

## ***Interactive comment on “The North American Carbon Program Multi-scale synthesis and Terrestrial Model Intercomparison Project – Part 1: Overview and experimental design” by D. N. Huntzinger et al.***

**Anonymous Referee #1**

Received and published: 24 September 2013

### General comments

This paper provides a description of the MsTMIP effort and general design of the experimental protocol. As part of a multipart publication, it is a useful document to allow readers a more efficient way to focus on aspects of a large intercomparison such as this instead of wading through a single large paper. It is well-written (increasingly rare!) and well-organized.

From intercomparison always comes the inevitable issue of assessing if some models are “better” or perhaps contain algorithmic components that can be deemed “better”

C1529

or “more complete”, etc. Comparison to observed quantities is often what many think of as the metric by which this is determined, but this isn’t always clear because sometimes observations and model simulated quantities are not directly comparable. There may also be underlying diagnostic tests that, though not directly tied to an observable, are considered metrics to test model “realness”. In atmos models, things like interhemispheric gradients of a conserved quantity might qualify. There may be similar quantities in TBMs that are emergent yet identify lack of “realness” or suggest internal inconsistencies. Worth thinking about and if the authors have such metrics, worth writing about in this paper (maybe just a para).

The authors chose to share “preliminary results” in this manuscript. Though interesting, it might be unclear what the rationale for this is, especially given that the paper has the stated aim of providing an overview and description of experimental design. I do appreciate prelim results, but the question is where to draw the line? Why this set of preliminary results? Why not more or different? Are they important to the narrative about the design? It might help to motivate why you are showing these here and why this particular suite of results.

My general recommendation is for publication with minor revisions. I list some of these minor concerns below.

### Specific comments

I don’t see how the quantification of structural differences as contributors to across-model variability will contribute to evaluation and feedback to improve the state of the art (page 3982, lines 20-25, paraphrased). It seems to me that you should be a bit more upfront about the differences between understanding the variability that emerges (tracing it to structural differences or parameter choices) and determining what choices, approaches, parameterizations are “better”. The latter is challenging in more ways than one (note under general comments), but I think intercomparisons have to be somewhat “tough” in this regard, particularly since you will avail of observations that will evaluate

C1530

the output of the models (and that is key). In other words, don't be too shy or overly diplomatic about that fact that some models will not perform as well as others against observations. This doesn't always mean some are better and some are worse, but it is this upfront comparison alongside the structural understanding that will lead to model progress, I believe.

Page 3985, lines 5-14: It can also lead to observing systems better-tuned or optimized to test models – not to be underestimated as an added benefit of model-obs comparisons. Vice-versa is true: ensuring models generate variables that can be directly compared to observable quantities is an important goal in this subsection text.

Parameter variation is mentioned quickly on page 3984 (near top) and one wonders if there will be some form of systematic exploration of the parameter space in addition to structural space? In this way you can generate an RMS of the internal uncertainty and the external uncertainty (model spread).

Technical corrections

Page 3984, line 18: I think the spelling should be “rationale”?

Page 3984, line 23: “the” at the end of the line should be removed.

Page 3990, line 3: the sentence starting on this line is run-on. Break into two?

Page 3990, line 23: grammar, perhaps “the” should be removed?

Page 3992, line 1: perhaps “but inter-model differences remain, particularly. . .”?

---

Interactive comment on Geosci. Model Dev. Discuss., 6, 3977, 2013.