Response to anonymous referee 1

April 18, 2025

We would like to thank referee 1 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. In general, I feel the authors do not clearly state the status in the field of parameter estimation in climate models and the motivation of the study. For example, authors should clearly state the rationale behind the proposed framework and the problems being solved. Why do you need a long DA window for parameter estimation? Why do you think the new framework have any benefits? In the current formulation, the parameter estimation also requires a sensitivity matrix to the parameters. Would this be difficult to be obtained in complex climate models?

The abstract has been modified and the introduction has been expanded taking into account this comment and those of referee 2. For an answer to the questions raised here please see point 4 of the detailed comments.

2. The methodology section needs significant restructure. The authors should first briefly introduce the synchronisation method as a generic method instead of its L63 formulation. Then, the cost function of variational method should be introduced along with the gradient of the cost function. Finally, the authors should introduce the Lorenz 63 model, the exact formulation of the synchronisation model with the Lorenz 63 model, and terms in the cost function and its gradient Lorenz 63 model. The paper might also be benefited from an experiment setup (sub)section, which provide details of the choice of nudging strategy, the chosen value of observation noises/twin experiments setup, the metrics used etc. One of the given benefits of synchronisation approach is the possibility of using long DA window. However, the DA window of 100TUs is given at the very end of Sect. 3.2.

The methodology section has been reorganised following the above suggestions. A separate experimental setup section has also been added to introduce the Lorenz '63 model specific details.

3. The authors need to check Eq. (7 - 15). The adjoint equations presented in this study are normally obtained when the cost functions are temporal integral because it involves integration by parts. Yet, the cost functions are given as discrete time summations. In the SFDA section, the cost function is given as the misfit between observations and x_a, presumably x_a has 3 elements. However, given Eq (8), the gradient of the cost function in Eq. (12) does not hold as Eq. (12a) is a differences between a 3 element vector and 6 element vector as $M*_{SFDA}$ is a 6 x 6 matrix. Do authors define $x_a = (x_f x_a)$, or is $M*_{SFDA}$ given incorrectly? Moreover, it is unclear to me why do authors decide to provide the gradient of cost function (Eq. 13, Eq. 15) with respect to the parameter theta as an element-wise multiplication while it is supposed to be a matrix-vector product of dM/dtheta and lambda(t).

All equations have been double-checked and modified where required, and integrals are now used in place of sums to be consistent with the text book terminology on the adjoint method.

4. More interpretation is needed for results section. In Figure 4 and 7, with small alpha, the solid line does not look like the median of the ensemble, and I can only guess that all results lead to increased errors. Is this the case? Also, why does the parameter estimation perform better with increased alpha? Authors also need to provide better comparison and interpretation for the differences between the single, SFDA and HDA

approach. What causes the need for different alpha? The lack of interpretation leads to very similar Sect. 3.1 and 3.2.

We followed this advise by adding more interpretation of the presented results and by expanding our discussion about optimal values of α.

5. It will be good to look at the impact of different length of DA window on the parameter estimation, or Lyapunov exponents of the system. It may also be useful to check the performance when components of the syncrhonised system have different parameter values.

The Lyapunov exponent of the Lorenz '63 model is directly dependent upon its parameters. By varying the parameter through the optimisation, we investigate a wide range of scenarios. For perfect models with additive noise, the precision of the recovered parameters will improve with increasing window length, as a result of the law of large numbers. Comparing estimates obtained over long windows with averages from estimates over short windows, for which no synchronisation is necessary, would be a very valuable addition to the paper. However, short window assimilations could be done in various ways. Windows could be treated independently and results could be averaged, or (probably more naturally) optimisations could be performed sequentially using the results from the previous window as background information. This would require more detailed work, which would expand the paper quite a bit. However, we added respective discussions (see answer to detailed comment 2 below.) We covered the case of an imperfect model (similar to systems having different parameter values) by perturbing the equations with extra terms. Since all parameters are optimised, they can only be different if we reduce the number of optimised parameters, and make the remaining parameters different. As the system has only three parameters, we did not want to make optimisation even simpler.

2 Detailed comments:

1. "a sequential data assimilation scheme (Bertino et al., 2003) and the variational approach (Le Dimet and Talagrand, 1986)." → "...sequential and variational data assimilation schemes..."

In paragraph 2 of the introduction: 'There are two common assimilation approaches typically used to incorporate observations into a model: sequential and variational data assimilation schemes (Wunsch, 1996).'

