

Biogeosciences Discuss., author comment AC3
<https://doi.org/10.5194/bg-2022-109-AC3>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Reply on RC2

Cédric Bacour et al.

Author comment on "Assimilation of multiple datasets results in large differences in regional- to global-scale NEE and GPP budgets simulated by a terrestrial biosphere model" by Cédric Bacour et al., Biogeosciences Discuss.,
<https://doi.org/10.5194/bg-2022-109-AC3>, 2022

We thank the reviewer for the rather positive feedback and the helpful remarks/corrections. Please, see our responses below to the points raised by the reviewer.

Summary of manuscript:

- **The manuscript describes a set of data assimilation experiments applied to the ORCHIDEE Terrestrial Biosphere Model. Three data streams were used to optimize model parameters: NEE measurements from flux sites, atmospheric CO₂ observations, and satellite derived NDVI values (at select pixels). These three data streams were used for data assimilation separately, in pairs, and all together (either simultaneously or in a step-wise procedure). Parameters related to photosynthesis, phenology, and respiration were optimized in various combinations appropriate to the data streams being employed. Broadly, the authors found that simultaneous data assimilation resulted in better bias reduction and potentially less overfitting, and that calibrating the initial soil carbon stocks has a large influence on the data-model agreement (particularly for observed CO₂ concentrations).**

To clarify the summary made by the reviewer:

- the study focuses on assessing how different combinations of data-streams within a simultaneous approach may result in different optimized C fluxes. The results of the stepwise mostly derive from a previous work (Peylin et al., 2016) although we have now expanded the temporal coverage of the atmospheric CO₂ data assimilated to be comparable to our simultaneous set-ups (we have updated the results accounting now for 10 years of data instead of 3 years as in the previous version of the paper - cf response 2 to Reviewer #1);
- it is the assimilation of the whole data-streams considered that reduces the risk of overfitting, not the simultaneous data assimilation in itself.

General comments:

- **Data assimilation in the context of terrestrial biosphere models is challenging—this manuscript is a detailed attempt at meeting this challenge, which seems valuable. I think this manuscript could benefit from clearer overarching justification. Is the data assimilation exercise presented here purely for technical exploration? How will it serve broader goals related to C cycle forecasting and model uncertainty analysis? What are the next steps after this technical advance? Other broad questions might be worth briefly addressing:**

-To what extent could optimization disguise biases that result from missing processes or the underlying model structural assumptions (e.g., how soil carbon is simulated, see below)?

-To what extent is overfitting identifiable? Is there a way to quantify overfitting quantitatively in this context, or is it being defined qualitatively here?

-How might this data assimilation procedure compare to other procedures (e.g., MCMC)?

The first comment on the need to clarify objectives and take home messages addresses a concern that was also raised Reviewer #1. We are sorry if we have not met these goals. As detailed in our responses to points 1 and 3 of Reviewer #1, we have restructured the introduction and conclusion sections to make these aspects clearer.

We also recall that this study addresses not only the Data Assimilation community, but mostly the broad community using process-based TBMs, which may be less aware about the challenges and caveats associated with the joint assimilation of multiple data-streams in the context of optimizing net and gross carbon fluxes at the global scale. There are only a few similar studies addressing those questions from a global carbon cycle perspective.

On the specific points raised by the reviewer:

- Aspects of this study that could be of more direct benefit to the wider research community are 1) the use of the different metrics that we have explored here to assess and optimize our DA set-up and 2) the outcome on the importance of global scale atmospheric CO₂ data to improve the representation of the soil carbon imbalance in land surface models (how this is best achieved remains debatable and will be addressed in our further responses to the Reviewer).

- The analysis of the impact of model-data biases (including those related to incorrect process representations in the model), although important in our study, has been more extensively quantified and detailed in the paper of MacBean et al. (2016). To make our paper more concise, we have chosen to mostly refer to this study in our Discussion section rather than repeating its different outcomes.

- A quantification of the overfitting could be undertaken using synthetic experiments, and also other independent data-sets. This investigation is however out of the scope of this study. The notion is employed in a qualitative way to indicate that the assimilation of one specific data-stream may result in a set of optimized parameters that is not appropriate for other data-streams (or even for the same observed variable, but for other sites/pixels or time period), i.e. the calibrated model loses its genericity.

