Response to anonymous referee 2

April 18, 2025

We would like to thank referee 2 for their extremely constructive comments. During the revision process, we took all of them into consideration, following a colour coding system explained below. The manuscript has improved considerably thanks to these comments and we hope that it is now ready for publication.

Colour code:

- Anonymous referee comment.
- Comment agreed and resolved.
- Response to comment and reasoning for not making the suggested change.

1 Major comments:

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 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.
The formulation of the problem follows the strong constrained formulation described in Chapter 4 of Evensen et al. (2022) where the primary goal is to estimate unknown model parameters, and the secondary goal is to improve its state through said parameters. Therefore we are using joint state-parameter estimation problem without a joint vector z as in Evensen et al. (2022). Different from their formulation no background term is employed. Even though prior information is available and it would guarantee a well-posed problem, we consider the limit of infinitely small weights on the background term, such that it does not affect the estimation of the parameters. We don't think this is a problem as we are confident in the observability of the parameters. We added this detail.
- 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.

References to parameters have been changed to only reference model parameters. Following this comment and others from the first anonymous referee a generic definition synchronisation, the cost function, and the adjoint are introduced in the methodology section. Specific Lorenz '63 details are now only found in the experimental setup section. We also added the following paragraph to section 3.5: 'After parameter estimation, the optimised parameters are used to initialise a free unsynchronised run of the model. The attractors are plotted against the attractor of the true model. In all cases with synchronisation greater than or equal to the optimum value, the attractors' KDE shows precise and consistent agreement with that of the true model. Results are not displayed as there is no differences which are merit discussion. Thus, our focus in the subsequent results is to compare how the examined setups differ in terms of accuracy and precision of optimised parameters recovered. '

- 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?

  We now include a discussion on Lyapunov exponents and have introduced a generic ODE in the methodology to discuss synchronisation, the cost function, and the adjoint. This is found throughout the new form of Section 2.

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.

Paragraphs two and three of the introduction now read: 'There are two common assimilation approaches typically used to incorporate observations into a model: sequential and variational data assimilation schemes (Wunsch, 1996). Sequential data assimilation (Bertino et al., 2003) involves the application of a filter, most commonly Kalman filters (Kalman, 1960; Evensen, 1994, 2003; Tippett and Chang, 2003; Houtekamer and Mitchell, 2001). This technique merges a predicted state with observations at each analysis time step by estimating a joint probability distribution between the two by taking into account their respective modelling and observational uncertainties. Variants of the Kalman filter technique include: extended Kalman filters, ensemble Kalman filters, and square-root filters (Bar-Shalom et al., 2004; Simon, 2006; Evensen, 2003; Van Der Merwe and Wan, 2001; Tippett et al., 2003). They all share a similar basic procedure while differing in case specific variations of the methodology. The strength of all filtering techniques is that the sequential procedure allows for real-time assimilation of observations, for example in initialised numerical weather forecasting.

In contrast, variational data assimilation (Le Dimet and Talagrand, 1986) estimates a joint probability distribution over an extended period of time by minimising a scalar cost function, defined as the quadratic misfit between the model trajectory and all available observations within a defined time window. The most common approaches include four-dimensional variational assimilation (4D-var.) (Rabier and Liu, 2003), three-dimensional variational data assimilation (3D-var.) (Gustafsson et al., 2001), weak and strong constraint 4D-var (Tremolet, 2006; Fisher et al., 2011), and ensemble variational filters including 4DEnVar (Desroziers et al., 2014). Variational data assimilation is a useful technique for solving both initial value and parameter estimation problems (Evensen et al., 2022; Goodliff et al., 2015; Ruiz et al., 2013; Zou et al., 1992). It will be exclusively used in this study.'

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.

Changed to 'Tandem data assimilation'.

2 Detailed comments:

1. 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..."
   'Earth system models (ESMs) can be used to forecast future states of the system'

2. 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.
   We intended to discuss both accuracy and precision as is shown in our plots and will be valuable for real-world applications. The relevant references to precision listed above now include a reference to accuracy. For example in the introduction we state: 'Therefore, the objective of this paper is to investigate the accuracy and precision of such a synchronised data assimilation approach.'

3. In the first sentence of Sec. 2.1, please cite the original paper Lorenz (1963) in addition to the Yang et al. (2006) paper.
   This is now found in Section 3.1: 'In this study, we use the Lorenz '63 system for all our experiments (Lorenz, 1963).'

4. 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)?
   Section 3.1 now includes: 'This system of equations will be subsequently referred to as the true model with the parameters $\vec{\theta_t} = (10, 28, 8/3)$. This true model is used to generate pseudo-observation which will be used for synchronisation, data assimilation, and parameter estimation. Noise is included in these pseudo-observations by adding random values from a Gaussian distribution centred at zero relative to the true trajectory. The random noise magnitudes are bounded to 25% of the Lorenz '63 system's standard deviation. These pseudo-observations will be labelled as $\vec{x}_o = (x_o, y_o, z_o)$.'

5. In Eq. (2): please define the vector x; I assume this means 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 x0 and xa in Eq. (7) and xf in Eq. (9).
   Section 3.1 now includes: '...where $\vec{x} = (x, y, z)$ are the state variables...'

6. 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?

We changed the opening to Section 3.3 to: 'A multi-model tandem technique is now considered, which consecutively synchronises two forward models before running the adjoint of the second model backward in time.'

7. Regarding Fig. 1, the figure looks nice, however making the linewidths thicker will make it a bit more readable.
   The figure has been rescaled.

8. 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?
   Section 3.3.2 now contains a clarification of this point: 'In TDA we assume that both models may differ in resolution or numerical formulation but are governed by the same continuum dynamics.'

9. 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.
   This reorganisation was undertaken and the new corresponding subsections can be found in the experimental setup section 3, in the form suggested by this comment.

10. In the second paragraph of the Results, why do the authors plot the 68% percentile? A brief sentence justifying this choice would be useful.
    We added the following sentence to Section 3.5: 'The PIs are included to illustrate the statistical spread of the results and reproducibility, not to explicitly indicate uncertainty. Hence, we choose 68% for our PIs to give a concise visualisation of the central 1σ of results.'

11. 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.
    Section 4.1 now included the statement: 'Fig. 5 shows the results of two fits carried out for data with an applied noise of 25% relative to the systems' standard deviation.'

12. 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.
    This question has been addressed in the results section 4 as the authors agree that this does suggest an optimal choice for α.

13. 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.
    This sentence in Section 4.3 now reads: 'In this section, the tandem data assimilation (setup 2) is be used with different forward and adjoint models that share common physics to examine the impact of introducing model discrepancies.'