Review #3

Summary of manuscript

The authors explore the assimilation of atmospheric CO_2 concentration data for parameter calibration of the ORCHIDEE model using the 4DEnVar data assimilation method. Through carefully designed data assimilation experiments, they demonstrate the capability of 4DEnVar in assimilating atmospheric CO_2 concentration for parameter calibration, and highlighted its superiority over the ε-4DVar method in terms of computational efficiency, parameter recovery, and fitting to CO_2 concentration observations.

We would like to thank the reviewer for proofreading the manuscript, for their insightful comments that helped improve the article, and for pointing out numerous typos. (Please note that changes to the text are referenced by the line number in the new manuscript.)

General comments

Due to the continuous evolution of land surface models (LSMs), the use of 4DVar and gradient descent method for calibrating LSM parameters faces significant challenges in maintaining the tangent linear and adjoint models. Thus, it is necessary to explore adjoint-free variational methods. However, as the authors mentioned, the results obtained here for the ε-4DVar are not equivalent to a standard 4DVar, and no conclusions can be drawn regarding the comparison between the 4DEnVar and standard 4DVar methods. In light of this, to what extent can this study provide insights and practical guidance for the application of 4DEnVar to other LSMs and the assimilation of real, multi-source observations? I believe the manuscript would benefit from a clearer articulation of its research significance.

We thank the reviewer for this comment and agree that better stating the significance of the research will help the paper have a strong impact. While we do not directly compare the results to 4DVar, we do compare to ε-4DVar, which has been used as a surrogate for 4DVar in our community due to the difficulty in maintaining the tangent linear/adjoint of the model. We show in this study that 4DEnVar can be used to assimilate atmospheric CO2 data, and we compare the ε-4DVar method in order to strengthen this message. We agree that this message may be missed in the article and propose the following addition text L 534:

> **The assimilation of atmospheric CO2 concentration data using 4DVar has been implemented with a tangent linear model, as in Castro-Morales et al., 2019, or an adjoint model, as in Scholze et al., 2007. In these cases, the tangent linear or adjoint model was developed alongside the forward model. However, the ε-4DVar method was used in experiments where obtaining the tangent linear or adjoint model proved too difficult, such as in Peylin et al., 2016, and Bacour et al., 2023. Although ε-4DVar is not equivalent to standard 4DVar, a comparison of 4DEnVar with ε-4DVar demonstrates the strong performance of 4DEnVar, making it a promising candidate for this application.**

We also modify the following paragraph L580 by:

> **Moreover, the 4DEnVar method was used to assimilate several types of data using either simple carbon model (Douglas et al., 2025) or more complex LSM as the JULES LSM (Pinnington et al., 2020, 2021; Cooper et al., 2021). This new application in the ORCHIDEE LSM shows that this method is model-independent. By adding different observation terms (one term per data flux) to the cost function, the method should be able to perform multi-flux data assimilation, which would help to reduce the equifinality**

problem.

The manuscript focuses on the introduction, application, and evaluation of the 4DEnVar and ε-4DVar methods throughout the methodology, results, and discussion sections. However, this focus is not well reflected in the title. Perhaps the authors could consider revising the title in light of related works, such as Yaremchuk et al. (2016).

We thank the reviewer for this comment, however, we believe the title does accurately reflect the content and outcome of our paper. This work was conducted to prepare for the assimilation of atmospheric CO_2 concentration data in a Land Surface Model (LSM), which offers numerous advantages as presented in the introduction. However, this type of assimilation is not straightforward and requires the development of a robust data assimilation (DA) system. Our article focuses specifically on this DA system, which is why the title begins with 'Toward the.' While we do not present actual assimilation of atmospheric CO_2 concentration data, we propose promising methodologies for future application. We particularly emphasize the use of Adjoint-free Variational Methods, as many complex LSMs like ORCHIDEE, JULES, CLM cannot rely on adjoint models. We believe it is important for the community to know that this type of assimilation is possible with a system that is easier to implement and requires only forward simulations. The comparison with the ε-4DVar method was included because it has been previously used in ORCHIDEE (Peylin et al., 2016; Bacour et al., 2023), making it a relevant benchmark to strengthen our message. Unlike the significant work of Yaremchuk et al. (2016), our study focuses exclusively on Adjoint-free Variational Methods (as ε-4DVar does not require an adjoint).

