RESPONSE TO REVIEWER COMMENTS

We thank the Editor and reviewers for taking the time to thoroughly read through our paper and offer useful suggestions for improving the writing. Our paper updates and responses are given below in bulleted, blue, and bold text.

Best regards,
Joshua Fisher
Munish Sikka
Deborah Huntzinger
Christopher Schwalm
Junjie Liu

REVIEWER 1

The authors present a product of downscaled monthly global net ecosystem exchange from 15 terrestrial biosphere models, which I think is a timely and valuable contribution to the carbon-cycle modelling community, not just for atmospheric modelling community that they mentioned in their paper, but also for land surface modeling community. This paper is well structured as a technical note, but in some sections I could not quite quickly follow how the downscaling was done. While the whole processing appears robust, there are several minor points could be improved and that would be useful for helping understand some of processing in their product.

- **We thank the reviewer for pointing out the value and timeliness of our work. We edited the broader utility sentence (L38-39) to include the land surface community: "These data were created initially for NASA's carbon Monitoring System (CMS), and are useful to the broader land surface and atmospheric scientific community (Fisher et al., 2011; Fisher et al., 2012)." We have addressed the other minor points below.**

- In the calculation of NEP, the authors use subtracting GPP from Re. This is quite different from the definition in terrestrial ecosystem models that they use equation of NPP minors Re. The assumption in this paper is improved compared to Olsen and Randerson (2004), but is still not close to the assumption in terrestrial ecosystem models.

- **Here, NEP is the difference between GPP and Re (i.e., heterotrophic plus autotrophic respiration), not between NPP and Re. If the latter, then Ra is double-counted. We have added clarifying text to define Re (e.g., L86).**

- The authors have taken the other fluxes from disturbances (e.g. fires) into consideration by balancing their downscaled NEE with fluxes from terrestrial ecosystem models, which spread the difference equally within months. This could work fine at monthly step but I could imagine it would smooth the temporal pattern of NEE for product at hourly time step if there are fires occurred.

- **The reviewer makes a very good point. We have updated the text to clarify the ramifications of this assumption (L94-96): "Nonetheless, this assumption smooths what could otherwise be punctuated fire or disturbance effluxes, so caution should be given when assessing these effluxes at 3-hourly time steps (e.g., relative to observations)."**

- The authors have validated their results with FLUXNET observations. What I'm interesting is if they have compared their results with independent estimates using aircraft datasets which represent the fluxes at large scale. Because we notice the resolution of this product (>0.5 degree) doesn't really match the resolution of FLUXNET (~1km).

- **This is a very good suggestion. Nonetheless, such an uncertainty assessment with complex data such as airborne CO_2 concentration measurements run through an atmospheric inversion to compare to net fluxes is beyond the scope of this technical note with the purpose of describing the downscaling methodology. We hope that the reviewer or others in the scientific community take up this suggestion with our data, and we would be happy to collaborate.**

REVIEWER 2

In the technical note '3-hourly temporal downscaling of monthly global terrestrial biosphere model net ecosystem exchange' Fisher and co-authors describe the methodology employed to downscale monthly output from 15 comprehensive land-surface model and 4 ensemble products. Results and methodology are of interest for atmospheric forward and inverse modeling as well as land-surface model evaluation. The manuscript summarizes the approach and reads well. Nevertheless, I think this paper needs some clarification that have to be addressed first, and which prevent me of accepting this paper in its present form. Therefore, I recommend acceptance of this manuscript after some minor revisions.

- **We thank the reviewer for noting the broad interest of our work, and highlighting the quality of our writing.**

General Comments: 1- The authors have developed an approach on two major assumptions: (1) one forcing set, CRU-NCEP, which is downscaled at 3-hour timestep while other forcing set are available at 3-hour timestep over the time-window (2004- 2010) like the Princeton Global Forcing set, WFEI, WATCH and so on. . . (2) they use monthly output from various land-surface models that are downscaled independently from their skill. These models are able to produce daily output which can also be exploited. I wonder if a higher output frequency could improve 3-hour NEE estimates?

