EGUsphere, author comment AC1
https://doi.org/10.5194/egusphere-2022-641-AC1, 2022


# Reply to Referee #1

Vivek K. Arora et al.

---

Author comment on "Towards an ensemble-based evaluation of land surface models in light of uncertain forcings and observations" by Vivek K. Arora et al., EGUsphere, https://doi.org/10.5194/egusphere-2022-641-AC1, 2022

---

**We thank Referee #1 for their helpful comments. Our replies to their comments are shown in bold below.**

Summary

The authors present an evaluation of their updated land model, CLASSIC. The primary update from a previous model is a new nitrogen cycle, although new land cover reference data have been implemented also. Notably, they evaluate the model using two different meteorological driving data sets, and also comparing new vs old land cover data and with/without the new nitrogen cycle. Their evaluation system effectively compares model scores to benchmark scores that are based on observational uncertainty. They conclude that their new model is reasonable compared to both observation and other models, and also that present-day land atmosphere co2 flux is independent from the initial land carbon state, with respect to the variations included in this experiment.

Overall review

I appreciate the authors' more comprehensive approach to evaluating their updated land model. The evaluation system provides clear results. However, in the end it isn't clear that the model is better, but it appears that the GSWP3 forcing gives better results than the CRU forcing. I also think that the conclusions regarding the independence of flux from initial land state are overstated, mainly because this is a highly constrained case where the land start and end points are shifted by a similar amount and the land change trajectory is nearly identical (but shifted) between the two cases. I recommend the following main revisions (see below for additional details):

**Thank you for your overall positive review of our manuscript.**

1) The different land cover cases need to be redefined. They are not different reconstructions. They just reflect an update in the present-day reference land cover data that are used to anchor the model's land trajectory. While this is a reasonable update, it does not represent the uncertainty of land use/cover change.

**Thank you for noting this. Yes, it is correct that the change in crop area over the**

**historical period in both land cover data sets in our study is the same. In this sense, we agree, that it is not entirely correct to call these land cover data sets two reconstructions. The distinction here is between land use change (LUC) and land cover. We treat LUC similarly despite the differences in the two land cover cases. If we are given the opportunity to revise our manuscript, we will clarify this distinction and redefine the two land cover cases.**

2) Qualify your conclusions regarding the robustness of model fluxes under different initial carbon states. This is a very specific, highly constrained comparison and there are many factors and uncertainties, particularly in the land space, that are not considered here but have substantial impacts on carbon flux and storage estimates.

**The response of the terrestrial biosphere over the historical period is driven by four primary global change drivers – increasing $CO_2$, changing climate, land use change (LUC), and N deposition. We agree that in our current framework we haven't taken into account the uncertainty associated with LUC and yes there are several other uncertainties as well. However, our simulations do allow us to evaluate how the response to the three other global change drivers is dependent on two driving meteorological data, two land cover cases, and two model variations (with and without an interactive N cycle). We will clarify this point when revising our manuscript.**

Please also note that our statement about little dependence on the initial land carbon state is in the context of the NET atmosphere-land CO2 flux. The reason why this happens is that the model is first spun up to equilibrium conditions and then forced with time-variant forcings. So while the absolute fluxes (Gross primary productivity and respiratory fluxes) are different, the NET atmosphere-land CO2 flux is similar across simulations in that the net flux from all simulations lies within the uncertainty range from the Global Carbon Project.

3) Complete your background and comparisons with literature on land data uncertainties. Some suggestions are below, but there is more out there showing the complexity of this problem.

**Thank you for pointing to the additional references that highlight the uncertainty associated with LUC emissions. These additional references will help us highlight and clarify that our framework does not account for uncertainty associated with LUC.**

4) Swap the more useful appendix figures for the unreadable paper figures. Try to make the figures more readable.

**We were not sure if the figures showing the spread across the simulations or the figures showing the effect of land cover, meteorological data, and the inclusion or the absence of the N cycle separately were more useful. We will swap the figures between the appendix and the main text.**

Specific comments/suggestions

Abstract

line 6: I would not call these two land cover sets different historical reconstructions. You just replaced your current day reference land cover with newer data. There are many other factors that affect the reconstruction, most notably the land use data and the assumptions used to apply land use to land cover.

