlated through the forecast error covariance matrix: these correlations in the state estimation approach provide constraints to the estimation problem. In this context, the number of ensemble members required to obtain accurate error statistics is of the order of magnitude mentioned by the Reviewer (30 to 50 members). In Part II of this series of papers, a state estimation approach has been indeed developed for wildfire spread forecasting; it is shown that the number of members in the ensemble required to reach convergence is considerably lower than for that of the present parameter estimation approach due to the reasons listed above.

It is worth noting that running 30 to 50 members in the ensemble is already an expensive task in operational frameworks. This motivates the investigation and development of a reduced-cost ensemble Kalman filter both for a parameter estimation approach and for a state estimation approach (this approach is

presented in Part II of this series of papers). The reference to the first author's PhD thesis could be added in the corresponding paragraph (page 3297).

(2) “Why do you assume that there is no correlation between the observational errors?”

► We acknowledge that the classical assumption of uncorrelated observational errors is questionable for the current application. While not preventing the data-driven simulator prototype from providing efficient corrections, this aspect needs to be further investigated. Future plans include addressing the correlations of observational errors along the fireline at a given time, similarly to spaceborne data along the pass of polar-orbiting satellites (Brankart et al., 2009; Gorin and Tsyrlunikov, 2011). The authors propose to add a comment on this aspect in the paper. However, note that the procedure of selecting front markers with a low number of markers ($N_{fr}^o \ll N_{fr}$) can be regarded as a filtering procedure that tends to reduce the spatial correlations between the observed front markers. We therefore believe that this assumption does not significantly deteriorate the results in the experiments presented in the paper.

(3) “What does it mean for beta to be optimum?”

► β corresponds to the fuel layer packing ratio, which is an input parameter to the Rothermel-based model of rate of spread. It is important to mention that the value of β has a direct impact on the combustion processes and thereby, on the flame reaction intensity. These processes are parameterized in the Rothermel's semi-empirical model. In this context, the optimum packing ratio is also a model parameter characterizing the optimum arrangement of the biomass fuel that produces the most effective mixing between air and fuel gas reactants for combustion given a fuel particle size. We propose to clarify the meaning of the optimum value of β in order to avoid confusion with the idea of optimization behind the data assimilation procedure presented in the paper (page 3301).

(4) “The flame is the region where c takes values between 0 and 1. Could you explain this approach?”

► As explained in Sect. 2.2.2., the prognostic variable of the FIREFLY simulator is the two-dimensional progress variable $c = c(x,y,t)$. The location of the simulated fire front is then defined as the contour line $c_{fr} = 0.5$. Thus, a single front is defined through the application of a “simple iso-contour algorithm” after the integration of Eq. (7) up to the next observation time. We propose to remove the general sentence “The flame is the region where c takes values between 0 and 1” that might confuse the reader. A diagnostic (that was not shown in the paper but that can be found in Rochoux et al. 2013a and that is detailed in Rochoux 2014) has been performed to check the thickness of the front over time; the thickness of the front $c_{fr} = 0.5$ was shown to remain thin, constant and negligible with respect to the size of the fire.

(5) “I do not understand the way the ensemble Kalman filter seems to be applied. There is a clear need for clarification.”

► As noted by the Reviewer, the control vector x_t only includes the parameters to be controlled, not the model state c_t . This is mentioned at the beginning of Section 3.1 (p. 3306). The full answer to this comment is divided into 3 parts presented below.

(5.1.) “In equation 10 where is the state in this equation? [...] What about the position of the front at (t-1)? ”

► We agree that Eq. (10) does not include all the elements required for the time-integration of the forward model but this was for clarity purposes. Since this formulation seems to introduce confusion in the definition of the control vector, we propose to modify the formulation of Eq. (10) in the paper (p. 3307), consistently with Eq. (8) and with all the dependencies, as follows:

$$y_t = G_t(x_t) = H_t \circ M_{[t-1,t]}(c_{t-1}, \lambda', x_t)$$

where c_{t-1} is the initial condition of the progress variable at time (t-1), where λ' are the input parameters of the Rothermel-based model that are not controlled, and where x_t is the vector of controlling parameters.

(5.2.) “You update only the control parameters in the analysis step and I do not understand how this is a legal application to the filter”

► As noted by the Reviewer, the ensemble Kalman filter only corrects the control parameters, the position of the fire front being indirectly modified by integrating again the model with the new set of parameters following Eq. (8). This is a standard application of the EnKF for parameter estimation as done by Moradkhani et al. (2005) or Durand et al. (2008, 2010) for hydrology applications. Note that these references are already mentioned in the paper.

(5.3.) “it is not clear what is the initial front at time (t-1) for each ensemble member? I believe the analysis front positions should be computed by the filter at time t, along with the analysis control vectors at time t. If not, explain how you can skip this.

► The idea underlying parameter estimation is to obtain more accurate statistics for the control parameters over the time period [t-1,t] starting from an initial condition of the progress variable c at time (t-1). Here are the main steps over the assimilation cycle [t-1, t]:

- (1) build an ensemble of forecast control parameters based on Eq. (23), starting from the progress variable field corresponding to the mean analysis field obtained at time (t-1);

- (2) integrate the observation Eq. (10) that includes the model integration from time $(t-1)$ to time t to obtain the model counterparts of the observations at time t ;
- (3) apply the Kalman filter equation at time t for each member of the ensemble based on Eqs. (26-28);
- (4) re-integrate the model Eq. (8) with the analysis parameters over the time period $[t-1,t]$ to obtain the corrected locations of the fire front and the updated progress variable field at time t .

Restart step (1) for next assimilation cycle $[t, t+1]$, the integration of the model starts again from the analysis locations of the fire front at time t with the modified control parameters following the random walk model (see Eq. (23)). In this context, the evolution of the state variable (i.e., the progress variable field) is a consequence of the update of the control variables.

We believe that with the clarification of equation (10), this algorithm will be explained with enough details in the paper.

(6) “Figure 6: The error bars are narrower for lower numbers of members. Why?”

- It is indeed possible that the standard deviation computed with a low number of members is not reliable due to non-converged error statistics. This is why it is recommended in the paper to include at least 40 members in the ensemble. The authors propose to add a comment in the paper (p. 3319)