In the comparison of the assimilation results between the 4DEnVar and ε-4DVar methods, the authors repeatedly attribute the poorer performance of the ε-4DVar method to the fact that it falls into a local minimum. However, for the 4DVar method, whether the parameter iteration converges to a local minimum undoubtedly depends on factors such as the a priori parameter vector. This study employed only one a priori parameter vector, and its generation process was not clarified. This raises concerns about the reproducibility and generalizability of the findings. In other words, would different conclusions be reached if a different a priori parameter vector was used?

No other prior was used as the kind of assimilation are expensive therefore only one optimisation is performed (see for example Peylin et al., 2016; Schürmann et al., 2016; Castro-Morales et al., 2019; Bacour et al., 2023). We clarify how the prior was generated L296:

> **A new vector of a priori parameters was generated manually, ensuring that it differed from the "real" parameter values while retaining physically meaningful values.**

However, we also note that the poorer performance of the e-4DVar is also due to certain regions that are less well monitored by atmospheric stations in L486, as shown in Figure 7 and A3.

> **We believe that the different spatial structure obtained by ε-4DVar is likely to be explained by the fact that the two PFTs TrBE and BoND are not well monitored, creating a dipole in the Amazonian and Siberian regions to compensate for the erroneous carbon flux in other regions.**

A more detailed description and presentation of the methods and results are needed. The manuscript currently lacks an explanation of the parameter set's value range and sampling approach. It would be beneficial to include formulas that demonstrate how the selected parameters influence ecosystem processes such as photosynthesis, respiration, and other carbon cycle components. Personally, I would appreciate seeing the distribution of the parameter ensemble and the spread of the ensemble simulations, as presented in Pinnington et al. (2020).

We completely agree with the reviewer that showing the prior and posterior parameter uncertainties is a vital part of this type of work, and indeed is a key strength of the 4DEnVar method. These aspects have also been noted by other reviewers, and we apologise for omitting the part of the methods used to generate the sample. Figures 3 and 6 have been modified to show the standard deviation of the *a priori* and *a posteriori* distributions for the 4DEnVar methods, as well as how the ensemble were generated in accordance with the work of Douglas et al. 2025. And added the following text and in L463:

> **Furthermore, Fig. 3 shows a significant decrease in the standard deviation of the posterior ensemble. This allows us to identify which parameters and therefore which PFTs appear more sensitive. In this case, it seems that the results for the TrNC3 and Crops C4 PFTs are the most uncertain.**

and in L473:

> **The posterior ensemble generated for the 4DEnVar also shows a reduction in uncertainty for all parameters. This uncertainty reduction is not equal for all parameters - a maximum reduction can be seen for the Q10 parameter(reducing the standard deviation by 94 %) and the lowest for the less sensitive $m_{maint.resp}$ parameter (with a 14% reduction for the NC4 PFT).**

We have chosen not to include the equations of the model involving the parameters to be optimised for the sake of clarity, especially given the complexity of the ORCHIDEE model. We have shared all previous work describing the model in detail in section 2.1.1. We have also explained the choice of calibrated parameters and their respective relationship with the main processes in which they participate, as detailed in sections 2.3.1 and 2.3.2. The impact of each parameter on atmospheric CO2 concentration is examined in section 2.4. We believe that describing the many model equations would not really improve the main messages of the article.

The authors may need to consider citing and discussing some recent studies, such as Douglas et al. (2025).

The paper of Douglas et al. (2025) was discussed in L86

> **This method was also successfully used by Douglas et al. 2025 to calibrate the parameters of a simple carbon model in a twin experiment.**

in L274:

> **A posterior ensemble can be obtained as it is described by Douglas et al 2025 by calculating Xa' where**
>
> $$X'_a = X'_b(I + (HX'_b)^T R^{-1} HX'_b)^{-\frac{1}{2}}$$

Specific comments

Line 34: “i.e.” to “i.e.,”.

Thank you for spotting this error. We corrected the text accordingly.

