Nonlin. Processes Geophys. Discuss.,
https://doi.org/10.5194/npg-2018-18-AC2, 2018

Nonlinear Processes in Geophysics
Discussions
Open Access

Interactive comment on "Application of ensemble transform data assimilation methods for parameter estimation in nonlinear problems" *by* Sangeetika Ruchi and Svetlana Dubinkina

Sangeetika Ruchi and Svetlana Dubinkina

s.dubinkina@cwi.nl

Received and published: 12 June 2018

Point by point comment.

This article presents a comparison between an EnKF and a particle filter based approaches for parameter estimation in a time independent model. I think that this comparison is relevant and can provide good insights about the performance of these two approaches in the context of parameter estimation. However I found many aspects that needs further clarification to support the conclusions made by the authors of this work. Major revisions are required to the paper.

Major points

- It is not clear if Importance Sampling and particle filters can be treated as synonyms. From my point of view particle filters can include different approaches for particle re-sampling to avoid the collapse of the filter and this is different from Importance Sampling which in principle does not include the resampling step.

Reply: Agree. We have changed the text accordingly.

-Page 2, 30 it is stated that particle filters do not update the uncertain parameters. This is not correct, many particle filters with different resampling approaches has been developed. These resampling steps introduce changes in the uncertain parameters so they get closer to the ones that produce the maximum observation likelihood, so the parameter ensemble evolves with time. It is true that the proposed technique performs this in a different way introducing a deterministic update of the parameter values (while usually resampling techniques in particle filters are stochastic). The difference between the implemented technique and previous techniques should be more clearly stated.

Reply: Agree. We have changed the text accordingly.

- In this work the implemented techniques are described as smoothers, however all the experiments performed are time independent. It is not clear for me what would be the difference between a filter or a smoother if there is no time involved. Please clarify this point. In the methodology I cannot find a difference between the filter implementation or the smoother implementation since there are no time index in the equations.

Reply: Agree, they are filters not smoothers. We have changed the text accordingly.

-Page 3, near 5: it is stated that ETKS does not employ the correlation in the estimation of the parameter. Filter equations are solved in the space defined by the ensemble members, but this implementation is basically equivalent to other EnKF which relies on the correlation between uncertain parameters and observed variables. Please clarify this point.

Reply: We have removed this sentence and also an iterative Kalman Smoother to avoid confusion. Both ETKF and ETPF considered in the paper solve equations defined in ensemble phase space.

-Page 8, 5 an iterative Kalman Smoother is mentioned here and shown in Figure 1, but detailed information about this technique is lacking. I suggest removing this technique since it has not been used in the experiments with the Darcy flow and also it has not been described in detail in the methodology section.

Reply: Agree, removed.

-In Figure 1, d, e and f a Gaussian prior produces a non-Gaussian posterior using ETKS. Since the EnKF relies on the linear and Gaussian assumption is it possible to obtain a non-Gaussian posterior from a Gaussian prior?

Reply: ETKF is able to give a non-Gaussian posterior due to the nonlinearity of the map between the uncertain parameters and observations.

-What is the motivation behind the functional introduced to define the observations in page 9, 15? What is r_l which appears in the definition of $L_l(P)$?

Reply: r_l denotes the location of the observation. This form of the observation functional and parameterization of the uncertain parameters given below guaranty the continuity of the forward map from the uncertain parameters to the observations and thus the existence of the posterior distribution as shown by Iglesias, M. A., Lin, K., and Stuart, A. M.: Well-posed Bayesian geometric inverse problems arising in subsurface

flow, inverse problems, 30, 114 001, 2014. (This text in added to the revised version.)

-Figure 6 shows the distribution for the first 3 modes of Z. Please clarify how these modes are obtained.

Reply: For the log permeability we use Karhunen-Loeve expansions of the form

$$\log(k(x)) = \log(5) + \sum_{i=1}^{n^2} \sqrt{\lambda_i}\nu_i(x)\mathcal{Z}_i,$$

where λ and $\nu(x)$ are eigenvalues and eigenfunctions of **C**, respectively, and the vector $\mathcal{Z}$ is of dimension n^2 iid from a Gaussian distribution with zero mean and variance one. Making sure that the eigenvalues are sorted in descending order $\mathcal{Z}_i \sim \mathcal{N}(0,1)$ produces $\log(\mathbf{k}) \sim \mathcal{N}(\log(\mathbf{5}), \mathbf{C})$. (This text in added to the revised version.)

