Reviewer 1

In this paper, the authors tested the impact of aggregation of soil texture data to coarser resolutions on model simulations, and the sensitivity of vegetation and soil carbon to soil texture in three popular Terrestrial Biosphere Models (TBMs). They demonstrate that vegetation and soil carbon are, for the most part, insensitive to soil texture. Data aggregation to coarse resolutions from TBMs hinder most of the soil texture spatial variability.

First, I would like to acknowledge the importance of studying the influence of soil texture in models. I believe it's a subject that hasn't received as much attention as it should lately. I was surprised to see the lack of sensitivity of models to soil texture, and I'm glad this paper is bringing this up.

We would like to thank reviewer #1 for their positive evaluation of our work. We agree that soil texture (and belowground processes in general) does not receive as much attention from the vegetation modelling community as it should and hence this study. We even added the following sentence in the discussion to insist on that point:

"In the context of increasing drought intensity and severity, a better representation of the impacts of soils on plant productivity and status is urgently needed. More generally, the belowground compartment should receive more attention from the vegetation modelling community."

Our detailed reply to the reviewer's comments can be found right below.

<u>Major comments:</u>

My most important concern is regarding the relevance of the aggregation of soil texture input data. I think the point the authors make is valid, however, it's an issue that models have with most data input, as the scale at which they work is coarse. What I mean is that although is good to point this out, I don't think it's groundbreaking or surprising. Therefore, I'm not sure it's worth having this result as the main finding of the paper. I think the other result about the sensitivity of the vegetation and soil C to soil texture is much more important and should be the focus of the paper. While the soil data aggregation could be a secondary finding. I liked the suggestions provided in the discussion on how to address this issue in models. However, in the abstract and conclusions, I think the framing should be more directed towards a place where models could improve by proving some measure of uncertainty with the selection of soil texture.

We totally agree with this comment. Indeed the impact of the soil aggregation of soil textural data is not the most important finding of the paper and the second aspect (the lack of sensitivity of most model outputs to soil texture) is much more interesting, also in our opinion. We addressed this comment by putting more emphasis on what used to be the second objective in the previous manuscript version. This includes: deleting and/or re-ordering the statements about soil aggregation in the abstract and conclusion, rephrasing and re-ordering the objectives of the paper, and shifting/re-balancing the results and discussion. Now the objective of the paper is clearly about testing the sensitivity of the vegetation and soil C to soil texture. We kept the results about the aggregation of soil texture input data because they were interesting to us but this is now just a secondary outcome of the study.

Regarding the same topic, I think it's worth reinforcing in the discussion or the abstract why it's important to consider the uncertainty in the influence of soil texture in vegetation activity and drought, particularly for large scale/ecosystem level studies.

We agree with the reviewer on the importance of including the uncertainty of soil texture in TBMs, in the context of increasing drought and model inaccuracy. We added multiple references to this issue throughout the paper, including (abstract, intro, discussion):

" (...) we suggest that better estimates of the soil texture uncertainty resulting from soil texture data aggregate should be considered in the future. Those steps forward are critical to properly account for future increasing drought stress conditions in tropical regions."

"Yet, most TBMs do not consider the uncertainty in the influence of soil texture in vegetation activity and drought when applied at large scales."

"The South-American tropics frequently suffer from (mega)droughts such as the 2015/16 El Niño event, which severely impacted ecosystems on the continent. These droughts result from the increasing environmental pressures, including climate change and deforestation (Stall et al., 2020). In a recent study, Yang et al. (2022) linked patterns of forest biomass changes with drought severity and duration as well as soil clay content, which indicates that both water demand and supply (or both climate and soil/roots) influence forest functioning."

About the biomass, GPP and soil carbon simulations' comparison with observed data: can you provide more information in how this comparison was done? The methods don't provide much detail on what was done and it's difficult to understand exactly what you did. I think it was a correlation analysis, but I'm not completely sure that was it.

