RESPONSES TO THE REVIEWER COMMENTS

Reviewer #2

This manuscript by Chen et al. “Pairing FLUXNET sites to validate model representations of land use/land cover change” aims at evaluating the performance of CLM and Noah-MP LSMs in simulating the impacts of LULCC on surface energy balance. Authors rely on observations from paired FLUXNET sites for model validation. The manuscript contains new and significant research, especially efforts to utilize the FLUXNET observation in a paired scheme for LULCC analysis. Also, the choice of the LSMs are very well justified and results could potentially help inform future model improvement. Writing, especially methods and results, could be improved by adding sufficient details for an unfamiliar audience. In its current form the manuscript is very hard to follow, especially, if the reader is not familiar with all the LSM lingo. Also, excessive referring of key information by pointing the readers to tables does not help either.

To an extent figures should be self-explanatory, especially when not restricted by page limit. I am not a big fan of figure caption stating: “same as figure x”, this caption is no better than a figure w/o caption.

Thank you for your thorough evaluation and thoughtful suggestion. Below we specifically respond to each of the individual comments, which have greatly improved our manuscript. And we are sorry for the confusion. We have changed the figure captions.

Major comments:
1) Provide some details on how point-scale models were implemented as this directly relates to foot print at which FLUXNET towers and model operate. Based on the limited information, it is hard to make sense of the differences between model and observation shown in Figure 2. FLUXNET towers are typically have bigger footprint, in some stances > 1 km, that may vary between open and closed canopies. Were these differences in spatial scale between model and observation accounted? Figure 2c suggests otherwise and diverging patterns could be driven by the scale. See Desjardins et al., 1992; Baker et al. 2003, and Griebel et al., for details.

That is a good point. We do admit that the tower footprints may bias the comparison of surface fluxes between the open and forest sites. In other words, the observed difference between the paired sites can only be partially attributed to land cover change because their environmental conditions may be different. However, it is relatively difficult to eliminate its effects on observed surface fluxes.

As most of current studies using paired sites to represent LULCC, we have assumed that the paired sites share the similar background atmospheric conditions, and any differences in surface climate conditions can be attributed to the the LULCC (e.g., Teuling et al. 2010; Luyssaert et al. 2014; Vanden Broucke et al 2015; Lejeune et al. 2017).

Meanwhile, we run the singe-point simulations with two types of meteorological forcings for each site (measurements at this site and measurement at the neighboring paired site), which can generate three types of simulated flux difference (difference derived from individual forcings, difference from identical “forest forcings”, and difference from identical “open forcings”). Because meteorological measurements at individual FLUXNET towers can be influenced by

their local environment, our experimental design (by switching the forcings) can effectively examine the effects of tower footprints in simulating surface fluxes and their difference. In figure 2a-b, the simulated latent heat fluxes are consistent between the two types of forcings (solid red/orange circles vs. triangles). Figure 2c also shows consistent signals from the three types of simulated differences. Therefore, our comparisons are robust and can effectively represent the LULCC-induced climate change, and the impacts of footprints at individual sites are probably trivial.

We have added more explanation and discussion to clarify this in the revised manuscript (P21-22, L471-482).

2) The inclusion of CLM-PFT and CLM-PFTCOL with CRUNCEP forcing makes no sense to me as you cannot directly compare the diurnal energy fluxes with other simulations and attribute the differences to LULCC. For direct comparison, all model simulations should be forced with similar climate forcings. At least, I will not try to use these simulations to explore mechanism as shown in Figure 3 and discussed between Line 231:248.

Sorry for the confusion. The CLM-PFT and CLM-PFTCOL simulations are included because the final goal of evaluating the LSMs at point scale is to using them to investigate the LULCC-induced climate change at global scale. Therefore, it is worthwhile to examine if the sub-grid results from global simulations are comparable and consistent with the single-point simulations.

