Review of *Long-window hybrid variational data assimilation methods for chaotic climate models tested with the Lorenz 63 system* by Kennedy, Banerjee, Köhl, and Stammer, submitted to NPG.

Kennedy and co-authors present a new and interesting method of parameter estimation for Earth system models, in particular to estimate parameters for a high-resolution model for which there is no adjoint, using a lower-resolution model of the same dynamics for which an adjoint is present. Their technique uses "synchronization" of these two dynamical models using pseudo-observations to estimate the parameters. The authors derive and test this technique on the Lorenz '63 model. The goal of synchronization is to improve the length of the time integration of the high-resolution model through improved parameter estimation, with applications to climate modeling, which run forecasts on longer time scales than numerical weather prediction.

The authors present a technique that may be of use for climate modeling applications, and I can easily see the extension of this technique to more complex models. The foundations of this paper are good, however, there are several parts of the paper that require further explanation, clarification, clear definitions, and contextualization in the current data assimilation literature. Addressing these questions and comments will significantly improve the manuscript and its contribution to the community. Therefore, I recommend major revisions for this manuscript before being considered for publication by NPG.

Major Comments

I begin with a series of major comments, which if addressed, can significantly improve the clarity and purposes of this paper.

1. My first major comment is necessary to address, because it will clarify the contributions of this paper. If I understand correctly, the goal of this paper is to estimate the parameters of a dynamical model that will be used for forecasting the states of this dynamical model for long time periods (i.e., on climate timescales). This dynamical model does not have an adjoint, therefore a variational data assimilation approach for estimating these parameters given observations cannot be done. However, a simpler, related dynamical model does have an adjoint, therefore optimization with this adjoint can be used to estimate these parameters, which is does through a process the author's call "synchronization." If this is correct, then this needs to be clarified in the introduction and Section 2. Below are a series of more specific details regarding this comment:

    - In the second paragraph of the introduction and first paragraph of Section 2.2, the authors refer to a "cost function," however, having an explicit formula for this cost function, particularly in Section 2, will help to clarify the author's intention. This will emphasize the need for an adjoint (as well as define the adjoint prior to its definition in Eq.(2)), clearly define the arguments of the cost function for which you intend to minimize, and contextualize this work within the existing variational data assimilation literature. In addition, it would be helpful to clarify whether you are also minimizing such cost function for the state estimate as well, therefore defining a joint state-parameter estimation problem. For example, Chapters 4 and 5 of Evensen et al. (2022), formulate weak constraint and strong constraint 4D-Var data assimilation for the joint state parameter vector $\mathbf{z}$. It would be very helpful to compare what you are doing with standard formulations, such as those presented in this book. With respect to the cost functions defined in 2.4.1-2.4.4, these cost functions look different than the standard 4D-Var cost functions in the data assimilation literature (e.g. like those presented in Desroziers et al., 2014; Evensen et al., 2022). The authors should explain the difference between these cost functions and the cost functions used in 4D-Var, which again will help to clarify the intentions of this work and contextualize it within existing data assimilation literature.

- This next comment is regarding the specific details of the experiments: In the first paragraph of Sec. 2.2, the authors use the phrase, "control parameters," however it is unclear if these are are these the model parameters σ, ρ, β or possible the state variables x, y, z. Definition of a cost function in this section would address this question. Second, is this set up correct: the assimilation window is 100 model time units, and over this window only the parameters of the Lorenz '63 are estimated (the state variables x, y, z are not), and this generates estimates of new parameters? What is the frequency of the pseudo-observation time series that is assimilated in this window? After the new parameters are estimated, do the authors perform a forecast of the state with these new parameters to compare with the true model to compute the RMSE? The content of Sections 2 and 3 can be expanded to address these questions, which will help the readers better understand the experiments. This will also help to clarify results presented in figures in Sec. 4.
- The authors introduce the idea of synchronization: in the abstract, there are facts about synchronization that are described in the abstract (such as reducing positive Lyapunov exponents to negative values) that should also be discussed in Section 2.3, and possibly in the introduction as further motivation for this technique. I suggest adding a more detailed description of synchronization in the beginning of Sec 2.3, particularly after the sentence "The problem can be mitigated by synchronization..." Are there any simple examples that can illustrate the synchronization technique one could describe here, before showing how it applies to the Lorenz '63 system?

2. The second major comment I will make is on the literature review and discussion of data assimilation, which begins in the first two paragraphs of the introduction and is discussed in various places throughout the rest of the manuscript. In order to correctly contextualize and understand the contributions of this work, the authors can expand their literature review on data assimilation. In the second paragraph of the introduction, the authors state that there are two common approaches to data assimilation, "sequential data assimilation" and "variational approach." This is correct, but can improved. Sequential data assimilation should be explained and contrasted with variational data assimilation: if by sequential data assimilation you mean Kalman filters and their variations (e.g, extended Kalman Filter, ensemble Kalman filters, square-root filters), please specify these and cite the appropriate references (for instance, but not limited to Kalman, 1960; Evensen, 1994; Tippett et al., 2003; Houtekamer and Mitchell, 2001) and note that these methods, by design, are adjoint free methods since they do not require minimizing a cost function. With respect to the variational methods, there are other variational schemes in addition to 4D-Var, such as 3D-Var, weak-constraint and strong constraint 4D-Var, and the ensemble-variational filters such as 4DEnVar (Desroziers et al., 2014, and references therein). A brief review of these methods and references should be included.

