Reply to Referee 1

This manuscript aims to compare performance of ensemble transform Kalman smoother (ETKS) and ensemble transform particle smoother (ETPS) in nonlinear parameter estimation problem. The authors conducted observing system simulation experiments and obtained reasonable results. The scope discussed in this manuscript suits well to Nonlinear Processes in Geophysics. I do not have any major concerns for the experiments presented in the manuscript. However, some discussions and descriptions are difficult to follow due to insufficient explanation. Here I list the concerns, which would be beneficial to improve the manuscript further.

[General Comments]
1. Scientific Significance: The authors addressed that they applied the ETPS for estimating a large number of uncertain parameters (P2L34). It can be a good motivation; however I could not understand the scientific significance that can be achieved by applying the ETPS and ETKS for the large-dimensional problem. Please address this point clearly in abstract and conclusion.

Reply: The large number of uncertain parameters is of a particular interest for subsurface reservoir modelling as it allows to parameterise permeability on the grid. The most reliable methods of MCMC are computationally expensive and sequential ensemble methods such as ensemble Kalman filers and particle filters provide with a favourable alternative.

2. Lack of explanations: I could not follow several logics of the manuscript, therefore, my major comments includes many "whys" and "reasons". Most of the issues should be solved by adding sufficient explanations.

Reply: we have added more explanations (please see point-by-point answer).

3. Results (Figures): Some figures were discussed insufficiently. It is better to remove figure(s) if they are not needed.

Reply: we have adopted the revised version accordingly.

4. Methods: The author compared the ETKS and ETPS. I am wondering the difference between the ETPS used in this study and a nonlinear ensemble transform filter by Tödter and Ahrens (2015) (Tödter, J., and B. Ahrens, 2015: A second-order exact ensemble square root filter for nonlinear data assimilation. Mon. Wea. Rev., 143, 1347–1367).

Reply: The paper by Tödter and Ahrens (2015) addresses an important issue of ensemble Kalman (square root) filter being biased for nonlinear models and makes a correction for that. The resulting algorithm does not attempt to estimate the full analysis pdf in contrast to the ETPF.

Also, it is better to compare the localization methodology with local particle filters (Penny and Miyoshi 2016; Poterjoy 2016). Please add more discussion on difference from existing methods.

Penny, S. G. and T. Miyoshi, 2016: A local particle filter for high-dimensional geophysical systems. Nonlin. Processes Geophys., 23, 391-405. Poterjoy, J. (2016). A localized particle filter for high-dimensional nonlinear systems. Monthly Weather Review, 144(1), 59-76.

Reply: The localization methodology considered in this manuscript was particularly developed for ETPF by S. Reich and C. Cotter (2015). The same holds for localization for ensemble transform Kalman filter by Hunt at al (2007). Therefore we keep this comparison. However, we mention Penny and Miyoshi 2016; Poterjoy 2016 in the localization section to give a flavour of an ongoing research in local particle filters.

[Major Comments]

1. P1L13: Please add reason(s) why ETPS is very sensitive w.r.t. the initial ensemble.

Reply: ETKF is very robust while ETPF is very sensitive with respect to the initial ensemble due to a sampling error.

2. P1L15: Please add reason(s) why the localization deteriorated the posterior estimation.

Reply: An issue of an increase in the root mean square error after data assimilation is performed in ETPF for a high-dimensional test problem is resolved by applying distance-based localization, which however deteriorated the posterior estimation of the first mode by largely increasing the variance, which is due to a combination of less varying localized weights, not keeping the imposed bounds on the modes via the Karhunen-Loeve expansion and the main variability explained by the first mode.

3. P7L15: Isn't it possible to apply the localization between variables?

Reply: Localization can be defined for variables that depend on space. Thus it cannot be applied to geometrical parameters, for example.

4. P8L1: I could not understand the sentence " is made such that $y_{\text{obs}} = 48$". Please rephrase this sentence.

Reply: The true parameter u^{true} gives $h(u^{\text{true}}) = 48$.

5. P10L14: Please explain more about reason(s).

Reply: It is interesting to note that ETKF overestimates the tails of the pdfs while ETPF underestimates them, which indicates that there is not enough spread in the ensemble.

6. P11L5: Why? Does it relate to the resampling issue discussed later?

Reply: ETPF provides ensemble members that stay within the original bounds, while ETKF—outside the bounds. Moreover the optimal transport problem solved by ETPF results in some particles being almost identical. Therefore ETPF gives smaller spread than ETKF.

7. Fig.4, Fig. 8 (b) and (c), : I did not understand why this figure is needed because they were not discussed.

Reply: In Fig. 4 we plot the true parameters $\mathbf{u}^{\text{true}}$, the mean $\bar{\bar{\mathbf{u}}}^a$ and the spread $\bar{\bar{\mathbf{u}}}^a \pm \bar{\mathbf{u}}^a_{\text{std}}$ of estimated parameters averaged over 10 simulations.
In Fig. 8 (in the revised version it is Fig. 6) we plot mean, minimum and maximum over 10 simulations after data assimilation for the data misfit (a), RMSE (b), and variance (c). ETPF is shown in blue and ETKF in red. We observe that ETPF in underdispersive compared to ETKF (c). This is due to the linear transformation, as it results in some ensemble members being nearly identical. Misfit (a) given by ETPF is smaller than the one given by ETKF for almost all simulations at ensemble sizes greater than 150. The RMSE(b) on the contrary is larger.

8. P15L5, perturbation of ensemble member: In generic PF, the resampling (or inflation) method is very important to avoid the particle convergence. Could you explain why you did not need to consider this issue?

Reply: The model is time independent thus the issue of collapse does not rise here.

9. Fig. 8 (b): I was confused why the ETKS outperforms the ETPS if RMSE is used for the metric.

Reply: The first 3 modes $\mathcal{Z}$ are better estimated by ETPF than by ETKF, thus the permeability field defined only by those modes gives better resemblance to the true permeability when approximated by ETPF (please see Fig.13 in the revised version). ETKF, however, provides a better estimation of higher order modes, thus ETKF outperforms the ETPF if RMSE is used for the metric taking into account all modes.

10. Fig. 10: It is helpful to add RMSEs on the figure.

Reply: Agreed, it is added.

11. Table 1: Could you discuss why the optimal radius for the ETKS is larger than that of the ETPS?

Reply: It was also observed by Cheng and Reich (2015) that the localization radius for ETKF is larger than for ETPF. This is probably related to more noisy approximation of the posterior by ETPF than by ETKF.

Y. Chen and S. Reich, Data assimilation: a dynamical system perspective. Frontiers in Applied Dynamical Systems: Reviews and Tutorials Vol 2, 75-118, 2015

12. P16L15: Please discuss why the localization degrades the posterior estimation.

Reply: The posterior estimation of the first mode $\mathcal{Z}_1$ degraded, while of others improved. The Kullback-Leibler divergence for the first mode is 0.73 (compare to 0.21 without localization), and for second and third is 0.2 and 0.18,

correspondently (compare to 0.42 and 0.6 without localization). Variance of the posteriors is larger when localization is applied for both ETPF and ETKF. The localized weights given by Eq. 11 vary less than the non-localized weights given by Eq. 3. Therefore the localized pdf is less noisy than the non-localized. However, localization applied in the form of the Karhunen-Loeve expansion given by Eq. 14 does not retain the imposed bounds on the modes $\mathcal{Z}$ as we need to invert a matrix product of eigenvalue and eigenvector matrices to obtain the modes. By increasing the localization radius to 1.2 we get the Kullback-Leibler divergence 0.65 for the first mode, and 0.14 and 0.12 for the second and third, correspondently, thus the posterior approximation improved only slightly.

13. Conclusion: It would be helpful to add findings and limitation further in this section.

Reply: MCMC methods remain the most reliable methods for estimating the posterior distributions of uncertain model parameters and states. They, however, also remain computationally expensive. Ensemble Kalman filters provide computationally affordable approximations but rely on the assumptions of Gaussian probabilities. For nonlinear models even if the prior is Gaussian the posterior is not Gaussian anymore. Particle filtering on the other hand does not have such an assumption but requires a resampling step, which is usually stochastic. Ensemble transform particle filter is a particle filtering method that deterministically resamples the particles based on their importance weights and covariance maximization among the particles.

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 consistently 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 localization. At large ensemble sizes ETPF performs as well as ETKF, while at small ensemble sizes 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.

Reply to Referee 2

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 resampling 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 $\mathbf{C}$, 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 consistently 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.

Application of ensemble transform data assimilation methods for parameter estimation in nonlinear problems

Sangeetika Ruchi[1] and Svetlana Dubinkina[1]

[1]Centrum Wiskunde & Informatica, P.O. Box 94079, 1098 XG Amsterdam, The Netherlands

Correspondence: Sangeetika Ruchi (s.ruchi@cwi.nl)

Abstract. Over the years data assimilation methods have been developed to obtain estimations of uncertain model parameters by taking into account a few observations of a model state. ~~However, most of these computationally affordable methods have assumptions of Gaussianity, e. g. an~~ The most reliable methods of MCMC are computationally expensive and sequential ensemble methods such as ensemble Kalman filers and particle filters provide with a favourable alternative. However, Ensemble Kalman Filter has an assumption of Gaussianity. Ensemble Transform Particle Filter does not have ~~the assumption of Gaussianity~~ this assumption and has proven to be highly beneficial for an initial condition estimation and a small number of parameter estimation in chaotic dynamical systems with non-Gaussian distributions. In this paper we employ Ensemble Transform Particle ~~Smoother (ETPS~~Filter (ETPF) and Ensemble Transform Kalman ~~Smoother (ETKS~~Filter (ETKF) for parameter estimation in nonlinear problems with 1, 5, and 2500 uncertain parameters and compare them to importance sampling (IS). The large number of uncertain parameters is of a particular interest for subsurface reservoir modelling as it allows to parameterise permeability on the grid. We prove that the updated parameters obtained by ~~ETPS~~ ETPF lie within the range of an initial ensemble, which is not the case for ~~ETKS~~ETKF. We examine the performance of ~~ETPS and ETKS~~ ETPF and ETKF in a twin experiment setup, where observations of pressure are synthetically created based on the know values of parameters. ~~The numerical experiments demonstrate that the ETKS provides good estimations of the mean parameters but not of the posterior distributions and as the ensemble size increases the posterior does not improve. ETPS provides good approximations~~ For small number of uncertain parameters (1 and 5) ETPF performs comparably to ETKF in terms of the mean estimation and outperforms in terms of the posterior ~~and~~ estimation as the ensemble size increases~~the posterior converges to the posterior obtained by IS with a large ensemble. ETKS is very robust while ETPS is very sensitive~~ . For large number of uncertain parameters (2500) ETKF is robust with respect to the initial ensemble while ETPF is sensitive due to a sampling error. An issue of an increase in the root mean square error after data assimilation is performed in ~~ETPS~~ ETPF for a high-dimensional test problem is resolved by applying distance-based localization, which however deteriorated the posterior estimation ~~.~~ of the first mode by largely increasing the variance, which is due to a combination of less varying localized weights, not keeping the imposed bounds on the modes via the Karhunen-Loeve expansion and the main variability explained by the first mode. A possible remedy is instead of applying localization to use only first modes that are well estimated by ETPF. This approach is fully Bayesian but demands a knowledge at which mode to truncate.

1 Introduction

An accurate estimation of subsurface geological properties like permeability, porosity etc. is essential for many fields specially where such predictions can have large economic or environmental impact, for instance prediction of oil or gas reservoir locations. Knowing the geological parameters a so-called forward model is solved for the model state and a prediction can be made. The subsurface reservoirs, however, are buried thousands of feet below the earth surface and exhibit a highly heterogeneous structure, which makes it difficult to obtain their geological parameters. Usually a prior information about the parameters is given, which still needs to be corrected by observations of pressure and production rates. These observations are, however, known only at well locations that are often hundreds of meter apart and corrupted by errors. This gives instead of a well-posed forward problem an ill-posed inverse problem of estimating uncertain parameters, since many possible combinations of parameters can result in equally good matches to the observations.

Different inverse problem approaches for groundwater and petroleum reservoir modelling, generally termed as history matching, have been developed over the past years, e.g. Oliver et al. (1997) implemented Markov chain Monte Carlo methods with different perturbations and tested it on a 2-D reservoir model; Reynolds et al. (1996) obtained reservoir parameters estimations using Gauss-Newton method; Vefring et al. (2006) used Levenberg–Marquardt method to characterize reservoir pore pressure and permeability. A review of history matching developments is written by Oliver and Chen (2011).

For reservoir models the term data assimilation and history matching are used interchangeably, as the goal of data assimilation is the same as that of history matching, where observations are used to improve a solution of a model. Ensemble data assimilation methods such as Ensemble Kalman filters (Evensen, 2009) have been originally developed in meteorology and oceanography for the state estimation. Now it is one of the frequently employed approaches for parameter estimation in subsurface flow models as well (e.g. Oliver et al., 2008). ~~The model state is augmented with uncertain parameters and the correlations between uncertain parameters and predicted data are used to correct the parameters together with the state. The simultaneous update of the model state and parameters, however, results in a unbalanced (unphysical) model state. Therefore Evensen (2000) introduced an ensemble smoother, where only uncertain parameters are estimated and the model state is computed by solving the forward model with corrected parameters.~~ A detailed review of ensemble Kalman filter developments in reservoir engineering is written by Aanonsen et al. (2009). An ensemble Kalman filter efficiently approximates a true posterior distribution if the distribution is not far from Gaussian, as it corrects only the mean and the variance. For nonlinear models with multimodal distributions, however, an ensemble Kalman filter fails to correctly estimate the posterior, as shown by Dovera and Della Rossa (2011).