We agree that the methods were not fully detailed on this aspect in the previous version of the manuscript. We added in the material and methods:

"To evaluate the model performance, we averaged the model outputs (GPP, AGB, soil carbon) for each grid cell and vegetation model/scenario over the period of observation of the remote sensing products (see previous section). We compared the resulting maps

through a correlation analysis to compare their spatial distribution, and compared their density distribution through standard metrics (mean, root mean squared error).”

We also added in the results section:

“The spatials correlation of AGB between the map of Avitabile et al. (2016) and the models varied between 0.35 (LPJ-GUESS) and 0.82 (ORCHIDEE), with an intermediate performance for ED2 (0.67).”

“The spatial correlations of soil C between SoilGrids250m and the models were systematically low: 0.10 for ORCHIDEE, 0.25 for ED2, and 0.26 for LPJ-GUESS.”

Minor comments:

I provided some minor comments in the attached pdf file.

We would like to thank the reviewer for this detailed feedback, which was very useful. We implemented all the suggested changes in the new version of the manuscript.

Reviewer 2

In this study, the authors perform a sensitivity analysis to different soil texture properties (from the global SoilGrids250m dataset) on the carbon cycle in three Terrestrial Biosphere Models (TBMs), namely LPJ-GUESS, ED2 and ORCHIDEEv2.2. They evaluated the aboveground biomass spatial distribution, ecosystem Gross Primary Productivity (GPP), soil carbon content and drought stress simulated by the three models over the Amazon rainforest region, using model default pedotransfer functions. They found that the model outputs were mainly insensitive to soil texture change, showing the poor representation of the soil-vegetation coupling in the TBMs.

Overall I find the topic very interesting and important to produce accurate simulations in the land surface models. There are some points that I think need to be improved/clarified in the manuscript to be suited for a publication.

We thank the reviewer for the positive evaluation of our work and the detailed feedback below.

Major comments:

1. I think the soil texture is also connected to the land use / land cover over a region. Even more, the root depth plays a key role in the water uptake and therefore in GPP and ET processes. Given that the Amazon basin has gone through deforestation activity for more than three decades, it is important to take into account the land cover, the root depth and the soil texture to have a realistic effect in the change in biomass.

We totally agree with the reviewer that to realistically model spatial (and temporal) patterns of biomass and GPP, land cover (and land use changes) need to be taken into account on top of soil and root properties. However, the main purpose of our paper was to

test the intrinsic sensitivity of all the investigated models (ED2, LPJ-GUESS, and ORCHIDEE) to soil texture, not to improve the overall accuracy of those TBMs for the Amazon. Hence, the model performance analyses are short because they can somehow be found in the original publications describing the different models. Therefore, we added in the new version of the manuscript (intro/discussion):

“Over the last three decades, the Amazon tropical forest has been facing an increase in environmental pressures, including the severity and length of drought events (Spinoni et al. 2014). This trend is projected to aggravate by the end of the century (Duffy et al. 2015), also resulting from the rapid deforestation rates and the regional precipitation recycling (Stall et al., 2020).”

“The South-American tropics frequently suffer from (mega)droughts such as the 2015/16 El Niño event, which severely impacted ecosystems on the continent. These droughts result from the increasing environmental pressures, including climate change and deforestation (Stall et al., 2020).”

“Other critical processes are impacting this region and require as much attention. The Amazon basin has gone through intense deforestation activities for more than three decades. Among the consequences of such environmental pressure, deforestation increases drought (Stall et al., 2020) and modifies soil properties (Veldkamp et al., 2020). (...) These processes and their interactions should be accounted for in the next generation of TBMs.”

“We used those datasets for qualitative comparison only and not with the objective to improve model accuracy and/or precision.”

Because deforestation and soil properties are actually interconnected (Veldkamp et al., 2020), the comment of the reviewer inspired us to do an additional analysis. Even though

there is not an agreement on the exact impact of land cover change on the soil texture, the literature suggests that the impact of land cover change on soil texture is of the same order of magnitude as the changes that we applied in our soil scenarios. We thank the reviewer for pointing out this important aspect and we adapted Figure 1 so that it now also compares the changes of soil texture we considered in the scenarios with the observed changes of soil texture due to land cover changes in a specific study.