**Yes, we agree as mentioned above and we will redefine the two land cover cases.**

line 12: awkward sentence transition. probably do not need the beginning of this sentence; just start with "Simulated area burned…"

**Thanks for your suggestion.**

Introduction

lines 90-94: there are other studies on this topic. the most relevant one is probably this one because it addresses uncertainty in land cover in conjunction with the effects of co2, nitrogen deposition, and climate:

A.V. Di Vittorio, J. Mao, X. Shi, L. Chini, G. Hurtt, and W.D. Collins, "Quantifying the effects of historical land cover uncertainty on global carbon and climate estimates", Geophysical Research Letters. doi: 10.1002/2017GL075124.

This one looks at land change emissions across several land cover representations:

Peng, S., P. Ciais, F. Maignan, W. Li, J. Chang, T. Wang, and C. Yue (2017), Sensitivity of land use change emission estimates to historical land use and land cover mapping, Global Biogeochem. Cycles, 31, 626–643, doi:10.1002/2015GB005360.

but this one is also relevant:

A.V. Di Vittorio, X. Shi, B. Bond-Lamberty, K. Calvin, A. Jones, 2020, "Initial land use/cover distribution substantially affects global carbon and local temperature projections in the integrated Earth system model", Global Biogeochemical Cycles. doi: 10.1029/2019GB00683.

**Thank you for mentioning these references that will help us highlight and clarify that our framework does not account for uncertainty associated with LUC.**

CLASSIC modeling framework

lines 178-180: Some folks may disagree here. Different types of trees have different leaf/canopy shapes, orientations, and colors that may affect interception and also radiative processes.

**This statement was made in the context of current formulations used in land surface models which typically only use leaf area index (or plant area index) and a PFT-dependent parameter to calculate the storage capacity of leaves for calculating how much precipitation is intercepted. Such an approach is used in CLASSIC. In this context, the PFT-dependent parameter accounts for leaf shape and orientation but not the underlying deciduous or evergreen phenology of the leaves. We will reword our statement to clarify this.**

Driving data

line 230: This may be true in some cases, but the later step of creating the reconstruction by applying the land use trajectory to this static cover map can generate greater uncertainty, not to mention the additional uncertainty in the land use data. See the papers above. See comment below. And also these:

Di Vittorio, A.V., L.P. Chini, B. Bond-Lamberty, J. Mao, X. Shi, J. Truesdale, A. Craig, K. Calvin, A. Jones, W.D. Collins, J. Edmonds, G.C. Hurtt, P. Thornton, and A. Thomson

(2014). From land use to land cover: restoring the afforestation signal in a coupled integrated assessment - earth system model and the implications for CMIP5 RCP simulations, Biogeosciences, 11:6435-6450, 2014, doi: 10.5194/bg-11-6435-2014.

Meiyappan, P., and A. K. Jain (2012), Three distinct global estimates of historical land-cover change and land-use conversions for over 200 years, Front. Earth Sci., 6(2), 122–139, doi:10.1007/s11707-012-0314-2.

**We take your point. In fact, there are two sets of uncertainties here. The first is converting 20-40 land cover classes to a much smaller set of plant functional types (PFTs) that a model simulates, and as we showed in our manuscript this affects the pre-industrial state of vegetation and soil carbon (899 vs 1171 Pg C in our case) but also the magnitude of the current terrestrial sink. The second set of uncertainties is related to incorporating LUC data into a model's land cover over the historical period (as you highlighted) which is what leads to uncertainties in LUC emissions and therefore also affects the terrestrial sink. These two sets of uncertainties affect model behaviour differently. This latter set of uncertainties is not taken into account in our framework and we will revise our manuscript to clarify this.**

lines 266-286: Figure 1 indicates that changing your reference cover map does not generate the greatest uncertainty. The range of vegetation across the other models is much greater than the difference you show for your data. What is the nominal year for your data sets? What about for the other models? What is actually driving the variability in these data across trendy? Are they all using different reference rs data? Or are other factors contributing?