~~Particle filtering (Doucet et al., 2001), also known as~~ Importance Sampling (IS) ~~,~~ is quite promising for such models as it does not have any assumptions of Gaussianity. It is also an ensemble based method in which the probability density function is represented by a number of ~~particles (also called samplesor ensemble members). One particle~~ samples. One sample corresponds to one configuration of uncertain model parameters. The forward model is solved for each ~~particle~~ sample and predicted data is computed. The weight is assigned to ~~particles~~ samples based on the observations of the true physical system and the predicted data. The drawback of IS is that it does not update the uncertain parameters but only their weight. Therefore a family of

particle filters (Doucet et al., 2001) has been developed where IS is supplied with resampling and a sample is called particle. A significant work for parameter estimation using particle filtering has been done in hydrology. Moradkhani et al. (2005) used it to estimate model parameters and state posterior distributions for a rainfall-runoff model. Weerts and El Serafy (2006) compared an ensemble Kalman filter and a particle filter with different resampling strategies for a rainfall-runoff forecast and obtained that as the number of particles increases the particle filter outperforms the ensemble Kalman filter. Guingla et al. (2012) employed particle filtering to correct the soil moisture and to estimate hydraulic parameters.

Particle filtering The resampling in particle filtering is, however, does not update the uncertain parameters only their weight. Ensemble transform particle filter stochastic. Ensemble Transform Particle Filter (ETPF) developed by Reich and Cotter (2015) is a particle filtering method that *deterministically* resamples the particles based on their weights and covariance maximization among the particles. Therefore it has an IS advantage of predicting the correct posterior but does not have its disadvantage of resampling lacking. Ensemble transform particle filter has ETPF has been used for initial condition estimations and for parameter estimations in chaotic dynamical systems with a small number of uncertain parameters (Lorenz 63 model). It has not been applied, however, in subsurface reservoir modelling for estimating a large number of uncertain parameters. In this paper we employ it for estimating uncertain parameters in subsurface reservoir modelling. We call it ensemble transform particle smoother (ETPS) in analogy with ensemble smoothers, where the model state is computed from the forward model with corrected model parameters. ETPS does not use correlations between predicted data and uncertain parameters. On the one hand, the correlations should improve the estimation but on the other hand, it is not always clear how to compute them. ETPF provides the equations that are solved in the space defined by the ensemble members. Therefore for comparison we employ Ensemble Transform Kalman Smoother (ETKSFilter (ETKF) developed by Bishop et al. (2001) also does not employ the correlations in its estimation. It that also transforms the state from the model space to the ensemble space, minimises the uncertainty in the ensemble space and transforms the estimation back to the model space.

In this paper we investigate the performance of ETPS and ETKS ETPF and ETKF for parameter estimation in nonlinear problems and compare them to IS with a large ensemble. This paper is organized as follows: in the section 2 we describe IS, ETPS, and ETKS ETPF, and ETKF for parameter estimation. We apply these methods in Sect. 3 to a one parameter nonlinear test case, where the posterior can be computed analytically, and in Sect. 4 to a single-phase Darcy flow, where the number of parameters is 5 and 2500. In Sect. 5 we draw the conclusions.

2 Data assimilation methods

We implement an ensemble transform Kalman smoother filter and an ensemble transform particle smoother filter for estimating parameters of subsurface flow. Both of these methods are based on Bayesian framework. Assume we have an ensemble of M model parameters $\{\boldsymbol{u}_m\}_{m=1}^{M}$, then according to this framework, the posterior distribution, which is the probability distribution $\pi(\boldsymbol{u}_m|\boldsymbol{y}_{\text{obs}})$ of the model parameters $\boldsymbol{u}_m$ given a set of observations $\boldsymbol{y}_{\text{obs}}$, can be estimated by the pointwise multiplication of the prior probability distribution $\pi(\boldsymbol{u}_m)$ of the model parameters $\boldsymbol{u}_m$ and the conditional probability distribution $\pi(\boldsymbol{y}_{\text{obs}}|\boldsymbol{u}_m)$

of the observations given the model parameters, which is also referred as the likelihood function,

$$\pi(\boldsymbol{u}_m|\boldsymbol{y}_{\text{obs}}) = \frac{\pi(\boldsymbol{y}_{\text{obs}}|\boldsymbol{u}_m)\pi(\boldsymbol{u}_m)}{\pi(\boldsymbol{y}_{\text{obs}})}.$$

The denominator $\pi(\boldsymbol{y}_{\text{obs}})$ represents the marginal of observations and can be expressed as:

$$\pi(\boldsymbol{y}_{\text{obs}}) = \sum_{m=1}^{M} \pi(\boldsymbol{y}_{\text{obs}}, \boldsymbol{u}_m) = \sum_{m=1}^{M} \pi(\boldsymbol{y}_{\text{obs}}|\boldsymbol{u}_m)\pi(\boldsymbol{u}_m),$$

which shows that $\pi(\boldsymbol{y}_{\text{obs}})$ is just a normalisation factor.

2.1 Ensemble Transform Kalman ~~Smoother~~Filter

~~We employ an ensemble Kalman smoother based on a transformation of an ensemble from the model phase space to the ensemble space—the ensemble transform Kalman smoother Bishop et al. (2001).~~

Assume we have an ensemble of M initial model parameters $\{\boldsymbol{u}_m^b\}_{m=1}^M$, where b refers to a background (prior) ensemble, which are sampled from a chosen prior probability density function, then the ensemble Kalman estimate (or analysis) $\{\boldsymbol{u}_m^a\}_{m=1}^M$ is given by:

$$\boldsymbol{u}_m^a = \sum_{l=1}^{M} \text{diag}\left(s_{lm} + q_l - \frac{1}{M}\right)\boldsymbol{u}_l^b, \quad m = 1, \ldots, M,$$

where diag is a diagonal matrix, s_{lm} is the (l, m) entry of a matrix $\mathbf{S}$

$$\mathbf{S} = \left[\mathbf{I} + \frac{1}{M-1}(\mathbf{A}^b)^T\mathbf{R}^{-1}\mathbf{A}^b\right]^{-1/2}, \tag{1}$$

and q_l is the l-th entry of a column $\boldsymbol{q}$

$$\boldsymbol{q} = \frac{1}{M-1}\mathbf{1}_M - \mathbf{S}^2(\mathbf{A}^b)^T\mathbf{R}^{-1}(\bar{\boldsymbol{y}}^b - \boldsymbol{y}_{\text{obs}}).$$

Here $\mathbf{I}$ is an identity matrix of size $M \times M$, $\mathbf{1}_M$ is a vector of size M with all ones, $\bar{\boldsymbol{y}}^b$ is the mean of the predicted data defined by

$$\bar{\boldsymbol{y}}^b = \frac{1}{M}\sum_{m=1}^{M} \boldsymbol{y}_m^b,$$

$\mathbf{A}^b$ is the background ensemble anomalies of the predicted data defined as

$$\mathbf{A}^b = \left[(\boldsymbol{y}_1^b - \bar{\boldsymbol{y}}^b) \quad (\boldsymbol{y}_2^b - \bar{\boldsymbol{y}}^b) \quad \ldots \quad (\boldsymbol{y}_M^b - \bar{\boldsymbol{y}}^b)\right],$$

and $\mathbf{R}$ is the measurement error covariance. To ensure that the anomalies of analysis remain zero centered we check whether ~~$\mathbf{A}^a\mathbf{1} = \mathbf{A}^b\mathbf{S}\mathbf{1} = \mathbf{0}$, given $\mathbf{S}\mathbf{1} = \mathbf{1}$~~ $\mathbf{A}^a\mathbf{1}_M = \mathbf{A}^b\mathbf{S}\mathbf{1}_M = \mathbf{0}$, given $\mathbf{S}\mathbf{1}_M = \mathbf{1}_M$ and $\mathbf{A}^b\mathbf{1}_M = \mathbf{0}$. The model parameters $\boldsymbol{u}_m^b$ and the predicted data $\boldsymbol{y}_m^b$ are related by $\boldsymbol{y}_m^b = h(\boldsymbol{u}_m^b)$, where h is a nonlinear function and here we assume that the function h is known.

2.2 Ensemble Transform Particle ~~Smoother~~Filter

In particle filtering we represent the probability distribution function using ensemble members (also called particles) as in ensemble Kalman filter. We start by assigning prior (background) weights $\{w_m^b\}_{m=1}^M$ to M particles and then compute new (analysis) weights $\{w_m^a\}_{m=1}^M$ using the Bayes' formula and observations $\boldsymbol{y}_{\text{obs}}$

$$w_m^a = \frac{\pi(\boldsymbol{y}_{\text{obs}}|\boldsymbol{u}_m^b)w_m^b}{\pi(\boldsymbol{y}_{\text{obs}})}. \tag{2}$$

~~It is important to note that particle filters *do not* change the parameters $\boldsymbol{u}$, they only modify the weight of the particles. Therefore a sophisticated perturbation needs to be implemented for parameter estimation. Instead particle filtering has been modified using a coupling methodology which resulted in an ensemble transform particle filter Reich and Cotter (2015). Since at the data assimilation step we update only the parameters and not the states, with analogy to ensemble smoothers we will call this method ensemble transform particle smoother (ETPS).~~

We assume that initially all particles have equal weight, thus $w_m^b = 1/M$ for $m = 1, \ldots, M$, and that the likelihood is Gaussian with error covariance matrix $\mathbf{R}$, then from Eq. (2) w_m^a is given by

$$w_m^a = \frac{\exp\left[-\frac{1}{2}(\boldsymbol{y}_m^b - \boldsymbol{y}_{\text{obs}})^T\mathbf{R}^{-1}(\boldsymbol{y}_m^b - \boldsymbol{y}_{\text{obs}})\right]}{\sum_{j=1}^M \exp\left[-\frac{1}{2}(\boldsymbol{y}_j^b - \boldsymbol{y}_{\text{obs}})^T\mathbf{R}^{-1}(\boldsymbol{y}_j^b - \boldsymbol{y}_{\text{obs}})\right]}, \quad m = 1, \ldots, M. \tag{3}$$

In ~~a well-known~~ Importance Sampling (IS)~~data assimilation method~~, which will be used in this paper as a "ground" truth, these weights define the posterior pdf. The mean parameter for IS is then

$$\bar{\boldsymbol{u}}^a = \sum_{m=1}^M \boldsymbol{u}_m^b w_m^a.$$

~~ETPS~~ It is important to note that IS *does not* change the parameters $\boldsymbol{u}$, it only modifies the weight of the particles (samples). Therefore a resampling needs to be implemented for parameter estimation, which is usually stochastic. Instead particle filtering has been modified using a deterministic coupling methodology which resulted in an ensemble transform particle filter of Reich and Cotter (2

ETPF looks for a coupling between two discrete random variables B_1 and B_2 so as to convert the ensemble members belonging to the random variable B_2 with probability distribution $\pi(B_2 = \boldsymbol{u}_m^b) = w_m^a$ to the random variable B_1 with uniform probability distribution $\pi(B_1 = \boldsymbol{u}_m^b) = 1/M$. The coupling between these two random variables is an $M \times M$ matrix $\mathbf{T}$ whose entries should satisfy

$$t_{mj} \geq 0, \quad m, j = 1, \ldots, M, \tag{4}$$

$$\sum_{m=1}^M t_{mj} = \frac{1}{M}, \quad j = 1, \ldots, M, \tag{5}$$

$$\sum_{j=1}^M t_{mj} = w_m^a, \quad m = 1, \ldots, M. \tag{6}$$

An optimal coupling matrix $\mathbf{T}^*$ with elements t_{mj}^* minimizes the squared Euclidean distance

$$J(t_{mj}) = \sum_{m,j=1}^M t_{mj}||\boldsymbol{u}_m^b - \boldsymbol{u}_j^b||^2 \tag{7}$$

and the analysis model parameters are obtained by the linear transformation

$$\boldsymbol{u}_j^a = M \sum_{m=1}^{M} t_{mj}^* \boldsymbol{u}_m^b, \quad j = 1, \ldots, M. \tag{8}$$

Then the mean parameter for ~~ETPS~~ ETPF is

$$\bar{\boldsymbol{u}}^a = \sum_{m=1}^{M} \boldsymbol{u}_m^a \frac{1}{M}.$$

We use $FastEMD$ algorithm of Pele and Werman (2009) to solve the linear transport problem and get the optimal transport matrix.

