Nonlin. Processes Geophys. Discuss.,
https://doi.org/10.5194/npg-2018-6-RC2, 2018

Nonlinear Processes in Geophysics
Discussions
Open Access

Interactive comment on "Ensemble Variational Assimilation as a Probabilistic Estimator. Part II: The fully non-linear case" *by* Mohamed Jardak and Olivier Talagrand

M. Bocquet (Referee)

marc.bocquet@enpc.fr

Received and published: 12 March 2018

Please see my review of the companion manuscript (Jardak and Talagrand, 2018). I found the discussions of the numerical results of this part II paper equally well acute and clever. Section 4 on the weak-constraint EnsVAR is very nice, as well as Section 5 with a strong discussion and conclusion part.

Yet, most of the remarks that I made on the form of the companion paper still apply and need to be addressed.

Like for the first and companion paper, I believe that a minor revision of the manuscript is necessary to address a few flaws and a list of very minor points. In particular, a few references are missing.

Specific remarks, in connection, or not, to the previous remarks are:

1. Abstract, line 1: You should mention here that EnsVAR is equivalent to EDA. The main issue is the confusion that it may generate, and the fact that, because you fail to refer to EDA in the abstract, you will restrict your potential readership.
2. Abstract, lines 3-4: If you had cycled the analysis, you would have observed that QSVA is not as mandatory, expect maybe for very long windows. So I believe you should mitigate the statement.
3. Abstract, line 9: "without need to resort to QSVA" $\longrightarrow$ "without the need for QSVA".
4. line 19: "Kuramuto" $\longrightarrow$ "Kuramoto", as well as in both references by Yoshiki Kuramoto et al.
5. lines 24-25: "The performance of EnsVAR is compared with that of Ensemble Kalman Filter and Particle Filter in Section 3.": again, out of a specific context, this does not make much sense in the absence of cycling, proper tuning of the methods, and so on.
6. line 28: "successful in nonlinear as in linear conditions.": it always depends on how long the data assimilation window is. As any other method, EnsVAR is bound to fail for very large windows.
7. line 33: "twice a day": please mention that this corresponds to 0.10 time units since the Lorenz model is primarily defined in those units.

8. line 35: Fine with the "I." but the notation for referring to equations is not consistent throughout the manuscript and does not follow the Nonlinear Processes in Geophysics guidelines.

9. line 59: "the the" $\longrightarrow$ "the".

10. Section 2: Nice results. Similar and consistent results have been obtained, which should be briefly mentioned. Bocquet and Sakov (2013) have obtained very similar results with the iterative ensemble Kalman smoother (IEnKS) with the same window of 10 days, a time-interval of 1 day (as opposed to twice a day), an ensemble of 20 members and $\sigma = 1$: see Figure 4 of Bocquet and Sakov (2013). In particular the MDA IEnKS ($S = 1$), which is quasi-static, outperforms the SDA IEnKS which (in this reference) is not quasi-static. Other directly relevant references worth citing about quasi-static EnVar methods are Goodliff et al. (2015) and Carrassi et al. (2017).

11. lines 82-83: "This improvement must be due to the fact that more observations have been used.": above all this is due to the fact that the middle point is farther apart from the end of the window, so that fresher observations have a strong information content leveraged by the unstable modes of the dynamics. This has been shown in Bocquet and Sakov (2014).

12. line 105: "bayesianity" $\longrightarrow$ "Bayesianity".

13. line 106: "bayesian" $\longrightarrow$ "Bayesian".

14. lines 102-113: Part of this analysis coincides with that of H. Abarbanel and his collaborators. I believe you should at least refer to one of their paper, for instance Ye et al. (2015).

15. line 115: "shown on Figure 5" $\longrightarrow$ "shown in Figure 5".

16. line 129: "As the errors in the ensemble means..": I believe you mean "error standard deviations of the ensemble means".

17. lines 151-153: My experience is that, on the contrary, rank histograms of a deterministic EnKF (ETKF specifically) are $\cap$-shape (the ensemble is overdispersed). The difference might be due to the nature of the EnKF, the fact that your EnKF run is not long enough, or simply, that your inflation is insufficient. Moreover, once again, localisation is unnecessary with an ensemble of 30 members and may be detrimental to the quality of the ensemble. Anyway, the $\cup$-shape that you have obtained for the EnKF is note a generality.

18. line 167: $\mathbf{b}$ is a notation usually reserved for a possible bias in VarBC.