Regarding the rooting depth, while we fully agree with the reviewer that this is a key model parameter in all three models, investigating its impact was beyond the scope of this paper focusing on the effect of soil texture. Since root parameters are also linked to plant hydraulic representation, assessing the sensitivity to such parameters would also require an in-depth analysis of other plant functional type parameters, which will complexify the analysis and dilute the key message of this study. We added/rephrased in the manuscript:

"Soil water availability is intimately related to root distribution (De Deurwaerder et al., 2021), root and soil depth (Nepstad et al., 1997), as well as soil properties including texture"

"More generally, the belowground compartment should receive more attention from the vegetation modelling community. It is known for long now that root and soil depth critically influence drought tolerance (Nepstad et al., 1994; Fan et al., 2017), yet this knowledge has not been integrated in most TBMs (Verbeeck et al., 2011). In this study, we focused on soil texture only but further research should include other soil properties (e.g. soil structure see Fatichi et al., 2020) and root traits."

2. I think the change in soil clay should be carried out with a recalibration of other land surface-related parameters in the TBMs to have an accurate representation of land surface

processes. Maybe the low sensitivity observed in the simulations is in part related to this non parameter recalibration?

While we agree with the reviewer that any model structural change should be accompanied with a recalibration of the model parameters to match the observed behavior of the modelled biome, here we show that the model is (almost) completely insensitive to soil texture. This means that recalibrating the models for each of the soil texture maps would lead to (almost) the same sets of parameters. We think that the low sensitivity observed in the simulations is due to the model structure that oversimplifies the belowground component of the land surface rather than a model calibration issue.

Minor comments:

1. I suggest in the introduction to include also relevant literature in regards with Amazon deforestation. I think deforestation and subsequent land cover change has an effect in soil texture variability that is worth to mention.

We have now included these three references in the manuscript:

Eleftheriadis, A., Lafuente, F., and Turrión, M.-B. (2018). Effect of land use, time since deforestation and management on organic C and N in soil textural fractions. Soil Tillage Res. 183, 1–7. https://doi.org/10.1016/j.still.2018.05.012.

Staal, A. (2020). Feedback between drought and deforestation in the Amazon. Env. Res Lett 10.

Veldkamp, E., Schmidt, M., Powers, J.S., and Corre, M.D. (2020). Deforestation and reforestation impacts on soils in the tropics. Nat. Rev. Earth Environ. 1, 590–605. https://doi.org/10.1038/s43017-020-0091-5.

2. The authors cite literature about ORCHIDEE and ED2, but not about LPJ-GUESS. I suggest to include recent work done with this model.

We have included a new, recent reference to the LPJ-GUESS model:

Oberpriller, Johannes, Christine Herschlein, Peter Anthoni, Almut Arneth, Andreas Krause, Anja Rammig, Mats Lindeskog, Stefan Olin, and Florian Hartig. 2022. "Climate and Parameter Sensitivity and Induced Uncertainties in Carbon Stock Projections for European Forests (Using LPJ-GUESS 4.0)."

The older references (Smith et al., 2001; Smith et al. 2014) are still valid though.

3. Line 96 → Please specify sensitivity of what property/variable from the three TBMs to soil texture?

See next comment

4. Line 96-98 → Please rephrase this idea "which occupy different positions along the vegetation representation abstraction continuum". Perhaps do the authors mean something like the three models have different levels of complexity?

The sentence now reads:

"In this study, we explored the sensitivity to soil texture of important model outputs (e.g. GPP, soil carbon) of three state-of-the-art TBMs with different levels of complexity"

5. Line 107 → Could you please explain your reasons to select the cohort mode over the other two options and the related implications?

Cohort mode is the default vegetation representation for simulations using LPJ-GUESS. Population mode is inherited from the original LPJ-model, which is essentially a big leaf model (Smith et al., 2001). For simulating this vegetation representation we would recommend to use a recent development of the LPJ model instead (e.g. von Bloh et al., 2018).