Remark: An important property of ~~ETPS is to retain the~~ ETPF is preservation of imposed interval bounds ~~of~~ on ensemble members. Consider an ensemble of parameters $\{\boldsymbol{u}_m^b\}_{m=1}^M$ given by

$$\boldsymbol{u}_m^b = (a_m^b \; b_m^b \; c_m^b)^T, \quad m = 1, \ldots, M,$$

where we assume all the parameters $\{a_m^b\}_{m=1}^M$, $\{b_m^b\}_{m=1}^M$ and $\{c_m^b\}_{m=1}^M$ are bounded between 0 and 1. Therefore, the following inequalities hold:

$$0 < a_{\text{min}} \leq a_m^b \leq a_{\text{max}} < 1, \quad m = 1, \ldots, M,$$
$$0 < b_{\text{min}} \leq b_m^b \leq b_{\text{max}} < 1, \quad m = 1, \ldots, M,$$
$$0 < c_{\text{min}} \leq c_m^b \leq c_{\text{max}} < 1, \quad m = 1, \ldots, M.$$

Now we assume two discrete random variables B_1 and B_2 have probability distributions given by

$$\pi(B_1 = \boldsymbol{u}_m^b) = 1/M, \quad \pi(B_2 = \boldsymbol{u}_m^b) = w_m^a,$$

with $w_m^a \geq 0$, $m = 1, \ldots, M$ and $\sum_{m=1}^M w_m^a = 1$. As ~~ETPS~~ ETPF looks for a matrix $\mathbf{T}^*$ which defines coupling between these two probability distributions, each entry of this coupling matrix satisfies the conditions given by Eq. (4)–(6). These conditions assure that each entry of the coupling matrix will be non-negative and less than 1. Since the analysis given by Eq. (8) is

$$\boldsymbol{u}_m^a = \begin{bmatrix} a_1^b(Mt_{1m}^*) + a_2^b(Mt_{2m}^*) + \cdots + a_M^b(Mt_{Mm}^*) \\ b_1^b(Mt_{1m}^*) + b_2^b(Mt_{2m}^*) + \cdots + b_M^b(Mt_{Mm}^*) \\ c_1^b(Mt_{1m}^*) + c_2^b(Mt_{2m}^*) + \cdots + c_M^b(Mt_{Mm}^*) \end{bmatrix}, \quad m = 1, \ldots, M,$$

these conditions lead to

$$0 < a_{\text{min}} \leq a_m^a \leq a_{\text{max}} < 1, \quad m = 1, \ldots, M,$$
$$0 < b_{\text{min}} \leq b_m^a \leq b_{\text{max}} < 1, \quad m = 1, \ldots, M,$$
$$0 < c_{\text{min}} \leq c_m^a \leq c_{\text{max}} < 1, \quad m = 1, \ldots, M.$$

Thus the coupling matrix bounds the analysis ensemble members to be in the desired range. This is not observed in ~~ETKS~~ ETKF as the matrix $\mathbf{S}$ given by Eq. (1) does not impose any of the non-equality and equality constraints, so it results in values outside the bound.

2.3 Localization

All variations of ensemble Kalman filter and particle filter are limited by the ensemble size. Since, even if the dimension of the problem is just up to a few thousands, a large ensemble size will make each run of the model computationally very expensive. This limit of a small ensemble size introduces a sampling error. To deal with this issue localization for ~~ETKS~~ ETKF was introduced by Hunt et al. (2007) ~~. We use a distance based localization method. More advanced methods such as wavelet-based approaches of Chen and Oliver (2012b) are outside the scope of this paper.~~ and for ETPF by Reich and Cotter (2015). More recent approaches to particle filter localization include Penny and Miyoshi (2016) and Poterjoy (2016)

For the local update of a model parameter $\boldsymbol{u}_m(X_i)$ at a grid point X_i, we introduce a diagonal matrix $\hat{\mathbf{C}}_i \in R^{N_y \times N_y}$ in the observation space with an element

$$(\hat{\mathbf{C}}_i)_{ll} = \rho\left(\frac{||X_i - r_l||}{r_{\text{loc}}}\right), \tag{9}$$

where $i = 1, \ldots, n^2$, $l = 1, \ldots, N_y$, n^2 is the number of model parameters, N_y is the dimension of the observation space, r_l denotes the location of the observation, r_{loc} is a localisation radius and $\rho(\cdot)$ is a taper function, such as Gaspari-Cohn function Gaspari and Cohn (1999)

$$\rho(r) = \begin{cases} 1 - \frac{5}{3}r^2 + \frac{5}{8}r^3 + \frac{1}{2}r^4 - \frac{1}{4}r^5, & 0 \le r \le 1, \\ -\frac{2}{3}r^{-1} + 4 - 5r + \frac{5}{3}r^2 + \frac{5}{8}r^3 - \frac{1}{2}r^4 + \frac{1}{12}r^5, & 1 \le r \le 2, \\ 0, & 2 \le r. \end{cases}$$

Then the estimated model parameter at the location X_i is

$$\boldsymbol{u}_m^a(X_i) = \sum_{l=1}^{M} \text{diag}\left(s_{lm}(X_i) + q_l(X_i) - \frac{1}{M}\right) \boldsymbol{u}_l^b(X_i), \quad m = 1, \ldots, M, \tag{10}$$

where diag is a diagonal matrix, $s_{lm}(X_i)$ is the (l, m) entry of the localized transformation matrix $\mathbf{S}(X_i)$

$$\mathbf{S}(X_i) = \left[\mathbf{I} + \frac{1}{M-1}(\mathbf{A}^b)^T(\hat{\mathbf{C}}_i\mathbf{R}^{-1})\mathbf{A}^b\right]^{-1/2}$$

and $q_l(X_i)$ is the l-th entry of the localized column $\boldsymbol{q}(X_i)$

$$\boldsymbol{q}(X_i) = \frac{1}{M-1}\mathbf{1}_M - \mathbf{S}(X_i)^2(\mathbf{A}^b)^T\mathbf{R}^{-1}(\bar{\boldsymbol{y}}^b - \boldsymbol{y}_{\text{obs}}).$$

Localization ~~at ETPS introduced by Reich and Cotter (2015)~~ of ETPF modifies the likelihood and thus the weights given by Eq. (3) are computed locally at each grid X_i

$$w_m^a(X_i) = \frac{\exp\left[-\frac{1}{2}(\boldsymbol{y}_m^b - \boldsymbol{y}_{\text{obs}})^T(\hat{\mathbf{C}}_i\mathbf{R}^{-1})(\boldsymbol{y}_m^b - \boldsymbol{y}_{\text{obs}})\right]}{\sum_{j=1}^{M}\exp\left[-\frac{1}{2}(\boldsymbol{y}_j^b - \boldsymbol{y}_{\text{obs}})^T(\hat{\mathbf{C}}_i\mathbf{R}^{-1})(\boldsymbol{y}_j^b - \boldsymbol{y}_{\text{obs}})\right]}, \quad m = 1, \ldots, M, \tag{11}$$

where $\hat{\mathbf{C}}_i$ is the diagonal matrix given by Eq. (9). Then the estimated model parameter $\boldsymbol{u}_j^a(X_i)$ at the grid X_i is given by

$$\boldsymbol{u}_j^a(X_i) = M \sum_{m=1}^{M} t_{mj}^* \boldsymbol{u}(X_i)_m^b, \quad j = 1, \ldots, M,$$

where t_{mj}^* is an element of an optimal coupling matrix $\mathbf{T}^*$ which minimizes the squared Euclidean distance at the grid point X_i

$$J(t_{mj}) = \sum_{m,j=1}^{M} t_{mj} [u_m^b(X_i) - u_j^b(X_i)]^2, \tag{12}$$

which reduces the localized ~~ETPS~~ ETPF to a univariate transport problem. It should be noted that localization can be applied only for grid-dependent parameters.

~~**Remark:** A so-called $R0$ approximation of Reich and Cotter (2015) consists of computing the weights according to Eq. instead of Eq. but solving the optimization problem for each grid (or each parameter) defined by Eq. separately. The $R0$ approximation has an advantage of being a solution of a parallel and computationaly less expensive univariate transport problem. For parameter estimation of independent parameters as the ensemble size increases the $R0$ approximation defined by Eq. converges to the full approximation defined by Eq. since the sampling noise reduces (not shown). Therefore for computing the posterior distributions with large ensembles we use the $R0$ approximation.~~

3 One parameter nonlinear problem

First we consider a one parameter nonlinear problem from Chen and Oliver (2013). The prior distribution is Gaussian distribution with mean 4 and variance 1. ~~An observation of a function~~

$$h(u) = \frac{7}{12}u^3 - \frac{7}{2}u^2 + 8u$$

~~is made such that $y_{\text{obs}} = 48$. Observation~~ The nonlinear forward model is

$$h(u) = \frac{7}{12}u^3 - \frac{7}{2}u^2 + 8u.$$

The true parameter u^{true} gives $h(u^{\text{true}}) = 48$ and the observation error is drawn from a Gaussian distribution with zero mean and variance 16. In Fig. 1 we plot the posterior probability density functions estimated by ~~ETPS~~ ETPF (top), ~~ETKS (middle~~ETKF (bottom) with ensemble sizes 10^2 (left), 10^3 (center), and 10^4 (right). The prior distribution is shown in red and the posterior estimated by IS with ensemble size 10^5 is shown in black. We can see that ~~ETPS provides with~~ ETPF provides better approximation of the true probability density function~~. We have also implemented ensemble Kalman Smoother with perturbed observations (SEnKS) developed by Burgers et al. (1998), that is used in an iterative ensemble Kalman smoother called multiple data assimilation of Emerick and Reynolds (2013). The posterior probability density function obtained by SEnKS is shown at the bottom of Fig. 1. The posterior is less skewed than the one provided by ETKS but otherwise is still~~

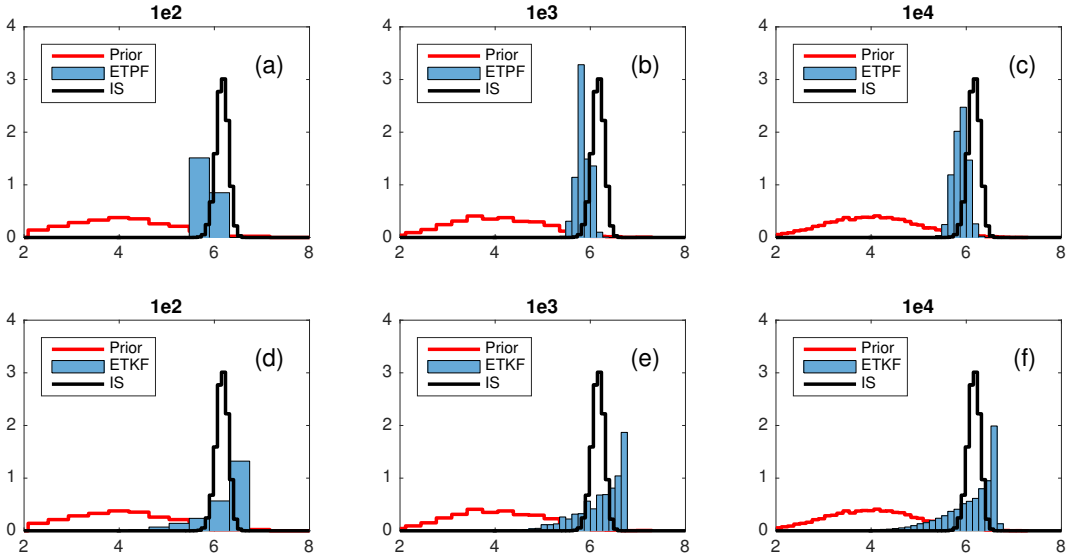

Figure 1. Probability density functions for the one parameter nonlinear problem. Top: ~~ETPS~~ETPF, ~~middle: ETKS,~~ bottom: ~~SEnKS~~ETKF. Left: ensemble size 10^2, center: ensemble size 10^3, right: ensemble size 10^4. Prior is in red. True pdf obtained by IS with ensemble size 10^5 is in black.

~~not a good approximation of the true pdf. We will not use SEnKS for parameter estimation as it does not provide substantially different results than ETKS and both ETKS and ETPS are based on a transformation, while SEnKS is not~~, while ETKF gives a skewed posterior. It should be noted that ETKF is able to give a non-Gaussian (though wrong) posterior due to the nonlinearity of the map between the uncertain parameters and observations.

4 Single-phase Darcy flow

We consider a steady-state single-phase Darcy flow model defined over an aquifer of two-dimensional physical domain $D = [0,1] \times [0,1]$, which is given by,

$$-\nabla \cdot (k(x,y)\nabla P(x,y)) = f(x,y), \quad (x,y) \in D$$
$$P(x,y) = 0, \quad (x,y) \in \partial D$$

where $\nabla = (\partial/\partial x \; \partial/\partial y)^T$, $\cdot$ denotes the dot product, $P(x,y)$ the pressure, $k(x,y)$ the permeability, $f(x,y)$ the source term, which we assume to be $2\pi^2 \cos(\pi x)\cos(\pi y)$, and ∂D the boundary of domain D. The forward problem of this second order elliptical equation is to find the solution of pressure $P(x,y)$ for given $f(x,y)$ and $k(x,y)$. We, however, are interested in finding permeability given noisy observations of pressure at a few locations.

