Nonlin. Processes Geophys. Discuss.,
https://doi.org/10.5194/npg-2017-7-AC1, 2017

Nonlinear Processes
in Geophysics
Discussions
Open Access

Interactive comment on "Ensemble Kalman filter for the reconstruction of the Earth's mantle circulation" *by* Marie Bocher et al.

Marie Bocher et al.

marie.bocher@ens-lyon.fr

Received and published: 31 October 2017

We thank the anonymous reviewer for his/her very complete review of our paper. We really appreciate the time investment that it must have been, and hope that the answers provided here, as well as the modifications proposed in the paper will be satisfactory. Please find below the list of comments, each associated with our answer and details on the associated modifications of the manuscript.

Printer-friendly version

Discussion paper

1 Major Comments

1. It's fine to introduce an established method to a new field, but there should be a number of additional references to literature in other areas of geophysics where all of the tuning, methodology, and evaluation techniques have been more carefully documented.
   ⇒**We added reference to the litterature when deemed necessary (10 new data assimilation references have been added). If you feel that a specific reference is missing, please let us know.**

2. The method used to create the initial ensemble is unclear and not common in the ensemble filtering literature for geosciences. The discussion of 'continuous' versus sample P matrices seems to be part of the confusion. Normally, one wouldn't really know P outside of the ensemble filter context, except perhaps in a loose 'climatological' sense. General discussion of P's versus ensemble sample P's is fuzzy. The discussion starting at the end of p. 7 seemed particularly confusing. Apparently there are 400 'climatological' samples. These will only span a phase subspace of at most 399 dimensions. However, an eigenvector analysis somehow produces 1928 distinct eigenvalues. An SVD analysis, the common way to filter a sample covariance in general, would give 399 or fewer. After the eigenanalysis is completed, an initial analysis step is somehow done (continuous, not ensemble?) and finally the analysis covariance is somehow sampled to generate an initial ensemble. Additional clarity is needed here. Also, doing anything more than filtering the sample covariance from the 400-member sample seems inappropriate. Can you relate this to a similar procedure in the literature from a more mature ensemble field?
   ⇒ **We are not simply computing the sample correlation matrix, but use the symmetries of the problem to better estimate the covariance matrix. That is why we do not do an SVD but an EVD instead and obtain more than 399**

eigenvalues. We rephrased the paragraph to make this point clear. We also cite the paper where the procedure is explained in more details (Bocher et al., 2016).

3. The Kalman Filter and ensemble variants basically depend on exponentially growing directions in the model phase space to work effectively. For novel applications, it is important to know something about the growth of error. Many geophysical applications will include an experiment where ensembles are evolved without assimilation to demonstrate that there is ensemble error growth and to show that the assimilation improves on this control case. I suggest including results from such a case. Easiest way would be just to use one or more of the existing ensemble initial conditions, but it can also be done with smaller ensembles to just explore error growth.
   ⇒ **This work has already been done and discussed extensively in Bello et al. (2014). We reworked section 2.1 on the mantle convection model, adding more explanations on the choice of the model, and adding a paragraph on the chaotic nature of mantle convection, the thermal turbulence that affects it, and the result of twin experiments measuring error growth for our model.**

4. Although it happens far too often in the literature, looking at analysis innovations is simply bad practice. It is almost impossible to interpret the results as you demonstrate with your figure 7. There is, however, a simple alternative that is much easier to interpret: look at forecast innovations. The forecast observations are independent of the forecast and so should give results consistent with comparisons to truth (although noisier of course). You can, of course, also easily simulate withheld (not assimilated) observations and compare your analysis to these. However, given the dearth of available observations it is unlikely that this is what you would choose to do in a real data experiment.
   ⇒ **We are not looking at the analyzed innovation, but forecast innovation, as defined in equation 21 of the original manuscript, or equation 44 of the**

revised version. The x-axis title of figure 2 "number of analyses" might have been confusing, so we changed it to "forecast number".