First, the PFT-level comparison from CLM-PFT or CLM-PFTCOL can be considered as the impacts of LULCC. The PFTs in a single grid box share the exactly same meteorological forcings, but biogeophysical and biogeochemical processes are calculated for each PFT independently. It is analogous to the comparison between the paired-site simulations.

Second, we definitely agree that it would be better if the direct comparison can have all the simulations with the same forcings. However, it is impossible to have single-point and global simulations with identical forcings because of their scales are different. The climatological simulations (CLM-PFT and CLM-PFTCOL) are forced with CRUNCEP forcings in 1991-2010, which covers the observational period of most of the paired sites. Meanwhile, the PFT-level results are extracted based on the geographical location of the paired sites, to ensure the single-point and global simulations have the similar climate. Based on the results, we can also find that the single-point and global simulations are comparable and consistent in most of the cases, especially when the shared-soil-column issue is fixed (CLM-PFTCOL).

To clarify this, we have added to the revised manuscript: "*The paired PFTs are identified based on the locations and land cover types of the FLUXNET paired sites, to ensure the single-point and global simulations comparable.*" (P8, L175-176).

3) I do not see the point of including the FLUXNET data with energy balance closure correction when it is not being discussed after Figure 2. This only makes the figures crowded and confusing. Suggest comparing the uncorrected and corrected observations in the beginning, or may in the supplemental, and then using one of the two as a reference for further comparisons with model simulations [which you have already done for some figures].

Thank you for the suggestion. We have removed the balance-closure corrected fluxes from Figure 4~6. A supplementary figure is added to compare the corrected and uncorrected observations (Figure S2 and P13, L275-276).

4) Considering the large difference in LE between some of the paired sites (in particular 3, 7, 12, and 15) I would suggest setting a threshold for inclusion. These differences in LE and H within paired sites are comparable to the corresponding changes under deforestation and cannot be overlooked.

That is a good point. Figure R1 (below) shows the changes in LE, H, G, and R_{net} from forest to open (open – forest) land excluding the pairs 3, 7, 12, and 15. The exclusion of these pairs shows very consistent patterns with the results including all sites (Figure 2-9), even though there is a slight influence on the magnitude of changes in fluxes (e.g., daytime LE). Therefore, large changes in surface fluxes within some pairs (or "outliers") do not affect the robustness of our results. We have included this figure in the supplementary information (Figure S1), and added some discussion in the revised manuscript (P10, L204-206).

5) As of now the analysis is mostly focused on validation with very little emphasis on the sources of over- and under-estimation in energy fluxes. The discussion section is very speculative and mostly hand waving. Authors should put more emphasis on mechanistic model diagnosis that goes beyond forcing.

The step from validation to diagnosis is large. For CLM in particular, there is a mechanism through the NCAR Land Model Working Group (LMWG) to illuminate and contribute to model development. The authors actively participate in the LMWG and collaborate with colleagues at NCAR to mechanistically diagnose and improve the model. We do have some discussion about the possible reason for the biases (L416-426; L439-464), but taking the next step to systematic diagnosis will be part of our collaborative project with NCAR colleagues.