3. The use of the word "hybrid data assimilation" particularly in the title and the method introduced in Sec. 2.4.2, either needs to be changed or properly contextualized, as it may be a bit misleading. The term "hybrid data assimilation" exists in the data assimilation literature, first defined in Hamill and Snyder (2000), where the term "hybrid" arose from combining technique from 3D-Var and the EnKF to incorporate flow-dependence in the background

estimation error covariance in the 3D-Var update step. To my understanding, the authors here are not performing hybrid data assimilation, but rather introduce a hybrid of their two models for the parameter estimation. To avoid confusion, I request that the authors change this terminology or clearly distinguish what they mean by hybrid from the terminology that already exists in the data assimilation literature.

Minor Comments

- In the second sentence of the introduction that beings with "ESMs can be used …" Even though ESM is defined in the abstract, it also needs to be redefined in the text and should not start a sentence. This can be rewritten as "Earth system models (ESMs) can be used …"
- The word "precision" is used throughout (e.g., last sentence of the abstract, fourth paragraph of the introduction, twice on page 4, once on page 11, twice on page 12). I believe the authors mean to use "accuracy" instead, since the authors are looking at errors compared to a truth run. Precision is defined as repeated experiments yielding the same result, though that result may not be accurate.
- In the first sentence of Sec. 2.1, please cite the original paper Lorenz (1963) in addition to the Yang et al. (2006) paper.
- Regarding the pseudo-observations in Sec. 2.2, (1) how often are these observations saved in the time series, and (2) when defining the additive Gaussian noise, can you please specify the mean and variance for the white noise (this may be defined in the last sentence of this paragraph, however this is a bit unclear)?
- In Eq. (2): please define the vector $\boldsymbol{x}$; I assume this means $\boldsymbol{x} = (x, y, z)$, however this should be defined explicitly either before, along with, or immediately after Eq. (2). The same goes for the vectors $\boldsymbol{x}_0$ and $\boldsymbol{x}_a$ in Eq. (7) and $\boldsymbol{x}_f$ in Eq. (9).
- The beginning of Sec. 2.4: what do you mean by "This" in the first sentence? Can you briefly summarize what you mean by this?
- Regarding Fig. 1, the figure looks nice, however making the linewidths thicker will make it a bit more readable.
- Section 2.4.2, the sentence "…may differ in resolution or numerical formulation by are governed by the same equations" By "same equations" do you mean same continuum dynamics?
- Regarding Eqs. (8) and (10), it is unclear where these adjoints appear. To improve this, I suggest the authors reorganize section 2.4 by combining sections 2.4.1 with 2.4.4 and 2.4.2 with 2.4.5 so that these equations are adjacent in the text and better explain your methodology to the readers.
- In the second paragraph of the Results, why do the authors plot the 68% percentile? A brief sentence justifying this choice would be useful.
- In the first second paragraph of Sec. 3.1, the authors say "results of two fits carried out with noise of 25%" can you clarify what it is meant by "noise of 25%" namely what is this noises added to, and what is the 25% being applied to calculate the noise.

- This is a question that may be of interest: based on your experiments, is there an optimal choice of α? For example, Fig. 7 is slightly concave up, indicating a value of α that can minimize the error. This may be worth exploring.
- The first sentence of Sec. 3.3 begins with the acronym "HDA," I suggest that the authors change this sentence so that it does not start with an acronym.

References

Desroziers, G., Camino, J., and Berre, L. (2014). 4DEnvVar: link with 4D state formulation of variational assimilation and different possible implementations. *Q. J. Roy. Meteor. Soc.*, 140:2097–2110.

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, 99:10143–10162.

Evensen, G., Vossepoel, F., and van Leeuwen, P. (2022). *Data Assimilation Fundamentals*. Springer.

Hamill, T. M. and Snyder, C. (2000). A Hybrid Ensemble Kalman Filter–3D Variational Analysis Scheme. *Mon. Weather Rev.*, pages 2905–2919.

Houtekamer, P. L. and Mitchell, H. L. (2001). A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Weather Rev.*, 129:123–137.

Kalman, R. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82:35–45.

Lorenz, E. (1963). Deterministic nonperiodic flow. *J. Atmos. Sci.*, 20:130–141.

Tippett, M., Anderson, J., Bishop, C., Hamill, T. M., and Whitaker, J. (2003). Ensemble Square Root Filters. *Mon. Weather Rev.*, pages 1485–1490.