We perform numerical experiments with synthetic observations, where instead of a measuring device a model is used to obtain observations. We implement a cell-centered finite difference method to discretize the domain D into $n \times n$ grid cells X_i of size Δx^2 and solve the forward model with the true parameters. Then the synthetic observations are obtained by

$$\boldsymbol{y}_{\text{obs}} = \mathbf{L}(\mathbf{P}) + \eta,$$

where with an element of $\mathbf{L}(\mathbf{P})$ is being a linear functional of pressure, namely

$$L_l(\mathbf{P}) = \frac{1}{2\pi\sigma^2} \sum_{i=1}^{n^2} \exp\left(-\frac{||X_i - r_l||^2}{2\sigma^2}\right) P_i \Delta x^2, \quad l \in 1, \ldots, N_y$$

where $n = 50$, $\sigma = 0.01$, r_l denotes the location of the observation and $N_y = 16$, which is the number of observations. The observation locations are spread uniformly across the domain D and η denotes the observation noise drawn from a normal distribution with zero mean and standard deviation of 0.09. 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 et al. (2014).

4.1 Five parameter nonlinear problem

For our first numerical experiment with Darcy flow, we consider a low-dimensional problem where the permeability field is defined by mere 5 parameters similarly to Iglesias et al. (2014). We assume that the entire domain $D = [0,1] \times [0,1]$ is divided into two subdomains D_1 and D_2 as shown in Fig. 2. Each subdomain of D represents a layer and is assumed to have a permeability function $k(\mathbf{X})$, where an element of $\mathbf{X}$ is defined by X_i for $i = 1, \ldots, n^2$. The thickness of a layer on both sides a and b, correspondingly, defines the slope of the interface and a parameter c defines a vertical fault. The layer moves up or down depending on $c < 0$ or $c > 0$, respectively, and its location is assumed to be fixed at $x = 0.5$.

Further, for this test case we assume piecewise constant permeability within each of the subdomains, hence $k(\mathbf{X})$ is given by

$$k(\mathbf{X}) = k_1 \delta_{D_1}(\mathbf{X}) + k_2 \delta_{D_2}(\mathbf{X}),$$

where k_1 and k_2 represent permeability of the subdomain D_1 and D_2, respectively, and δ is Dirac function. Then the parameters defining the permeability field for this configuration are

$$\boldsymbol{u} = (a \; b \; c \; \log(k_1) \; \log(k_2))^T.$$

We assume that the true parameters are $a^{\text{true}} = 0.6$, $b^{\text{true}} = 0.3$, $c^{\text{true}} = -0.15$, $k_1^{\text{true}} = 12$ and $k_2^{\text{true}} = 5$. These parameters are used to create synthetic observations. Figure 2 shows the true permeability with dots representing the observation locations. Next, we assume that the five uncertain parameters are drawn from a uniform distribution over a specified interval, namely $a, b \sim \mathcal{U}[0,1]$, $c \sim \mathcal{U}[-0.5, 0.5]$, $k_1 \sim \mathcal{U}[10, 15]$ and $k_2 \sim \mathcal{U}[4, 7]$.

As it was pointed out in Sect. 2.2, ETPS ETPF updates the parameters within the original range of an initial ensemble, while ETKS ETKF does not. Therefore a change of variables has to be performed for ETKS ETKF so that the updated parameters

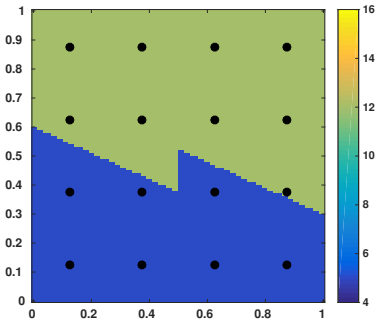

Figure 2. True permeability of the 5 parameter nonlinear problem with dots representing the observation locations.

are physically viable. In order to be consistent we perform the change of variables for ~~ETPS~~ ETPF as well. As the domain D is $[0,1]\times[0,1]$, the parameters a and b should lie within the interval $[0,1]$. To enforce this constraint we substitute a according to

$$a' = \log\left(\frac{a}{1-a}\right), \quad a' \in R$$

and similarly b is substituted by b'. Thus the uncertain parameters are now $\boldsymbol{u}' = (a'\ b'\ c\ \log(k_1)\ \log(k_2))^T$.

In Fig. 3 we plot probability density functions for parameters a (a)–(d), c (e)–(h) and $\log(k_2)$ (i)–(l), as the parameters b and $\log(k_1)$ show similar results. The posterior obtained by IS with ensemble size 10^6 is plotted as a black line and the true value of parameters is plotted as a black line with crosses. The posterior of ~~ETPS~~ ETPF is shown at the top and the posterior of ~~ETKS~~ ETKF at the bottom. ~~ETPS and ETKS~~ ETPF and ETKF used 10^3 (odd columns) and 10^4 (even columns) ensemble members. It is interesting to note that ~~ETKS~~ ETKF overestimates the tails of the pdfs while ~~ETPS underestimates them~~ETPF underestimates them, which indicates that there is not enough spread in the ensemble. While for the parameter c pdf shown in Fig. 3(h) this is an advantage of ~~ETKS~~ETKF, for the parameter $\log(k_2)$ pdf shown in Fig. 3(l) it is most certainly a disadvantage. ~~ETKS~~ ETKF optimises for the mean (and variance), which is better approximated by ~~ETKS than by ETPS~~ETKF than by ETPF, as seen in Fig. 4(e). However this comes at a price of incorrect posterior shown in Fig. 3(k–l).

In order to avoid any bias due to an initial ensemble we perform 10 simulations based on a random draw of an initial ensemble from the same prior distributions. We conduct the numerical experiments for ensemble sizes varying from 10 to 10^3 with an increment of 50. In ~~figure~~ Fig. 4 we plot the true parameters $\boldsymbol{u}^{\text{true}}$, the mean $\bar{\bar{\boldsymbol{u}}}^a$ and the spread $\bar{\bar{\boldsymbol{u}}}^a \pm \bar{\boldsymbol{u}}^a_{\text{std}}$ of estimated parameters averaged over 10 simulations

$$\bar{\bar{u}}^a_i = \frac{1}{10}\sum_{r=1}^{10}\bar{u}^{a,r}_i, \quad \bar{u}^a_{\text{std}} = \frac{1}{10}\sum_{r=1}^{10}\sqrt{\frac{1}{M-1}\sum_{m=1}^{M}(u^{a,r}_{i,m}-\bar{u}^{a,r}_i)^2}, \text{ where } \bar{u}^{a,r}_i = \frac{1}{M}\sum_{m=1}^{M}u^{a,r}_{i,m},\ r=1,\ldots,10,$$

M is ensemble size, $i=1,\ldots,5$ is parameter index, and the superscript a is for the analysis. We observe that both data assimilation methods perform comparably in terms of mean estimation. The spread from ~~ETPS ishowever~~ ETPF is, however, smaller than from ~~ETKS~~ ETKF for each parameter. ETPF provides ensemble members that stay within the original bounds, while ETKF—outside the bounds.

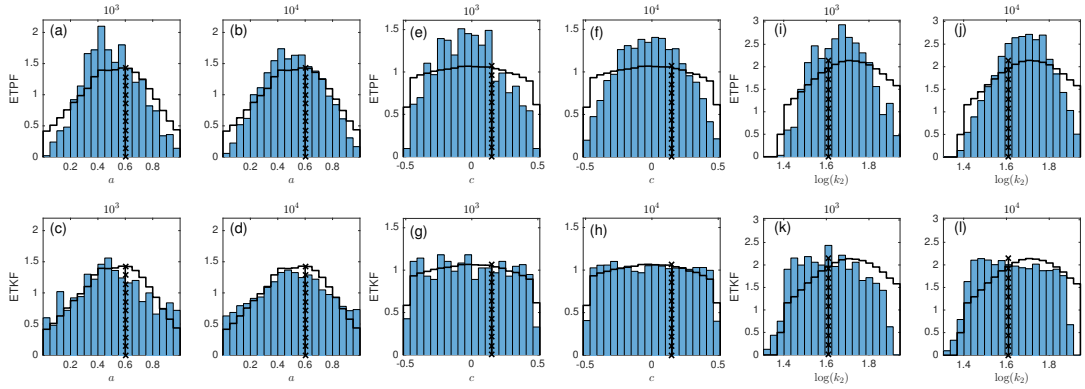

Figure 3. Probability density functions for the parameters a (a)–(d), c (e)–(h), and $\log(k_2)$ (i)–(l). The posterior obtained by IS with ensemble size 10^6 is plotted as a black line and the true values of parameters are plotted as black crosses. The posterior of ~~ETPS~~ ETPF is shown at the top and the posterior of ~~ETKS~~ ETKF at the bottom. ~~ETPS~~ ETPF and ~~ETKS~~ ETKF used 10^3 (odd columns) and 10^4 (even columns) ensemble members.

We compute an average of the relative error over all parameters

$$\text{RE}^{a,r} = \frac{1}{5}\sum_{i=1}^{5} \frac{|\bar{u}_i^{a,r} - u_i^{true}|}{|u_i^{true}|},\ r = 1,\ldots,10,$$

and the data misfit

$$\text{misfit}^{a,r} = (\bar{\boldsymbol{y}}^{a,r} - \boldsymbol{y}_{\text{obs}})^T R^{-1} (\bar{\boldsymbol{y}}^{a,r} - \boldsymbol{y}_{\text{obs}}),\ r = 1,\ldots,10 \tag{13}$$

after data assimilation. The same metrics are computed before data assimilation and denoted by a superscript b. In Fig. 5(a)–(b) we plot $(\text{misfit}^{a,r} - \text{misfit}^{b,r})$ and $(\text{RE}^{a,r} - \text{RE}^{b,r})$, respectively, for each simulation r as a function of ensemble size. ~~ETPS~~ ETPF is shown in blue and ~~ETKS~~ ETKF in red. Black line is at zero level. Positive values of the differences mean an increase of either data mismatch or relative error after data assimilation. We observe a data misfit decrease for both ~~ETPS and ETKS~~ ETPF and ETKF except at an ensemble size 10. RE does not always decrease for ~~ETPS~~ETPF: for some simulations ~~ETPS~~ ETPF is at zero level or slightly above it, while for ~~ETKS~~ ETKF the sole exception is at an ensemble size 10.

4.2 High-dimensional nonlinear problem

Next, we consider a high-dimensional problem where the dimension of the uncertain parameter is $n^2 = 2500$. The domain D is now not divided into subdomains. However, unlike in the previous test case here we implement a spatially varying permeability field. We assume the log permeability is generated by a random draw from a Gaussian distribution $\mathcal{N}(\log(\mathbf{5}), \mathbf{C})$. Here $\mathbf{5}$ is an n^2 vector with all 5. $\mathbf{C}$ is assumed to be an exponential correlation with ~~maximum correlation along $3\pi/4$,~~ an element of $\mathbf{C}$ ~~is~~ being

$$C_{i,j} = \exp(-3(|h_{i,j}|/v)),\ i,j = 1,\ldots,n^2.$$

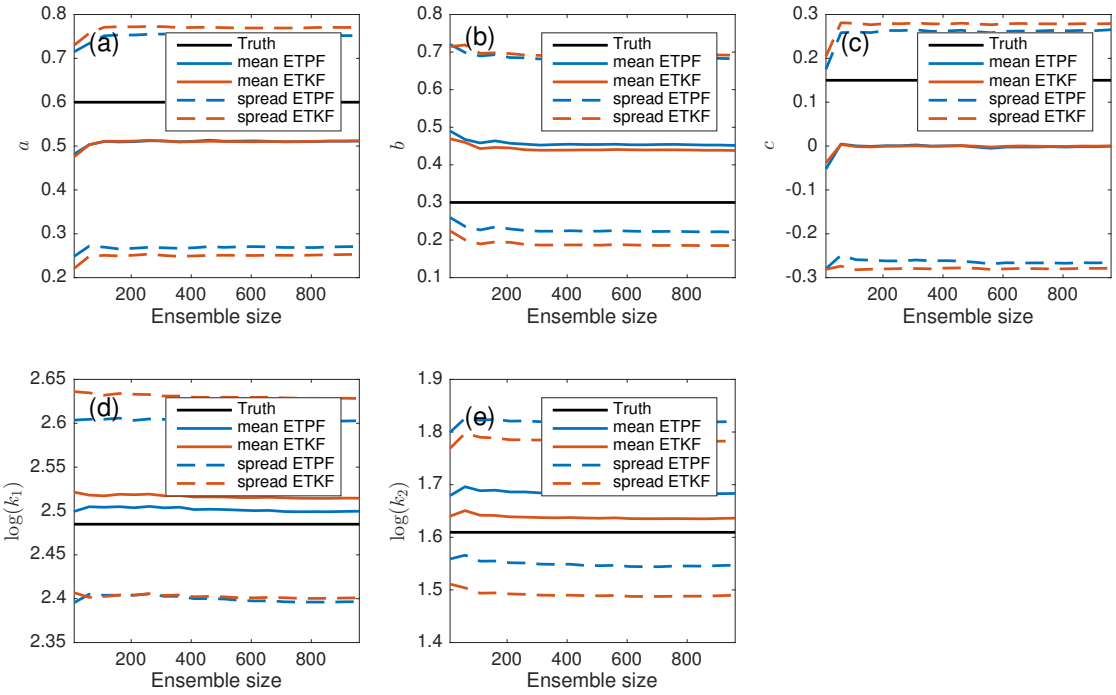

Figure 4. $\bar{\boldsymbol{u}}^a$ and $\bar{\boldsymbol{u}}^a \pm \bar{\boldsymbol{u}}^a_{\text{std}}$ w.r.t ensemble size: (a) for the parameter a, (b) for b, (c) for c, (d) for $\log(k_1)$, (e) for $\log(k_2)$. ~~ETPS~~ ETPF is shown in blue, ~~ETKS~~ ETKF in red and the true parameters are in black.

