Dear Dr. Exbrayat, anonymous reviewer, Dr. Alan

Thank you for sharing your time to review our manuscript. And I apologize to editors, and reviewers for delay in the submission.

We add the study motivations and discussions to the revised manuscript to define the significance of this study's findings. For the revision of statistical analysis (ANOVA), we avoid imbalance in the combination of RCPxGCMxGVM by cutting into dataset. We'd like to answer one by one in the followings;

The manuscript is relatively short and the language needs to be clarified in places, especially the abstract. This study would also benefit from some more details in the methods. For example, an equation for the cophenetic correlation coefficient is missing. Perhaps the authors could consider including a table with each GVM's main features that can be related to the results presented here: e.g. is Ra a fixed fraction of GPP? How is phenology handled? What are the reference NPP, VegC and SoilC to which projections are compared? etc. . . I would also like the authors to discuss their results in the context of their findings in Nishina et al. (2014) that use the same ensemble.

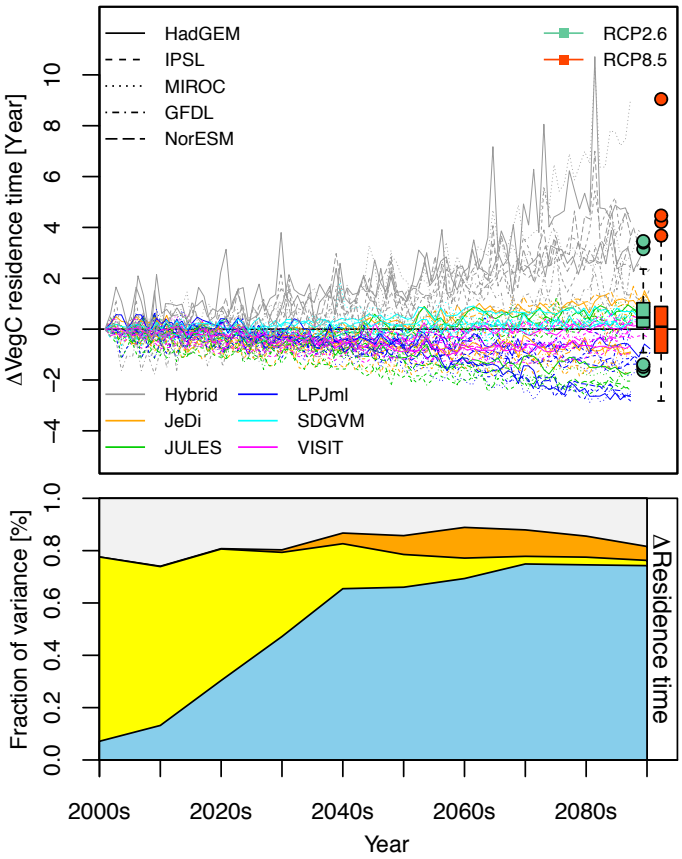
My main suggestion is that, although I agree with the main findings, I believe that some more analyses would strengthen the paper, such as reporting the changes in VegC and SoilC turnover times. The parameterization of turnover time is highly model- dependent and controls pool sizes at equilibrium (e.g. Exbrayat et al., 2014) because VegC transfer to litter / soil pools, and SoilC decomposition are represented using first- order kinetics. Therefore, (initial) pool sizes control the absolute response of C release following a relative change in turnover time in response to changes in environmental conditions. Therefore, part of the uncertainty in VegC and SoilC shown in Figure 1 may be partly attributed to differences in historical VegC and SoilC pools reported in Nishina et al. (2014). Studying changes in turnover times would enable to compare GVMs with a new approach.

Thank you for your kind suggestions. In revised manuscript, we add more detail information about ISIMIP protocols and the used models in this study. We agree your suggestions are lack of point of views in our manuscript.