- The computational expenses of our DA set-up using a process-based model like ORCHIDEE, in particular the simulations at the global scale over a 10 year period that are required for the assimilation of atmospheric CO₂ data, preclude (at least for now) the use of ensemble methods, as MCMC, to explore the space of the parameters to be optimized.

We have compared in previous studies (Santaren et al., (2014) and Bastrikov et al. (2018)) the performance of gradient-based (BFGS) versus random search algorithms (genetic algorithms - GA) for the assimilation of *in situ* flux measurements. Although BFGS is more likely to be stuck into local minima, the studies showed that the assimilation of data at multiple sites (hence similar to what is done here) somehow regularizes the problem and makes the solution found by BFGS more consistent with GA results. The improvement of computing capacity and the enhancement of the model calculation times (emulators of some of more computationally demanding processes of the model are being developed) should enable the use of ensemble methods for the optimization of the ORCHIDEE model parameters in the future.

- **In addition to these broad questions I have a major technical concern related to the treatment of soil carbon. Optimizing the initial soil carbon pools without any actual constraints on soil carbon seems far from ideal. Why not use actual observations of soil carbon stocks or respiration? The link between soil carbon and the variables that are being used for data assimilation (e.g., atmospheric CO₂) is very indirect.**

We agree with the reviewer's comment about the non-ideal treatment of the soil carbon pool in our set-up; this is a point that is actually discussed in §4.2 of our paper.

Actually, the main point of this study (and this was not formulated with enough clarity in our paper) is not to have the initial carbon pools right, but the soil carbon imbalance correctly modeled. Indeed, we demonstrate in this study that if we do not have correct soil C disequilibrium, we are not able to represent the atmospheric CO₂ trend. The approach followed in this study (optimization of a scaling factor of the soil carbon pools) is similar to what was proposed in Carvalhais, et al. (2008). Because our study addresses the regional and global scales, we need to optimize the soil C disequilibrium at those spatial scales (hence the choice of regionalized scaling parameters) and we can not use local scale measurements for the task. Actually there are no direct observations/measurements of soil carbon stocks or respiration that are available for the required spatial scale. Corresponding products exist at the global scale (Nave et al. (2016) or Jian et al. (2021) for instance), but they are derived based on assumptions that may not correspond to those of the ORCHIDEE soil carbon model (as for instance the depth of the soil carbon profile). Their use to constrain the soil carbon pools simulated by ORCHIDEE would likely impose to deal with their own model-data errors and biases (the differences between such datasets and TBMs is discussed in the study of Todd-Brown et al., (2013) suggested by the Reviewer in the next point) in the process. Even if the resulting optimized soil C disequilibrium would be more consistent, the correction of the soil carbon disequilibrium would still be needed to match global scale measurements related to heterotrophic respiration (atmospheric CO₂ data for now as a surrogate of a better observational constraint).

As emphasized now in the discussion, the use of a longer transient period (following spin-up), as it is done in the TRENDY protocole (a spin up for 1800 condition and then a transient simulation with Land Cover changes and with CO₂ increase and possibly climate change), would have created a "reasonable" sink and more consistent soil C pools (with respect to the model overall structure) and would have therefore decreased the initial trend bias with respect to atmospheric CO₂ time series data. However, the use of the ERAI meteorological forcing data does not go back in time before 1989 and the cycling over the years available for the transient simulations can not account for the actual climate history of the pixels considered which, in addition to land use and management, has also an impact on soil carbon balance (Carvalhais et al., 2008). Again, the correction of the soil carbon disequilibrium would still be necessary with a more optimal set-up with respect to

the transient period.

It is worth noting that, since the developments of this study, the soil C module has improved recently in ORCHIDEE and so the way the spin-up and transient is performed (now following the TRENDY like protocols as a standard approach). The use of global datasets related to soil carbon and respiration has become a more credible prospect, at least for model evaluation if not for calibration, in future studies.

