Answers to referees: npg-2018-5, 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. Lines 54-56: I am not aware of the paper by le Gland et al, 2009, and I am not sure it is generally available as it appears to be an internal research memo of a specific institution. Further, I do not think it is actually necessary in this context, if we take the view that the EnKF converges to the KF for large ensemble size and the KF is a consistent bayesian estimator for linear dynamics and gaussian errors. We have given a different, more accessible, reference. And, since we are concerned with the Bayesianity of ensemble estimates, we think it is legitimate to consider the Bayesianity of the EnKF, and in particular to stress that it cannot be Bayesian in the general nonlinear case.

2. Lines 64-65: "They exist in numerous variants, many of which have been mathematically proven to achieve bayesianity in the limit of infinite ensemble size". Please provide relevant references.
We have added a reference which deals with the general question of asymptotic convergence of Particle Filters.

3. Line 102-105: The Bardsley et al. 2014 reference appears to be missing. Some further discussion of their method would be useful here, as the response of the EnsVAR method to nonlinearities is the central issue of this paper. The Bardsley et al. paper was actually referenced in the manuscript, but at the wrong alphabetical place. And the question of further discussion has also been raised with some emphasis by Referee 2. See his review, especially his 'weak point' 1, and our response to it.

4. Lines 177-179: I do not understand this remark and the implied derivations behind it. Can the Authors please expand? Since Bayesianity is ensured in the conditions of linearity and Gaussianity, it is legitimate to consider the case when linearity holds, but not Gaussianity. As for "the implied derivations", if the referee means how the results we state can be obtained, that is actually trivial (and we have given a reference anyway).

5. Lines 4337-438: "We have evaluated the Gaussian character of the ensembles...by computing their negentropy". I suspect the Authors have verified the Gaussianity of some marginals of the full pdf, not the Gaussianity of the full multivariate distribution. Can the Authors be more specific on this point? Correct, we never evaluated the Gaussianity for the full multivariate distribution. We have now made that clearer.

6. Lines 444-447: Can the lack of sensitivity of the analysis pdf to the pdf of the obser- vations be considered a consequence of the Central Limit Theorem, or do the Authors have an alternative explanation?
The fact, which has been mentioned just before, that the ensembles produced by the assimilation are close to Gaussianity is certainly a consequence of the Central Limit Theorem. That the ensembles are insensitive to the shape of the pdf of the errors in the observations seems rather natural in these conditions, but we cannot really say more at this stage.

7. I think it should be made clear that the comparison with the EnKF and PF is only qualitative, as the EnKF/PF results are known to be very sensitive to localization/inflation and there does not appear to have been a lot of work in this paper aimed at finding the optimal values. We give numerical results, and the comparison is quantitave. But it is certainly not exhaustive. We had alredy stressed that point in the paper, which has also been raised by the other referee.

8. Regarding the EnsVAR and EnKF comparison, I would expect the two systems to give equivalent results in the purely linear case. Have the Authors verified that this is the case, or if it is not why?
Yes, one could expect similar results in the linear (and also Gaussian) case from EnsVAR and EnKF, since both are then bayesian (as so is PF). But there is actually a slight difference. EnsVAR produces a set of independent realizations of the conditional pdf for any ensemble size, while EnsVAR or PF do not. We have not made the comparison. The reason for that is that our paper is concentrated on EnsVAR, and we have used the linear and Gaussian case, not for evaluating it per se, but as a benchmark for evaluation of the nonlinear case

2 To referee 2 :

We thank M. Bocquet for his comments and suggestions, *particularly for the references he has mentioned* . These are printed below in black, and our responses in red.

1. An insufficient discussion in the introduction of the fact that we already know that naive RTO / EDA / EnsVAR is not (perfectly) Bayesian and proper references to Oliver et al. (1996); Bardsley et al. (2014); Liu et al. (2017) should be added or better referred to.
Thanks. We have mentioned these papers and discuss them in the perspective of our two papers.

2. An abstract that – very surprisingly given how well the paper is written – does not faithfully reflect the findings of the paper. We do not fully understand what the referee means. But see our response below to his specific remark 3.

3. Some technical issues such as:

   - The use of capital letters in the titles of the sections in an inconsistent manner; please check all titles.
   - Improper use of *citet* and *citep* Please check this throughout the manuscript.
   - Many inconsistencies in how you refer to equations. This should be made consistent throughout the whole manuscript. Note that there is a recommended notation for Nonlinear Processes in Geophysics (see the guide- lines).
   - One speaks of and writes "the ensemble Kalman filter", not "Ensemble Kalman Filter" (which could also let the reader think that you are not very familiar with the EnKF literature).
   - Spaces in between text and ":" need to be removed (especially in the figure captions).