Here $h_{i,j}$ is the distance between two spatial locations and v is the correlation range which is taken to be 0.5. ~~As the covariance matrix is symmetric, we factorize it in upper and lower triangular matrices using Cholesky decomposition and denote the upper triangular matrix by **G**. Next, we generate a~~ For the log permeability we use Karhunen-Loeve expansions of the form

$$\log(k_j) = \log(5) + \sum_{i=1}^{n^2} \sqrt{\lambda_i}\nu_{i,j}\mathcal{Z}_i, \quad \text{for} \quad j = 1,\ldots,n^2 \tag{14}$$

where λ and ν are eigenvalues and eigenfunctions of $\mathbf{C}$, respectively, and the vector $\mathcal{Z}$ is of dimension n^2 iid from a Gaussian distribution with zero mean and variance one~~, and create a permeability field $\log(\mathbf{k}(\mathbf{X}))$ across the domain D according to Oliver et al. (2008):~~

$$\log(\mathbf{k}) = \log(\mathbf{5}) + \mathbf{G}^T\mathcal{Z},$$

. Making sure that the eigenvalues are sorted in descending order $\mathcal{Z}_i \sim \mathcal{N}(0,1)$ produces $\log(\mathbf{k}) \sim \mathcal{N}(\log(\mathbf{5}), \mathbf{C})$. The uncertain parameter is ~~$\boldsymbol{u} = \mathcal{Z}^T$. Thus the dimension of the uncertain parameter is~~ thus $\boldsymbol{u} = \mathcal{Z}$ with the dimension $n^2 = 2500$.

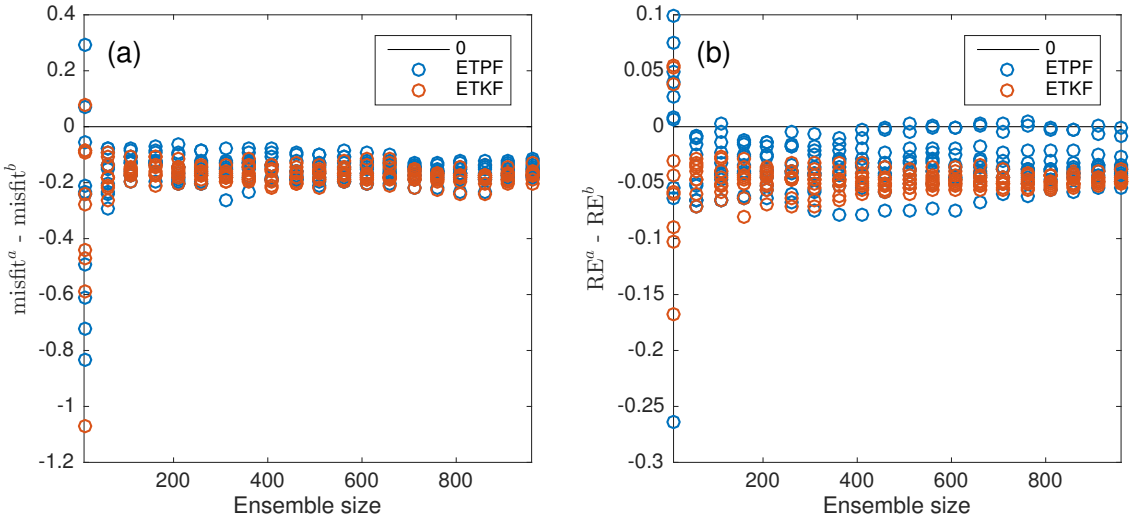

Figure 5. misfit^{a,r} − misfit^{b,r} (a) and RE^{a,r} − RE^{b,r} (b) w.r.t ensemble size. ~~ETPS~~ ETPF is shown in blue, ~~ETKS~~ ETKF in red and the zero level in black. A circle is for one simulation.

~~In Fig. 11 we plot the posterior pdf of first three modes $\mathcal{Z}_1$ (left), $\mathcal{Z}_2$ (center), and $\mathcal{Z}_3$ (right) obtained by IS with ensemble size 10^6 and by ETPS (top) and ETKS (bottom) with ensemble size 10^4. The posterior of these modes is roughly approximated by ETPS as shown in Fig. 11(a)–(c). This means that an ensemble size 10^4 is not sufficient for ETPS to correctly estimate the posterior, when the total number of modes is large (2500). ETKS provides a skewed posterior of the modes shown in Fig. 11(d)–(f), which was also observed in the one parameter nonlinear problem, see Fig. 1(f). tThe posterior probability density function of parameters $\mathcal{Z}_1$ (left), $\mathcal{Z}_2$ (center), and $\mathcal{Z}_3$ (right). The posterior obtained by IS with ensemble size 10^6 is plotted as a black line and the true parameter as a black cross. The posterior of ETPS is shown at the top and the posterior of ETKS at the bottom. Both ETPS and ETKS used 10^4 ensemble members.~~

We perform 10 different simulations based on a random draw of an initial ensemble from the prior distribution. We conduct the numerical experiments for ensemble sizes varying from 10 to 10^3 with an increment of 50. ~~In Fig. 4 we plot mean and spread for $\mathcal{Z}_1$ (a), $\mathcal{Z}_2$ (b), and $\mathcal{Z}_3$ (c) averaged over 10 simulations in blue for ETPS and in red for ETKS. The true variables~~

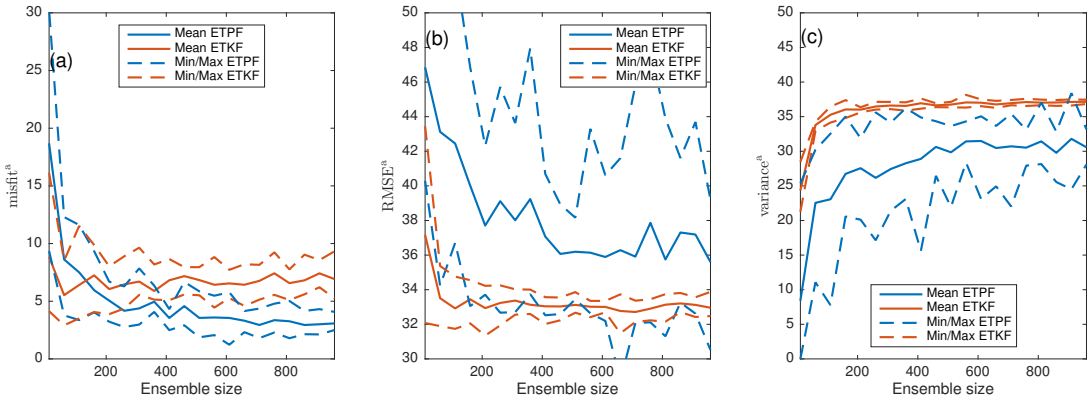

Figure 6. Mean, minimum and maximum over 10 simulations after data assimilation for the data misfit (a), RMSE (b), and variance (c). ~~ETPS~~ ETPF is shown in blue and ~~ETKS~~ ETKF in red.

~~are shown in black. We observe that in terms of the mean estimation of the first three modes ETPS outperforms ETKS. The spread provided by ETPS is smaller than the one provided by ETKS, as in the previous test cases. tMean and spread for $\mathcal{Z}_1$ (a), $\mathcal{Z}_2$ (b), and $\mathcal{Z}_3$ (c) w.r.t ensemble size. ETPS is shown in blue, ETKS in red and the true parameters are in black.~~

We compute the root mean square error (RMSE) of the log permeability field

$$\text{RMSE}^{r,a} = \sqrt{\frac{1}{n^2}(\overline{\mathcal{Z}}^{a,r} - \mathcal{Z}^{\text{true}})^T \mathbf{C} (\overline{\mathcal{Z}}^{a,r} - \mathcal{Z}^{\text{true}})} \sqrt{\left(\log(\overline{\mathbf{k}}^{a,r}) - \log(\mathbf{k}^{\text{true}})\right)^T \left(\log(\overline{\mathbf{k}}^{a,r}) - \log(\mathbf{k}^{\text{true}})\right)}, \quad r = 1, \ldots, 10,$$

and variance

$$\text{variance}^{r,a} = \frac{1}{M-1} \sum_{m=1}^{M} \left((\mathcal{Z}\log(\mathbf{k}_m^{a,r}) - \log(\overline{\mathbf{k}}^{a,r}) \right)^T ~~\mathbf{C}~~ \left(\log(\mathcal{Z}\mathbf{k}_m^{a,r}) - \log(\overline{\mathbf{k}}^{a,r}) \right), \quad r = 1, \ldots, 10.$$

We also compute the data misfit for each simulation after data assimilation by Eq. (13). In Fig. 6 we plot mean, minimum and maximum over 10 simulations after data assimilation for the data misfit (left), RMSE (center), and variance (right). ~~ETPS~~ ETPF is shown in blue and ~~ETKS~~ ETKF in red. We observe that ~~ETPS in~~ ETPF is underdispersive compared to ~~ETKS. This is due to the linear transformation, as it results in some ensemble members being nearly identical. Therefore a perturbation of ensemble members is needed, which could be performed based on random walk. This is, however, out of the scope of this paper.~~ ETKF as particle filters are highly degenerative compared to Kalman filters. Misfit given by ~~ETPS~~ ETPF is smaller than the one given by ~~ETKS~~ ETKF for almost all simulations at ensemble sizes greater than 150. The RMSE on the contrary is larger. In Fig. 7(a)–(b) we plot $(\text{misfit}^{a,r} - \text{misfit}^{b,r})$ and ~~$(\text{RE}^{a,r} - \text{RE}^{b,r})$~~ $(\text{RMSE}^{a,r} - \text{RMSE}^{b,r})$, respectively, as a function of ensemble size for a simulation $r = 1, \ldots, 10$. The superscript b is for the metrics before data assimilation and the superscript a is for the metrics after data assimilation. ~~ETKS~~ ETKF always provides a decrease in both the data misfit and RMSE except at ensemble size 10. ~~ETPS~~ ETPF gives a decrease in the data misfit though an increase in RMSE, which indicates that ETPF overfittes the data. However, as the ensemble size increases this happens less often as can be seen in Fig. 7(c), where we plot

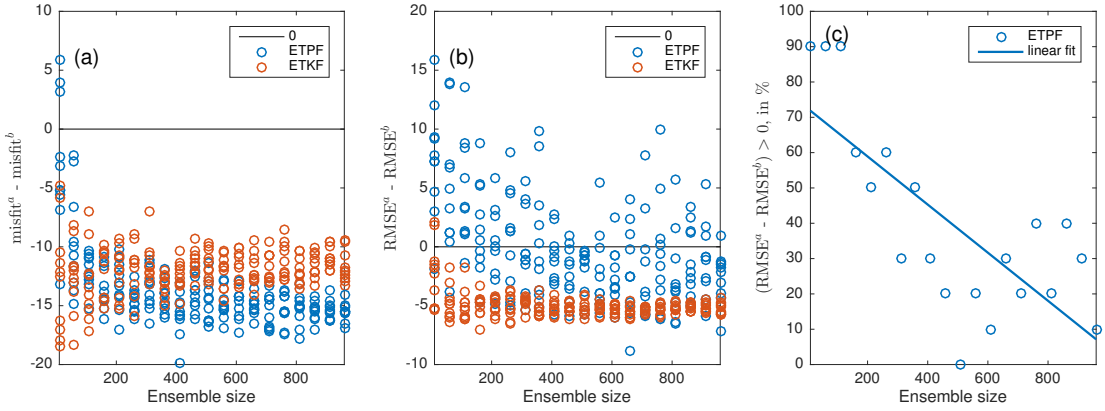

Figure 7. misfit^{a,r} − misfit^{b,r} (a) and ~~RE^{a,r} − RE^{b,r}~~ RMSE^{a,r} − RMSE^{b,r} (b) w.r.t ensemble size. ~~ETPS~~ ETPF is shown in blue, ~~ETKS~~ ETKF in red and zero level in black. One circle is for one simulation. For ~~ETPS~~ ETPF % of simulations that result in (RMSE^a − RMSE^b) > 0 and a linear fit as a function of ensemble size are shown in (c).

for ~~ETPS~~ ETPF a percentage of simulations that result in (RMSE^a − RMSE^b) > 0 and a linear fit as a function of ensemble size.

