

## **REVIEWER 1**

### General Comments:

This analysis of carbon cycle uncertainty for Alaska is an extension of previous research. Its strength lies in bringing together 40 models that contributed to earlier NACP, TRENDY, and WETCHIMP analyses, and to a more limited extent subjecting a few of those models to site-specific comparisons against flux tower data from sites at Barrow and elsewhere.

As a result there are two useful lessons to be learned. First, the sensitivity of NEE to CO<sub>2</sub> fertilization and climate is compared with the conclusion reached that climate is the more uncertain of the two. It is not clear, however, what component of climate might be dominating this dynamic.

Second, is that mean annual NEE shows no consistent spatial pattern among the various models evaluated. It was surprising not to see this mentioned in the Abstract. My preference would be to see this result described in the Abstract and not the site-specific comparisons to AmeriFlux. This is especially true since the Discussion section concludes with an observation that these analyses could be used to guide future field data collection efforts. This would be a valuable contribution, but it must be mentioned how this could be done and/or what the recommendations might be coming from that analysis.

- **We thank the reviewer for an accurate assessment and appreciation of the strengths of this paper and the lessons learned.**
- **The reviewer suggested that we mention the inconsistent spatial pattern in NEE among the models in the Abstract. We had included that sentiment in the Abstract already with, “The spatial patterns in regional carbon stocks and fluxes varied widely with some models showing NEE for Alaska as a strong carbon sink, others as a strong carbon source, while still others as carbon neutral.” However, we modified that sentence to clearly indicate the “no consistent spatial pattern” as worded by the reviewer.**
- **We appreciate the reviewer’s keen recognition that this paper can be a valuable contribution for field data collection efforts. The reviewer suggested that we mention how this could be done and/or what recommendations we could offer. We had already included the recommendation: “For CARVE, ABoVE, and NGE Arctic, in particular, these campaigns must sample the geographic regions that encompass both the greatest representativeness and the greatest uncertainties. Our uncertainty maps alone provide a guide for campaign sampling location strategy.” We added a more discussion to help guide these recommendations with more quantification and specific geolocation.**

### Specific Comments:

Abstract: Although it is mentioned that an analysis of “structural and parametric uncertainty” was conducted, this is not how results are presented in this section. It is possible that the multi-model standard deviation against the mean is a measure of total uncertainty for a given quantity and not structural and parametric separately.

- **We did not mean to say that we were evaluating parametric uncertainty among individual models, but that model output was a manifestation in part due to parametric uncertainty. Thus, while we focus primarily on structural uncertainty, parametric uncertainty was integrated but not delineated. We are sorry for the confusion, and we have clarified the language on this in the text.**

Also, it is unclear whether these flux and stock estimates are the only sources of uncertainty evaluated or were there others.

- **We added “that follows” at the end of the second sentence to clarify that these were the quantities we assessed (and not others).**

It should also be emphasized that baseline means results from a transient simulation from 1991 to 2009, not uncertainty estimate for 2100 or beyond.

- **We had indicated the “analysis was conducted of 20<sup>th</sup> century...” to indicate that we did not assess 21<sup>st</sup> or 22<sup>nd</sup> centuries. However, following the reviewer’s suggestion, to clarify we added the exact years (1901-2010) to the abstract.**

The title and Abstract will lead readers to believe this is a new and unique analysis. Rather, while it is interesting, it is a future compilation and analysis of previous (NACP, TRENDY, and WETCHIMP) model inter-comparisons.

- **Agreed. We originally had language in the Abstract specifying NACP, TRENDY, and WETCHIMP, but journal word limitations required us to cut that text. We added that text back in, and will leave it to the editors to decide whether or not to keep the text.**

Introduction: This is a reasonable presentation of previous research and results that come from other model inter-comparisons. Thus, this section sets a good stage for readers. It does seem to capture uncertainty in broad terms, but hesitates to dip into the details of sensitivity analyses, uncertainty quantification, etc.

- **We thank the reviewer for an accurate assessment of this section. True, we avoided background discussion literature on uncertainty quantification and sensitivity analyses. This was done to streamline the text and allow the reader to more quickly move through the paper. We added text to better tie the objectives to the sensitivity analyses.**

Materials: Rather than a new community coming together for this model inter-comparison, it appears that it compiles previous results from NACP, TRENDY, and WETCHIMP. The rationale for this is not exactly given, but should be stated in the Introduction. It will be interesting to see if and where the larger analysis differs from the three individual NACP, TRENDY, and WETCHIMP efforts.