- **We thank the reviewer for noting the other potential forcing datasets, already available at 3-hourly time steps. We added in those references (L151-153). The reason why we selected CRU-NCEP is because all of the models were originally run with CRU-NCEP, so we remained with CRU-NCEP for consistency. We added this clarification in to the manuscript (L149-151): "All models were driven by CRU-NCEP meteorological forcing data, hence our use of the same data source for the downscaling approach applied here. We note that there are other meteorological forcing datasets also available at 3-hourly time steps for those interested in applying our downscaling approach with different data (Sheffield et al., 2006; Weedon et al., 2011; Weedon et al., 2014)."**
- **We provide a model ensemble product weighted based on skill (see L149-153).**
- **The reviewer is correct in that indeed some models can output at sub-monthly time steps. The standard MsTMIP model output is at the monthly time step, however. Certainly, any model providing output at a higher temporal resolution would have improved estimates at high temporal resolutions. We added this clarification into the manuscript (L153-155): "Although some models are capable of output at sub-monthly time steps, the standard MsTMIP output is at the monthly time step."**

2- At the end of the manuscript, the authors state "A full uncertainty analysis of the approach is beyond the scope of this technical note intended to describe the methodological detail of the downscaling.". I first would mention that this statement should appear earlier in the manuscript. Then, while I agree with the authors, I think the authors miss at indicating how they assess the uncertainty in their methodology or clearly state caveats and potential sources of errors. I think this kind of information is fundamental in technical note because this manuscript will support a dataset, software and the companion test case (described herein). For example, the quantile-

quantile plot (Figure 4) can be extended to GPP ad Re (available in most of the FluxNet stations). Readers will also need to see how the mean diurnal cycle for GPP, Re and NEE derived from 3-hourly downscaled estimates of these fluxes (the distribution can be accurate but with the wrong timing).

- **For succinctness and brevity, we go right into describing the downscaling approach. The title, abstract, and first sentence of the main text clearly define what the technical note is about. Remember, it is a technical note allowing the readers to explore a dataset (e.g., for uncertainty assessments against FLUXNET data, aircraft data, etc.); this is not a full evaluation or analysis of the dataset. We sympathize with the Reviewer in wanting to see a full uncertainty assessment, as these are the types of full papers we normally read in the scientific literature. Our team has published extensively on uncertainty assessments, so we know well the rigor necessary to do this properly, and how this falls outside the scope of this technical note. The analysis we did include in the technical note was to illustrate that we performed the downscaling method correctly, not to provide uncertainty numbers with the product. This assessment is not even necessary in the technical note, so it is something extra we include at the very end.**

3- While I understand the approach to use TRENDY / MST-MIP model output, I think the authors have to clearly mention the limits of these models, especially for the Re which aggregate processes that are not well simulated by the current generation models (errors in Ra and Rh often compensates). Details on models that compose the dataset are required to my point of view.

- **We agree with the reviewer. The limits and details of these models are well documented in the cited references. As this technical note focuses on the downscaling approach, rather than the model physics, we point the readers to the full papers describing the individual models in further detail.**

Specific comments: L41: please detail what means 'ecosystem respiration' L58-79: please provide further details on the methodology of Olsen and Randerson (2004). It is unclear whether the same scaling is applied to the various model results or for each individual model (with simulated surface temperature or leaves temperature instead of air temperature).

- **Here, ecosystem respiration (Re) means autotrophic respiration (Ra) plus heterotrophic respiration (Rh), though there are nuances from model to model, as can be gleaned in the cited references that detail the differences in model physics and assumptions. We include further discussion on Re in the paragraph starting on L86.**
- **We provided the relevant details of the Olsen and Randerson (2004) approach, and the differences between their approach and our updates.**
- **We added "(2 m)" for clarity to distinguish the near surface air temperature from surface/skin/leaf temperature.**