For the SOC turnover time, regrettably, there are no soil respiration outputs in ISIMIP database due to the protocol. So, we can't conduct further analysis in this study. We add more discussion about SOC as follow;

> Moreover, differences in the initial SOC stock resulting from different spin-up procedures among GVMs critically contributed to the incoherence in SOC dynamics. In a CMIP5 study, Nishina et al. (2014) demonstrated that microbial decomposition processes are a dominant factor de- termining the amount of global SOC stock rather than C input from photosynthetic products. Determination of the initial SOC stock is important for future soil carbon stock and land surface fluxes (Exbrayat et al., 2014). In our results, there was no regional and ecosystem type (climatic divisions) dependency on GVM contributions to uncertainty in SOC changes. Therefore, to reduce GVM uncertainties in SOC projection, improvement of spin-up procedures and microbial decomposition will be effective for reduction of SOC uncertainties at both local and global scale.

For vegetation turnover time, we add a new analysis for vegetation VegC residence time according to the definition of Friend et al (2014). So, we rewrote the results and discussion secessions for this new analysis. The following figure shows the results for VegC residence time. We add the results and discussion about this results.



Specific comments:

p. 1200 l. 6: “Potsdam”

I revised it.

p. 1201 l. 14: Could the authors detail why they only use 70 simulations? From 6 GVMs, 5 GCMs and 4 RCPs, there should be 120 simulations available.

There are not available for all combinations of GVM, GCM, and RCP. In ISIMIP first track, the minimum simulation set is defined as 2 scenarios for RCP (RCP2.6 and RCP8.5).

From p. 1201 l.26 to p. 1202 l. 7: Isn’t SDGVM a DGVM? Could the authors

also indicate which fixed land cover data they used in the GVMs? How comparable are DGVMs and GVMs with fixed land cover?

We add the new table for general properties in this study's simulation settings. In this study, we didn't pay insufficient attention to the differences in land cover among GVMs and DGVMs. But, in the VegC residence time, we cannot see the any common trends among GDVMs (HYBRID4; increase, LPJmL: decrease).

p. 1205 ll.3-5: please rephrase this long sentence.

We divided two sentences here. Thank you.

p. 1206 ll. 2-3: historical simulations are not reported here.

We suggested the figure 1 in the end of study. But, as you suggested, we didn't applied the statistical treatments in this period. It's just judge on the appearance. So, we also added "seemingly" in this sentence.

p. 1206 ll. 8-9: "in previous inter-comparison of models"

We revised it.

I struggle to read Figure 3. It is a very complicated figure that deserves a more detailed description than the short paragraph 3.3. In a dendrogram, the definition of a cluster relies on the choice of a threshold in the similarity value, a subjective choice that has to be indicated and justified here. I agree that the four GVM-based clusters for SoilC are fairly obvious, but the number of clusters for NPP and VegC can vary several folds following slight variations in the threshold and it is hard to relate the description in paragraph 3.3 to Figure 3. Figure 4 would benefit from using another colour scheme with more contrast better extreme values (classical RGB?).

According to Reviewer #2 suggestions, we removed these analysis from the revised manuscript.

Dear Reviewer 2

First, we appreciate all of your comments and suggestions for this manuscript. We reconsidered the terms and English expressions through the entire manuscript.

I agree with the previous reviewer's general criticism that the paper is often vague and could do with more precise language and definition of concepts. This applies across the whole manuscript. Also, the methods are insufficient to understand and reproduce the work (for example but not limited to: how have the simulations have been implemented, presumably the clustering is applied to global annual data but this is not described, the authors say that 70 simulations have been made but by my count it's 72). I also agree that the manuscript lacks depth. For example, there has been very little examination of the causes of differences between the models and there is ample room to expand the discussion.

We had another look at our study to clearly define our goals and to deepen our understandings by expanding discussion part. First, we re-wrote in our aim of this study, as follow;

> In ecosystem climate impact assessments, how the uncer- tainties of climate impacts matter is still a challenging is- sue, in part due to the lack of standardized impact evaluation protocols. The Inter-Sectoral Impact Model Intercompari- son Project (ISI-MIP) is the first attempt to apply ensembles of both impact and climate models to obtain robust future assessments (Warszawski et al., 2014). In assessments of climate impacts on ecosystem functions, regionality is ex- tremely important for the severity and timing of impacts ow- ing to the different types of climate change in each region and the presence of different ecosystem types in different areas (Warszawski et al., 2013; Friend et al., 2014). For compre-

hensive climate impact assessments in ecosystems, it is necessary to possess spatial temporal information for which uncertainty sources can be chosen or ignored, for which some processes contributed to uncertainty, and for which how the contribution of each uncertainty source changed with time is known. Separation of the different sources of uncertainty in projections of ecosystem models in various aspects can be used to comprehend the uncertainties and risks in climate impacts on ecosystem conditions and C cycling.

