Authors' Response to Comments of Reviewer #2

Thanks very much for the valuable comments. We've seriously considered each comment and taken actions to address them. Below are details of our responses:

1. Response to "*The authors need to characterize the driver data prepared for the global domain and the North America domain in detail. According to Huntzinger et al. (2013), one of the purposes for designing two simulations at different domains is to test the influence of both spatial resolution and changing driver data on model estimates. For the purpose, it would be informative if the manuscript includes descriptions about similarities and dissimilarities between the two driver data and their potential impact to model results. Specifically, I recommend to include comparisons of the long-term trend and spatial distribution of air temperature, precipitation, and shortwave radiation of the two driver data."*

   Thank you for the comment. In order to address the reviewer's comment, in Section 3.1.3 and Supplement 2: Comparison of global and North American climate data, we compared (1) the monthly time-series and trend and (2) spatial distribution of long-term monthly mean of surface air temperature, precipitation, and downward shortwave/longwave radiation between global and North American climate data. Specifically, we analyzed the "decrease-increase" trend in global climate data (NCEP-based), original NARR climate data (before MTCLIM calibration), and improved NARR climate data (after MTCLIM calibration). We found that both original and MTCLIM-rescaled NARR downward shortwave radiation demonstrates a decreasing trend in the 1980s and increasing trend after 1990, which agrees with the findings reported in Wild et al., 2005 and Pinker et al., 2005. However, this "decrease-increase" trend pattern was not observed in the CRU-NCEP data.

   The detailed results of this analysis are provided in Supplement 2.

2. Response to "*The authors need to discuss similarities and dissimilarities in environmental driver data compared with other model intercomparison activities, such as VEMAP, ISI-MIP, and TRENDY GCP. I'm surprised that the authors completely neglected to mention how driver data were prepared in predecessor projects, VEMAP and NACP. Discuss lessons learned and improvement in data preparation from those projects. Also, please check how driver data were prepared in Trendy GCP, for which the data choice is very similar to MsTMIP.*"

   According to the reviewer's suggestion, we revised the manuscript so that in each environmental driver data category, we not only discuss available data sources, but also discuss those data that have been used in past and on-going MIP activities, including ISI-MIP, Postdam NPP MIP, and GCP-TRENDY. We also briefly discuss why we chose or didn't choose these data for MsTMIP (e.g. paragraph 1&2 in section 3.1.1, paragraph 1 in section 3.8, and paragraph 1 in 3.9.1). We feel this is a way to showcase (and acknowledge) how other MIP activities inspired MsTMIP on its environmental driver data selection and preparation.

   This manuscript is a companion paper to a manuscript published in GMD last year (Huntzinger et al., 2013), which provides the general framework for and reasoning behind the MsTMIP experimental design. In the introduction of that paper,

we discuss how the MsTMIP activity was built off of, and designed to complement, past, recent, and ongoing synthesis or MIP efforts. The development of the MsTMIP experimental design and its environmental driver data sets relied heavily on findings from past MIPs. VEMAP was a pioneer MIP activity started in the 1990s and many of its ideas and findings have become de facto standards (e.g. common driver data sets, spatial/temporal/physical consistency among driver data, and sensitivity analysis) for later MIPs. Because part I (Huntzinger et al., 2013) of this companion paper set talked about the influence of past and ongoing MIPs on the design of the MsMTIP activity, here, in the revised manuscript, we have focused on adding text that show how knowledge from past MIPs helped guide the choice of, and processing steps applied, to the environmental driver data (e.g. paragraph 2 and 5 in section 1).

3. Response to "*Though I understand the importance of sharing experience, Lesson learned section is not very informative in this context. There would be other suitable opportunities to share your opinions. Instead, a thorough characterization of the driver data would strengthen the aim of the paper. So, I recommend removing or shortening this section (if you insist).*"

   We think the "lessons learned" are valuable for data-intensive and multi-partner modeling activities like MsTMIP, especially for data compilation, management, and distribution. VEMAP experienced schedule delays due to always seeking create "better" data products. MsTMIP also had this issue even though it was aware of experiences from past MIPs. Preparing, managing, and distributing data, especially for a project with 20 participating modeling teams, is a time-consuming task and requires a well-designed data management framework. The "lessons learned" section highlights some important ideas, from the data management perspective, that may help guide future activities so that they can be conducted in a more efficient manner. However, we improved this section by condensing it and adding text (1st paragraph on page 23) to better justify this section.

4. Response to "*Reference for CRU TS3.2 needs to be update to Harris et al. (2013). See http://badc.nerc.ac.uk/view/badc.nerc.ac.uk__ATOM__dataent_1256223773328276. "*

   We've updated the reference for CRU TS3.2 to "Harris, I., Jones, P.D., Osborn, T.J., and Lister, D.H.: Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset. Int. J. Climatol.,34: 623-642. Doi: 10.1002/joc.3711, 2014". This is the most recent reference we could find for CRU TS data.

5. Response to "*If CRUNCEP mentioned in the manuscript is different from "CRU-NCEP" which is maintained by Nicolas Viovy, the authors need to explicitly state so.*"

   The "CRUNCEP" data is the same as what's maintained by Nicolas Viovy. We changed the name of this data to "CRU-NCEP" to be consistent with other naming conventions. Although the CRU-NCEP data has been used in previous MIP activities like GCP-TRENDY, it has never been officially published. As such, it is included in this manuscript and in the published environmental driver data products for MsTMIP.

Nicolas Viovy is a co-author of this manuscript, as well as the published data products.

6. Response to "*According to the data processing described, only the magnitude of the 3-hourly NARR precipitation data were calibrated with the monthly GPCP product by linear rescaling. Did you perform any calibration to the frequency of rainfall events?"*

   The linear rescaling applied to the 3-hourly NARR precipitation data with monthly GPCP product only altered the magnitude of the NARR precipitation, while the frequency of rainfall events remained unchanged.

7. Response to "*Recently, Bohn et al. (2013) validated the performance of MT-CLIM at the global scale. You may want to cite their work to support a reliability of MT-CLIM simulation.*"

   Thank you for the recommendation. We've cited Bohn's work on MTCLIM assessment in the manuscript.

8. Response to "*I believe that you are aware of fPAR3g and LAI3g products derived from GIMMS NDVI3g data (in conjunction with neural network and MODIS products) (Zhu et al., 2013). The spatio-temporal specification of the fPAR3g and LAI3g (15 day-quarter degree data from 1981-2011) is suitable for being used in diagnostic model runs. Instead of relying on a simple method from Sellers et al. (1996), you may want to consider switching to these observation-based products.*"

   Yes, we are aware of these data products. However, they were not publicly available when we prepared the MsTMIP model driver data. These will be great data sources for MsTMIP follow-on activities.