-Figure 8 shows that the RMSE associated with ETKS is always lower than the RMSE for ETPS, however the first 3 moments of Z are better estimated by ETPS than for ETKS. Does this mean that ETKS provides a better estimation of higher order modes?

Reply: The first three moments were averaged over 10 simulations and thus it was misleading to show and draw conclusions based on that figure. Instead we now plot a figure that shows an error of first three moments. We observe that in terms of the estimation of the first three modes ETPF outperforms ETKF. We explore the estimation based on only those modes further. We use only first three modes in the Karhunen-Loeve expansion when computing the estimated log permeability keeping the number of uncertain parameters the same, namely 2500. In Fig. 12 we observe that ETPF outperforms ETKF for large ensemble sizes independent of an initial sample. Moreover, ETPF is not overfitting the data anymore since RMSE always decreases after data assimilation except at small ensemble sizes. In Fig. 13 we show the mean fields for the best and worst initial samples of 10^4 size. ETPF gives RMSE at the best sample 31.1 and the worst sample 32.98. By comparing it to 30.51 and 39.2 obtained using the full Karhunen-Loeve expansions, we observe that the maximum RMSE over simulations decreased substantially, while the minimum RMSE only slightly increased. ETKF gives RMSE at the best sample 32.27 and the worst sample 33.23. (Compare to 32.48 and 33.9 using the full Karhunen-Loeve expansions). Thus ETKF slightly decreases both maximum and minimum RMSE over simulations.

-IS and ETKS provide spatially smother solutions than ETPS (Figure 10), however ETPS seems to provide a better representation of the spatial variability and patters of the parameter. The explanation provided by the authors is not convincing for me. IS with a large number of particles should provide a very good estimation of the parameters (this approach is used as a benchmark by the authors).

Reply: The spacial variability is indeed a result of sampling error. (This text in added to the revised version.)

-Also the distribution for the first 3 moments of Z are relatively similar between ETPS, ETKS and IS (but the spatial variability shown in Figure 10 are very different). This point is very important and I think it should be explored and discussed in more detail.

Reply: The distributions shown in Figure 6 (Fig. 11 in the revised version) are different between ETPF, ETKF and IS.

-The authors show that in many cases ETPS improves the fitting to the observations but degrades the RMSE of the parameter. Can this be due to an over fitting of the observations?

Reply: It is indeed due to an overfitting of the observations.

-For the experiments including localization, the authors do not show the spatial distribution of the estimated parameters. This is very important since using localization can significantly improve the small scale details in the estimated parameter field. This figure should be included in order to better evaluate the impact of localization.

Reply: New figures: Figure 15 (mean field) and Figure 16 (variance) are added to the revised version.

-It is also strange that there is almost no improvement between the global and local implementation of the ETKS algorithm. With such a large number of variables and for the smaller ensemble sizes a larger positive impact would be usually expected. The degradation of the ETKS with a small ensemble size using localization is unexpected. The authors indicate that better localization approaches should be used but previous studies usually indicates that the impact of localization is stronger for smaller ensemble sizes. Are other works that shows this kind of behavior with localization degrading the performance of the filter for small ensemble sizes?

Reply: At small ensemble sizes ETKF is less robust than at large ensemble sizes due to the sampling error and the localization radius was chosen based on 1 simulation and fixed for the remaining 9, which should not have been done. In the revised version the optimal localization radius was obtained over all 10 simulations. The results are shown in Fig. 12. At small ensemble sizes both ETKF and ETPF with localization give smaller misfit and RMSE and larger variance than without localization but ETKF still outperforms ETPF. For large ensemble sizes ETPF performs now comparably to ETKF. Moreover, for ensemble sizes greater than 150 all simulations result in the RMSE decrease after data assimilation (not shown).