- **We added clarifying text to the Introduction as suggested to note that these MIPs were not focused solely on Alaska.**
- **Indeed, we also thought it would be interesting to see if the different MIPs cluster in results. We included that in the 3<sup>rd</sup> paragraph of the Discussion.**

Results: Although expectations were otherwise, the breadth of the uncertainty analysis conducted here is rather limited to carbon flux and stocks. Some interesting results were presented that were not emphasized or mentioned in the Abstract.

- **We have the results also for quantities other than carbon—i.e., water and energy. We did not include those results because we thought that this would be an overwhelming amount for the readers to assimilate; thus, we streamlined the paper only to the carbon cycle, which is still very lengthy!**
- **We tried to capture the wide range of results within the word limitations of the Abstract, but apologize if we could not fully capture everything. If there is some result in particular that the**

**reviewer would appreciate seeing added to the Abstract, then please add another online comment and we will try to accommodate that request.**

Discussion: Good discussion of soil carbon stocks, their uncertainties, and comparison to field measurements and previous summaries. These differences do seem to be striking leaving it an open question as to the explanation.

- **Again, the reviewer provided an accurate assessment of this discussion, and we thank the reviewer for the comment. Soil carbon modeling is an active field of research right now, and hopefully we will see improvement and convergence in model estimates in the near future.**

It is interesting in that given all the directions that this analysis could go, the emphasis was placed on the relevance of these data to future field campaigns. Not sure what exactly motivates that decision especially given the poor agreement among models in spatial representation of processes. However, assuming that the analysis of spatial patterns of carbon flux and stock uncertainties would yield useful information for field studies, the question becomes why are those maps not presented and where are those recommendations? Those alone could be quite valuable.

- **The poor agreement in model estimates is exactly what motivates that decision in informing field campaigns to help constrain those divergent estimates. We added text to better contextualize this analysis relative to other analyses. As mentioned above, we included the recommendation: "For CARVE, ABoVE, and NGEE Arctic, in particular, these campaigns must sample the geographic regions that encompass both the greatest representativeness and the greatest uncertainties. Our uncertainty maps alone provide a guide for campaign sampling location strategy." The maps were presented in Figures 2-5 as well as in the supplementary information.**

Conclusion: Good summary of the major objectives of this project and its findings.

- **Thank you.**

## **REVIEWER 2**

This paper compares the results of a range of models for Alaskan carbon cycling. The goal is to set a benchmark, quantifying the predictive uncertainty in current models.

This paper very closely duplicates the results from a previously published paper: McGuire, A., Christensen, T., Hayes, D., Heroult, A., Euskirchen, E., Kimball, J., Koven, C., Lafleur, P., Miller, P., Oechel, W. C., Peylin, P., Williams, M., and Yi, Y.: An assessment of the carbon balance of arctic tundra: comparisons among observations, process models, and atmospheric inversions, *Biogeosciences*, 9, 3185-3204, 10.5194/bg-9-3185-2012, 2012. The authors of this submission argue that McGuire et al. includes only the results of 3 models. But this is not the case: the McGuire et al. paper includes many other model outputs from the TRENDY project, which are included in this new submission (see their table 7). This duplication is highly problematical. The McGuire et al. paper is a more complete product, as it summarises the results from inversion models, and from flux observations, alongside a multi-model comparison; it also covers the entire pan-Arctic, with a breakdown to sub-regions that includes North America. The McGuire et al paper includes detailed discussions of uncertainty. This paper needs to be completely rethought and rewritten to emphasise its novelty.

- **Both the McGuire et al 2012 paper and this paper were written at the same time (this paper has had a delayed review process), and with pre-draft manuscripts sent to both groups of authors, as well as with multiple authors included in both papers. The papers were constructed to be complementary, not redundant. The focus of the McGuire et al paper was on pan-Arctic carbon budgets. The focus of the current paper in review is on Alaskan carbon uncertainties, sensitivities, and spatial patterns. While certainly converging on some methods (i.e., terrestrial biosphere models) and broad topics (Arctic carbon cycling), the two papers are distinct and both add value to the scientific literature. Inasmuch as one could argue that the McGuire et al paper is redundant with any other global scale paper that focused on carbon budgets, the McGuire et al paper is still valuable in its regional application.**
- **The current paper in review is even more applied still: current and planned field campaigns for Alaska, as described in both the Introduction and Discussion, require a more direct quantification of carbon budget uncertainties across the carbon cycle for Alaska as a single domain. With a different focus, McGuire et al provide uncertainties only for net CO<sub>2</sub> and CH<sub>4</sub> exchange. The current paper in review provides uncertainties for 4x as many carbon cycle variables. To more fully encapsulate model uncertainty, the current paper includes a much larger community-inclusive 4x number of terrestrial biosphere models by incorporating the work of both the NACP Regional and Site Syntheses, in addition to TRENDY and WETCHIMP. The current/upcoming field campaigns require justification for geographic sampling. The current paper in review provides clear maps of uncertainties. The McGuire et al paper provides no maps (other than the pan-Arctic domain boundaries). To understand future uncertainties with respect to both CO<sub>2</sub> and climate change, the current paper in review provides such an analysis for century time scales; the McGuire et al paper does not. Finally, to understand uncertainties in model estimates further, the current paper in review provides comparison to *in situ* measurements at AmeriFlux sites; the McGuire et al paper does not. Our point of noting what the McGuire et al paper is lacking is not to imply any deficiencies in the McGuire et al paper; rather, it is simply to illustrate that the two papers are different and complementary.**
- **While we feel we have made our argument here in review response, we understand that we could have made these points clearer in the paper so as to avoid the understandable confusion that the reviewer was confronted with when reviewing our paper. Hence, we provided a new**

