

Answers to referees: npg-2018-6, 2018

Mohamed Jardak & Olivier Talagrand

1 To referee 1:

We thank M. Bonavita for his comments and suggestions. These are printed below in black, and **our responses in red**.

1. In Sect. 3 the Authors compare results of the QSVA EnsVAR, EnKF, PF. Not surprisingly, QSVA EnsVAR shows better results as a probabilistic estimator and also for more standard resolution measures. This is unsurprising, to my mind, because this comparison is not fair. As the Authors noted, the costly QSVA extension is needed to keep EnsVAR assimilation in an approx. linear error evolution regime and thus guarantee good behaviour in this long-window assimilation set-up. To compare apples with apples the Authors should directly compare results of the standard EnsVAR algorithm at the end of the window with those of the EnKF and PF. Additionally, it would also be of interest to compare results of QSVA EnsVAR with those of an EnKF whose assimilation is run on shorter assimilation windows, to guarantee linear behaviour, and then cycled.

We have actually exactly done in Section 3, relative to strong-constraint assimilation, what the referee requests, viz., performing the comparison between the results of EnsVAR, EnKF and PF at the final times of the assimilation windows. That is now made more explicit.

Now, concerning Section 4 of the paper, relative to weak-constraint assimilation, it is true that our comparison has been made, not at the final time of the assimilation windows, but on the last 13 days of the windows (see Fig. 12). In that sense, the comparison is not fair. A perfectly fair comparison may be included in a future work.

The referee also mentions a ‘cycled’ EnKF. We do not understand what he means. To us, the EnKF, which is sequential, is cycled by construction.

2. In Sect. 4 on weak-constraint assimilation, I understand that the model error perturbations are drawn from the same error distribution whose covariance is used in the 4D-Var cost function. If this is correct, this is a significant limitation on the potential applicability of the results, as the difficulty in obtaining realistic characterizations of Q is probably the most important cause of the limited success of weak-constraint 4D-Var in realistic applications.

The referee raises a very important question, which applies equally to the EnKF and PF. But that question goes well beyond the scope of our present papers. An assimilation algorithm must first be evaluated in conditions where the errors affecting the data follow the same statistics (first- and second-order moments) that are used in the assimilation. That is what we have done. The same question arises but may be more critical concerning observation errors. We have mentioned explicitly that our twin experiments are fully ‘consistent’ as concerns the errors on the data.

3. In the last paragraph of Sect. 4, the Authors explain that the performance of EnsVAR, EnKF, PF in the weak-constraint case appears in terms of reliability measures (e.g., rank histograms). This could depend on localization used in the EnKF, for example. Have the Authors explored this parameter space?

This remark is appropriate. It would be of course impossible to explore the full space of parameters for the three algorithms. Now, concerning localisation of the EnKF, and at the explicit request of referee 2 (see his comment 36 on paper 1), we have performed experiments without localisation in the setup of paper 1. The results are summarised in the new version of the latter. They show changes (some features of the assimilated fields, including the rank histograms, are improved, while others, such as the RMS errors, are degraded). These results, as interesting as they are, cannot however be studied in detail in the frame of our present work.

4. Lines 310-311: “...many possibilities exist for the reducing the cost of EnsVAR, through simple parallelization or ...”. I am puzzled on how parallelization can reduce the computational cost of EnsVAR. Maybe the Authors meant clock time?

Yes, you’re right. Thanks for the remark. correction done.

5. Lines 319-320: “On the other hand, EnsVAR is largely empirical, with the consequence that, should difficulties arise, conceptual guidelines may be missing to solve these difficulties.” I struggle to see what these difficulties might be. In the linear case, EnsVAR (aka EDA) is constructed so as to be a consistent statistical estimators assuming the input data errors are correctly sampled. In the nonlinear case, its behaviour will depend on the amount of nonlinearity and the ability to track the true global solution. In this respect, EnsVAR is as empirical as the EnKF.

Well, one must always be ready to encounter unexpected difficulties. And, yes, EnKF is also empirical, and the remark we make about EnsVAR also applies to EnKF (with the additional difficulty that the latter contains many more arbitrary parameters than our EnsVAR). What we want to stress is that it is easier to interpret the results produced by a method built on a solid theoretical basis, and to correct its possible weaknesses. We have mentioned that our remark applies as well to the EnKF and PF.

