Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2020-211-RC3, 2020

Geoscientific Model Development Discussions
Open Access

Interactive comment on "Combining Ensemble Kalman Filter and Reservoir Computing to predict spatio-temporal chaotic systems from imperfect observations and models" *by* Futo Tomizawa and Yohei Sawada

Anonymous Referee #3

Received and published: 21 September 2020

1 Major issues

In spite of the obvious care taken to write this manuscript, it has quite a few issues, mainly but not only related to the presentation of the theory and to the choices made for the numerical experiments:

- I don't see the point in presenting the Kalman filter. Why not go straight to the introduction of the (L)ETKF? The introduction of the Kalman filter is really unnecessary.

- The introduction of the LETKF can be perfected (see a few points below).
- Localisation is never explicitly mentioned, which is at serious odds with the use of the LETKF.
- The description of the main algorithm is not very precise. I believe that it could be significantly improved.
- The statements about the applicability of the method in the discussion and the conclusion are much too strong, and should either be retracted or significantly mitigated.
- The numerical experiment choices show that localisation is totally unnecessary here which significantly undermines the manuscript. Using the ETKF would have yielded results at least as good. In particular any claim of applicability to higher dimensional models the authors make in the discussion and conclusion is seriously undermined by this issue (see specific points below). In this vein, it is rather disappointing that the authors did not apply the *local* RC (together with the LETKF) which would have been a significant added value to the paper. This is what I initially expected from the title and abstract of the paper.
- I believe that the English of the manuscript could be significantly improved.

As a consequence, I believe the manuscript requires major revisions before being acceptable for publication.

2 Typos, remarks and suggestions, some related to the main issues:

1. l.46: “have been receiving” $\longrightarrow$ “have received”

2. l.69: Brajard et al. 2020a and Brajard et al. 2020b correspond to the same paper. You can safely remove Brajard et al. 2020a (which is the arXiv preprint of Brajard et al. 2020b). However, there is another 2020 paper from these authors which is directly relevant to your manuscript, see below in the references (Brajard et al., 2020b).

3. l.71-72: "However, their method needs to iterate the data assimilation and training, which is computationally expensive and infeasible toward the real-world problem.": Not really. Actually it depends on the number of iterations. One can use only one iteration for instance – this is actually what you do. Please mitigate your statement.

4. l.67-75: Bocquet et al. (2020) have recently proposed to use a local EnKF (LETKF for instance) coupled with machine learning. This must be cited since this is very relevant to your paper.

5. l.94-97: Your definition of the periodic nature of the L96 model is not precise enough, since x_0 or x_{m+1} are not defined for instance. Please improve the wording of your definition.

6. l.100: The characteristic time of the L96 model has been discussed in the original paper by Lorenz and Emanuel (1998). If don't believe that you should cite Miyoshi et al. 2005 here in the text.

7. l.108: "and identically distributed on Gaussian distribution" $\longrightarrow$ "and identically distributed from a Gaussian distribution"

8. l.119: "in some countries": this is too vague a statement; which ones for instance?

9. l.126: "and the time width of k corresponds to the assimilation window": This statement is not clear to me.

10. l.129: I suggest the change “and $N(\mathbf{0}, \mathbf{Q})$ means the Gaussian distribution” $\longrightarrow$ “and $N(\mathbf{0}, \mathbf{Q})$ means the multivariate Gaussian distribution” in order to contrast this definition with your previous definition for the univariate $N(0, \epsilon)$.
11. l.134: “extracted from” $\longrightarrow$ “sampled from”
12. l.136-141: It is not clear at this point why you would assume the model to be linear.
13. l.149-150: “If either the model operator M or observation operator H is nonlinear, we cannot directly use this method.”: that is a bit of an exaggeration. If the model is mildly nonlinear, the extended Kalman filter can be used. Please reformulate.
14. l.153: “is EnKF. EnKF uses” $\longrightarrow$ “is the EnKF. The EnKF uses”
15. l.161: Your statement is misleading; the covariances matrices in state space are actually never explicitly computed in the (L)ETKF! Please correct this.
16. l.171, Eq.(10): The second equation is incomplete; x_k^f is missing.
17. page 10: How come you don’t ever mention localisation for the LETKF?
18. l.180: Why use the dedicated name “LETKF-Ext”? Forecasting from the analysis is just standard proceeding. I may have missed something here.
19. l.184: “shows the architecture.” $\longrightarrow$ “shows its architecture.”
20. l.185: “in the Section 1, the previous works” $\longrightarrow$ “in Section 1, previous works”
21. l.189: “$\mathbf{u}_k \in \mathbb{R}^m$”: is it m or M as displayed in l.191?
22. l.194: “is extracted from uniform distribution” $\longrightarrow$ “is sampled from the uniform distribution”

23. l.197: I guess v should be bold, as a vector.

24. l.200: “is the operator for nonlinear transformation.“ $\longrightarrow$ “is an operator of nonlinear transformations.“

25. l.202: “Therefore, the computational cost required to train RC is small”: You do not explain why! In particular you do not mention that the problem to optimise is linear (the loss function is quadratic).

26. l.213: “matrix” $\longrightarrow$ “matrices”

27. l.240-241: It seems to me like the first iteration of Brajard et al. 2020’ loop.

