# Reviewer 2

Raoult describes a model parameter tuning exercise using the ORCHIDEE model and Fluxnet + CO2 FACE experiments. It builds on recent work comparing model simulations to global change experiments by directly tuning the model to FACE experiments for two PFT types.

We would like to thank the reviewer for taking the time to read and comment on our manuscript.

The take-home messages are that the parameter tuning with the FACE studies seems to improve some aspects of the model and not others and that tuning to the flux data alone can alter the predicted response to rising CO2 (I did find it interesting that the prior and the Flux+FACE predictions in Figure 5 were similar to the priors, suggesting knowledge of how the model responded to CO2 likely influence past parameter sets).

Yes, this is interesting but we cannot be sure whether this is due to prior knowledge or just a coincidence - an artefact of this model version - since posterior parameter values are different. It speaks more to the fact that there are a lot of possible parameter sets, which may give similar response in some parts of the model and very different response in other parts.

I appreciate the inclusion of the FlxGN-AMB simulation because without its inclusion it would be hard to separate the influence of the type of data at the FACE sites (NPP, LAI) from the use of the experiments.

I would challenge the authors to make the manuscript have more impact beyond users of the ORCHIDEE model. What can other modeling groups learn from the manuscript? As an example, the manuscript discusses how the inclusion of the nitrogen cycle influenced the capacity to tune the model to the FACE studies. However, this is not directly tested in the results in a standardized framework. Adding those results would alert other modeling groups to be cautious with optimizing the C + N model using a similar approach as a C-only model. Related, are there any suggestions about how to update the parameter tuning to work better with N cycle models. It seems the KSoil parameter was an issue because it adjusted both C and N pools. Should there be a KSoil for each soil pool so that pools that have high N mineralization rates are adjusted differently than ones with lower N mineralization rates?

We agree that we do need to better highlight the broader perspectives of this paper and how the wider modelling community can benefit from this study. As discussed in the

conclusions, we present not only methodology for optimisation against these data but also a warning to other modelling groups looking to calibrate their model, especially since most land surface models are now starting to include a nitrogen cycle. We have added the following to L420:

> These results highlight the fact that optimising an LSM with the nitrogen cycle is more difficult and complex than with a carbon-only LSM, given the increased model feedbacks. **In particular the dependence of plant productivity on soil nitrogen availability and in reverse the dependence of soil N content to litter input (hence productivity) induce strong positive feedbacks. This provides a warning to other modelling groups looking to calibrate the carbon and nitrogen cycles in their model.**

We also highlight a new use for FACE data, which importantly can be used to identify structurally deficiencies in the model (L444):

> Finally, structural changes do need to be made to the model to better capture the inter-annual variability of simulated NPP and LAI. **This highlights how we can use FACE data to identify structural issues in models providing an important tool for model development.**

One of the biggest lessons from the study was the added complexity of getting the initial carbon pools right now that we need to maintain C/N ratio. We choose to implement one multiplicative scalar (KSoil) to the different pools, applying it to the slow, labile and active pools for both C and N content, thus preserving the C:N ratio of the pools. Using a different KSoil for each pool or a different KSoil for C and N, although technically feasible, would likely be associated with adverse effects. If one pool decreases more than the others (both for C and N content) or one pool becomes more depleted in N than the others at the initial time step, it is likely that the model will enter a transient phase with possibly strong compensating fluxes (i.e., possibly a large net carbon flux) in order to reach again the internal model consistency. Ideally, we would need to optimize the different model parameters controlling the turnover time of each pool and their C:N ratio over the full duration of the spin-up, which is practically not feasible. We have added the following to the conclusions (L424):

> However, the parameter now also changes the initial nitrogen stock and hence the mineralisation flux in the soil, which impacts GPP. **Another approach would have been to have several multiplicative factors each changing different pools or indeed one for each C and N. However, this would likely lead to more complications given the strong feedbacks observed. If one pool declines more than others in terms of C and N content, or if one pool becomes more depleted in N, it is probable that the model will enter a transient phase with potentially**

**strong compensating fluxes, such as a large net carbon flux, to restore internal consistency. Ideally, we would optimize various model parameters that govern the turnover time and C:N ratio of each pool throughout the entire spin-up period. However, achieving such optimization is not currently feasible.**

Specific comments:

1. More description of the ORNL and Duke FACE data is needed. How did you handle the split-plot design at Duke FACE? Where specifically did the data come from? Pulled from the table of a manuscript or from a data repository?

   We have added the following to Section 2.3 "In situ data"

   > "**The data for these sites came from the FACE-MDS project (Walker et al., 2018a and 2018b; https://facedata.ornl.gov/facemds/).** For each site, we used the data from two experimental plots **(with their associated error bars)**; one with unaffected atmospheric $CO_2$, i.e. ambient (AMB), and one with elevated atmospheric $CO_2$ (ELE). **Although the DUKE experiment also has ammonium nitrate treatments at half of its plots from 2005 onwards (Feng et al., 2010), we only consider the data from the plots without nitrogen fertilization.**"

   References:

