Review of *Ensemble Variational Assimilation as a Probabilistic Estimator. Parts I and II (revision)* by M. Jardak and O. Talagrand Sumitted to Nonlin. Processes Geophys. A. Carrassi Editor.

Marc Bocquet

July 18, 2018

I thank the authors for their answers and their revision. I recommend acceptance of both manuscripts. There are a few points about which the authors might want to give it a second thought. In the following, I am not commenting on the choices made by the authors, even when I would have done otherwise. I only comment when I believe they did not get something significant. If need be, any technical correction that the authors would like to make can be made prior to uploading the final draft, under the control of the Editor.

1 Ensemble Variational Assimilation as a Probabilistic Estimator. Part I

- 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 absence 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.

  You did not really answer to this point. You now mention both Bayesian estimator and deterministic estimator. Hence, my remark is still valid. In the absence of cycling of the EnsVar, you have not shown that the EnsVar outperforms the EnKF, at least as a deterministic estimator. To be clear: the comparison that you make in the abstract is the outcome of your experiments. But these experiments do not reflect the general goal of data assimilation for the geofluids, which require cycling over a long period of time. I think this is totally acceptable for the Bayesian estimator, but not for the deterministic estimator. There is nothing wrong with your results but you should contextualise them and mention their limitation.

- 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).) More- over, 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.

  You should mention the name RTO here too, since it is used by several authors.

- 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.

  My remark is still valid. Your presentation is unclear and unconventional. The reader has no way to know at this stage what this time-step is.

- 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'.

  This is not Lorenz's definition of a day for his model, and not the definition used by all of your colleagues. This is very confusing.

- 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'.

  You have to contextualise your results. Your comparison between the EnsVar and the EnKF is established in the restricted context of your numerical experiments. It turns out that your experiments are not that general because you do not cycle EnsVar. Hence you cannot make general claim about the EnKF. A the very least, you have to recall the absence of cycling of the EnsVar. Again, the primary goal of data assimilation in meteorology/oceanography is the long-term tracking of the state of a geofluid. You have not defined an algorithm suited for this (since it cannot extend beyond a certain time horizon). I am fine with your choice. But you have to discuss this and contextualise.

- 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.

  Yes, 'stochastic' only means that the data are perturbed at some stage in the algorithm. This algorithmic connection is very strong, and very well know. Try comparing with a deterministic EnKF: you will understand then!

2 Ensemble Variational Assimilation as a Probabilistic Estimator. Part II

- 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.

  Please see my response to the first paper authors' response. It is the sentence "and its performance is as least as good as" that causes trouble because it is not put into the rather narrow context of the experiments of the paper. No, 'proper tuning of the methods' is not an endless task (but certainly tedious), otherwise we would not be bench-marking EnKF methods. This is usually a requirement.

- 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 ?

  As opposed to the EnKF, nobody has ever been able to set up a (single analysis) 4D-Var for chaotic models over very long windows. For the strong constraint 4D-Var, it is due to the dynamical instability of the model. For the weak-constraint 4D-Var, this is due to the numerical cost (storage, computational). Even for stable dynamics (atmospheric chemistry for instance), it is customary to split the 4D-Var analyses over several segments of a few weeks each.

- 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.

  The day has been defined by Lorenz as 0.20 time unit. Most colleagues stick to this (not so arbitrary actually) definition. Your choice may generate a lot of confusion.