28. l.236-242: What you intend to do is not clear enough to me. Please be more specific. For instance, give the corresponding algorithm. What is the model used by the LETKF supposed to be?

29. p.15: A state space dimension of $m = 8$ is small while the ensemble size $N_e = 20$ is quite large for such state space size. That makes localisation totally unnecessary if not useless. This is really problematic as a demonstration of the method.

30. l.253: “F term in” $\longrightarrow$ “The F term in”

31. l.253-254: “F term in the model was changed to represent the model bias.”: This is only one degree of freedom. Any good data assimilation method can handle this without resorting to machine learning, see for instance Bocquet and Sakov (2013). This undermines your demonstration to some extent.

32. l.262, Eq.(15): the work/concept of “localisation” has not been mentioned once. Has it?

33. l.267: “The configuration of RC used in this study was similar” $\longrightarrow$ “The configuration of RC used in this study is similar”

34. l.267, Eq.(16): The words “odd“ and ’even” need not be italicised.
35. From line 271 till the end of the numerical section, please use the present tense, not the past tense.
36. page 16: You could also use an (L)ETKF that accounts for model error, for instance that of Raanes et al. (2015). All NWP data assimilation methods have model error correction steps.
37. l.295: What you refer to is the “forecast lead time”.
38. l.312: It seems to me that you cannot refer to figures of the supplement material. Your paper has to be self-contained. Please incorporate the figures or remove the paragraph. You may want to discuss this issue with the Editor.
39. l.324: “under imperfect models”: do you mean “perfect model“? This is very confusing because you just discussed perfect model experiments and will do so again in the next sentence of the same paragraph (figure 5). It is only in the next paragraph that you report experiments with imperfect model. Please improve the text.
40. l.334-335: To which values do you change F?
41. Figure 2: Where are the tags (a), (b), ..., (e) in the panels?
42. Figure 4: You could use a log scale for the y-axis (RMSE) to see what happens for the shorter forecast lead times.
43. l.336-338 and Figure 6: The comparison is visually difficult. I suggest to merge Figure 6a and 6b which should facilitate the comparison.
44. l. 351: You cannot refer to figures of the supplement material. Your paper has to be self-contained. Please incorporate the figures or remove the paragraph.

45. l.388-390: No, in their method one can limit the number of iterations, to 1 for instance, just as you do. These authors also have a sequel to this paper, where they use a physical imperfect model as a first iteration, just as you do and this is potentially applicable to high-dimensions (Brajard et al., 2020b). By the way, you are using m=8 while they are using m=40, so they are a little closer to high-dimension than you are.

46. l.390: “contrary” $\longrightarrow$ “By contrast”

47. l.406: “The advantage of our proposed method is that we allow both models and observation networks to be imperfect.”: So do a lot of other methods that use other ML technique than RC (Bocquet et al., 2019; Brajard et al., 2020a; Bocquet et al., 2020) for instance.

48. l.413-416: When reading the abstract of your paper, I honestly thought that you would use the local RC, since you were claiming to use the LETKF. Doing so would have make your paper much stronger and more consistent (using a local RC with a local ETKF). Why did you not try it?

49. l.421-422: “These results imply that our proposed method can be applicable to various realistic problems.”: Testing the method on a 8-dimensional problem with a global RC does not make it applicable to various realistic problems! Please mitigate this too bold statement.

50. l.431-433: “Our new method is robust to the imperfectness of both models and observations so that it is feasible to apply it to the real NWP problem.”: I really don’t believe you can make such statement, from an 8-dimensional L96 model. Please remove that statement, which could be shocking to many colleagues working in numerical weather prediction and data assimilation.

References

Bocquet, M., Brajard, J., Carrassi, A., Bertino, L., 2019. Data assimilation as a learning tool to infer ordinary differential equation representations of dynamical models. Nonlin. Processes Geophys. 26, 143–162. doi:10.5194/npg-26-143-2019.

Bocquet, M., Farchi, A., Malartic, Q., 2020. Online learning of both state and dynamics using ensemble kalman filters. Foundations of Data Science 0, 00–00. URL: https://arxiv.org/pdf/2006.03859.pdf. accepted for publication.

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.

Brajard, J., Carrassi, A., Bocquet, M., Bertino, L., 2020a. Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations: a case study with the Lorenz 96 model. J. Comput. Sci. 44, 101171. doi:10.1016/j.jocs.2020.101171.

Brajard, J., Carrassi, A., Bocquet, M., Bertino, L., 2020b. Combining data assimilation and machine learning to infer unresolved scale parametrisation. Philosophical Transactions A 0, 0. URL: https://arxiv.org/pdf/2009.04318.pdf. submitted.

Lorenz, E.N., Emanuel, K.A., 1998. Optimal sites for supplementary weather observations: simulation with a small model. J. Atmos. Sci. 55, 399–414. doi:10.1175/1520-0469(1998)055<0399:OSFSWO>2.0.CO;2.

Raanes, P.N., Carrassi, A., Bertino, L., 2015. Extending the square root method to account for additive forecast noise in ensemble methods. Mon. Wea. Rev. 143, 3857–38730. doi:10.1175/MWR-D-14-00375.1.

Printer-friendly version

Discussion paper