Lines 50-51: Pay attention to the spacing before or after the paragraph.

We corrected the text following the reviewer’s suggestion.

Line 54: “4DVar” to “four-dimensional variational (4DVar)”. Please check the use of abbreviations in the manuscript to ensure they are correct.

We have included this to the manuscript.

Lines 56-58: It is necessary to add references here, such as Talagrand and Courtier (1987).

We agree with the reviewer and added the references

Line 83: The citation format is incorrect and needs to be changed from “Pinnington et al. (2020)” to “(Pinnington et al., 2020)”.

We have corrected the manuscript, thank you for spotting this.

Line 92: The space between 'approaches' and ',' is extra.

We have corrected the manuscript, thank you for spotting this.

Line 92-94: The sentence is not concise and clear. It is recommended to revise it as follows: “Although tangent linear or adjoint models are not required for methods such as GA, MCMC, or emulator-based approaches, these methods necessitate defining a large ensemble, making them unfeasible for use in this study due to the time-consuming nature of model simulations.”

We agree with the reviewer. We revised the text to include this modification L92

Line 123: The period currently at the beginning of the line should be placed at the end of the previous line.

We have edited the manuscript to correct all typographical errors and rephrase certain sentences.

Line 151 and Figure 1: You mentioned the stations are selected according to their 6-month averaged sensitivity. Which six months were chosen? Given seasonal variations, it would seem more reasonable to select a full year or multiple years. Additionally, you may need to clarify whether any climate pattern, such as ENSO or IOD, occurred during the sensitivity analysis period and the simulation period. Please provide a more detailed description.

We acknowledge that the station selection could have been explained in more detail. First we selected continental stations operating during the 2000-2001 period. For those stations, we evaluated the daily average sensitivity over the last six months for each month of the 2 year assimilation window. Then, we calculated the average sensitivity over the last six months (which corresponds to an average of 24 maps, each map representing the average sensitivity over the last six months). We have added in the legends of the Fig.1

> **Monthly mean sensitivity map of atmospheric CO2 concentrations to land carbon fluxes at the 21 stations considered over the 2000-2001 period. The average sensitivity map is obtained by deriving, for each atmospheric station and each of the 24 months, the map of the average daily sensitivity of the atmospheric concentration of CO2 to surface carbon flux (in ppm/GtC) over the last six months, and then calculating the average of the 24 maps.**

To our knowledge no ENSO or IOD events occurred during this period.

Line 160: The version of the Global Fire Emission Database used in the study is outdated, or why used this one?

We acknowledge that this dataset is not up-to-date. However, because this study relies on simulated data only, we believe that using a more recent version of the database would not affect the study's conclusion or it's key message . However, we have been working on an updated configuration of the data assimilation framework aiming to assimilate real data. In doing so, we are considering updated datasets for the component of the surface CO_2 fluxes other than the biospheric one (including GFED) as well as an updated version of the atmospheric transport model.

Line 171: Please verify that the equations are correctly written. For example, vectors should be in italics, while matrices should not.

Thank you for spotting this. We have corrected all incorrect notations.

Line 278: It is suggested to consider organizing the default parameter values in a table and placing them in the supplement.

We have added a table to the appendix in order to show the True and Prior parameter values.

> We simulate the NBP fluxes at the global scale using the ORCHIDEE LSM with the default parameter values **(see Tab. A2)**

Line 284: It is necessary to specify how the a priori parameter vector was obtained.

The new a priori parameter vector was manually perturbed in order to find a new set of parameters giving a simulation of the a priori atmospheric concentration consistent with the observations.. We added to the text L296:

> **A new vector of a priori parameters was generated manually, ensuring that it differed from the "real" parameter values while retaining physically meaningful values.**

Lines 289-290: "Vcmax" and "℃" should not be italicized.

We have corrected all incorrect notations in italics in the manuscript.

Line 312: It is recommended to provide some references regarding this setup.

Lines 313-314: A more detailed explanation of the parameter range settings and the rationale behind them is needed.

We added to the text some references and justification in L334.