In Fig. 8 we plot log permeability fields. In Fig. 8(a) the true permeability is shown with dots representing the observation locations, and in Fig. 8(d) the mean permeability field obtained by IS with ensemble size 10^5. The RMSE provided by IS is 32.62. In Fig. 8(b–e) and Fig. 8(c–f) we display mean permeability fields obtained with ensemble size 10^3 by ~~ETPS and ETKS~~ETPF and ETKF, respectively. In Fig. 8(b–c) we plot the mean log permeabilities for the smallest RMSE over simulations, which is 30.51 for ~~ETPS~~ ETPF and 32.48 for ~~ETKS~~ETKF. In Fig. 8(d–e) we plot the mean log permeabilities for the largest RMSE over simulations, which is 39.2 for ~~ETPS and 33.9 for ETKS~~ETPF and 33.87 for ETKF. We observe that ~~ETKS~~ ETKF as well as IS provide smooth mean permeability fields that have smaller absolute values than the true permeability. ~~ETPS~~ ETPF gives higher variations of the mean permeability field and is in an excellent agreement with the true permeability for a good initial ensemble shown in Fig. 8(b). ~~However, it remains unclear how to choose a good initial ensemble~~This means that ETPF is sensitive to the initial sample is due to the sampling error and that the spatial variability of ETPF is a result of the sampling error. It should be noted that IS with ensemble size 10^3 and this good initial ensemble gives the RMSE 30.51 and the same mean log permeability field as ~~ETPS~~ ETPF shown in Fig. 8(b). ~~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.~~

In Fig. 9 we plot variance of the permeability fields obtained with ensemble size 10^5 by IS (d), with ensemble size 10^3 by ~~ETPS~~ ETPF (b–e) and ~~ETKS~~ ETKF (c–f) . ~~ETPS (b) and ETKS~~ Fig. 9(b–c) is for the smallest RMSE and Fig. 9(e–f) is for the largest RMSE. ETKF again provides smoother variance than ETPF.

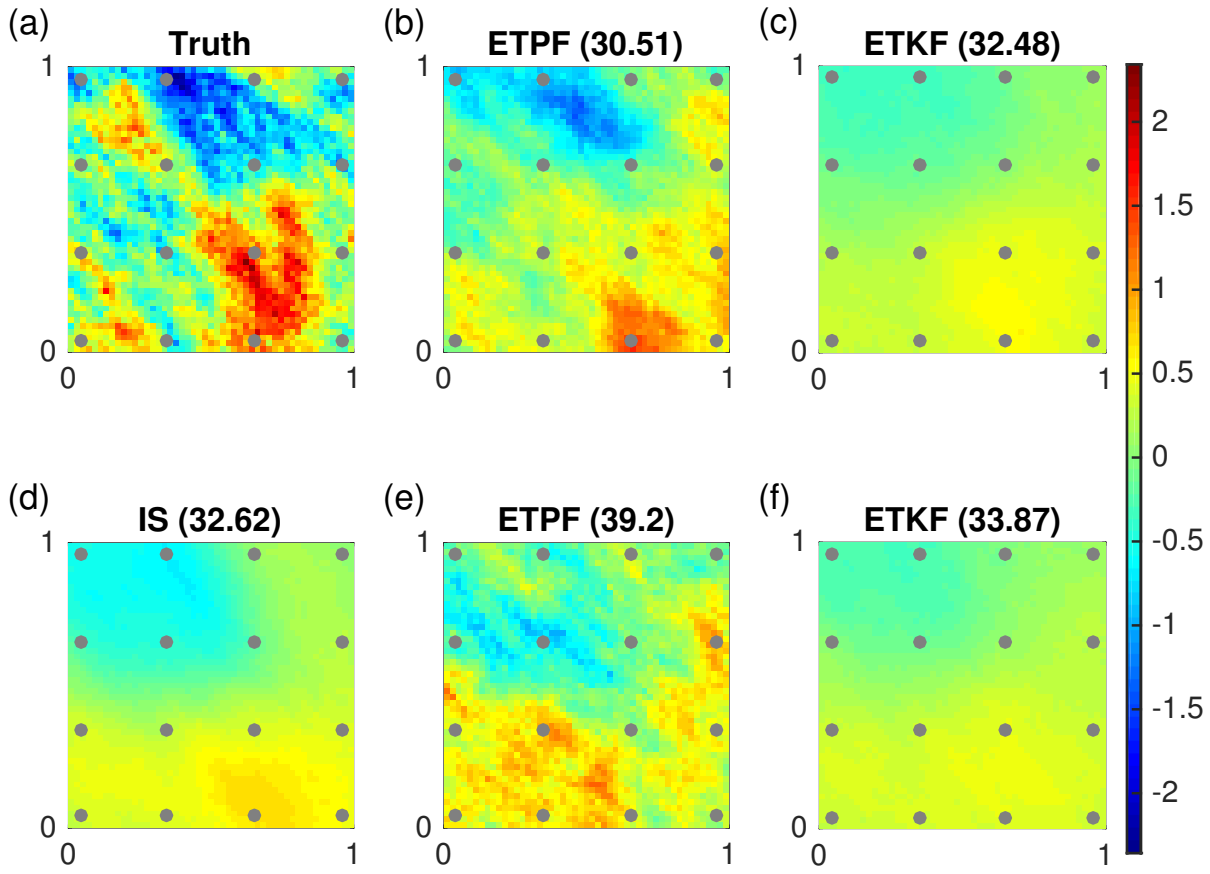

Figure 8. Log permeability field: truth with dots representing the observation locations (a), mean obtained with ensemble size 10^5 by IS (d), mean obtained with ensemble size 10^3 by ETPF (b–e) and by ETKF (c–f). Mean at the smallest RMSE (b–c) and at the largest RMSE (e–f) over simulations. The corresponding RMSE is given in brackets.

In Fig. 10 we show squared error $(\overline{\mathcal{Z}}^a - \mathcal{Z}^{\text{true}})^2$ in blue for ETPF and in red for ETKF for three first modes $\mathcal{Z}_1$ (a), $\mathcal{Z}_2$ (b), and $\mathcal{Z}_3$ (c)~~are for the smallest RMSE,~~ , where solid line is for median and shaded area is for 25 and ~~ETPS (e)~~ 75 percentile over 10 simulations. We observe that in terms of the estimation of the first three modes ETPF outperforms ETKF. In Fig. 11 we plot the posterior of $\mathcal{Z}_1$ (left), $\mathcal{Z}_2$ (center), and $\mathcal{Z}_3$ (right) obtained by IS with ensemble size 10^6 and by ETPF (top) and ETKF (bottom) with ensemble size 10^4. The posterior of these modes is roughly approximated by ETPF as shown in Fig. 11 (a)–(c). ETKF provides a skewed posterior of the modes shown in Fig. 11 (d)–(f), which was also observed in the one parameter nonlinear problem, see Fig. 1(f). In order to perform an objective comparison between the probabilities we compute the Kullback-Leibler

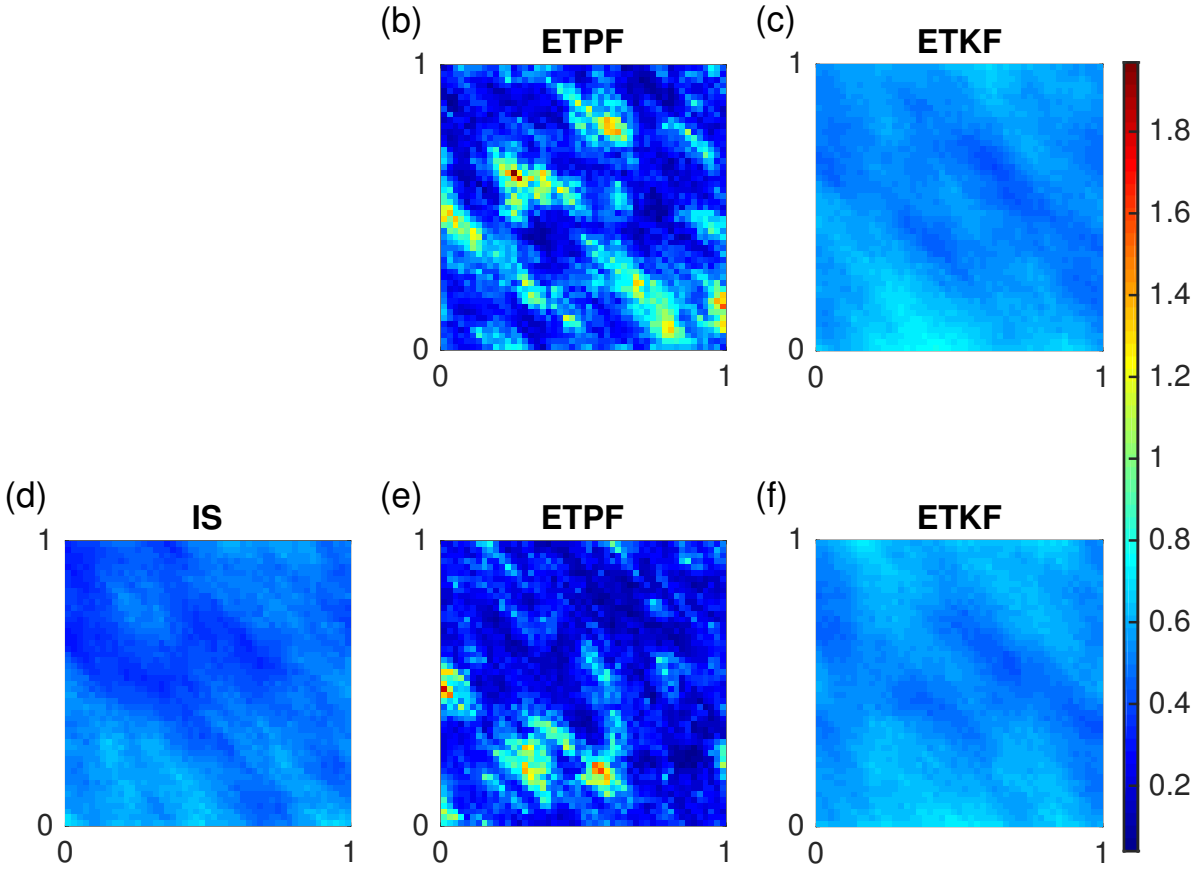

Figure 9. Variance of log permeability fields: obtained with ensemble size 10^5 by IS (d), with ensemble size 10^3 by ETPF (b–e), and ETKF (c–f). Variance at the smallest RMSE (b–c) and at the largest RMSE (e–f) over simulations.

divergence of a posterior π obtained by either ETPF or ETKF and the posterior π^{IS} obtained by IS

$$\mathrm{D}_{\mathrm{KL}}(\pi^{\text{IS}} \parallel \pi) = \sum_{i=1}^{N_{\mathrm{b}}} \pi^{\text{IS}}(u_i) \log \frac{\pi^{\text{IS}}(u_i)}{\pi(u_i)},$$

where $N_{\mathrm{b}} = 100$ is the number of bins. ETPF gives the Kullback-Leibler divergence 0.21, 0.42, and 0.6, while ETKF 0.16, 0.07, and 0.5 for the modes $\mathcal{Z}_1$, $\mathcal{Z}_2$, and $\mathcal{Z}_3$, respectively. Thus ETKF gives a better approximation of the true pdf. We use only first three modes in the Karhunen-Loeve expansion given by Eq. (14) when computing the estimated log permeability keeping the number of uncertain parameters the same, namely 2500. In Fig. 12(a) 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

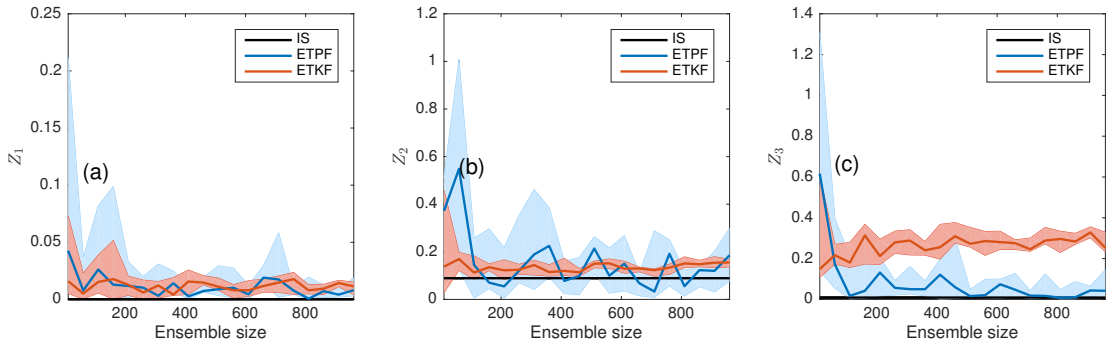

Figure 10. Squared error between the true and the mean estimated modes for $\mathcal{Z}_1$ (a), $\mathcal{Z}_2$ (b), and $\mathcal{Z}_3$ (c) w.r.t ensemble size. ETPF is shown in blue and ETKF in red with solid lines for median and shaded area for 25 and 75 percentile over 10 simulations. IS with ensemble size 10^5 is in black.

Table 1. Optimal localization radius for ETPS and ETKS at different ensemble sizes M.

M	10	110	210	...	910
ETPS	~~0.2~~ 0.4	0.6	~~0.8~~ 0.6	...	~~0.8~~0.6
ETKS	~~0.2~~ 0.6	1.2	1.2	...	1.2

always decreases after data assimilation except at small ensemble sizes shown in Fig. 12(b). 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 ~~ETKS (f) are for the largest RMSE over simulations. ETKS again provides smoother variance than ETPS.~~ 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.