On the other hand, the individual mode in LPJ-GUESS shares the vast majority of its code with cohort mode in the current version of the model. While individual mode would allow for variations between different members of a cohort (e.g. in establishment parameters), this is not currently implemented. Therefore we expect that the impact of choosing cohort mode over individual mode would not have any influence on the outcome of our study (except for computational time!). This sentence now reads:

" For this study, the model was run in cohort mode (default vegetation representation)"

6. Line 108 → I believe the acronym "PFT" has not been properly introduced nor explained so far.

The acronym is now defined after its first use

7. Line 110 → Which meteorological drivers? And where do you get this input data from?

The order of M&M was changed according to one of the next comments of the reviewer, so that it is clear from the beginning where the input data come from and which met drivers are used for each model (new section 2.1)

"To drive the model, we used the 6-hourly CRU-NCEP v7 meteorological forcing dataset (Viovy 2018). Climate variables include air temperature and humidity, incoming short-wave and long-wave radiation, precipitation rate, surface pressure and winds."

8. Line 117 → "Soil moisture in the top two layers (20 cm) is available for surface evaporation". Does this mean that deeper soil moisture (>20 cm) is not available for evapotranspiration processes? I believe this is not an accurate representation of the soil water uptake by the Amazon rainforest (see doi: 10.1038/372666a0). How can this affect the interpretation of your results?

We believe that there was a confusion here between the soil evaporation, and the overall evapotranspiration. While only the top layer is contributing to the former, plants can

access both top and bottom layers for transpiration. We did not modify this sentence which explicitly refers to surface evaporation and not plant evapotranspiration.

9. Line 117 → "Only two larger percolation layers are defined". Larger than what?

This refers to the previous sentence in the manuscript, where we describe that the soil profile is divided into 15 soil layers of 10cm thickness each. However, for simulating percolation, the model currently uses two larger (than 10cm) layers of 50cm and 100cm. This is a remnant of a previous version of the model, which only used these two larger layers for all soil processes. The 15 layers in the current version are mainly used for a more detailed calculation of soil temperature. We modified this sentence as:

"Yet, only two large layers are defined for percolation: excess water from the top layers (down to 50 cm) percolates into the bottom layers (remaining 100 cm), where it is distributed between the 10 layers depending on their water capacity."

10. Lines 121-122 → About the soil water content per grid cell, how does it change in time? Do you give soil moisture as input data to the model to compute the water content at filed capacity and at wilting point?

In all three models, soil water content is a dynamic state variable that is tracked during the water cycle resolution during the simulations. Hence, soil moisture is not provided as an input, but is a prognostic variable computed based on the different in- and outflows of water in the soil compartments. The water content at field capacity and wilting point are computed from soil information according to pedotransfer functions, as described in the manuscript (Supplementary section 1).

11. Lines 133-134 → I suggest to remove this information or move it to the introduction, as I do not see its relevance for the methods nor for the results.

We removed that sentence

12. Line 137 → Similar to my previous comment, where do you get the meteorological forcing from?

We reorganised the methodological section according to one of the reviewer's comments below so that the origin of the drivers is more clear.

13. Lines 152-153 → "Simulated sites are characterised by vertically uniform soil texture and hence hydraulic properties over the entire soil column". Did you mean: Simulated sites are characterised by vertically uniform soil texture and hence uniform hydraulic properties over the entire soil column?

This description was moved to new section 2.4 where it now reads:

"Soil texture and hence hydraulic properties were assumed vertically uniform for each model and simulation."

14. Lines 15-156 → I am confused here. If this model can use the vertically integrated soil water from the deepest soil layer (which I believe is 8 m depth based on line 148), how can you compare the results obtained from this model (ED2) with those obtained from the previous model (LPJ-GUESS) that only uses water available from the first 20 cm of the soil column? Also, the third model (ORCHIDEE v2.2) has 2 m depth soil profile, so the same question would apply for the results from the third model.

Indeed, the three models have different soil depths (and number of soil layers, and formulations of water uptake, and etc.). We are aware of those differences that we tried to summarise in Table 1 and this is the main reason why our study focuses on the sensitivity of those models to soil texture rather than model performances. The key aspect of such model intercomparison exercises, as commonly performed (see e.g. the LBA-MIP project, doi: 10.1111/gcb.13442), is precisely the fact that each model represents the same processes with different approaches and hypotheses. While we expected clear differences

in sensitivity because of their structural differences, the fact that all three models exhibit a very low sensitivity to soil texture is in our opinion the key finding of this study.