> **The configuration of the R and B matrices was based on previous data assimilation studies with ORCHIDEE and a simplified carbon model (Kuppel et al., 2012 and 2013; MacBean et al., 2016; Peylin et al., 2016; Bastrikov et al., 2018). These studies employed diagonal matrices for R and B to assimilate in situ observations, while Peylin et al. (2016) specifically used them for atmospheric CO_2 observations.**

Lines 364-365: Although RMSD and MAD are common statistical metrics, I still recommend that the authors provide their calculation formulas and explanations here. Since the observations have already been synthesized, are the simulation results involved in the calculation also synthesized? Furthermore, both RMSD and MAD, in terms of their form, resemble the cost function, as they include the critical term representing the difference between observations and simulations. In assimilation experiments, reductions in these metrics are expected. It would be valuable to explore additional metrics with distinct physical interpretations (e.g., coefficient of determination, R^2) to comprehensively assess method performance.

We added an Appendix named **Metrics calculation**

The RMSD and MAD are calculated as follows:

$$RMSD = \sqrt{\frac{1}{N_t}\sum_{t=0}^{N_t}\left(\boldsymbol{\mathcal{H}}(\boldsymbol{x_*})_t - \boldsymbol{y}_t\right)^2},$$

$$MAD = \frac{1}{n_{param}}\sum_{i=0}^{n_{param}} |\boldsymbol{x_*}_i - \boldsymbol{x_{true}}_i|,$$

where x* can be either xb or xa. The Pearson correlation coefficients were computed using the Numpy Python library with the 'corrcoef' function. The paired t-tests were computed using the 'stats.ttest_rel' function from the Scipy library.

We agree with the reviewer that a range of metrics is necessary to evaluate performance. This was also requested by the second reviewer.

In addition to the evaluation of the NBP already in the manuscript, we computed to Pearson correlation coefficients of the NBP in time and in space and add this to the manuscript in L 442:

The Pearson correlation coefficient between the 'synthetique' NBP and the prior NBP is 0.87 in time and 0.17 in space. The posterior NBP obtained by the 4DEnVar method shows a Pearson correlation coefficient against the 'synthetique' NBP of 0.99 in time and 0.98 in space. In comparison, the posterior NBP obtained by the e-4DVar method has correlation coefficients of 0.98 in time and 0.84 in space.

We added the figure of the GPP estimates in the appendix (Figure A3) and added this text in discussion L491

Fig. A3 shows the differences in spatial distribution of gross primary production (GPP) between the "synthetic" fluxes and the prior/posterior estimate of the two methods, as well as their global yearly budget. We can see that GPP obtained with the 4DEnVar method is slightly better than the e-4DVar method for

the global budget and better matches the spatial distribution of the synthetic flux.The e-4DVar method appears to compensate for the lack of change between the prior and posterior GPP across most of the Northern Hemisphere.

We have also added Figure A4, which shows the time series for the different stations. We calculated Pearson's correlation coefficients for each station, but they were all between 0.98 and 1. This high value is mainly due to the fact that we are in a twin experiment mode, where 'synthetic' observations are generated by the model. We felt that it was not necessary to include them in the manuscript.

Line 367: There should be a space between the number and the unit.

We have corrected the manuscript.

Line 370-372: I don't fully understand why configurations with more ensemble members (e.g., 350, 400) result in a smaller RMSD reduction. Could the authors provide an explanation?

We thank the reviewer for this very interesting question, which is not yet fully understood. We have added some hypotheses that we have formulated to the text L519.

But the performance of the 4DEnVar method seems dependent on the generated ensemble. As shown in Table 2, slightly lower performance is observed with larger ensembles, indicating that a bigger ensemble does not necessarily yield better results. This could be due to the increased dimensionality of the problem, making the iterative minimization more challenging.
Additionally, we generated a new ensemble for each experiment, which provides different information about the parameter space

and can lead to different optimal values. This shows the importance of the prior ensemble generated. Nevertheless, the reduction in RMSD remains satisfactory, with a reduction of more than 90%.

Line 395: The use of 'seem to' here makes the experiment appear insufficiently rigorous.

We agree and have removed ‘seem to’.

Line 412: Use exponential notation and change “GtC/year” to “Gt C year^{-1}”.

All notations (including those in the figure) have been modified.