5. I suspect that you will find that a deterministic ensemble filter will produce significantly better results for your problem with small (say less than 100) ensemble sizes. The additional sampling error may be a primary cause of the 96 member ensemble being significantly worse than the large ones.
   ⇒ **We actually started implementing a deterministic ensemble filter. However, as is said on page 10, second paragraph of the submitted manuscript, and validated on figure 6 for example, we need to apply localization on both the horizontal and vertical direction. Since the observations are only located at the surface, we need to apply localization in the state space. To do so with a deterministic filter would require choices on the resampling after analysis that we could not justify properly, which is why we did not implement and test it yet.**

6. Repeated claims are made that the 288-member ensemble (or an ensemble of size about 300) is optimal/optimum. These claims are unsupported. A norm is not established and it is clear that you are doing some intuitive combination of quality and cost. In addition, having only 3 ensemble sizes gives you no basis for claiming that the middle one is optimal. You would need to try additional cases.
   ⇒**We agree and rephrased the two occurences were we used the word optimal for the ensemble size.**

7. The adaptive inflation approach you are using is fairly simplistic. More robust methods that do an evolving Bayesian estimation for inflation have been described in papers like Miyoshi, T., 2011: The Gaussian Approach to Adaptive Covariance Inflation and Its Implementation with the Local Ensemble Transform Kalman Filter. Mon. Wea. Rev., 139, 1519-1535, and ANDERSON, J. L. (2009), Spatially and temporally varying adaptive covariance inflation for ensemble fil-

ters. Tellus A, 61: 72–83. doi:10.1111/j.1600-0870.2008.00361. It is clear that almost all of your ensembles are significantly under dispersed and improved performance should result from better inflation. This may be particularly important in improving the 96-member case, too.

⇒ **The adaptive inflation, although simplistic, corrects rather successfully the forecast variance of the temperature as soon as it is close to observations (see rank histogram of figure 7a of the revised manuscript: we have a slight bias of the ensemble towards colder temperatures, but the ensemble is not under-dispersed at the surface). However, in depth, the ensemble is under-dispersed and our interpretation is that, since observations are only at the surface, we do not correct the ensemble spread adequately in depth, and any adaptive inflation scheme based on the innovation statistics will not improve the spread of the ensemble in depth. We added a paragraph on the possible improvement of adaptive inflation in the discussion, and added a subsection on rank histograms in the section 4:A posteriori evaluation of the ensemble Kalman filter method.**

8. Doing all interpretations with normalized error and spread is not common and can make interpretation of relative capabilities complicated. Certainly, the discussion at the end of the paper referring to percentage errors is confusing and potentially misleading. I suggest at least including a few results that look at the unnormalized RMSE, etc.

   ⇒ **Our aim was to provide the reader with a reference point, since the temperature is non dimensional in our models. We agree that this formulation made the whole discussion on results confusing, so we changed all the figures to plot RMSE instead of the RMS of the normalised error. To provide a point of reference for the error, we plotted also on the figures 1 and 3 the error that we would have made if we had supposed a 1D temperature profile corresponding to the average temperature field computed from a very long**

free run.

9. You briefly look at the range of the ensemble and whether it bounds the truth (end of p. 22). This type of evaluation is misleading since how often the truth should be bounded is a function of the ensemble size. A rank histogram analysis would be more appropriate and would also have the potential to reveal more about challenges being faced by the ensemble assimilation.
   ⇒ **Indeed, this was a mistake on our part. We deleted sentences referring to this in the text. We performed a rank histogram analysis for the temperature, heat flux and velocities, the results are described in section 4.3 and shown in figures 6 and 7 of the revised manuscript.**

2 Minor Comments

1. p. 3, line 9: The reference to Evensen 1994 should include a caveat that it is not a correct derivation of the EnKF and that the 1998 Burgers et al presents the correct algorithm.
   ⇒**Text has been modified**

2. P. 6, line 8: "data that ARE"
   ⇒**Text has been modified**

3. P. 7, equation 11: Using an extended (or joint) state is okay. However, you don't really motivate why. The most standard practice would have been to include all the observation priors in the joint state. Include a brief discussion of why you made this choice.
   ⇒ **We rephrased the paragraph in question to clarify the use of the augmented state to avoid a nonlinear observation operator, added a reference**

to Evensen 2003, and a reference to the paragraph where the observation operator is discussed.