Next we employ localization for both ~~ETPS and ETKS. Optimal localization radius~~ ETPF and ETKF. The optimal localization radius was obtained in terms of the smallest RMSE ~~was obtained for one simulation (~~and shown in Table 1~~) and fixed for the remaining 9 simulations~~. It should be noted that smaller localization radius for ETPF than for ETKF was also observed

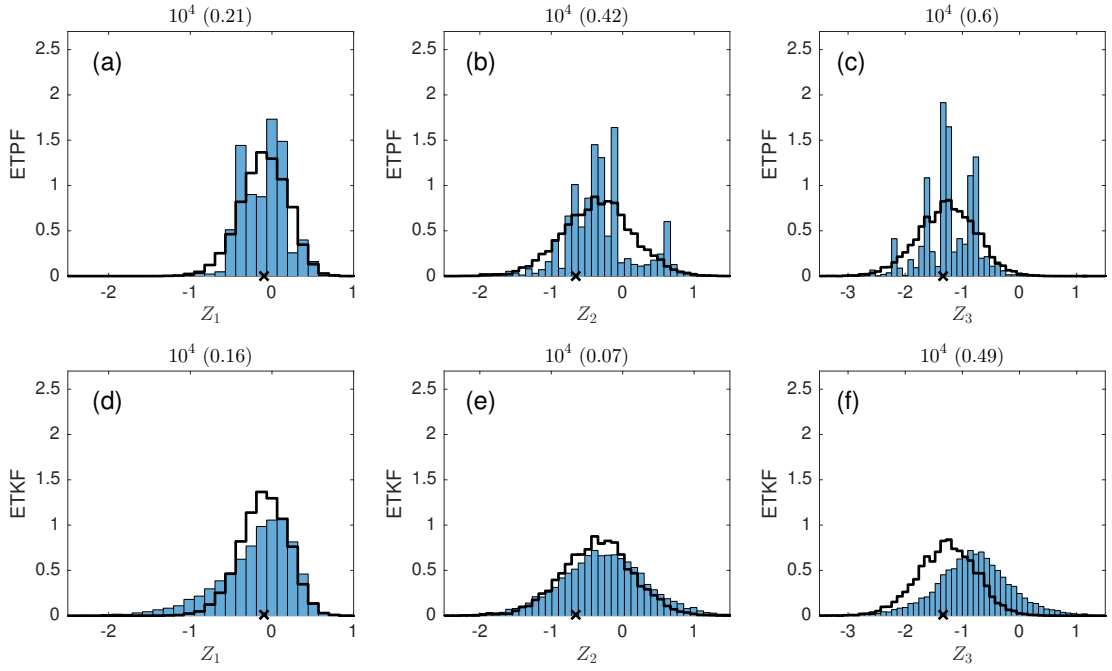

Figure 11. The posterior probability density function of parameters $\mathcal{Z}_1$ (left), $\mathcal{Z}_2$ (center), and $\mathcal{Z}_3$ (right). The posterior obtained by IS with ensemble size 10^6 is plotted as a black line and the true parameter as a black cross. The posterior of ETPF is shown at the top and the posterior of ETKF at the bottom. Both ETPF and ETKF used 10^4 ensemble members. The Kullback-Leibler divergence is in brackets.

by Chen and Reich (2015) for Lorenz 96 model and it is probably related to more noisy approximation of the posterior by ETPF than by ETKF. In Fig. **??**~~(a–c) we plot change in~~ 14 we plot misfit, RMSE and ~~the percentage of simulations for which RMSE of ETPS increased after data assimilation, respectively. ETKS with localization gives an equivalent performance as without localization at~~ variance.

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 ~~, which is to be expected as the optimal localization radius is large. For the small ensemble size $M = 10$ ETKS with localization performs worse than without localization. This could be related to a need of more advanced losalization methods based on wavelets rather than on distance, see Chen and Oliver (2012 ETPS on the other hand highly improved when localization was applied: for ensemble~~ ETPF performs now comparably to ETKF (by increasing the localization radius to 1.2 we do not see an improvement in ETKF). Moreover, localized ETPF overfits the data less often than non-localized: 40% agains 90% for ensemble size 10 and 0% agains non-zero% for ensemble sizes greater than ~~300 all simulations result in the RMSE decrease after data assimilation. The RMSE decreased and the variance increased for all ensemble sizes as~~ 150 (not shown).

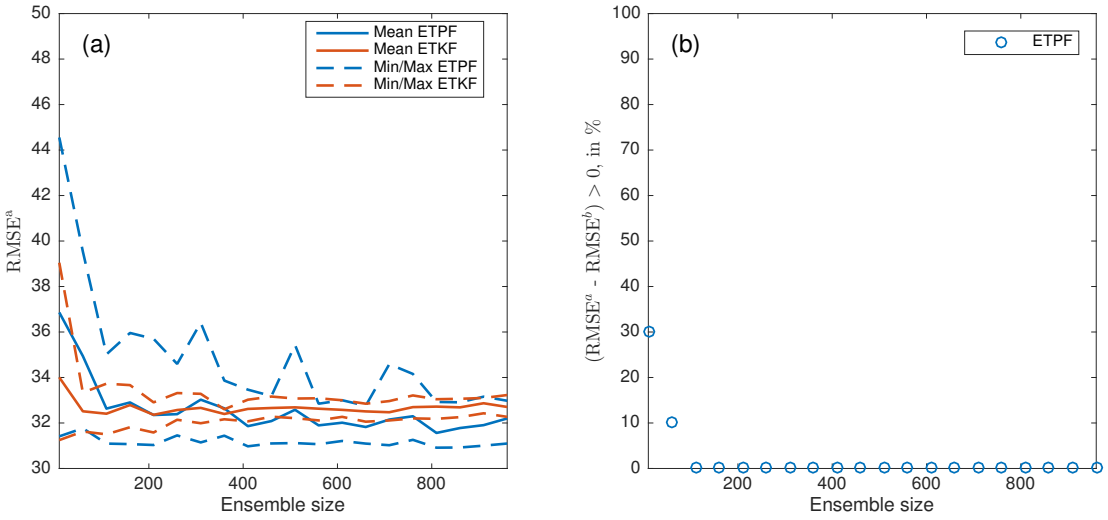

Figure 12. ~~Log permeability field: truth with dots representing~~ Using only three first modes in the ~~observation locations~~ KL expansion. Panel (a)~~, mean obtained with~~ : RMSE after data assimilation w.r.t ensemble size ~~10^5 by IS (d), mean obtained~~ with ~~ensemble size 10^3 by ETPS (b–e)~~ mean, minimum and ~~by ETKS (e–f)~~maximum over 10 simulations for ETPF shown in blue and ETKF in red. ~~Mean at the smallest RMSE~~ Panel (~~b–e~~b)~~and at the largest RMSE (e–f) over~~ : % of simulations that result in $(\text{RMSE}^a - \text{RMSE}^b) > 0$ for ETPF.

In Fig. 15–16 we plot mean and variance of the log permeability field at ensemble size 10^3 for ETPF (b)–(e) and ETKF (c)–(f) with localization at the smallest RMSE (b)–(c) and largest RMSE (e)–(f) over simulations, which are 32.29 and 34.08 for ETPF and 32.92 and 34.09 for ETKF, respectively. We observe that localization decreases the sampling noise and the spatial variability of the mean field obtained by ETPF at ensemble size 10^3 resembles IS at ensemble size 10^5. The variance obtained by ETPF with localization shown in Fig. ~~14(b–c) compared to Fig. 6(b–c). The posterior estimations~~16(b–e) has also improved.

The posterior estimation of the first mode $\mathcal{Z}_1$, however, degraded~~: the pdf shown in Fig. ??(a–c)with ensemble size 10^4 resembles the posterior obtained by ETKS shown in Fig. 11(d–e).Therefore on the one hand localization improved the RMSE behaviour for all simulations but on the other hand it deteriorated the posterior estimation.The increase in RMSE after data assimilation without localization could be related to a substantial adjustment of~~ , while of $\mathcal{Z}_2$ and $\mathcal{Z}_3$ improved. The Kullback-Leibler divergence for the first mode is 0.73 (compare to 0.21 without localization), and for second and third is 0.2 and 0.18, correspondently (compare to 0.42 and 0.6 without localization). Variance of the posteriors is larger when localization is applied for both ETPF and ETKF. The localized weights given by Eq. (11) vary less than the non-localized weights given by Eq. (3). Therefore the ~~uncertain parameters. A possible solution to this problem is an iterative approach to data assimilation as it has been done for ensemble Kalman smoothers (e.g Chen and Oliver, 2013; Emerick and Reynolds, 2013; Bocquet and Sakov, 2014) and will be a subject of our future study~~localized pdf is less noisy than the non-localized. However, localization applied in the form of the Karhunen-Loeve expansion given by Eq. (14) does not retain the imposed bounds on the modes $\mathcal{Z}$ as we need to invert a

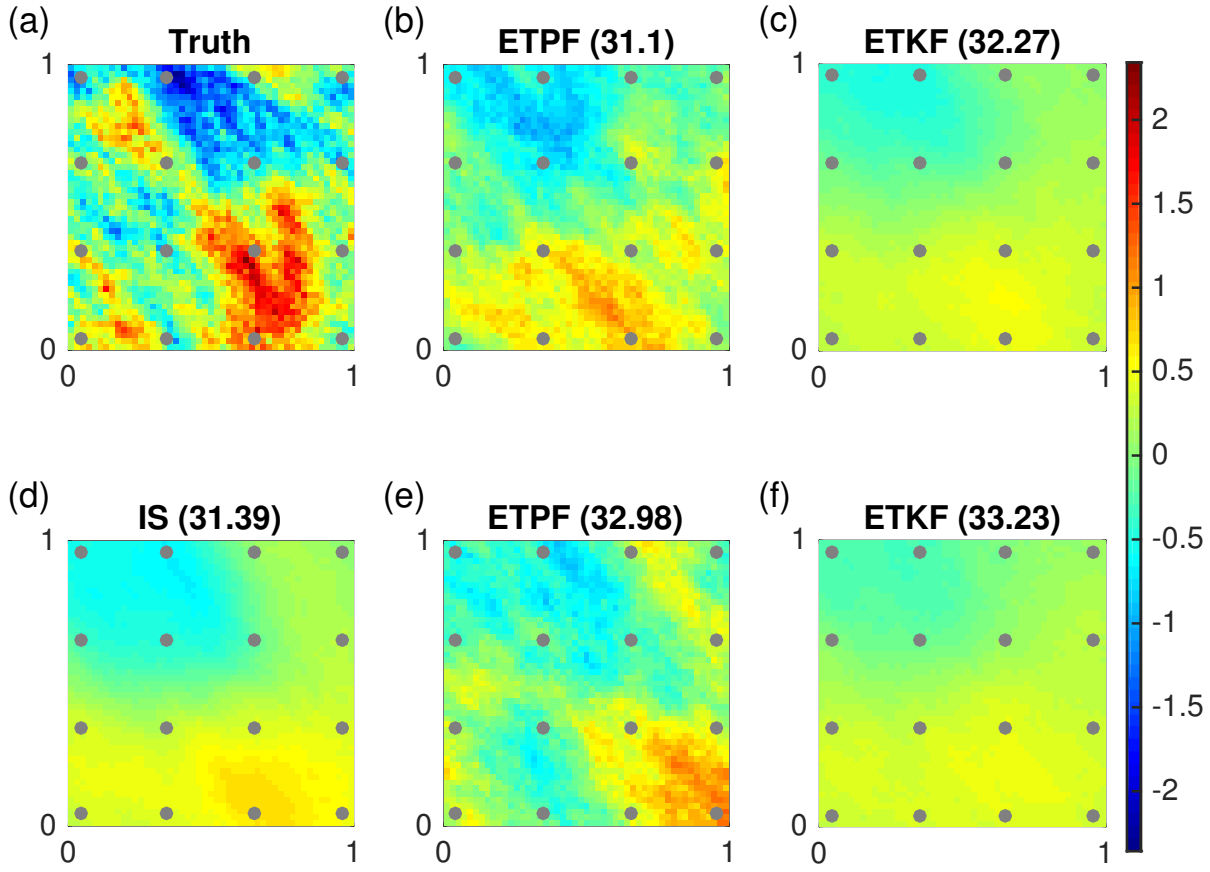

Figure 13. ~~Variance of log permeability fields: obtained with ensemble size 10^5 by IS (d)~~Same as figure 8, ~~with ensemble size 10^3 by ETPS (b–c), and ETKS (e–f). Variance at~~ but using only three first modes in the ~~smallest RMSE (b–c) and at the largest RMSE (e–f) over simulations~~KL expansion.

matrix product of eigenvalue and eigenvector matrices to obtain the modes. By increasing the localization radius to 1.2 we get the Kullback-Leibler divergence 0.64 for the first mode, and 0.13 and 0.11 for the second and third, correspondently, thus the posterior approximation improves only slightly.

5 Conclusions

MCMC methods remain the most reliable methods for estimating the posterior distributions of uncertain model parameters and states. They, however, also remain computationally expensive. Ensemble Kalman filters ~~(or smoothers)~~ provide computation-

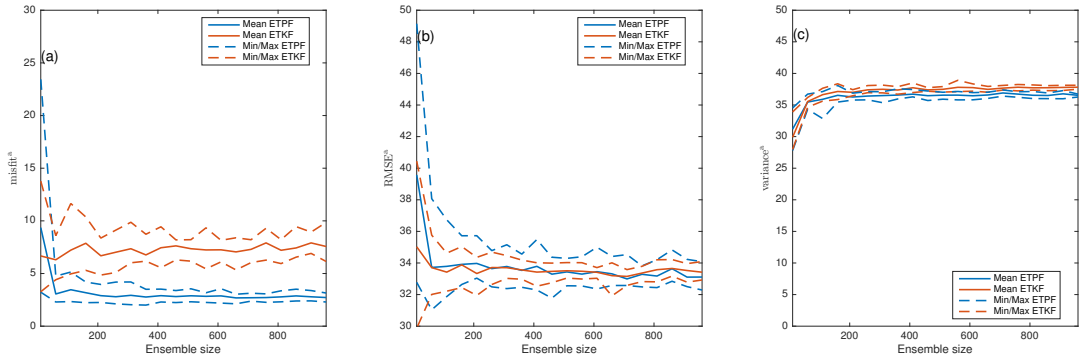

Figure 14. Same as figure ~~7~~6, but with localization.