Line 523: The line break in the link seems to be problematic.

We thank the reviewer and have corrected the manuscript.

Figure 1. The website should include the date of the last access.

We have corrected the manuscript.

Figure 3. “triangle” to “triangles”.

We have corrected the manuscript.

Figure 4: It is recommended to consistently retain two decimal places.

We thank the reviewers for these comments. We have modified Figure 4 to show only two decimal places.

Figure 7: The presentation of the last subplot can be improved, for example, the current color scheme does not match well with that of the other subplots.

We thank the reviewer for this comment

Figure A1: Revise unnecessary capitalization and add a comma at the end of the sentence.

We have corrected the manuscript.

References

Douglas, N., Quaife, T., and Bannister, R.: Exploring a hybrid ensemble–variational data assimilation technique (4DEnVar) with a simple ecosystem carbon model, Environmental Modelling & Software, 186, 106361, https://doi.org/10.1016/j.envsoft.2025.106361, 2025.

Pinnington, E., Quaife, T., Lawless, A., Williams, K., Arkebauer, T., and Scoby, D.: The Land Variational Ensemble Data Assimilation Framework: LAVENDAR v1.0.0, Geosci. Model Dev., 13, 55-69, 10.5194/gmd-13-55-2020, 2020.

Talagrand, O. and Courtier, P.: Variational Assimilation of Meteorological Observations With the Adjoint Vorticity Equation. I: Theory, Quarterly Journal of the Royal Meteorological Society, 113, 1311-1328, https://doi.org/10.1002/qj.49711347812, 1987.

Yaremchuk, M., Martin, P., Koch, A., and Beattie, C.: Comparison of the adjoint and adjoint-free 4dVar assimilation of the hydrographic and velocity observations in the Adriatic Sea, Ocean Modelling, 97, 129-140, https://doi.org/10.1016/j.ocemod.2015.10.010, 2016.

Bastrikov, V., Macbean, N., Bacour, C., Santaren, D., Kuppel, S., and Peylin, P.: Land surface model parameter optimisation using in situ flux data: Comparison of gradient-based versus random search algorithms (a case study using ORCHIDEE v1.9.5.2), Geoscientific Model Development, 11, https://doi.org/10.5194/gmd-11-4739-2018, 2018.

Peylin, P., Bacour, C., MacBean, N., Leonard, S., Rayner, P., Kuppel, S., Koffi, E., Kane, A., Maignan, F., Chevallier, F., Ciais, P., and Prunet, P.: A new stepwise carbon cycle data assimilation system using multiple data streams to constrain the simulated land surface carbon cycle, Geoscientific Model Development, 9, https://doi.org/10.5194/gmd-9-3321-2016, 2016.

Castro-Morales, K., Schürmann, G., Köstler, C., Rödenbeck, C., Heimann, M., and Zaehle, S.: Three decades of simulated global terrestrial carbon fluxes from a data assimilation system confronted with different periods of obse;rvations, Biogeosciences, 16, https://doi.org/10.5194/bg-16-3009-2019, 2019.

Kuppel, S., Chevallier, F., and Peylin, P.: Quantifying the model structural error in carbon cycle data assimilation systems, Geoscientific Model Development, 6, https://doi.org/10.5194/gmd-6-45-2013, 2013.

MacBean, N., Peylin, P., Chevallier, F., Scholze, M., and Schürmann, G.: Consistent assimilation of multiple data streams in a carbon cycle data assimilation system, Geoscientific Model Development, 9, https://doi.org/10.5194/gmd-9-3569-2016, 2016.

Bacour, C., Macbean, N., Chevallier, F., Léonard, S., Koffi, E. N., and Peylin, P.: Assimilation of multiple datasets results in large differences in regional-to global-scale NEE and GPP budgets simulated by a terrestrial biosphere model, Biogeosciences, 20, https://doi.org/10.5194/bg-20-1089-2023, 2023.

Scholze, M., Kaminski, T., Rayner, P., Knorr, W., and Giering, R.: Propagating uncertainty through prognostic carbon cycle data assimilation system simulations, Journal of Geophysical Research Atmospheres, 112, https://doi.org/10.1029/2007JD008642, 2007.