Review of egusphere-2022-623

Opening Pandora's box: How to constrain regional projections of the carbon cycle

By Teckentrup et al.

In this study, the authors analyze the impact of varying meteorological forcing obtained from the historical CMIP6 GCMs / ESMs on the historical carbon cycle. More specifically, they asses the impact of the selection of the simulated meteorological forcing on the response of the Australian carbon cycle using different strategies, e.g. bias correction, random-forest approach, ensemble averaging methods, as well as one dynamic global vegetation model, LPJ-GUESS. The authors compare the different methods and report their effect on carbon cycle simulation of LPJ-GUESS in space and time.

The analyzes presented by Treckentrup et al. are very interesting, comprehensive and useful in understanding the impact of different meteorological forcing on the carbon cycle. In some places, the manuscript seems a bit overloaded, making it somewhat more difficult to grasp the full scope of the analyses. Overall, the manuscript is well written. I have a few general comments and a short list of specific comments. Thus, I recommend minor revisions before publication.

Thanks for the positive review of our manuscript. The "bit overloaded" comment echoes Reviewer 1 and we will address this via improving the clarify and "take-home" messages in our revision.

General comments:

1. I like how the title reads and the scope of the study, but I find it a bit misleading. As far as I can judge, you are not looking into the projection of the carbon cycle, right? Projection, by definition, means simulating a potential future evolution of the system (e.g. boundary conditions are scenario-driven). Your analysis is based on historical simulations, where we have access to the boundary conditions, e.g. greenhouse gases, volcanic/anthropogenic aerosol loading, etc. So, I would use the word "regional simulations".

   We thank the reviewer for the suggestions and will adjust the title accordingly.

   Then I do not really see how you constrain the regional carbon cycle uncertainty. You show that one land model simulates different CC response dependent on different meteorological inputs. In your previous paper, you showed the uncertainty that is related to the variety of land models. So, I would say that you comprehensively demonstrate the entire uncertainty in simulating the carbon cycle related to the choice of models and choice of forcing, but I don't see really how you would go about in constraining this uncertainty. The proposed bias correction methods etc. do not really contribute to reduce the uncertainty, since, if we now ran all TRENDY models with your reanalysis-corrected / or "ensemble average weighting" meteorological forcing, we would end up with a similar uncertainty in the CC response. Bottomline is, maybe you should focus more on the “full uncertainty” aspect in communicating your analysis, than the “constraining” aspect.

This is a very insightful comment that we will address in a revised manuscript. The reviewer is of course correct, that to sample the full (or a “fuller”) uncertainty would necessitate driving the full TRENDY ensemble with corrected data, perhaps a future research direction – we will make this point clearly in revision. What we are doing is demonstrating how biases in climate forcing can be constrained and what some of the implications may be for the terrestrial carbon cycle.

2. Overall, I am very surprised that the effect of CO2 on plants, e.g. on water-use efficiency, or the direct stimulation of carbon assimilation, is not being discussed nor mentioned here at all. These effects are vital in simulating the carbon cycle under rising CO2. Were these effects accounted for in the LPJ-GUESS setup? I think so, since almost all runs show an increase in C_total, even those which received a decrease in precipitation and increase in temperature as forcing. How would Australian ecosystems accumulate more carbon under these circumstances? To estimate the impact of meteorological forcing, the CO2 effects might not be essential, but still, these effects need to be addressed and communicated.

   We thank the reviewer for this comment. Indeed, the simulations shown in the manuscript were forced with a transient CO2 (and nitrogen deposition) forcing, and we will add a paragraph in the discussion to clarify this.

   The climate forcing and other external drivers in LPJ-GUESS will always interact, and therefore the divergent carbon cycle response presented in this manuscript does not isolate the impact of climate forcing on carbon cycle uncertainty alone. However, given all LPJ-GUESS simulations have the same configuration apart from the climate forcing, i.e. the prescribed nitrogen deposition and atmospheric CO2 concentration are identical for all ensemble members, we argue that the experiment set-up is reasonable. Further, any future projections based on offline DGVM runs will be linked to a similar problem linked to inherent complex interactions, and we therefore were aiming to show the impact of climate uncertainty, and correction methods, in a default DGVM configuration. In our forthcoming work we plan to unpack these different drivers more clearly when we consider future simulations, the manuscript is about to be submitted.

3. I am hesitant to suggest more analysis, since this manuscript already contains a lot of analysis and is a bit over-loaded. So, it is difficult to grasp the entire scope of the manuscript. Are that many supplementary figures needed? I would suggest to assess whether one could reduce some parts in the manuscript, so that it becomes better accessible to the reader and the key messages come across.