15. Line 162 → For a given vegetation

Corrected.

16. Line 164 → I see that here you define for the first time the acronym PFT, but it was used several times before.

Corrected (see comment above)

17. Line 166 → A reference for Richards equation would be good here.

We added a reference for Richards equation.

18. Line 185 → With “current PFT distribution” do you mean a global ESACCI land cover map from 2015? From 2021? From 2022?

The land cover map corresponds to the year 2015 to be consistent with the maps from Avitabile et al. (2016) and MODIS data (2006-2016). We added:

“This land cover map corresponds to the year 2015, which is consistent with remote sensing products used for model evaluation described below (see section 2.5).”

19. Line 194 → What do you mean with “the most default”? Is it that you varied only few parameters from Table 1?

It means that we did not re-parameterize anything compared to the most recent version of the model, corresponding also to the model version commonly used by default in impact assessment or scientific studies that does not require new parameterizations. We rephrased to:

“Similarly, we did not change any model parameters (with the exception of the soil textural information) compared to the model default parameterization for the tropics.”

20. Lines 206-207 → I do not understand the sentence. Perhaps do you mean that you average either (i) the last ten years of the historical period (2006-2016) or (ii) the last year of the historical period (2016)?

We rephrased to:

"All the results from the vegetation model simulations presented below are either the averages of the ten last years of either the spin-up or the historical period (2006-2016) or the averages of the very last year of the historical period (2016)."

21. Lines 215-217 → The information of where to find the code is already in section Data and code availability, therefore, I suggest to remove it from the methods section.

Indeed. We removed it here and kept it in the code availability section.

22. I think Section 2 should be reorganized in a more straightforward way for understanding. For instance, you could start with a first subsection that contains a description of the study area (coordinates, land use/land cover, climatology, etc.). A second subsection could be the models' description. A third subsection could be a brief description of all datasets used in the study. In a fourth subsection you could integrate the simulation protocol, the soil scenarios and the model parameterization. And a fifth subsection could be the analyses.

We thank the reviewer for this suggestion. We re-organized this section as:

2.1 Study region and simulation protocol

2.2 Vegetation models

2.3 Soil scenarios and simulation protocol

2.4 Evaluation datasets

23. Figure 1B is not really showing the difference between the intra-grid cell and the inter-grid cell variability but both in the same plot. I would suggest either to rephrase this part of the caption (to something like inter (black) and intra (red) grid cell variability...), or to

really plot just one line that shows the difference between the two lines. Moreover, why does the legend include sd (mean) for the inter-grid cell variability and mean (sd) for the intra grid cell variability? What is “sd”? I presume is standard deviation, but the acronym is not defined.

We rephrased that part of the caption as:

“**Subplot B shows both the intra-gridcell (i.e. mean of the standard deviation (sd) of the clay content) and the inter-gridcell (the standard deviation (sd) of the mean clay content) variability as a function of the spatial resolution**”

24. Caption from Figure 1C. I suggest to remove the last sentence (“showing a clear shift toward larger clay contents in the Max. clay scenario”), as this should not be part of the caption, but part from the main text describing the results.

We removed that part of the sentence as it should indeed not be part of the caption.

25. Lines 238-239 → The sentence “The three soil scenarios were built on this intra-gridcell variability in soil texture...within each gridcell” is not part of the results but part of the methods. You should move this to the methods section.

We removed this sentence as it was already explained in the material and methods section.

26. Line 245 → Perhaps do you mean supplementary Figure S2?

Indeed, it was corrected.

27. Figure 2 and Figure S2 → I strongly recommend the authors to change these Figures. Instead of plotting the mean values of each model in the second row of the Figures, you could plot (for each model) the difference between the model output and the reference data. It perhaps will be interesting to see in which specific regions there is better/worse performance of the models. The way the Figures are displayed right now makes it very hard

to compare overestimation/underestimation of the models in regards with the reference data. Moreover, why is Figure S2 in supplementary if its results are as important as the results from Figure 2?