- **I also wonder if it might be more appropriate to optimize the rate parameters and transfer coefficients in the soil-carbon sub model within the TBM, rather than adjusting the initial stocks. Optimizing the initial stocks assumes that mismatch between the modeled and observed values is due to the fact that soil carbon pools are not at steady state. It certainly seems plausible that the steady state assumption doesn't hold, but there are plenty of other reasons why modeled soil carbon stocks might not match observations (e.g., the rate parameters and transfer coefficients in the soil carbon model aren't optimal, or more likely the DATCENT-type soil carbon model is only a crude approximation of soil carbon biogeochemistry). Optimizing initial stocks sweeps all of these issues under the rug, so to speak, implicitly making the argument that all of the error related to soil carbon is due to the fact that transience during the spin up period has been ignored. Carbon inputs and soil respiration might be at least as important (see for instance this paper: <https://doi.org/10.5194/bg-10-1717-2013>). The authors partly acknowledge these issues (lines 835-853), but do not fully explain why soil data can't be used for data assimilation, or outline the pitfalls associated with optimizing initial pools rather than inputs or rate parameters. If the approach used here is retained, it must be clearly explained. Also, it might be worth noting that DAYCENT-type first order soil carbon models (which are standard in terrestrial biosphere models) are all structurally deficient, in that they cannot capture feedbacks related to microbial activity. There has been a great deal of experimentation in the last decade aimed at addressing this deficiency (e.g., the MIMICS, MEND, and CORPSE biogeochemical models, which are all "microbially explicit"). At some level, DAYCENT-type models can only be right for the wrong reasons, no matter how they are optimized. While I understand that full optimization using soil carbon datasets would fall outside the scope of this paper, it would be a good idea to compare the optimized initial carbon stocks with observations at a regional scale as a qualitative reality check. The ISCN or ISRIC-WoSIS soil profile databases or the gridded Harmonized World Soil Database or Soilgrids data products might be useful for regional ground truthing.**

We have already addressed in the response above some of the points of the Reviewer.

The optimisation of turn over rates instead of scaling factors of the soil C pools has its own limitations. First, Carvalhais et al. (2008) compared with the CASA model the benefit of optimizing soil pools maximum turn over rates vs scaling parameters at the site scale; they found poorer fitting performances when the turn over rates were optimized. With our set-up, optimizing the parameters controlling the turnover times as well as the soil carbon input would require to include both the spin-up period (several 1000 years) and the transient period (about 200 years) in the iterative minimization process. This is currently not feasible with the optimization of several parameters, especially without the adjoint of the ORCHIDEE model (with our current set-up the assimilations of only atmospheric CO₂ data requires weeks of calculation).

We do not think that our optimization system could easily use soil maps (such as the HWSD data set) to force a global model like ORCHIDEE. First, with the "CENTURY" soil carbon model used in ORCHIDEE, the turnover time of the soil organic matter (for each reservoirs) together with the rate of organic matter input to the soil (litter and root turnover) determine a total soil carbon content that is in balance with all components of the model. It is thus difficult to optimize the soil carbon content with global estimates such as HWSD, while keeping the internal model consistency. First, as discussed above, the calculation times required to optimize the parameters controlling the turnover times and soil carbon input are currently highly prohibitive. Second the model does not represent yet high soil carbon content such as peat land or permafrost, while these soil types are usually taken into account in the available datasets. Finally, the HWSD soil C map corresponds primarily to the carbon content from 0 to 1 meter of soil; we would need first to adjust the observation so that it matches the total soil carbon content that is modeled. Overall, it is a rather complex process to optimize the soil carbon content in the case of ORCHIDEE if we want to keep the internal coherence of the model to improve its predictive skill. Hence, the use of correction factors of the soil C disequilibrium is a handy way to correct any issues related to the use of our Soil Organic C model.

We agree with the reviewer about the importance of such soil C and respiration datasets for improving ORCHIDEE soil C model and the representation of the spatial distribution of the soil C pools and soil C imbalance, and also for evaluating the model simulations. Thus, there are many avenues on this matter we would like to, and will, follow in future studies.

We added in §2.3.2

"The use of these [KsoilC] correction factors is a handy way to correct any issues related to the use of our soil organic C model and the soil carbon disequilibrium."

We modified the text in §4.3:

"The optimization of scaling factors of soil carbon pools is a handy alternative to the optimization of the parameters controlling the turnover times and soil carbon input of the ORCHIDEE soil C model. This would require that the spin-up (over at least one thousand year) plus the transient simulations to be included in the minimization process (hence an update at each iteration); The prohibitive calculation times precludes doing so for now. Exploiting in TBMs databases more directly related to regional soil carbon contents (**as the Harmonized World Soil Database (HWSD) (FAO/IIASA/ISRIC/ISSCAS/JRC, 2012)**, International Soil Carbon Network, Nave et al. (2016), or the global soil respiration database, Jian et al. (2021)) by TBMs is not straightforward because of **the errors associated these datasets (Todd-Brown et al., 2013)** and inconsistencies between the estimated quantities and the model state variables and underlying processes (**as for instance the depth of the soil carbon**). "

Detailed comments

- **Title: consider striking "different", since "multiple datasets" strongly implies different datasets.**

We agree with the reviewer and have removed "different" from the study title.