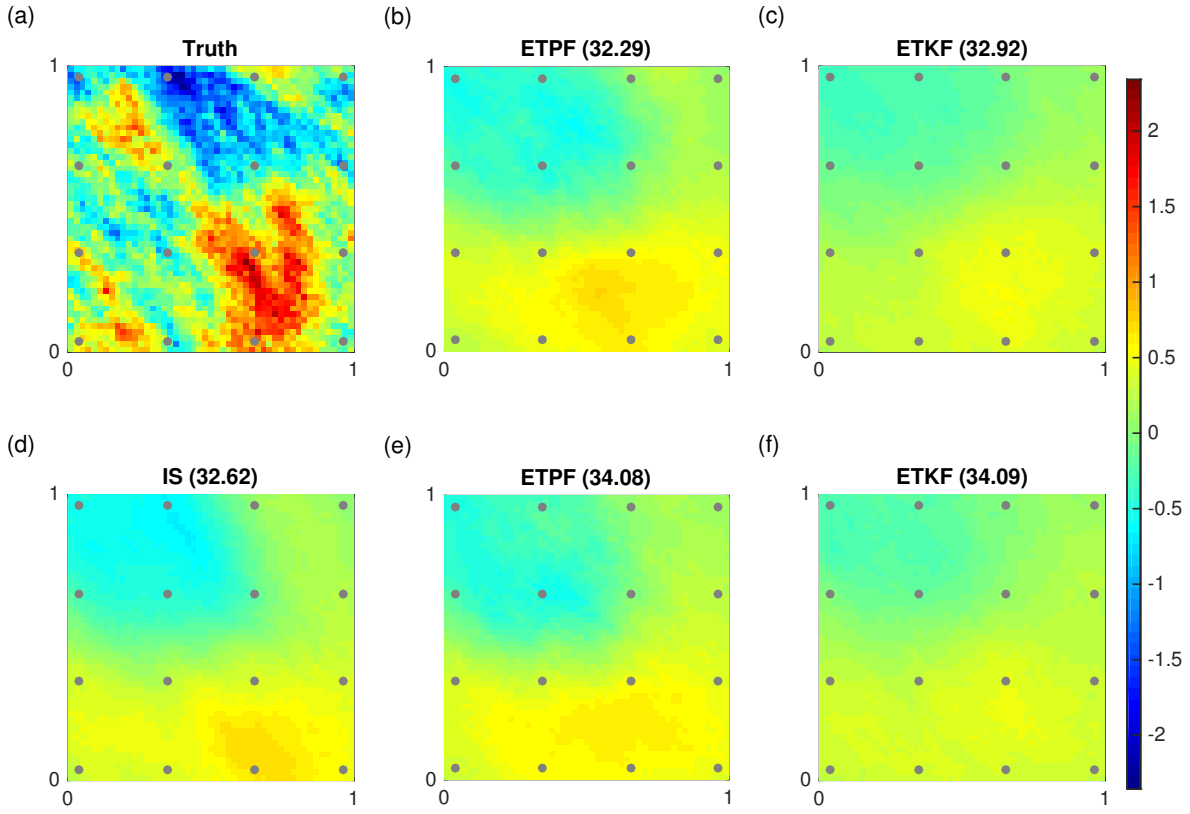

Figure 15. Same as figure ~~6~~8, but with localization.

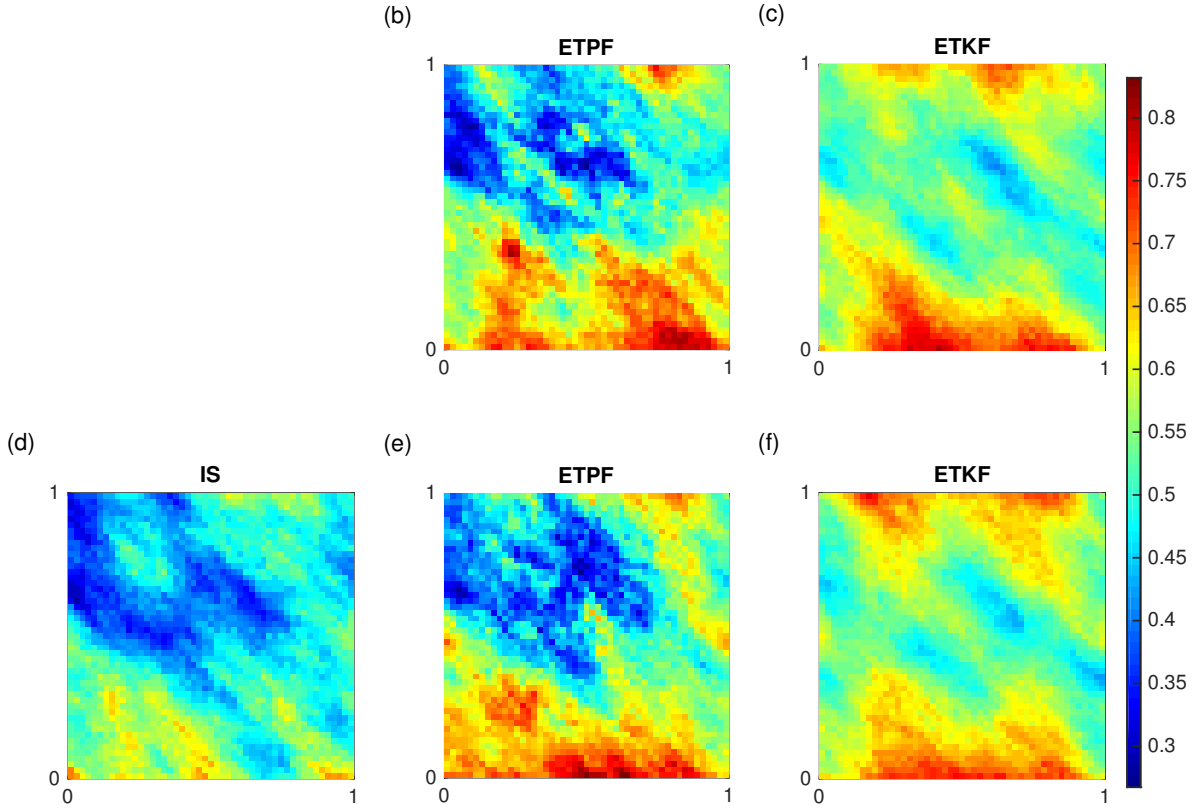

Figure 16. ~~The posterior probability density function of parameters $\mathcal{Z}_1$ (a)~~Same as figure 9, ~~$\mathcal{Z}_2$ (b), and $\mathcal{Z}_3$ (c). The posterior obtained by IS~~ but with ~~ensemble size 10^6 is plotted as a black line and the true parameter as a black cross. The posterior of ETPS with~~ localization~~used 10^4 ensemble members~~.

ally affordable approximations but rely on the assumptions of Gaussian probabilities. For nonlinear models even if the prior is Gaussian the posterior is not Gaussian anymore. Particle filtering ~~(or importance sampling)~~ on the other hand does not have such an assumption but requires a resampling step, which is ~~a challenge for parameter estimation~~usually stochastic. Ensemble transform particle filter ~~(or smoother)~~ is a particle filtering method that deterministically resamples the particles ~~(samples)~~ based on their importance weights and covariance maximization among the particles. ~~Therefore it has an importance sampling advantage of predicting the correct posterior but does not have its disadvantage of resampling lacking.~~

~~In this paper, we have shown that ETPS outperforms ETKS for the posterior estimations in both low- and high-dimensional nonlinear problems. Moreover as the ensemble size increases the posterior of ETPS converges to the posterior of IS with a large ensemble, while the posterior of ETKS remains unchanged. However, in the high-dimensional problem of~~ 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 consistently better than the ones obtained by ETKF. When the number of uncertain parameters is large (2500~~uncertain parameters for some simulations ETPS gives an increase in the RMSE after data assimilation. This issue is resolved once distance-based localization is applied, which however deteriorated the posterior estimation~~) a decrease of degrees of freedom is essential. This is performed by localization. At large ensemble sizes ETPF performs as well as ETKF, while at small ensemble sizes 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.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work is part of the research programme Shell-NWO/FOM Computational Sciences for Energy Research (CSER) with project number 14CSER007 which is partly financed by the Netherlands Organization for Scientific Research (NWO).

References

Aanonsen, S. I., Nævdal, G., Oliver, D. S., Reynolds, A. C., Vallès, B., et al.: The ensemble Kalman filter in reservoir engineering–a review, SPE Journal, 14, 393–412, 2009.

Bishop, C. H., Etherton, B. J., and Majumdar, S. J.: Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects, Monthly Weather Review, 129, 420–436, 2001.

Bocquet, M. and Sakov, P.: An iterative ensemble Kalman smoother, Quarterly Journal of the Royal Meteorological Society, 140, 1521–1535, 2014.

Burgers, G., van Leeuwen, P., and Evensen, G.: Analysis Scheme in the Ensemble Kalman Filter, Monthly Weather Review, 126, 1719–1724, 1998.

Chen, Y. and Oliver, D. S.: Multiscale parameterization with adaptive regularization for improved assimilation of nonlocal observations, Water Resources Research, 485, 15pp, 2012b.

Chen, Y. and Oliver, D. S.: Levenberg–Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification, Computational Geosciences, 17, 689–703, 2013.

Chen, Y. and Reich, S.: Data assimilation: a dynamical system perspective, Frontiers in Applied Dynamical Systems: Reviews and Tutorials, 2, 75–118, 2015.

Doucet, A., de Freitas, N., and Gordon, N.: Sequential Monte-Carlo Methods in Practice, Springer-Verlag, New York, 2001.

Dovera, L. and Della Rossa, E.: Multimodal ensemble Kalman filtering using Gaussian mixture models, Computational Geosciences, 15, 307–323, 2011.

Emerick, A. A. and Reynolds, A. C.: Ensemble smoother with multiple data assimilation, Computers & Geosciences, 55, 3–15, 2013.

Evensen, G.: Data assimilation: the ensemble Kalman filter, Springer Science & Business Media, 2009.

Evensen, G. van Leeuwen, P.: An ensemble Kalman smoother for nonlinear dynamics, Monthly Weather Review, 128, 1852–1867, 2000.

Gaspari, G. and Cohn, S. E.: Construction of correlation functions in two and three dimensions, Quarterly Journal of the Royal Meteorological Society, 125, 723–757, 1999.

Guingla, P., Antonio, D., De Keyser, R., De Lannoy, G., Giustarini, L., Matgen, P., and Pauwels, V.: The importance of parameter resampling for soil moisture data assimilation into hydrologic models using the particle filter, Hydrology and Earth System Sciences, 16, 375–390, 2012.

Hunt, B., Kostelich, E., and I., S.: Efficient data assimilation for spatialtemporal chaos: A local ensemble transform Kalman filter., Physica D, 230, 112–137, 2007.

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.

Moradkhani, H., Sorooshian, S., Gupta, H. V., and Houser, P. R.: Dual state–parameter estimation of hydrological models using ensemble Kalman filter, Advances in water resources, 28, 135–147, 2005.

Oliver, D. S. and Chen, Y.: Recent progress on reservoir history matching: a review, Computational Geosciences, 15, 185–221, 2011.

Oliver, D. S., Cunha, L. B., and Reynolds, A. C.: Markov chain Monte Carlo methods for conditioning a permeability field to pressure data, Mathematical Geology, 29, 61–91, 1997.

Oliver, D. S., Reynolds, A. C., and Liu, N.: Inverse theory for petroleum reservoir characterization and history matching, Cambridge University Press, 2008.

Pele, O. and Werman, M.: Fast and robust earth mover's distances, in: Computer vision, 2009 IEEE 12th international conference on, pp. 460–467, IEEE, 2009.

Penny, S. G. and Miyoshi, T.: A local particle filter for high-dimensional geophysical systems., Nonlin. Processes Geophys., 23, 391–405, 2016.

Poterjoy, J.: A localized particle filter for high-dimensional nonlinear systems., Monthly Weather Review, 144, 59–76, 2016.

Reich, S. and Cotter, C.: Probabilistic forecasting and Bayesian data assimilation, Cambridge University Press, 2015.

Reynolds, A. C., He, N., Chu, L., Oliver, D. S., et al.: Reparameterization techniques for generating reservoir descriptions conditioned to variograms and well-test pressure data, SPE Journal, 1, 413–426, 1996.

Vefring, E. H., Nygaard, G. H., Lorentzen, R. J., Naevdal, G., Fjelde, K. K., et al.: Reservoir characterization during underbalanced drilling (ubd): methodology and active tests, SPE Journal, 11, 181–192, 2006.

Weerts, A. H. and El Serafy, G. Y.: Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models, Water Resources Research, 42, 2006.