The ANOVA is unbalanced as two of the RCP factor levels (4.5 & 6.2) only have runs featuring the 6 GVMs with a single GCM factor level, not the full 5 GCM factor levels. Results of Type II sums of squares are sensitive to unbalanced data, biasing the results of the ANOVA. I suggest dropping these two RCP levels from the analysis to maintain a balanced design. Also I don't think that the variance partitioning equation presented in the methods holds using Type II sums of squares. Type II sums of squares for a main effect is calculated once variance for all other main effects has been calculated so variance shared between the main effects is not accounted for. Furthermore, there has been no discussion that for each factor (RCP, GVM, GCM) the factor levels are drawn from a 'population' of possible factor levels. For example, why were these 6 GVMs chosen? How representative are they of all GVMs? These types of questions have been extensively investigated in the literature with regards to parametric sensitivity analysis ().

Thank you for this comment. As Dr. Alan also suggested in the open discussion, we agree the application of ANOVA to the imbalanced design data make the parameter estimation in interaction terms tend to be inaccurate in part. Although, our study didn't focus on the interactions effect of the uncertainty sources, we revised the way to apply ANOVA by eliminating RCP4.5 and RCP6.0. For the selection of dataset, we add the detail database

information as a table in supplemental material and the descriptions.

While the clustering analysis is an interesting approach to analysing an ensemble of this nature, there has been no strong justification of why the particular method is nec- essary and why it is being used. I am unclear exactly what the clustering is doing. Exactly what similarity is the clustering based on? The analysis of Rouyer, cited by the authors for the wavelet clustering method, was used to analyse the temporal dy- namics of fish populations at multiple frequencies. Are the authors of this paper really interested in the temporal dynamics of NPP and carbon stocks at multiple frequencies? Which frequencies are this clustering technique picking out? Isn't the overall trend more important? The cluster analysis seems over-complicated with little sound justification and no real useful information pulled out and discussed by the authors.

We removed cluster analysis in revised manuscript. From cluster analysis, we aimed to reveal the time-series similarity of global trends of the variables (not only trend, but also episodic events). But, the almost message from this analysis in our results are common with the results of ANOVA and other visible trends. But, the visualizations of cluster analysis made little bit confusing. So, we removed this analysis in the revised manuscript.

I suggest much stronger definition of the research questions, collaborating with a statistician to help devise the appropriate statistical analyses to answer these questions, and base your methods more strongly in the literature.

Thank you for your suggestion. We clarified the aim of this study as follow;

Our objective was to explore the comprehensive uncertainties in future global and regional terrestrial C projections by decomposing the uncertainty sources to time, space, and processes.

p1199 ln15-17 I don't really understand what this means and I disagree with what I think you're trying to say. If GVM uncertainty dominates then we don't have sufficient understanding of terrestrial processes.

I re-wrote the abstract in the revised manuscript. I deleted this sentence. Instead of this sentence, we add the followings;

Our results suggest that to assess climate change impacts on global ecosystem C cycling among each RCP scenario, the long-term C dynamics within the ecosystems (i.e., vegetation turnover and soil decomposition) are critical factors rather than photosynthetic processes. The different trends in contribution of uncertainty source in each variable among climate divisions indicate that improvement of GVMs based on climate division or biome type will be effective. On the other hand, in dry regions, GCMs are the dominant uncertainty source in climate impact assessments of vegetation and soil C dynamics.

p1200 ln1 The climate system doesn't cycle C, the Earth System or the biosphere does. Also terrestrial ecosystems don't play a role in ecosystem services they provide them. This kind of imprecise language is used throughout the manuscript and the manuscript would really benefit from substantial editing for precision and clarity.