- **41-43: Is "correction of the initial carbon stocks" a general problem for TBMs,**

or a problem in this context given how spin-up was dealt with?

We corrected the "initial carbon stocks" to "soil carbon imbalance" for the reasons described above. The problem relates to our specific set-up but is shared among TBMs.

- **Line 47: Suggestion: replace "grown" with "increased".**

We agree with the reviewer and have corrected the text accordingly.

- **Lines 95-96: What does this mean: "considered potential incompatibilities between the model and the observations"?**

What is meant here is the fact that assimilating one data-stream may result in degraded performance with respect to other data-streams.

We corrected the text:

"However, many previous studies that assimilated multiple datasets hardly considered potential incompatibilities between the model and the observations (Bacour et al., 2015; Thum et al., 2017), **resulting in deterioration of model agreement with other observations not included in the assimilation.**"

- **Lines 109-110: This point about Gaussian errors seems important. How does it affect the analyses presented later in the paper?**

The assumption of Gaussian errors is indeed pivotal to the vast majority of data assimilation frameworks (including our own). The definition of the misfit function, of the posterior errors, and of the different metrics used in this study to characterize the impact of the observations, and hence the different analyses, rely on this assumption. Given the difficulty to characterize the distribution of the errors on parameters and model-data, the choice to assume Gaussian distributions is a practical solution, which also makes the inverse problem mathematically tractable and facilitates its resolution. With this assumption and using linearity assumption close to the optimal solution, the posterior uncertainties (both on the parameters and model-data sides) can be easily estimated, as well as the observation information content diagnostics.

- **Lines 130-132: Is it OK to tune priors or prior uncertainties? This seems antithetical to the definition of a prior.**

The point raised by the reviewer somewhat meets point 7 of Reviewer #1. We summarize here our previous response. The definition of a priori error statistics based on expert knowledge (i.e. how to set the parameter error statistics before assimilation) is difficult and hence imperfect (necessary assumptions are made). The approach of Desroziers was developed to check the consistency between the **R** and **B** covariance matrices, and hence ultimately allows improving the definition of the prior uncertainties.

- **Line 230: Were the pixels selected entirely at random given these constraints?**

Yes indeed, the selection is random given these constraints. More details on the pixel selection are provided in the paper of MacBean et al. (2015) which is cited in this section.

- **Line 256: The notation in the equation may be confusing to some readers: the superscript indicating transposition (t) might be mistaken for an exponent. Perhaps clarify what this superscript mean in the text, and/or use a different notation (the symbol ` or an upper case T might be less confusing).**

We agree with the reviewer and have now used T instead of t in equations 1 to 10.

- **Line 261: Does this approach really account for uncertainty in the model structure? Perhaps elaborate in another sentence, as it is not obvious why this would be the case.**

As detailed in our response 17 to Reviewer #1, the error covariance \mathbf{R} accounts by definition for the error in the observations and in the model (Bouttier and Courtier, 2002; Rayner et al. 2005 among others) when both observational and modelization uncertainties are Gaussian (Tarantola, 2005), which is consistent with our assumption.

- **Line 270: "the calculation of using" is confusing syntax, consider rephrasing**

A term is indeed missing (we have faced several similar issues with the disappearing of mathematical/symbols characters when generated the PDF for submission, as also pointed out by Reviewer #1): the correct sentence is "the calculation of $\square_x(Jx)$ uses..."