2. "defined as the quadratic misfit between the observational and model data within an assimilation time window" → Usually, 4DVar cost function has a background term making it equivalent to a maximum likelihood problem in the view of Bayesian theorem under Gaussian assumption. It would be useful to distinguish the cost function here compared to more common cost function formulation

It is true that a background term is usually employed, since prior information is usually available and since it is also required to guarantee well-posedness of the problem. However, since we are confident that the observability of the parameters are not a problem we consider the limit infinitely small weights on the background term such that the background term does not affect the estimation of the parameters.

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.

3. "Due to the nonlinearities within ESMs..." → "The use of adjoint models face several challenges. Due to..."

'In the context of a full non-linear ESM, the use of adjoint models face several challenges.'

4. "the problem can be mitigated by synchronisation which removes the non-linear or chaotic dynamics from the adjoint model leading to a smooth cost function" – I feel it might be better to phrase it as "...synchronisation which constructs a system with reduced sensitivity to initial conditions leading to ..."; Moreover, authors should discuss the benefits of long DA window especially for parameter estimations.

In the first paragraph of section 2.1 we now state: 'The non-linear or chaotic dynamics, which detrimentally effect the maximum likelihood estimate, can be removed by synchronisation (Abarbanel et al., 2010; Sugiura et al., 2014) which transforms the chaotic model into one with linear dynamics without positive Lyapunov exponents leading to maximum likelihood functions with one unique maxima.'

We also added the following text to the second paragraph of section 2.1 to address the benefits of long window DA: ‘According to the law of large numbers both with perfect models and in the presence of noise, the precision of the recovered parameters will improve with increasing window length since more data is integrated into the estimation. Similar benefits could be achieved by averaging estimates obtained over short windows, for which no synchronisation is necessary. However, underlying restrictions differ. For synchronisation, noise affects the state over the entire window, whereas for short windows noise effects are transported. Short window assimilation can be of benefit in perfect model settings from the error growth as suggested by the quasi-static variational assimilation (QSVA) framework (Pires et al., 1996) due to fact that sensitivities increase exponentially with time in chaotic models. The analogue of this QSVA effect in the Dynamical State and Parameter Estimation (DSPE) method (Abarbanel et al., 2009) is the attempt to reduce the synchronisation parameter as the optimisation progresses and parameters move closer to their true values. Since errors and sensitivities grow exponentially, feasible window lengths in QSVA have a maximum value due to limited numerical precision. Similarly, synchronisation parameters cannot approach zero for assimilation windows much larger than the predictability limit, because synchronisation will eventually fail if positive Lyapunov exponents exist (Quinn et al., 2009). We note that the reasoning for the need of long assimilation windows is somewhat different in the context of full ESM, for which it is essential to resolve long time scale physical mechanisms impacted by the specific choice of parameters, such as air-sea interactions of advection time scales in the ocean.’

5. "This method allows extension of" – "This method allows for the extension of"

In the fifth paragraph of the introduction we now say: ‘This method allows for the extension of the assimilation window beyond the predictability time-scale, provided that sufficient observations are available.’

6. "To mitigate both problems, we propose a novel framework where we use two climate models both coupled through synchronisation, one with a high resolution and the other with coarse resolution for which an adjoint exists. " – Here, is it the common that adjoint models of coarse resolution is available while the adjoint models of high-resolution models are not available? Authors should provide references and discussions on the existence of the issue. Further, authors should also discuss how this novel framework could mitigate the problem of smoothness and dimensionality.

We added the following information to paragraphs six and seven of the introduction: ‘The creation of an adjoint model code from the forward code usually requires considerable effort. Automatic differentiation tools, such as Giering and Kaminski (1998); Hascoet and Pascual (2013) were developed to aid in this step. But substantial changes to the forward model code are required unless it was already developed with the adjoint modelling in mind. Stammer et al. (2018) created the first adjoint of an intermediate complexity fully coupled earth system model that is automatically created from the forward model by automatic differentiation using the TAF compiler, called the Centrum für Erdsystemforschung und Nachhaltigkeit (CEN) Earth System Assimilation Model (CESAM). The adjoint of this intermediate-complexity model is intended to be utilised for tuning more complex CMIP-type models through parameter estimation since the basic underlying physics is very similar. Otherwise this is a manual process with considerable ambiguity in the choice of parameters (Mauritsen et al., 2012).