6. Lines 339-340: “EnsVAR has been implemented here on a small dimension system. It has to be implemented on larger dimension, physically more realistic models.”.

I suspect the Authors mean QSVA-EnsVAR in this context. Standard EnsVAR has been running at ECMWF and Météo-France for a number of years.

Yes, our statement was not correct in the sense that EnsVAR-EDA is operationally running at ECMWF and Météo-France on large models. But, it has not been systematically evaluated as a probabilistic estimator on a physically realistic large dimensional model. That is what we consider necessary, either in the form of QSVA-EnsVAR or otherwise. Indeed, the fact that EnsVAR-EDA, although largely empirical, is run operationally with undisputably some success is one additional reason for studying its properties more deeply.

2 To referee 2:

We thank M. Bocquet for his comments and suggestions. These are printed below in black, and our responses in red.

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.

Thanks, we have made the appropriate changes in both abstracts.

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.

Yes, many possibilities can be considered. For instance, performing the assimilation over long enough overlapping successive windows (but that is not really different from QSVA). This is mentioned in the conclusion, but has not been used in the paper. We have modified the abstract to simply say that the problem of assimilation over long windows has been solved in our setting by using QSVA.

3. Abstract, line 9: “without need to resort to QSVA” → “without the need for QSVA”.

Thanks, done.

4. line 19: “Kuramoto” → “Kuramoto”, as well as in both references by Yoshiki Kuramoto et al.

Thanks, done.

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.

We have mentioned (ll. 305-310) that the comparison with EnKF and PF certainly cannot be considered as definitely conclusive. But it is certainly instructive,

for instance in that it suggests that there are no major differences between the results produced by the three methods that have been compared. And we do not understand why the referee considers that ‘this does not make much sense in the absence of cycling’ (see our response to his specific remark 34 about paper 1). And ‘proper tuning of the methods’ could be an endless task.

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.

We qualified our statement by saying that it is valid only for the conditions of our experiments. But is not clear to us why any method is bound to fail for very large windows. Failure is certainly to be expected for strong constraint assimilation implemented with an erroneous model. But why should it be in the case of weak constraint ?

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.

We have defined the “day” in paper I as equal to 0.44 time unit in equation I.12.

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.

Thank you, we went throughout the manuscript and made the corrections.

9. line 59: “the the” → “the”.

Thanks, done.

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).

Thanks for mentioning those references. We have included them in our conclusion. On the other hand, they do not seem to deal directly with what is the primary goal of our paper, namely the objective evaluation of algorithms for ensemble assimilation as probabilistic estimators.

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).

Thanks also for mentioning. We have added the reference and mentioned this pertinent remark in the paper.

12. line 105: “bayesianity” → “Bayesianity”.

13. line 106: “bayesian” → “Bayesian”.

Thanks, done for both.

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).

Thanks again. The reference has been added.

15. line 115: “shown on Figure 5” → “shown in Figure 5”.

Thanks, done.

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

Thanks, done.

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.

We take note of the fact that what we observe is not a generality, but cannot really say more at this stage. Concerning the length of the assimilation window, the top left panel of Figure 7 suggests that the EnKF is already stabilized after 3 days of assimilation, so that it seems unlikely that a longer window would significantly modify the histogram (the diagnostics which the referee comments have been performed at the end of the assimilation window).

18. line 167: b is a notation usually reserved for a possible bias in VarBC.

The notation b for bias is by no means a standard.

19. line 175, Eq. (2): Assuming H is linear, it should read H.

Thank you, done.

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

Thanks, done.

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

This value of 10 days results from numerical tests which were not shown in the paper. That is now briefly mentioned.

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

As said now, the model noise has been added as random noise in all model integrations.

23. line 263: “simple :” → “simple:”.

Thanks, done.

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.

We think that this point has been properly, if succinctly, explained in the abstract, which we do not want to overload.

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;

This seems to repeat the referee’s comment 5 above. See our response there.

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.

We thank again the referee for this information, which we will take into account in a possible future work.

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.

Thanks once more. We have added the references.

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

Thanks, done

29. line 347: “ROT” → “RTO”.

Thanks, done

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.

All right. Thanks. This is now mentioned.