19. line 175, Eq. (2): Assuming H is linear, it should read $\mathbf{H}$.

20. line 182: Dot missing at the end of the sentence.

21. 195: "which corresponds to a predictability time of about 10 days": interesting. Can you please develop?

22. lines 252-255: How did you implement model noise in the EnKF and the PF? This should be described.

23. line 263: "simple :" $\longrightarrow$ "simple:".

24. lines 277-279: Yes, that is the most important added value of this couple of papers and should be emphasised in the abstract of the first manuscript.

25. lines 294-296: In general, no claim can be made as to the accuracy of these methods (with the goal to estimate the truth) in the absence of cycling.

26. lines 305-310: I already know for a fact (Bocquet and Sakov, 2013, 2014) that proper cycling would very significantly reduce the number of iterations. This should be mentioned.

27. lines 328-330: “is cycling necessary at all, or can one simply proceed by implementing EnsVAR over successive, possibly overlapping, windows ?”: This question has already a detailed answer in (Bocquet and Sakov, 2013, 2014) and subsequent references. To anticipate a question: yes, many of the conclusions obtained with the IEnKS would apply to EnsVAR. In essence: no, it is not absolutely necessary, but it would numerically help a lot to cycle the background (fewer iterations) and would yield a better accuracy.

28. line 338: “(Carrassi, Bocquet, pers. com.)”: this has now been published (Bocquet and Carrassi, 2017).

29. line 347: “ROT” ⟶ “RTO”.

30. lines 348-349: “This defines a theoretical improvement on EnsVAR, based on an appropriate use of the Jacobian of the data operator.” Liu et al. (2017) have already shown on a higher dimensional example that RTO might become inefficient (it is likely to be ultimately subject to the curse of dimensionality) as reported in their experiments and conclusions. This could be mentioned.

References

Bardsley, J.M., Solonen, A., Haario, H., Laine, M., 2014. Randomize-then-optimize: A method for sampling from posterior distributions in nonlinear inverse problems. SIAM J. Sci. Comput. 36, A1895–A1910.

Bocquet, M., Carrassi, A., 2017. Four-dimensional ensemble variational data assimilation and the unstable subspace. Tellus A 69, 1304504. doi:10.1080/16000870.2017.1304504.

Bocquet, M., Sakov, P., 2013. Joint state and parameter estimation with an iterative ensemble Kalman smoother. Nonlin. Processes Geophys. 20, 803–818. doi:10.5194/npg-20-803-2013.

Bocquet, M., Sakov, P., 2014. An iterative ensemble Kalman smoother. Q. J. R. Meteorol. Soc. 140, 1521–1535. doi:10.1002/qj.2236.

Carrassi, A., Bocquet, M., Hannart, A., Ghil, M., 2017. Estimating model evidence using data assimilation. Q. J. R. Meteorol. Soc. 143, 866–880. doi:10.1002/qj.2972.

Goodliff, M., Amezcua, J., van Leeuwen, P.J., 2015. Comparing hybrid data assimilation methods on the Lorenz 1963 model with increasing non-linearity. Tellus A , 26928doi:10.3402/tellusa.v67.26928.

Jardak, M., Talagrand, O., 2018. Ensemble variational assimilation as a probabilistic estimator. Part I: The linear and weak non-linear case. Nonlin. Processes Geophys. Discuss. 2018, 1–39. doi:10.5194/npg-2018-5.

Liu, Y., Haussaire, J.M., Bocquet, M., Roustan, Y., Saunier, O., Mathieu, A., 2017. Uncertainty quantification of pollutant source retrieval: comparison of bayesian methods with application to the Chernobyl and Fukushima-Daiichi accidental releases of radionuclides. Q. J. R. Meteorol. Soc. 143, 2886–2901. doi:10.1002/qj.3138.

Oliver, D.S., He, N., Reynolds, A.C., 1996. Conditioning permeability fields to pressure data, in: ECMOR V-5th European Conference on the Mathematics of Oil Recovery, pp. 259–269.

Ye, J., Rey, D., Kadakia, N., Eldridge, M., Morone, U.I., Rozdeba, P., Abarbanel, H.D.I., Quinn, J.C., 2015. Systematic variational method for statistical nonlinear state and parameter estimation. Phys. Rev. E 92, 052901. doi:10.1103/PhysRevE.92.052901.

Interactive comment on Nonlin. Processes Geophys. Discuss., https://doi.org/10.5194/npg-2018-6, 2018.