-Page 16, before 5, it is stated that "However, IS does not change the parameters, only their weights, while ETPS does change the parameters. Therefore ETPS has an advantage of IS representing the correct posterior but does not have its disadvantage of resampling lacking". If the posterior is correct and tacking into account that there is no time evolution in this context, what would be the problem with the lacking of resampling in the IS? The results described in this section also suggest that the solutions provided by IS and ETPS are very similar given that the initial condition is the same (once again resampling does not seems to be an issue in this context).

Reply: This sentence is removed.

-Does ETPS with 10^5 ensemble members produce a smooth field like the one produced by IS? In other words, the spatial variability that we see in Figure 10 b is produced by sampling errors or is the result of a better estimation of the parameter field? Results mentioned in the previous comment suggests that spatial variability is just a result of sampling noise and because of that is extremely sensitive to the prior ensemble. If we have a "lucky" prior then we end up with good results, but if the prior is bad then the result is also bad. In this sense ETKF seems to be more robust (which is reasonable when we need to update a large number of parameters with a relatively small ensemble and when the posterior distribution is not too far from a Gaussian).

Reply: Agree, the spatial variability that we see in Figure 10 b is indeed produced by sampling errors.

-Conclusions, page 19, 5: It is stated that ETPS better fit the posterior. However if we look at Figure 6 we found that for 10^4 particles (which is a large ensemble for most applications), ETPS fit is very noisy. Can the authors perform and objective comparison between the posterior provided by IS and the posterior provided by ETPS and ETKS (for instance using the Kullback-Leibler divergence or other objective comparison between two distributions).

Reply: In order to perform an objective comparison between the probabilities we

compute the Kullback-Leibler divergence of a posterior π obtained by either ETPF or ETKF and the posterior π^{IS} obtained by IS. ETPF gives the Kullback-Leibler divergence 0.21, 0.42, and 0.6, while ETKF 0.16, 0.07, and 0.49 for the modes $\mathcal{Z}_1$, $\mathcal{Z}_2$, and $\mathcal{Z}_3$, respectively. Thus ETKF gives a better approximation of the true pdf.

-Conclusions: Conclusions are very optimistic with respect to the performance of ETPS, however the RMSE of ETKS is always better in the large parameter space experiments. This suggests that the mean of the posterior is better estimated by ETKS rather than ETPS. While the mean is usually used as the best estimator of the parameter value, this should be mentioned in the conclusions.

Reply: ETPF certainly outperforms ETKF for a one parameter nonlinear test case by giving a better posterior estimation. This conclusion also holds for the five parameter test case, however demands a substantially larger ensemble size. Moreover the mean estimations obtained by ETPF are not consistenly better than the ones obtained by ETKF. When the number of uncertain parameters is large (2500) a decrease of degrees of freedom is essential. This is performed by using localization. At large ensemble sizes (greater than 50) ETPF performs as well as ETKF, while at a small ensemble size of 10 ETKF still outperforms ETPF. Even though localized ETPF overfits the data less often than non-localized, localization destroys the property of ETPF to retain the imposed bounds. This results in deterioration of the first mode posterior approximation. Another approach to improve ETPF performance is instead of applying localization to use only first modes in the approximation of log permeabilty as they are better estimated by the method. An advantage of this approach is that it is fully Bayesian. However, one needs to know at which mode to make a truncation and this is highly dependent on the covariance matrix of the log permeability.

Minor points

Page 12, 5: It is stated that is assumed to be an exponential correlation with maximum correlation along 3pi/4 ... It is not clear for me the meaning of this sentence.

Reply: We removed this sentence.

Page 7, 20: It is stated that R0 approximation is used with large ensembles in the experiments presented in this work, but in the result section it is not clear if this approximation has been used or not.

Reply: We removed this approximation for consistency.

Figure 10, It would be nice to include grid lines or to include the observation location in all the panels just to have a reference to compare smaller scale details in the estimated parameters.

Reply: The observation locations are added to the plots.

Interactive comment on Nonlin. Processes Geophys. Discuss., https://doi.org/10.5194/npg-2018-18, 2018.