**Yes, while in terms of total vegetated area the GLC2000 and ESA-CCI based land covers are not that different, the difference is large for the area of grasses and this leads to the 899 vs 1171 Pg C soil carbon difference. This is noted in the manuscript. The data are averaged over the period 1992-2018 for CLASSIC and all TRENDY models. The reason for the variability in vegetated area, and area of trees and grasses across models, is the subjectiveness in the process of mapping/reclassifying 20-40 land cover classes in land cover products to a selected number of PFTs in land models and that land modelling groups use different land cover products. We will clarify this when revising our manuscript.**

lines 358-387: This section is unclear, particularly with respect to how the benchmark scores are calculated (the ones comparing the obs). You show the benchmark scores in figure 10, but it isn't clear how these are calculated. I do like this benchmarking system, though.

**The benchmarking scores and how they are calculated are explained in the following paper.**

Seiler, C., Melton, J. R., Arora, V., Sitch, S., Friedlingstein, P., Arneth, A., Goll, D. S., Jain, A., Joetzjer, E., Lienert, S., Lombardozzi, D., Luyssaert, S., Nabel, J. E. M. S., Tian, H., Vuichard, N., Walker, A. P., Yuan, W., and Zaehle, S. 2022. Are terrestrial biosphere models fit for simulating the global land carbon sink? Journal of Advances in Modeling Earth Systems, p.e2021MS002946. https://doi.org/10.1029/2021MS002946.

**Short of including the whole description in the main text we will include the details in our appendix when revising our manuscript so that a reader doesn't have to refer to the above paper.**

line 393: figures a2-a16 are much more useful than figures 3-9, which are unreadable.

**We will flip the figures in the main text versus the appendix when revising our manuscript.**

Results

lines 558-559: It is unclear how the benchmark scores are determined.

**An additional section in the appendix of the revised manuscript will explain the benchmarking process in more detail as mentioned above.**

Conclusion

lines 649-651: The key word here is "present-day flux." The cumulative emissions over time are dependent on the land cover change trajectory, which you do not alter in these scenarios. This is also why your previous statement regarding model response being independent of initial land state makes sense here; you do not have a different transient land path that would change the outcome. Your two land covers are both estimates of "present-day" cover, and as such are not that different from each other. And using the same land cover backcasting your initial state changes by a similar amount. So the flux is also constrained by similarly adjusted endpoints.

**Since we do not take into account different LUC trajectories, it is also implied that had we not taken LUC into account at all (i.e. simulations were driven with only increasing CO2, changing climate, and increasing N deposition) even then the net atmosphere-land CO2 flux would have been similar across different simulations. This suggests that the present-day net land-atmosphere CO2 flux indeed is largely independent of the pre-industrial land carbon state in so far as the response to three other global drivers is concerned (CO2, climate, and N deposition). As mentioned earlier, we agree that the caveat related to LUC emissions has to be made more clear in our manuscript, and we note your feedback that we have overstated our conclusion related to the independence of present-day net atmosphere-land CO2 flux. We will reword this conclusion and tone down the message.**

If we were to plot the cumulative emissions from 1960 onwards so that they can be compared to estimates from the Global Carbon Project (GCP), similar to Figure 9a in the manuscript, even then all eight simulations lie within the uncertainty of estimates from the GCP as shown below.

Note that there are grammatical typos throughout.

**Thank you for noting these. We will address all these when revising our manuscript.**

Figures and Tables

Figures 3-5:

I cannot determine which simulations are where. A couple of colors are clear, but the groups of lines are muddled together and I cannot tell which simulations are in which group. If you colored them by output group it would be easier to tell which sims have similar results. Using a temporal average may also help (without the annual values shown, which make it messy).

Figures 6-9:

These are difficult to read. Since output groupings are less apparent here, I suggest selecting colors that reflect the experiment groupings. Temporal averages may also help here (without the annual values shown).

**The purpose of these figures is to give the reader an idea of the spread across the eight simulations. Even if the colours are wisely chosen there will be lines that overlap other lines for some variables. As suggested by you we will move these figures to the appendix, and the appendix figures to the main text.**

Figures A2-A16:

make sure the axis scales match across all panels in each figure.

**Thank you for noting this. We will make the y-axis scale similar for all sub-panels of a figure.**