Thank you for your kind comments. We re-consider all of our expressions throughout our manuscript.

p1200 ln6 Potsdam, not Postsdom. Was this really the name of the Sitch inter-comparison?

I revised it.

p1200 ln 26 What do you mean “phases”? p1200 ln26&7 I really struggled with this sentence.

I revised this phrase as follow;

> Thus, various uncertainty sources may result in the divergence for projected C cycling.

p1201 ln1 What experiences might be beneficial, be precise, spell them out. How are they relevant to this study?

We add more explanation as follow;

> For example, recently, the likelihood of the occurrence of large Amazon dieback in this century has become lower in simulation studies (Cox et al., 2000; Sitch et al., 2008; Cook et al., 2012) because of reduction of uncertainties in the projected precipitation in Amazon regions among GCMs (Sitch et al., 2008; Poulter et al., 2010; Cook et al., 2012). However, the improvement of vegetation processes in this region could result in the improvement of local vegetation–climate feedbacks, which might contribute to changes in temperature and precipitation in this region (Shiogama et al., 2011). At the global scale, in earth system models in the CMIP5 study, the sensitivities in global land climate–carbon feedback varied considerably (Arora et al., 2013). The reduction of C budget uncertainties in ecosystem models

could serve to reduce climate change uncertainties, particularly regarding the climate sensitivity of earth system models.

p1201 ln9. Four of these GVMs were part of ESMs in CMIP5? Which ones, to my knowledge it's only JULES but I couldn't find out if VISIT was also used in a CMIP5 ESM.

We delete this phrase, because of my misunderstandings.

p1201 ln21. I think you examined changes in these variables, again be more precise.

Thank you. We suggested the "changes" in this sentence.

p1202 ln2. As far as I can tell there are not 5 GCMs x 4 RCPs. Looking at Fig 1, only

We clarified the available simulation results in the supplemental table. And we add the more description for the selection of results in material and methods. Furthermore, we introduced ISIMIP database and the URL in the revised manuscript. Because of minimum requirement in ISI-MIP study, there were incomplete sets of simulation data are only available. In the revised manuscript, we clarified this information in the supplemental table.

Dear Dr. A. Hewitt

Thank you for your critical comments to our manuscript. We changed the way to application of ANOVA for imbalance issues and internal variance of GVM by

I have some concerns about the ANOVA methodology applied in the paper.
1) The description of the ANOVA methodology within the paper is brief and it would not be possible for your method to be reproduced with only the information contained in the paper.

This is because available simulation sets in ISI-MIP. We add the supplemental table to show the simulated dataset in ISI-MIP data archive.

2) The methodology developed in Yip et al 2011 is only suitable for a balanced ANOVA. Assuming one run for each GCM*GVM*RCP combination this would require 120 simulations whereas 70 are used in this paper.

I add the more descriptions of ANOVA in the revised manuscript to make sure the use of ANOVA. Thank you.

3) Internal variance is a non-negligible term and needs to be included in your ANOVA analysis. In order to quantify this, you will need to include multiple runs of at least some of your GCM*GVM*RCP combinations. Internal variance has been long observed, even in a stationary climate (see Madden 1976; Karoly and Wu 2005).

We agree the internal variance of climate variables affected the estimated variance of each uncertainties source in our manuscript. Booth et al. (2013 in ESD) demonstrated internal variance of GCM contributed to global mean temperature up to 20% especially at early 21st century. To avoid the effect of inter-annual variance in climate variables to GVM, we used 10-years average values for ANOVA in the revised manuscript.

A recent publication (Hingray and Said 2014) contains an ANOVA approach for unbalanced data, which would appear to be suitable for your data. You may wish to develop an alternative ANOVA method for unbalanced data, in which case the onus would be upon you to prove it is a robust method.

Thank you for kind introduction. We recognized the estimation of parameters in interaction terms. So, in the revised version, to avoid imbalance issue, we removed RCP4.5 and 6.0 dataset for the application ANOVA.