Therefore, we propose a novel framework in which we use two climate models both coupled through synchronisation, one with a high complexity and the other of intermediate complexity for which an adjoint exists to address the second problem. The technique also has a much wider range of additional applications, since resolutions using the adjoint method lag behind those applications featuring simpler assimilation methods as variational methods are typically a factor of 100 more costly than running the associated forward model. For example, the global GECCO3 ocean reanalysis based on the adjoint method (Köhl, 2020) features only a nominal resolution of 0.4°, while for instance the GOFS 3.1 (Laboratory, 2016) based on 3D-Var (Cummings and Smedstad, 2013) features 1/12°resolution. Employing coarser versions of the adjoint while still running the forward model with full resolution could significantly reduce the cost of the assimilation effort. Therefore, the objective of this paper is to investigate the accuracy and precision of such a synchronised data assimilation approach. We perform this test using Lorenz ’63 model.’

7. "The objective of this paper is to quantify the precision and the benefit of such a synchronised data assimilation approach." – I believe this is what you have done instead of the objective of the study. A better objective would be to investigate the performance of the novel approach you proposed.

Combination of referee comments 1 and 2 in paragraph seven of the introduction: ‘Therefore, the objective of this paper is to investigate the accuracy and precision of such a synchronised data assimilation approach.’

8. "We perform this test conceptually using a Lorenz 63 model system." – The test is not performed "conceptually".

In paragraph seven of the introduction:‘We perform this test using a Lorenz ’63 model.’

9. "The advantage is that it can be used to quantitatively evaluate the parameter dependence of the system prior to application in a full model." – This sentence needs rephrasing. I guess the authors want to say "...quantitatively evaluate data assimilation schemes..." because the parameter dependence of a system will change for different dynamical systems. In fact, parameters of Lorenz 63 are non-dimensionalised numbers of a convection system. These parameters may not appear explicitly in a full climate model.

Paragraph eight of the introduction now states: 'The advantage is that it can be used to rapidly evaluate parameter estimation techniques in data assimilation schemes prior to their application in a full ESM with low computational resource requirements.'

10. "It can also be used in a wide range of other applications (Du and Shiue, 2021; Cameron and Yang, 2019; Pelino and Maimone, 2007). " – you might want to describe examples of these applications.

Paragraph eight of the introduction now states: 'It can also be used in a wide range of other applications including, but not limited to, data assimilation, stochastic modelling terms, and predictions (Du and Shiue, 2021; Cameron and Yang, 2019; Pelino and Maimone, 2007).'

11. Eq. (1) describes the classic L63 model, which I believe should have a reference to the original paper by Lorenz.

Section 3.1 now opens with: 'In this study, we use the Lorenz '63 system for all our experiments (Lorenz, 1963).'

12. "Sub-section" can be just "Section"

'Section'

13. "The random noise value magnitudes are bounded by a given percentage relative to the systems' standard distribution." — Please provide more details of your random noise choices. This is for the sake of reproducibility and credibility of the research.

Section 3.1 now includes: 'The random noise magnitudes are bounded to 25% of the Lorenz '63 system's standard deviation.'

14. Eq. (2) is technically not the adjoint model/TLM. The TLM is defined as $d\delta x/dt = M\delta x$. Also, matrices are conventionally given as bold capital letters and the matrix transpose operator should not be italic. The vector x as well as the dot operator is not defined here. In fact, I doubt the necessity of this equation as this study does not use this equation at all.

As recommended these equations are now removed.

15. In Eq. (3), x_a, y_a, z_a are not defined. Considering that authors discuss the nudging of the z variable, would it be good to have $\alpha(z_o - z_a)$ term in Eq. (3) first? Moreover, this is an equation of synchronisation strategy specifically for Lorenz 63 model. Could the authors provide a general description of synchronisation before case-specific description? Also, is this the single model approach mentioned in Sect. 3? If this is the case, authors should clearly state it.

All points raised here have been resolved by the re-ordering of the methodology and experimental setup sections.

16. Again, I doubt the necessity of having Eq. (4).

Removed.

17. "...hybrid data assimilation (HDA.)..." → "...hybrid data assimilation (HDA). ..."

'tandem data assimilation (TDA).'

18. Eq. (11) is it always a gradient with respect to x_a? Should this depend on SFDA or HDA?

Removed and equations are now only given for the specific experimental setups.