   FACE-MDS Phase 2: Meteorological Data and Protocols. Walker, A.P., Yang, B., Boden, T., De Kauwe, M.G., Fenstermaker, L.F., Medlyn, B., Megonigal, J.P., Oren, R., Pendall, E., Zak, D.R., Zaehle, S., Burton, A.J., Drake, B.G., Evans, R.D., Hungate, B., Johnson, D.P., Kim, D., LeCain, D., Lewin, K.F., Lu, M., Mueller, K.F., Nowak, R.S., Riggs, J.S., Smith, S.D., Tharp, L.M., Zelikova, T.J., Norby, R.J., 2018a. doi:10.15485/1480325

   FACE-MDS Phase 2: Data from Six US-Located Elevated $CO_2$ Experiments. Walker, A.P., De Kauwe, M.G., Fenstermaker, L.F., Hungate, B., Medlyn, B., Megonigal, J.P., Oren, R., Pendall, E., Talhelm, A.F., Zaehle, S., Zak, D.R., Boden, T., Brown, A.L., Burton, A.J., Butnor, J.R., Day, F.P., Drake, B.G., Dijkstra, P., Evans, R.D., Finzi, A.C., Iversen, C.M., Jackson, R.B., LeCain, D., McCarthy, H.R., Powell, T.L., Nowak, R.S., Riggs, J.S., Smith, S.D., Stover, D.B., Tharp, L.M., Warren, J.M., Wullschleger, S.D., Norby, R.J., 2018b. doi:10.15485/1480325

Feng, X., Simpson, A. J., Schlesinger, W. H., & Simpson, M. J. (2010). Altered microbial community structure and organic matter composition under elevated CO2 and N fertilization in the duke forest. *Global Change Biology*, *16*(7), 2104-2116.

2. The use of parameter priors isn't clear. In standard Bayesian statistics, priors have a prior distribution. I think this prior distribution is a combination of the mean in Table 1 and the B matrix. A description of how the B matrix is created seems to be missing.

   We define the prior distribution of each parameter to be a Gaussian spanning 40% of the prior range - which itself it elicited from expert knowledge. This description has been added to the text as follows (L149:

   "**We define the prior distribution of each parameter to be 40% of the prior range.**"

3. The posterior parameter values should be added to Table 1 for the parameters that were fit.

   We agree that the posterior parameter values are lacking in the manuscript. Instead of adding the values to Table 1 (which would complexify the table), we have added a figure illustrating how the parameters change, as well as a section discussing some of these changes.

4. How is the R matrix determined?

   As in most studies, we set the **R** matrix to be diagonal. We defined the observation error (variance) as the mean-squared difference between the observations and the prior model simulation so that this variance reflects not only the measurement errors but also the model errors. Added the following to the text (L149):

   " **We set both matrices to be diagonal. For B … For R, we defined the observation error (variance) as the mean-squared difference between the observations and the prior model simulation so that this variance reflects not only the measurement errors but also the model errors**."

5. I recommend adding a discussion of the results in the context of Rastetter, E. B. (1996). Validating Models of Ecosystem Response to Global Change. *BioScience*, *46*(3), 190–198. https://doi.org/10.2307/1312740.

   We thank the reviewer for sharing this paper - it will definitely help strengthen the discussion. The discussion has been expanded as follows (L406):

Manipulation experiments allow us to test the model under different $CO_2$ regimes and its capabilities to reproduce the ecosystem responses. **FACE sites in particular are an important tool in evaluating modelled ecosystem response to climate change. They can be thought of as space-for-time substitution experiments (Rastetter, 1996), but where the change in atmospheric $CO_2$ is controlled and even manipulated to exceed conditions naturally found around the globe currently.** By optimising model parameters against data from both ambient and elevated atmospheric

6. Line 436: the sentence talks about how the study reduces parameter uncertainty but the manuscript doesn't actually present the prior vs. posterior parameter uncertainty or any ensemble of simulations with different parameter values from the prior and posterior distributions. The study optimizes the parameter value to be more consistent with the observations but doesn't necessarily reduce the uncertainty.

   It is true that we do not present the posterior uncertainties in this manuscript. We have calculated them using the Hessian at the optimal and found that the uncertainty is significantly reduced for all parameters. However, we are wary to add this information to the manuscript since we know our optimisation setup is not perfect, and while the posterior parameter values will be relatively unaffected by this, this can lead to complications when interpreting the posterior uncertainty reduction. Most notably, we do not include error correlations in the **R** matrix (note **R** is used in the calculation of the Hessian). This is because these correlations are extremely hard to quantify. Instead, we usually inflate the variances to account for the fact we are overestimating the information content of the observations. For this proof of concept study, we did not do this step - nevertheless, we did use large variances (defined by the mean squared difference between the model and observations) to partly account for this. In future studies, we would want to tune this multiplicative factor to ensure that the information content of the observations was not overestimated (using a $\chi^2$ test) or, in an ideal world, be able to quantify the off-diagonals of the **R** matrix.

   We have changed the sentence to be more nuanced removing the mention of uncertainty:

   > "By optimising the model against a number of different constraints, we **gain confidence in our** parameter and hence in the projections"

   And added the following to the text (L444):

“Finally, structural changes do need to be made to the model to better capture the inter-annual variability of simulated NPP and LAI. This highlights how we can use FACE data to identify structural issues in models providing an important tool for model development. **Although not shown, our framework also allows us to compute the posterior parameter uncertainty, which again can be very informative for model development. We do not discuss them in this paper since our imperfect setup (i.e., diagonal R matrix) means the information content of the observations is overestimated in the optimisation, but we do find that the uncertainty parameters are strongly reduced in all cases.**”