Nonlin. Processes Geophys. Discuss.,
doi:10.5194/npg-2017-7-RC1, 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.

Anonymous Referee #1

Received and published: 12 March 2017

Summary This manuscript presents an observing system simulation experiment using an ensemble Kalman filter for an idealized model similar to those used to simulate mantle convection. Apparently this is one of the first applications of an ensemble filter in this field. The results suggest that there is potential for applying ensemble filters to mantle convection models. The novel issue is the behavior of this particular model and its relevance for more realistic models that could use real observations. Given this, more information about the error growth characteristics of this model would be a useful addition, as would additional references to more mature ensemble filter explorations in other areas of geoscience.

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.

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?

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.

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 demon-

strate 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.

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.

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.

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 filters. 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.

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.

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.

Minor points:

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.

2. P. 6, line 8: “data that ARE”

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.

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.

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.

6. P. 7, line 19: Note that the velocity forward operator is just a vector extraction (identity).

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.

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.

9. P. 10, sentence starting on line 25: I cannot understand this sentence. Not sure what “direct forecast error localization” means here.

10. P. 11, line 9: Replace “noise” with “add noise to”

11. P. 11, line 10: What is the root mean square of surface heat flux... Instantaneous, variation over model?

12. Figure 1 caption: “150 Myr OBSERVATION dataset” makes it clear that you are just using the synthetic observations.

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.

14. P. 12, line 6: Not sure what is meant by a “stabilization”

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.

16. P. 13, line 13: rms values over what sample?

17. P. 13, line 16: Not sure what “at first order” means here.

18. P. 13, line 20: An estimate of the uncertainty, not the error, would be more common usage to describe the spread.

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.

20. Figure 3: These are dangerously under dispersed in 3 cases.

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.

22. P. 19, line 13: Are the figure 8 results for the best localizations?

23. Figure 7 caption: What is ‘K’?

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

25. P. 25, line 13: This looks like very poor parallel behavior to me, unless this is somehow dominated by the model forecast times.

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.

27. P. 26, line 27: Why do you think this? Are there more parameters than state variables?

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.

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