   Thanks, the above points have been fixed

1. Abstract: lines 5-7: even though the authors have worked on this idea for many years, the ensemble of data assimilation (EDA) principle implemented at Météo-France and ECMWF predates the EnsVAR and conceptually coincides with it. Moreover, it could be confused with the general terminology of EnVar (Ensemble Variational) methods. That is why I do not believe that this terminology should be put forward, especially in the abstract. That said, this obviously does not lessen the findings of the paper in any way. So, it is up to the authors. Thank you, we have stated explicitly in the abstracts of both parts of the paper, and in the texts,that EnsVAR is the same thing as EDA

2. Abstract, line 9: "the standard variational procedure" seems too vague. I am not sure of its meaning. We meant variational assimilation, as described and used in numerous papers and in operational prediction in places like ECMWF and Météo-France. We do not think that is too vague. We do not want to overload the abstract.

3. Abstract, line 10-14: the emphasis is on the performance (accuracy) of the method compared to, e.g., the EnKF. I do not believe that this is wise in the ab- sence of a proper cycling with which the EnKF could shine. I do not understand why the emphasis is not on the discussion of the Bayesian (or not Bayesian) trait of the method and the quality of the updated ensemble, which is the strong point of this study. By writing performance, we did not mean specifically numerical accuracy. We meant global performance of the algorithms under comparison, and primarily their performance as Bayesian estimators. We have modified the wording so as to avoid any misunderstanding.

4. Abstract, line 13: It is customary to write "the ensemble Kalman filter" instead of "Ensemble Kalman Filter". Thanks, fixed.

5. lines 10-20: "of the observations proper": what do you mean? Thanks, we meant physical observations. We have modified the wording so as to avoid any misunderstanding.

6. line 41: The EnKF definitely uses a definite article and should read: "The ensemble Kalman filter...". This should be systematically checked throughout the manuscript. This is the same comment as comment 4 above. Thanks, fixed

7. line 42: "by of" → "by". Thanks, fixed

8. line 80: "The present work is devoted to the study of that algorithm, and of its properties as a Bayesian estimator..." Precisely! That is why the abstract should reflect this point instead of focusing on the accuracy. See comment 3 above, and our response to it.

9. line 103-104: The connection between RTO and EDA as used in geophysical data assimilation has first been made, put forward and discussed in Liu et al. (2017) (and much more1). This must be mentioned here. (Incidentally this is how the authors of the present manuscript became aware of Bardsley et al. (2014).) Moreover, Oliver et al. and were the first to discuss this problem in 1996 (Oliver et al., 1996), which is something that Liu et al. (2017) recalled. You must cite this reference as well. We have added all these references, and commented on their connection with the present work.

10. line 116: "succinctly": you could avoid the adverb, as it lets the reader think that it will not be enough information to get a clear idea on the numerical results. Thanks, done

11. line 124: "multi-dimensional/one-dimensional Gaussian" more precisely called "multivariate/univariate Gaussian". Thanks, done.

12. line 125, before section 2: As already discussed (and illustrated) in Oliver et al. (1996); Bardsley et al. (2014); Liu et al. (2017), RTO/EDA (hence called naive RTO in Liu et al. (2017)) produces a biased nonlinear sampling. This should be briefly mentioned in the introduction as this is an established and published fact, important to your paper. Thanks, references added and commented on as it can be seen in the introduction.

13. line 133: "data operator": why invent a new name when there exist "observation / forward / Jacobian / source-Receptor operator" in geophysical data assimilation? We want to stress here (see ll. 136-140) that the data vector $\mathbf{z}$ contains all the information to be used for estimating the state vector $\mathbf{x}$ (physical observations, complete or partial background(s) or a priori estimates, 'balance' conditions or dynamical equations to be verified to some degree of accuracy by the final estimate,

. . .). The expressions suggested by the referee are usually more restrictive than that. And the expression "data operator" has indeed been used in the literature.

14. line 149: this statement only makes sense in an infinite numerical precision context. In practice, it of course depends on the condition number of operators built with Γ and Σ. Yes, of course. But strict mathematical results are always useful to start with, before considering numerical aspects.

15. line 175: "there is of course no reason to think...": again this has been settled in Oliver et al. (1996); Bardsley et al. (2014); Liu et al. (2017). So why not be more straightforward and factual here? Such as: "In general, this procedure does not lead to an unbiased Bayesian estimation, but can nonetheless provide a very useful approximation (Oliver et al., 1996; Bardsley et al., 2014; Liu et al., 2017)". All right. Corrected.

16. line 195: since it is here linear, you should use a bold upright character for H. See also line 291. Thanks, fixed.

17. line 240-251: nice remark! Thanks.

18. line 266: "bayesianity" → "Bayesianity". Thanks, fixed.

19. line 291: It is worth mentioning that this time-step is 0.05 time unit. We have introduced the "day" as equal to 0.44 time unit in equation 12.

20. line 293: Since the results do not depend on σ , why not resort to the very commonly used (in data assimilation) value of 1 ? Moreover, it is not the one used for the weakly nonlinear case. That may seem odd to the reader. Even though the results are independent of the value σ, one value must be used for the numerical computations. We mention the value we have used. And we write now in Section 5, relative to the nonlinear case, that it is necessary to scale the value of σ with the varaibility of the model.

