

Anonymous Referee #1

Review

This paper addresses a data assimilation problem related to flood forecasting. It shows that with a linear 1D flood wave propagation model, and under certain conditions of stationarity, a simplified EnKF can be used instead of a full EnKF, with almost no loss of performance. In the simplified EnKF, a surrogate of the (ensemble-based) covariance matrix is formed based on smart theoretical arguments. The paper is quite well written, although the abstract is a little bit too long and redundant with summaries given in the introduction and the conclusion sections. I strongly recommend to shorten the abstract and replace the summary in the introduction by more relevant material, see below.

Answer 1:

We have shortened the abstract and we propose a new version of the abstract and introduction sections that avoids redundancy and contain more relevant material as recommended.

My main concerns are related to the motivations and the novelty of this work.

The authors argue that the main motivation for this work is to reduce the computational burden of a full EnKF. This is surprising to me. First, as mentioned page 4, line 24, the ultimate models to be used for realistic flood forecasting are 1D or 2D shallow-water models. As far as I know, running an ensemble of $O(10)$ members is not very challenging with such models.

Answer 2:

This study is oriented toward the application of data assimilation to real-time flood forecasting. In the context of operational flood forecasting running $O(10)$ members is not always feasible because of real-time constraints and computational cost.

In order to answer Referee #1's remark about the $O(10)$ number of members, we now specify in the manuscripts that $O(10000)$ members are necessary to obtain less than about 1% accuracy compared to the asymptotic limit of an infinite number of members.

Also, the largest part of the computational complexity comes from the ensemble propagation, not from the analysis. The proposed method may help to reduce the computation time by a few percent, but this should be negligible. An argument, used page 16 line 1, is that for such problem a single forecast is carried out. That indeed answers the question.

Answer 3:

The use of $O(N)$ members is N times more expansive than the computational cost of one forecast model run, not a "few percent". Indeed, the EnKF model requires the integration of the model during a time separation to subsequent observations. Once the background error covariance matrix is parametrized we use (in the EEnKF) only one model run. We now state this point more clearly.

But this is again surprising, since such deterministic approach does not enable to identify the risks of extreme events, which floods are by nature. But if the assertion of the single forecast is actually true, it should be supported with scientific justification and appropriate references.

Answer 4:

As far as the risks of extreme events are concerned, both the EnKF and our EEnKF methods provide the matrix of analysis error correlations, which give valuable statistical information for the validity of the forecast.

Moreover, the parameterisation of the background error covariance matrix we exhibit here is relevant as long as the standard deviation of the error of the upstream forcing is steady.

Most of the introduction is dedicated to the description of the work presented in the paper, missing the essential introduction of the context: How data assimilation and forecast are performed in other flood forecasting systems, with a particular emphasis on those using the EnKF; The usual methods to form covariance matrices in Optimal Interpolation systems - with an emphasis on the method chosen in this paper. Using the asymptotic behaviour of the covariance matrix to simplify the Kalman filter implementation is indeed not new. The lack of references to past works on this subject is astonishing, especially because it is at the core of the present work. I cite Fukumori et al (1993) and Gelb (1974) for example.

Answer 5:

The simplification of Kalman filters through asymptotic analysis of the covariance is indeed not new and we have added more reference in the manuscript to emphasise this statement and described the various techniques that were used.

In fact, I wonder what the novelty of this paper is. Implementing an EnKF with a low dimensional linear model is not challenging, and the method described here to make it more efficient is not new. If the real novelty is the parametrization of the matrix, particularly using the diffusion operator, this must be put upfront in the abstract and the introduction. The introduction must also describe what the other authors do to form their covariance matrix.

Answer 6:

The novelty of our work is indeed the parameterisation of the covariance matrix summarised in the abacus of Figure 7, as well as a thorough validation of this method. Such a validation has never been done on this advection-diffusion problem. The generic nature of this model makes this parameterisation and this validation useful for more realistic and applied systems.

The use of a diffusion operator to compute the covariance matrix given its parameterisation is not a novelty in the present paper and is only presented as technical detail.

Moreover the main problem with an invariant Kalman filter algorithm is to define covariance functions consistent with the model because those functions spread the information brought by the observation over the domain. Many studies are about the application of data assimilation for flood forecasting purposes (Neal *et al.* (2007), Shiiba *et al.* (2000) or Madsen and Skotner (2005) for example) but none of them mention nor characterise the shape of the covariance functions. In

Madsen and Skotner (2005) the authors exhibit (without justifications) a panel of covariance functions that they use in their invariant Kalman filter algorithm and compare which one performs the best with data assimilation. One of those functions has a particular shape that looks like an anisotropic gaussian function and performs well with assimilation. Hence one goal of our study was to justify the shape of this function in the special case of the flood wave propagation model.

Other minor comments

- Page 4 line 7: Localization can indeed help in reducing the computational complexity of the stochastic EnKF, but not the deterministic one. Saving time is not the main purpose of localization. And again, the analysis remains cheap compared with the ensemble propagation.

Answer 7:

Again, the analysis computation requires integrating the model during a time separating two observations times in order to be able to propagate the errors covariance matrices, which is at the root of Kalman filter principles.

- In section 2.2, the authors seem to discover that the linear wave model preserves Gaussianity. This fact is well known and is at the basis of the Kalman filtering theory. I do not say that this section is useless, because the derived formulas are necessary for the following sections, but it should be presented slightly differently.

Answer 8:

Actually, an advection-diffusion process preserves Gaussianity only in the asymptotic limit for which advection is weak compared to diffusion. We demonstrate rigorously this result through a Fourier decomposition and an asymptotic expansion.

- Page 6 line 7, define U_m .
- Page 6: I think recalling the Runge-Kutta 4 scheme is not essential here, since it is not at the core of the work.
- Page 7, line 12: Explain that we switch to a spectral representation for the purpose of the theoretical derivation. Now, the reader expects the setup of an EnKF and may wonder why the authors do not simply sample the Gaussian distribution.
- Page 9 and at other places: I think L_p should be written L_p . There are a few other typos of the kind throughout the text.
- Beginning of section 2.2.2: Recall what N is.

- Equation 17: Denominator should be N_e^{-1} .

Answer 9:

We thank Reviewer 1 for these useful remarks that we have taken into account.

- Page 12: I do not understand the point of the last sentence of Section 3.1.

Answer 10:

We now state more clearly the fact that $O(10000)$ rather than $O(10)$ members are necessary for an EnKF in order to obtain accurate statistics, which are important when risks estimations are concerned like for flood forecasting.

References

Fukumori, I.; Benveniste, J.; Wunsch, C. & Haidvogel, D. B. Assimilation of sea surface topography into an ocean circulation model using a steady-state smoother *Journal of Physical Oceanography*, 1993, 23, 1831-1855

Gelb, A. Applied optimal estimation M.I.T. press, 1974

Answer 11:

These two references are now added to the manuscript bibliography.

Neal, J. C., Atkinson, P. M., & Hutton, C. W. (2007). Flood inundation model updating using an ensemble Kalman filter and spatially distributed measurements. *Journal of Hydrology*, 336(3), 401-415.

Shiiba, M., Laurenson, X., & Tachikawa, Y. (2000). Real-time stage and discharge estimation by a stochastic-dynamic flood routing model. *Hydrological processes*, 14(3), 481-495.

Madsen, H., & Skotner, C. (2005). Adaptive state updating in real-time river flow forecasting—a combined filtering and error forecasting procedure. *Journal of Hydrology*, 308(1), 302-312.