4. P. 7, line 12 and other places: "expectancy" To the best of my knowledge, this word is not being used correctly here. "expected value" might be better.
⇒**Text has been modified**

5. P. 7, line 12: This sentence is confusing since you never really compute the P's, why even refer to them? They are unknown and unknowable in some sense.
⇒ **We corrected the sentence, accounting for the fact that we actually compute the covariance matrices, but just for the initialization step. As discussed in major comment 2, P_1^f is the background covariance matrix, computed from a free run (the "climatology"), so we know it, if we consider the model to be perfect.**

6. P. 7, line 19: Note that the velocity forward operator is just a vector extraction (identity).
⇒**Text has been modified**

7. P. 8, line 20: What does 'efficient' mean here. Should become irrelevant anyway if this is behaving like a KF/EnKF; the initial ensemble choice should lose any qualitative impact as the filter proceeds. If this is not the case, then a Kalman filter class algorithm may not be a particularly good choice.
⇒ **Efficient in the sense that the errors are smaller at the beginning of the assimilation and decrease faster. This is important for our problem, since the spin up time of the assimilation is of the same order as the total timespan for which we have observations. We rephrased the sentence accordingly.**

8. P. 10, line 12: This statement is too strong. The Janjic approach can be significantly more efficient in some parallel computation situations. In the simple PDAF

implementation, this may not be the case.
⇒**Text has been modified**

9. P. 10, sentence starting on line 25: I cannot understand this sentence. Not sure what “direct forecast error localization” means here.
⇒ **we meant we apply localization directly on the forecast error covariance matrix, as opposed to the domain localization already implemented in PDAF, text has been modified accordingly.**

10. P. 11, line 9: Replace “noise” with “add noise to”
⇒**Text has been modified**

11. P. 11, line 10: What is the root mean square of surface heat flux... Instantaneous, variation over model?
⇒ **the root mean square of surface heat flux and velocity are long term averages computed from the results of a free run. They are characteristic of the dynamics of the system. Text modified with these precisions.**

12. Figure 1 caption: “150 Myr OBSERVATION dataset” makes it clear that you are just using the synthetic observations.
⇒**Text has been modified**

13. Figure 1: Note that the error seems to be going back up towards the end of the time series. This probably merits a comment in the text. I suspect it is due to the insufficient spread.
⇒ **The text has been modified to acknowledge the error growth at the end of the time series. Additionally, we show now in Figure 1 the evolution of the error of the surface velocity, as suggested by reviewer # 2. It shows that, contrary to the temperature, the error on velocity does not grow at the end of the time series. We also discuss in more details the reliability of**

the ensemble forecast in section 4.3 of the revised manuscript, using rank histograms.

14. P. 12, line 6: Not sure what is meant by a "stabilization"
⇒ **We deleted the word stabilization and changed the description of figure 1: we identify 2 phases: rapid decrease of error and then slow growth of error (which we described as stabilization in the former version)**

15. Start of p. 13, discussion of parallel efficiency. This discussion is inappropriate without lots more detail about the computing resources used. A good parallel implementation of an EnKF should scale very well (embarrassingly parallel) for a problem like this, so I was surprised that the time wasn't very nearly constant with a sufficient number of cores.
⇒ **The time is indeed nearly constant provided we have a sufficient number of cores, we meant here CPU time and not real elapsed time. We rephrased the sentence to make it clear that we evaluate the quality of the data assimilation against its computational cost.**

16. P. 13, line 13: rms values over what sample?
⇒ **We added precisions in the text (see also minor comment 11)**

17. P. 13, line 16: Not sure what "at first order" means here.
⇒ **We meant that the cumulative mean innovation check is not a comprehensive test, but allows only a partial check of consistency. We rephrased the sentence.**

18. P. 13, line 20: An estimate of the uncertainty, not the error, would be more common usage to describe the spread.
⇒ **Text has been modified**

19. Figure 2: I don't see how these can be consistent with figure 1 which shows N=96 much worse by the 10th assimilation time.