paragraph summarizing these points in the Introduction. We also corrected the number of models McGuire et al used.

### **REVIEWER 3**

The main aim of this paper is to estimate the recent carbon balance and its uncertainty of the Arctic in Alaska. To achieve that a multitude of bottom-up model results, mainly from prognostic terrestrial biosphere models have been used.

In the present form I cannot agree with a publication of the manuscript based on two main points A and B: A) Comparing the manuscript with the already published carbon balance and uncertainty estimation by D. McGuire and colleagues (Biogeosciences 9, 3185-3204). I cannot see the added value of the manuscript that would merit a publication. My reasoning is explained in more details in points 1 and 2 below.

1. When focusing on a small region like the Alaskan Arctic it makes no sense to me to include as many models as possible into a statistical analysis of result regardless of the processes that are represented by these models. In fact, results by models that include permafrost-specific processes should be more valid. Hence, I would strongly recommend grouping model results according to process representations. McGuire et al. also include TRENDY model results into their analysis but differences to the more appropriate models for this region are visible and discussed.

- **This is a good and valid point, and we had considered it previously. The reason why we did not create model physics-skill groupings is because those group cut-offs would be subjective. However, to continue this thought process further, we can do a simple comparison of the few models that claim to have advanced permafrost processes—i.e., TEM6, CLM4-CN, and ORCHIDEE. If we look at Figure 2, we see that there is no convergence among any of the three models: TEM6 shows a carbon sink in the northeast and source in the southeast, ORCHIDEE shows the opposite of TEM6, and CLM4-CN has Alaska as largely carbon neutral. We have now included that thought exercise in the Discussion.**

2. Using both, bottom-up and top-down approaches, McGuire et al. provide a more comprehensive analysis of the Arctic carbon balance and its uncertainty. McGuire et al. provide a break-down into regions with North America being the smallest kind of scale. It needs to be much more motivated what do we need a new paper with a particular focus on Alaska only?

- **Please see response to Reviewer 2 on comparison to McGuire et al 2012.**

B) There is a general mismatch between different sections of the manuscript (abstract, introduction, results, discussion, conclusion) in terms of aims of the paper and analyses done. See my comments 1-3 below.

1. The objective of the paper was to identify structural versus parametric uncertainty of the models (introduction). Maybe I overlooked a substantial part of the results but I cannot find this kind of analysis. I would assume that parameter uncertainty is assessed by a kind of Monte-Carlo simulation run, maybe at site level? Instead, the authors discuss the uncertainty coming from different forcing data and using a different spin-up procedure in comparison to structural differences.

- **We apologize for this confusion, as pointed out also by Reviewer 1. We did not mean to say that we were evaluating parametric uncertainty among individual models, but that model output was a manifestation in part due to parametric uncertainty. Thus, while we focus primarily on structural uncertainty, parametric uncertainty was integrated but not delineated. We have clarified the language on this in the text.**

2. The beta-gamma-analysis of results is not motivated in the introduction and is not included into the discussion, e.g. compared to already done analyses of this kind.

- **We have added text to the Introduction and Discussion tying the beta-gamma analysis in more centrally to the objectives.**

3. The aim of the site-level comparison is unclear. Was the aim to show the reliability of different concepts and assumptions (model structure) or different parameter values? Was the aim to use site-level evaluation results for a weighted average of regional-scale carbon balance results?

- **We have added text to clarify how the site level evaluation integrated into the objectives and how to understand the results within the larger context.**

4. Most of the conclusion is a repetition of aims and methods but no conclusion about the main objectives of the paper is given.