   Again, this comment reflects on our need to improve the clarity. We will do this in a revised manuscript, including being stronger in removing supplementary figures.

4. But I have to suggest at least one additional analysis point: You only use one realization (r1i1p1f1) of each model. To really get an idea of how the specific GCM compares to reanalysis and other GCMs, one should analyze as many realizations as possible. I would even suggest to get meteorological forcing from grand / large ensembles and one can identify real biases in the model. One realization is not representative for the model, except when some data-assimilation / nudging is conducted (e.g. as in reanalysis). I know it would be too much work for this study, but one should think about it.

   We thank the reviewer for the comment and agree. Indeed, we were initially considering using the CESM large ensemble. However, as the reviewer suggests, the additional DGVM runs would mean significant increase in computation time, which we already aimed to reduce by only applying bias correction methods on a subset of GCMs defining the ensemble spread. It may not be practical to add additional runs and note that the results should be viewed as indicative, rather than aiming to define an exact number that defines carbon cycle uncertainty. Our preference therefore is to address this comment via additional material in the discussion.

Specific Comments:

L18: What does "and above" mean here? and above global scale?

We thank the reviewer for pointing out this mistake and will remove 'and above' in the revised manuscript.

LL85-89: Rather long sentence containing many aspects - can you split it up in at least two separate sentences?

We thank the reviewer for the suggestion, and we will break up the sentence.

LL92-94: I don't understand the logic of this sentence. TRENDY models use the identical meteorological forcing and show a large difference in the response of the carbon cycle to the forcing. So, this calls for reducing uncertainty in the land-surface model predictions, rather than the meteorological forcing, no?

We thank the reviewer for the suggestion and will remove the sentence.

LL97-98: Can you provide more detail on what first generation and second generation DGVMs refer to?

First generation DGVMs typically simulate plant communities using a single area-averaged representation of each plant functional type (compare Fisher et al., 2018) while second-generation DGVMs simulate vegetation by individual plants with similar properties, such as age, size, or functional type, together. We thank the reviewer for pointing out that this needs clarification and will update the introduction accordingly.

L103: What simulation? Please be more specific. It is probably the “historical” simulation, but there are others, like esmHist, where the carbon cycle is fully coupled, etc.

We thank the reviewer for pointing out the lack of clarity and will include the information about the simulation (the historical simulation of CMIP6 is used here).

L104: What about the information about atmospheric humidity, i.e. VPD?

As a legacy of its development and (lack of) availability of humidity data, LPJ-GUESS does not use VPD as an input forcing. Stomatal conductance is based on an empirical boundary layer parameterisation following Huntingford and Monteith, 1998. This parametrisation expresses large-scale evapotranspiration as a hyperbolic dependency on surface resistance (i.e the inverse of stomatal conductance). Therefore, humidity as an input driver is not needed for LPJ-GUESS (compare Smith et al., 2014).

LL106-7: Can you really do that? Shouldn't you recycle all the inputs consistently then? You can have strong precipitation with simultaneous high shortwave radiation - what does LPJ-GUESS make out of these physically implausible inputs?

We apologise for the mistake and will rerun the simulations affected following this suggestion. Given incoming shortwave radiation does not limit vegetation growth in Australia, we expect that the results will not change.

LL108-109: This means you are doing some heavy down-scaling the input variables to a quite high resolution in comparison to the native resolution of the GCMs. Maybe better to remap to a common 1x1 degree grid, no? Or maybe it'd be better to use downscaled CMIP6 output, e.g. https://eartharxiv.org/repository/view/2646/

We remapped the relatively coarse GCM output to a 0.5 degree given that is the native grid of LPJ-GUESS. While we are aware that dynamically downscaled data exists, such as the CORDEX dataset, or ISIMIP. We chose the CMIP6 forcing given it was the newest climate simulation dataset available and has the largest number of ensemble members. Further, output from the CMIP6 ensemble is commonly used as input drivers for both regional, and global studies of terrestrial ecosystems.

L125: I think, that is not true. ERA5 is in 0.5x0.5 grid and there is a derivative that is at 0.25x0.25, but 0.05 seems extremely high resolution for reanalysis.

We apologise for the mistake and will correct that ERA5-Land is on a 0.1 spatial grid.

Figure 1: I think it would benefit the understanding of Figure 1, if you provided a slightly more elaborate figure caption. At least, you could specify the acronyms used in the figure, so the figure is readable without searching in the text for the acronym definitions.

We thank the reviewer for the suggestion and will update the figure legend accordingly.