We agree with the reviewer that representing these data as the difference with observation is also interesting. But doing so would prevent us from showing the density distributions between soil scenarios which is actually the key finding of this study (low to no sensitivity of the TBMs to soil texture). We therefore preferred to keep those in the main text and add the maps of the differences as supplementary (supplementary Figures S2 and S4). Figure S2 is supplementary because the conclusions that we can draw from it are redundant with those that we draw based on Figure 2. So we would prefer to keep it in supplementary as well to avoid adding an additional Figure to the main text.

28. Line 264 → The first sentence does not specify differences in what.

We added:

"**Large differences in all investigated model outputs existed between models for the same soil scenario.**"

29. Line 266 → Which correlation coefficient did you compute? The methods section does not mention any correlation analysis, so this came as a surprise in the results section.

This was derived from the quantile regression analysis. We added:

"..., as illustrated by the quantile regression analysis (Figure 3)".

And in the material and methods sections, we now also added the reason why we choose to perform a quantile regression analysis:

"We used a quantile regression given the nature of the vegetation productivity response to SDI: for SDI close to 1, the GPP variability is high (other resources can limit GPP) while for SDI to 0, the GPP is necessarily low."

30. Line 266-271 → Can you please explain how can we infer these numbers from Figure 3? Or is this information not shown in the Figures and you computed it elsewhere?

We clarified those sentences by (i) moving the references to Figure 1 to its proper location (section 3.1), and adding references to the vertical and horizontal boxplots of Figure 3:

"Across the three scenarios, we observed that increasing clay content slightly increased drought stress (i.e. decreased SDI) by 2.6, 0.7 and 1.5% (change of the drought stress index from the Min. clay to the Max. clay scenario) for ORCHIDEE, ED2 and LPJ-GUESS, respectively (horizontal boxplots in Figure 3). This increase in simulated drought stress was accompanied by a decrease in productivity for all three models, respectively by 2.7, 1.9 and 3.2% (vertical boxplots in Figure 3)."

31. Line 272 → Please indicate the location of these gridcells and the reasons to select them.

The location of these few grid cells is spread over the Amazon basin and differs for each model. We included the locations of those gridcells in supplementary Figure SXX..

32. Line 277 → You mention here Soil moisture index. Is it the same of soil drought stress index? If so, you should refer to it in the same way throughout the document.

Indeed, we meant the Soil drought stress index, which we now corrected throughout the document.

33. Line 278 → After the sentence "aboveground biomass" add: (supplementary Figure S3).

We added this reference to supplementary Sigure S3.

34. Line 284 → Did you perform any significance test to say this? If not, you should write "We observed some substantial impacts…"

We did not. So we now use the word "substantial" rather than "significant".

35. Line 290 → Change from "most important" to "the most important".

Corrected.

36. Figure 6 → Could you provide the R2 for the goodness of fit and the slope of the fitted line? Also, I strongly recommend to change the markers for the scenarios, it is really hard to differentiate one from the other.

We added the equation of the fitted lines (and the order of magnitude of the r^2) and changed the markers for the different scenarios to better differentiate them.

37. Line 304 → I would omit the cross-reference here of Figure 1 as I do not see a strong reason to use it in the context of the sentence.

We removed the cross-reference to Figure 1.

38. Lines 315-316 –> Explicitly indicate that the supplementary Figure S6 is from Poggio et al. 2021. Perhaps something like: see supplementary Figure S6 from Poggio et al. 2021.

We modified the sentence as suggested.

39. Figure S5 → Explicitly indicate in the caption that Ksat is saturated hydraulic conductivity.

Added.

40. Overall, I think that the discussion section should not present cross references to Figures. All the Figures (both from the main text and from the supplementary material) should be properly described in the results section and the main message derived from them can be included in the discussion. Specially, Figure S5 is mentioned for the first time in the discussion, so please move it to the results section.

In the new version of the manuscript, all figures are now mentionded in the results sections.