- **We have ensured that the conclusion summarizes the main findings as they pertain to the objectives.**

### **SHORT COMMENT BY ANANTA DAS**

Overall Comments: The analysis of carbon cycle uncertainty for Alaska is based on the outputs of 40 models. It brings together multiple models which were a part of NACP, TRENDY, and WETCHIMP analyses. One of the fundamental flaws of this study is, it attempts to compare different carbon flux variables that were generated from heterogeneous data sources. The forcing datasets were different for different models and this will be reflected in the carbon flux variables. Model intercomparison is valid when all the models are simulated with a homogeneous set of data and at the same spatial and temporal resolutions.

- **We thank Ananta Das for providing unsolicited comments and suggestions. The forcing datasets were common for the TRENDY models and the NACP Site Synthesis models, with some variability for the NACP Regional Synthesis models. The 3<sup>rd</sup> paragraph of the Discussion questions whether or not this is indeed an issue, and, at least for this study, it appears not to be an issue.**

Abstract: The authors mentioned autotrophic respiration and ecosystem respiration separately. To my knowledge, ecosystem respiration covers both autotrophic and heterotrophic respirations. This is confusing. Either authors need to mention about the results of heterotrophic respiration and autotrophic respiration separately or it should be combined into ecosystem respiration.

- **We apologize for this confusion. We could certainly remove Re from the Abstract, but Re is an important quantity for many scientists. This is not unlike the other combined carbon fluxes, such as NEE and NPP, which are important integrating quantities in and of themselves.**

The authors did not mention any detail on the spatial resolution of the individual models and what kind of forcing data were used to run those models. Despite the authors used the model outputs only, but this is important for the readers to know at least the basic information on what are the spatial resolution of the individual models and what forcing datasets were used to generate the carbon fluxes. Otherwise it appears extremely vague study.

- **There were many models used in this study, and we needed to strike a compromise between providing detail on each model vs. the length of the paper and providing references for readers to look up for more information on each individual model. We discuss spatial resolution in the Methods when describing the spatial downscaling, and again in the Results when we specify the range in spatial resolutions. The 3<sup>rd</sup> paragraph of the Discussion includes text on the forcing data (e.g., CRU+NCEP for TRENDY).**

The uncertainty analysis is not clear. Only simulated outputs from different sources were used. How the authors introduce uncertainty? This is misleading.

- **We introduce uncertainty, for example, in the 2<sup>nd</sup> sentence of the Abstract: "...uncertainty, defined as the multi-model standard deviation ( $\sigma$ ) against the mean ( $\bar{x}$ ) for each quantity."**

In section 2.1 it is mentioned that the model outputs were downloaded from some sources, but in page 2896 (line 15 – 20), it is mentioned that the TRENDY models are forced with CO<sub>2</sub> alone and forced with varying CO<sub>2</sub> plus climate. There is a mismatch between section 2.1 and here. Section 2.2 (Page 2897, line 1 – 5). Contradictory. Here it is said that the model outputs were downloaded. Here again it says that the models were forced with in-situ measurements. Given the nature of analysis it appears that the CO<sub>2</sub> flux outputs were used from different sources. How the models were run with the in-situ data in that case? Section 3.5 (Site level evaluation). Here you said that the subsets of models were run using in-situ

forcing data. This implies you have actually run all the models. But in section 2.1 you said that you used model outputs and the outputs were at different spatial resolutions. This is contradictory.

- **We downloaded all model output as contributed to the MIPs. We explain some of the model run details for those MIPs and provide references for further information. We did not run all the models ourselves (though this would definitely be an ambitious aspiration).**

Figure 2. Very poor spatial scaling. It does not make any sense. How did you scale the CO<sub>2</sub> flux outputs to 0.5 degree from different resolutions? Downscaling is a different technique. From the Figure, it appears that a linear scaling was done. As a result the appearance of the spatial distribution does not agree among models.

- **Downscaling was not done for Figure 2. It was done for Figure 3. This was also done for any other map or analysis that blended models together.**

Some models show CO<sub>2</sub> sink where as some shows CO<sub>2</sub> source over the same areas. This is not appropriate.

- **This is a good observation, and one of the main conclusions of the paper!**

The reasons behind showing 2003 results in Figure 3 and 5 are not understood.

- **We showed results for 2003 for example in the maps because the maps are similar for other years.**

Figure 4. Very poor.

- **This is another good observation, and another conclusion of the poor agreement among models.**

Figure 6. How many years of mean were taken? Spatial standard deviation does not make any sense here. In my opinion this paper lacks the quality that is required for Biogeosciences.

- **Interannual soil carbon does not change significantly within a model relative to among models. Figure 6 shows the annual mean for 2003 for consistency with the maps. The spatial standard deviation is interesting because of the extremely wide variability for such a small area.**
- **We hope that our revisions have improved the quality of this paper sufficient for Biogeosciences.**