L140: Can you provide more information on this estimator?

Table 2: The definition of the the summation notation would need more information to be mathematically correct, but I guess it is understandable as it is. https://en.wikipedia.org/wiki/Root-mean-square_deviation

We thank the reviewer for the comment and will update the equation accordingly.

L143: Well, these models historically evolved and they share code and concepts. It's hard to define which models are independent. Also, the models that are used to create the reanalysis e.g. IFS for ERA5 share code with CMIP6 models.

We agree with the reviewer and note that we are aware that this is just one of the many ways to define GCM independence. We will also clarify that dependence to reanalysis datasets can exist.

L147-148: Also, models that are highly dependent might not "correlate more" on monthly time-scale as the atmosphere is chaotic and highly dependent on the initial state etc.; I would assume that correlation of the spatial pattern in the climatological mean would provide more information. So, I think similar spatial bias matching would give you an idea whether models are similar or not, but maybe you do that, I did not fully understand.

We thank the reviewer for the suggestion and apologise for the lack of clarity in the methods description. We derive the correlation in the bias over all timesteps, and grid points, and therefore account for spatial patterns in bias.

L166: "Let us define" ?

We apologise for the lack of clarity and will update the sentence in the revised manuscript.

LL170-173: Does this part connect to any paragraph?

We included a brief description of univariate vs multivariate bias correction methods (205-210) in the methods description to remind the reader what they are.

L180: If you used temperature in Kelvin scale (so no negative values), one could only use this function for scaling consistently for all variables, no?

We apologise but do not quite understand this comment and would appreciate if the reviewer clarified it.

L191: "Let us denote" ?

We apologise for the lack of clarity and will update the sentence in the revised manuscript.

L205: Not sure how this fits in the structure of the paragraph.

We thank the reviewer for the suggestion and will remove both sections on the general difference between univariate and multivariate correction methods.

LL231-232: Then I really wonder why some representation of atmospheric humidity is not an input to LPJ-GUESS.

Given LPJ-GUESS focuses primarily on tracking carbon and nitrogen fluxes, the stomatal conductance in simulated vegetation is driven by CO2 and soil moisture rather than VPD. We agree that this is a shortcoming of the model and note that current model development efforts are looking to account for VPD when stomatal conductance is simulated (see Belda et al., 2022).

LL276-277: Can you explain why you include non-physical parameters such as longitude and latitude in the random-forest approach. Especially for a regional study, I would advise against this practice.

Testing the performance of the RF model out-of-sample with versus without including geolocation information have shown that including longitude and latitude information improve prediction performance, this is likely because including geolocation predictors enables RF to capture spatial dependencies.

Figure 2: b,d,f are the same - but I saw the uploaded corrected figure.

We apologise, and will include the correct figure in the revised manuscript.

LL305-onwards: Would it make sense to compare carbon fluxes from the actual CMIP6 models to get an estimate for carbon cycle uncertainty? Not all models (e.g. MPI−ESM1−2−HR), but most have some representation of the carbon cycle and the carbon fluxes? I also understand if you only wanted to focus on the effect of the selection of the meteorological forcing.

We agree that it is important to also consider carbon cycle projections from coupled simulations in CMIP6. However, we here aimed to focus exclusively on uncertainty in the meteorological forcing in offline runs given bias correction methods can be applied which is not possible for coupled runs. We will add this in the discussion.

Figure 3: “PPT” is a rarely seen abbreviation for precipitation, better pr?

We thank the reviewer for the suggestion and will update the abbreviation for precipitation.

LL446-447: In the context of Australia, I would assume one can also add “improved prediction of fire risk”, as fire depends largely on the fuel load thus vegetation / carbon cycle.

We agree with the reviewer and will add the suggestion in the manuscript.

LL589-590: Counter-argument: One should not only rely on using one DGVM for studies on ecosystem/carbon cycle impact. Maybe you can make the point, that we should use multiple DGVMs and multiple GCMs forcings.

We agree with the reviewer and will make this point in the discussion.

References

Huntingford, C., Monteith, J.L. The behaviour of a mixed-layer model of the convective boundary layer coupled to a big leaf model of surface energy partitioning. *Boundary-Layer Meteorology* **88**, 87–101 (1998). https://doi.org/10.1023/A:1001110819090

Smith, B., Wårlind, D., Arneth, A., Hickler, T., Leadley, P., Siltberg, J., and Zaehle, S.: Implications of incorporating N cycling and N limitations on primary production in an individual-based dynamic vegetation model, Biogeosciences, 11, 2027–2054, https://doi.org/10.5194/bg-11-2027-2014, 2014.