⇒ **Figure 1 represents the error on the whole temperature field. Figure 2 shows statistics on the innovation, so the difference between observed and forecast <u>surface</u> velocities and heat fluxes. This means that although the forecast and the observed data at the surface are close, the estimated temperature field at depth differs from the true temperature field much more for N=96 than for N=288 or 768. We added this remark to the paragraph commenting on figure 2. We also reorganized the whole description of figure 2, for more clarity.**

20. Figure 3: These are dangerously under dispersed in 3 cases.
⇒ **We added a remark in the result section, study more precisely where the ensemble is biased/underdispersed with rank histograms, and rediscuss underdispersion in the discussion.**

21. P. 17, line 2: It's the ensemble covariance that matters, not the scales of spatial variability. These may or may not be the closely related.
⇒ **We agree. However, the ensemble covariance will be affected by the way small perturbations evolve and grow in the system. In our system, a slight temperature perturbation in the upper boundary layer can lead to the developpment of a new plate boundary. This links our discussion of spatial variability due to the structure of plate boundaries to the ensemble covariance matrix. Text has been complemented to make this link clearer.**

22. P. 19, line 13: Are the figure 8 results for the best localizations?
⇒ **Yes, we added this precision in the legend of Figure 8.**

23. Figure 7 caption: What is 'K'?
⇒ **K=16, legend updated.**

24. Figure 9 caption: Need more caption info. What assimilation? Max and min temps from ensemble members? How many of them?

⇒**Text has been modified**

25. P. 25, line 13: This looks like very poor parallel behavior to me, unless this is somehow dominated by the model forecast times.
⇒ **We already stated that, indeed, "during the assimilation of a dataset, most of the computational time is dedicated to the forecast step".**

26. P.26, line 7: This is too simplistic. In a smoother for a problem with things that are advecting/convecting, an observation at the current time will have largest correlation with a point upstream at an earlier time. The localization needs to be shifted away from the observation as a function of time lag and the maximum value should be less than 1.
⇒**Yes, we agree that a potential smoother could benefit from shifting the localization away from the observation. However, we might already gain some information by applying a simple localization for the smoother. Nerger et al. (2014) obtain encouraging results with this type of localisation in a large-scale ocean circulation model for example.**

27. P. 26, line 27: Why do you think this? Are there more parameters than state variables?
⇒**Not necessarily, but, as explained in the following sentence, the relationship of mantle dynamics to different rheological parameters is highly nonlinear: most likely, we will need very large ensembles to determine accurately the parameters.**

28. P. 26, line 34: You think you are not converged? Plots look like you've bottomed out and error is increasing as a function of assimilation time.
⇒**We agree, we deleted the sentence.**

References

Léa Bello, Nicolas Coltice, Tobias Rolf, and Paul J Tackley. On the predictability limit of convection models of the earth's mantle. *Geochemistry, Geophysics, Geosystems*, 15:2319–2328, 2014.

M Bocher, N Coltice, A Fournier, and PJ Tackley. A sequential data assimilation approach for the joint reconstruction of mantle convection and surface tectonics. *Geophysical Journal International*, 204(1):200–214, 2016.

Lars Nerger, Svenja Schulte, and Angelika Bunse-Gerstner. On the influence of model nonlinearity and localization on ensemble kalman smoothing. *Quarterly Journal of the Royal Meteorological Society*, 140(684):2249–2259, 2014.

Interactive comment on Nonlin. Processes Geophys. Discuss., https://doi.org/10.5194/npg-2017-7, 2017.