- **Lines 291-292: Land use is important, but this statement is unreferenced and there are certainly other sources of variation (and other potential drivers of model-data mismatch).**

We agree with the reviewer and have changed the sentence:

"The size of soil carbon pools drives the magnitude of the net carbon fluxes exchanged with the atmosphere to a large extent and is closely related to the land use history."

to

"The size of soil carbon pools drives the magnitude of the net carbon fluxes exchanged with the atmosphere to a large extent; **Soil carbon** is closely related to **soil texture, climatic (temperature and moisture), disturbance history (including land use and fires), as well as ecosystem and edaphic properties (Schimel et al., 1994; Todd-Brown et al., 2013).**

- Schimel, D. S., Braswell, B. H., Holland, E. A., McKeown, R., Ojima, D. S., Painter, T. H., ... & Townsend, A. R. (1994). Climatic, edaphic, and biotic controls over storage and turnover of carbon in soils. *Global biogeochemical cycles*, 8(3), 279-293.
- Todd-Brown, K. E., Randerson, J. T., Post, W. M., Hoffman, F. M., Tarnocai, C., Schuur,

E. A., & Allison, S. D. (2013). Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations. *Biogeosciences*, 10(3), 1717-1736

▪ **Line 366: "of and of" ... missing word?**

Yes indeed (similar issue than in point 16). The correct sentence is "needs to be performed iteratively for successive values of x^a and of the corresponding residuals.."

▪ **Lines 515-517: Could this indicate overfitting? Tuning the soil carbon pools could be masking other sources of disagreement between model and data.**

We agree on the last remark of the reviewer. Our point here is related to the larger weight of the optimisation of the multiplicative C soil factors in the minimization process which inhibits any significant improvement relative to the other parameters/processes.

▪ **Line 588: Why no transient simulations in this study? Is this a computational constraint?**

The part stating that there is no transient simulation performed for the global scale simulations was an error. Actually, we spun-up the model recycling the 1989–1998 ERAI meteorology and then used a transient simulation from 1990 to 1999. We corrected the text at the end of §2.1.2 and added (in bold): "Each spin-up simulation was then followed by a transient simulation (starting from the first year of measurement for each data stream) and accounting for the secular increase of atmospheric CO₂ concentrations; **for the global simulations, only a short transient simulation from 1990 to 1999**".

As indicated earlier (point 3 above), the ERAI meteorological data does not go before 1989. We made the choice to recycle the available data for the spin-up but not for the transient (as discussed above, doing so would have not made it possible to account for the actual climate history and would have nevertheless resulted in incorrect soil carbon imbalance). We have relied on the tuning of the scaling factors of the soil C pools to optimize the soil carbon imbalance. This is actually what happened but we were not expecting such a high impact of this correction on the overall fitting performances. Having only a short transient period was therefore more a deliberate choice (which has turned out to be not so well-advised) than resulting from a technical/computational constraint. In our more recent studies, we have corrected this set-up using a TRENDY-like protocole.

▪ **Lines 832-834: This point needs some elaboration. How does correcting the CO2 trend bias hinder evaluation of photosynthesis?**

The point of the reviewer relates to the sentence "A consequence of correcting the trend bias is that the model improvement with respect to other processes (photosynthesis, phenology) is hindered".

This statement is supported in the text by several analyses and statements in the text (what will be modified in the manuscript is written in bold below):

- 3.1.2 (L536): “These results **and the low reduction in NEE and FAPAR RMSDs following the assimilation atmospheric CO₂ data described above** highlight the predominance of the correction of the trend in atmospheric CO₂ time series through the fitting of the carbon pool parameters, over the tuning of the other model parameters related to photosynthesis and phenology **(see Figure 3)”**;
 - 3.1.3 (L555): “When assimilating atmospheric CO₂ data, most of the parameter correction aims at reducing this bias.... Compared to the improvement in the bias in the trend, the improvements (reduction in bias) in the amplitude of the CO₂ seasonal cycle and in the length of **the carbon uptake period (CUP)** (Figures 3b and c) are marginal.”
 - 3.3 (L692): “... This reflects the lower constraint brought by the CO₂ assimilation experiment on photosynthesis and phenology related processes, **as already pointed out in §3.1.2.**”
- **Lines 847-848: The steady state or the non-steady state approach causes bias? This sentence is ambiguous.**

We changed the text to: “Going beyond the steady state assumption following model spin-up has been discussed already (Carvalhais et al., (2010); MacBean et al., 2022), as **steady state** results in biased estimates of soil carbon reservoirs (Exbrayat et al., 2014).”

- **Lines 849-851: What are the inconsistencies that limit application of soil carbon data? This point needs a far stronger justification and should be made earlier in the paper, given that soil carbon pools are being tuned without reference to soil observations.**

These inconsistencies have been discussed earlier in addressing points 3 and 4 and the reviewer. We recall that we improved §4.3 in the Discussion to justify why we do not rely on soil C products.