21. line 294: "(however, ...impact)": no need to place the statement within parentheses. In my opinion, your remark is legitimate and important. Thanks, parentheses removed.

22. line 364: "in non-linear and non-Gaussian situations where Bayesianity does not hold" is not a consistent statement with your current introduction (you never mentioned this explicitly even when referring to Bardsley et al. (2014)). You should have referred to Oliver et al. (1996), Bardsley et al. (2014) (with the main result clearly mentioned) and Liu et al. (2017) for this statement to fully make sense. All right, corrected.

23. line 369: "where one day is equal to 0.44 time unit...": I would rather say 0.20 time unit(?). That is actually our definition of a 'day'.

24. line 375: "succesive" → "successive". Thanks fixed.

25. line 392: "effect ." → "effect.". Thanks fixed.

26. line 392: It is certainly a nonlinear effect, but do you have an explanation for it? No, we do not have a clear explanation for it. We nevertheless mention now that, since the conditional expectation is the deterministic estimator which minimises the variance of the estimation error, it must lead in general to a smaller error variance than the deteministic assimilation performed on the raw observations.

27. line 402: "(top, middle and bottom p and accuracy" should be removed. Thanks, fixed.

28. lines 411-412: "The degradation of reliability in the lower two panels may therefore be due here to non-linearity.": this implies that you believe that naive RTO samples are not perfect draws from the conditional pdf. But, again, your introduc- tion should have discussed this point more clearly. As already said, this is now done.

29. lines 413-414: "is much larger for decrease of spatial density than for decrease of temporal density (middle and top panels respectively)." → "is much larger for the decrease of spatial density than for the decrease of temporal density (middle and top panels respectively)." We are not convinced, and leave that to the text editor.

30. line 416: "panelsrespectively" *rightarrow* "panels respectively". Thanks, fixed

31. lines 415-416: "consistent with the top two panels of Figure 4, which suggest that the model fields are more correlated in time than in space...": yes, a very well known fact about the L96 model (with this configuration), which is actually what E. Lorenz wanted to achieve with this low-order model. Thank you. We didn't know that this point is a well known fact.

32. . lines 440-442: You might want to slightly improve this discussion since comparing $10^{-}3$ to $10^{-}6$ could let the reader think the ensemble is actually very non-Gaussian (the appendix clarifies this point, but the text should not require the appendix to be clear). Thank, your comment has been taken into account and the text has been modified for more clarity.

33. line 449: "We present in this Section comparison..." → "We present in this Section a comparison...". Thanks, done.

34. lines 450-451: "Fair comparison is therefore possible only at the end of the assimilation window.": yes, but not only. A fair comparison of DA methods would also imply cycling, which is not the case of EnsVAR here. I am very fine about your using the EnKF and particle filter to compare the ensemble qualities; but not really when it comes to comparing RMSE at the end of the window. At the very least this should be briefly discussed. We do not really understand what the referee means here. What we mean is that it is only at the end of the assimilation window that the three algorithms have used the same amount of information, and that it is only at that time that comparison is fair, in terms of Bayesianity as well

as RMSE. Having a form of cycling for EnsVAR within the overall assimilation window would define another algorithm, which could also be compared to what we have obtained. But we do not understand in what that would be ‘fairer’.

35. line 459: “...is the one described by Evensen (2003)”: which one? G. Evensen’s book describes both stochastic and deterministic EnKFs. (Of course I know the answer, you just need to improve the statement.) By the way, you should, from time to time, insists on the fact that you picked up the stochastic EnKF since the deterministic is now more popular. Moreover, choosing the stochastic EnKF makes sense in this study as the EnsVAR is also stochastic. You could mention this as this would strengthen your choice for the stochastic EnKF. Yes, we have used a stochastic EnKF. This has now been added in the paper. Now, we do not see any necessary connection between the stochastic character of our two algorithms, if ’stochastic’ only means that the data are perturbed at some stage in the algorithm.

36. line 461: With a fully observed system and an ensemble of size $N = 30$, you do not need to use localisation. In the present context it could actually be detrimental to the quality of the EnKF ensemble! (e.g., Bocquet and Carrassi, 2017). I would recommend that you do not use localisation here. Following comments from both referees, we have made a few experiments not using localisation in the EnKF. The RMSE and the RCRV are significantly degraded, while the rank histogram and the resolution component of the Brier score are improved. The reliability component of the Brier score remained the same. All that is true for both assimilation and forecast. These results, only mentioned but not actually presented in the paper, would deserve further studies which are postponed for a future work.

37. lines 479-481: It is fine to report these numbers in here, but not allude to them in the abstract, where, out of context, they do not make much sense. Actually, it was not our intention to allude to these numbers in our abstract (see our response to point 6).

38. line 503: “Kuramuto” → “Kuramoto”, as well as in both references by Yoshiki Kuramoto et al. Thanks, fixed.