[Note that the reviewer's comments are shown below in regular text and the authors' response to the reviewer's comments are shown in **bold**]

Comments from Reviewer #2

This paper is the first in a series describing the MsTMIP. It describes the MsTMIP experimental design, the participating models, and some preliminary results. Overall, the paper is well structured and easy to read.

I disagree with the first reviewer regarding the inclusion of preliminary results. It was very clear to me from the text why the authors chose to include these specific preliminary results. The results show that the introduction of a more standardized simulation protocol significantly reduces but does not eliminate inter-model spread.

The basic experimental design itself is also well-thought through. As this is the focus o his paper, I have few complaints and recommend that it be published with only minor revisions.

Specific comments:

p 2986, line 8 - While not global, there are gridded observations of aboveground biomass carbon covering the tropics:

> Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T., Salas, W., ... & Morel, A. (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. Proceedings of the National Academy of Sciences, 108(24), 9899-9904.

> **We added "regional, gridded observations (e.g., aboveground biomass)" to the list of possible observations on page 2986. The Saatchi et al. (2011) product, as well as the IPCC-Tier 1 Biomass product are part of the MsTMIP benchmarking datasets.**

Follow-on papers will compare the strengths and weaknesses of the various models and relate these to structural differences. Why have the authors decided to not include at least some parametric sensitivity analyses? Or for that matter, model comparisons against experimental observations (FACE, throughfall exclusion, irrigation/fertilization)?

> **As mentioned in response to a comment made by Reviewer # 1, parameter value variation was not controlled for in the MsTMIP experimental design, thus a systematic exploration of parameter space is not feasible. In addition, very few, if any, common parameters or parameter values are shared across all participating models, further making an across-model parameteric uncertainty analysis challenging (if not impractical). However, if a modeling team chose to do so, a parametric sensitivity analysis could be done for a given model to provide information on within-model sensitivity. This exercise, however, would fall outside of the formal MsTMIP experimental design.**

> **This is partially discussed in the original manuscript, in the second paragraph of Section 2.1:**

"Building off of lessons learned from the NACP interim synthesis activities, one of the primary goals of the MsTMIP activity is to quantify and assess the impact of TBM structural variability, and therefore uncertainty, by examining how inter-model differences influence variability among model results. Structural uncertainty results from differences across models in their representation (or lack of representation) of biogeochemical and biophysical processes. Although structural uncertainty can be quantified, in part, for a given model through a series of sensitivity simulations, it is best quantified through a MIP. To do so, however, a large ensemble of models is needed to span the range of biogeochemical and biophysical process representations, and the simulation protocol must isolate structure, at least to the extent possible, by holding constant as many aspects of simulating the terrestrial carbon cycle as feasible except for the models themselves. Using a MIP in this way cannot truly separate parametric and structural uncertainty, as different model structures will have different parameters, and models with the same process representations might use different parameter values. Nevertheless, if properly designed, observed inter-model differences can be representative of structural difference in model representation, and thus complement parametric uncertainty analyses applied to a single model."

While we feel that the paragraph above adequately addresses the reviewer's comments, we did add some clarifying wording to the end of Section 2.3, which reads:

"Parameter values are not specified as part of the MsTMIP experimental protocol, thus all models are run with their model-specific parameterizations."

In terms of the comparing model estimates to FACE experiments, all of the MsTMIP simulations were performed using ambient CO_2 concentrations, and therefore comparison with FACE experiments with the current MsTMIP output is not possible. Similarly, it is unclear how the MsTMIP output could be meaningfully compared to observations from throughput exclusion and irrigation/fertilization experiments without additional sensitivity runs/simulations that would fall outside of MsTMIP's current experimental design.

What is the data management plan for the MsTMIP simulations? Will there be a central repository?

Data management support for the MsTMIP output, along with information on simulation output repository location were mentioned in the Acknowledgements of the original manuscript:

"Data management support for preparing, documenting, and distributing model driver and output data was performed by the Modeling and Synthesis Thematic Data Center at Oak Ridge National Laboratory (http://nacp.ornl.gov), with funding through NASA ROSES Grant # NNH10AN68I. Finalized MsTMIP data products will be archived at the ORNL DAAC (http://daac,ornl.gov)."

In addition, we added the following wording to the end of the first paragraph of Section 3:

"Finalized MsTMIP data products will be archived at the ORNL DAAC (http://daac,ornl.gov)."