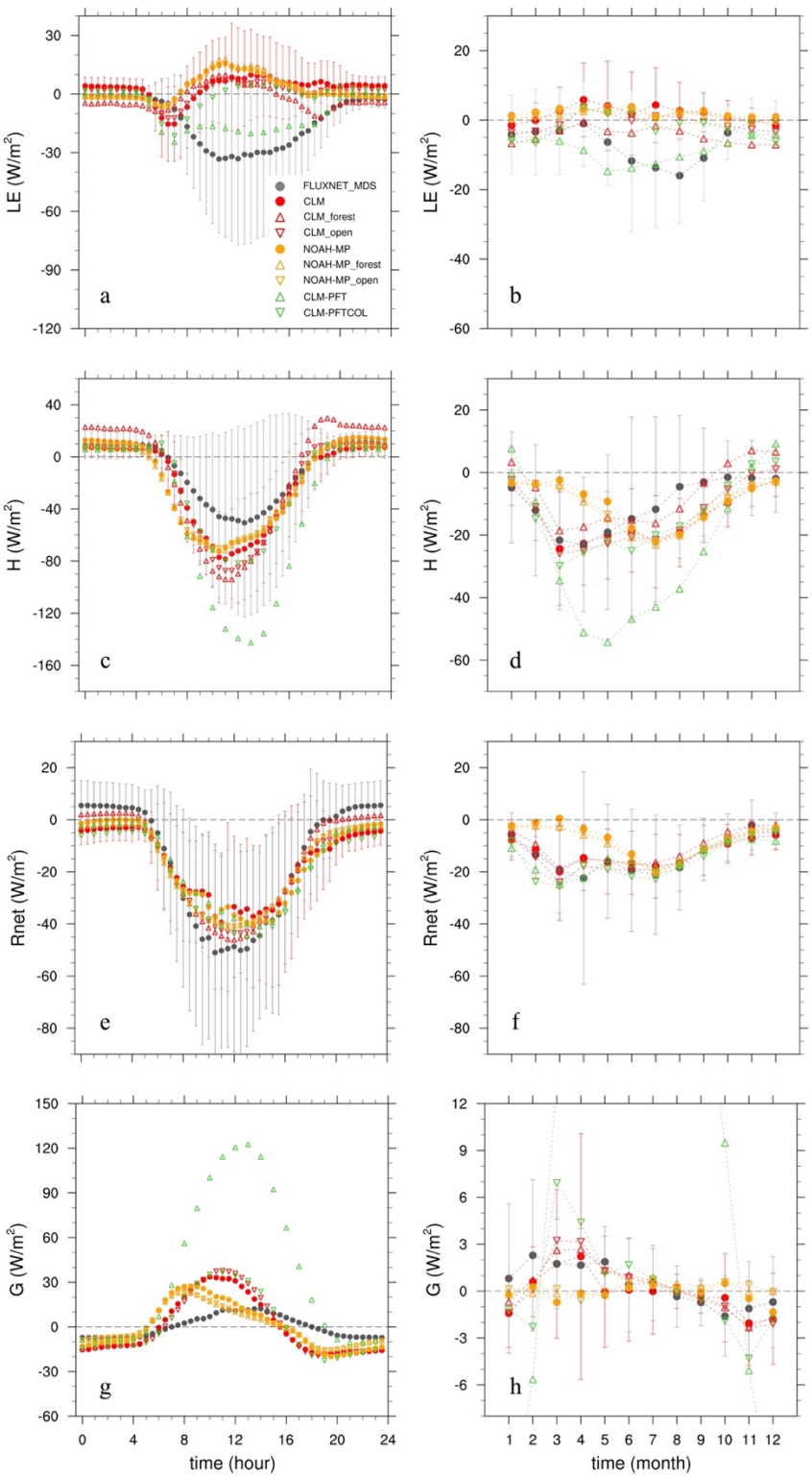

Figure R1. Change in the diurnal (left) and seasonal (right) cycle of LE (W/m^2, a-b), H (W/m^2, c-d), Rnet (W/m^2, e-f), and G (W/m^2, g-h) from forest to open (open – forest) land excluding the pairs 3, 7, 12, and 15.

Minor Points:

L35: what do you mean by deficiencies over forest land-cover type?

It means greater bias over the forest land-cover types. The models show greater biases in estimating the *LE* and *H* over forest, thus cannot capture the observed decrease in *LE* after deforestation. We have changed this sentence to make our statement clearer: "*These deficiencies are mainly associated with models' greater biases over forest land-cover types and the parameterization of soil evaporation*" (P2, L34-35).

L58 which were associated?

Sorry for the confusion. It indicates "the different climatic responses". We have reframed this sentence: "*... Brovkin et al. (2013) also found different climatic responses to LULCC among the participating models, and the diverse responses are associated with different parameterizations ...*"(P3, L58).

L130-133: I do not think this statement is supported by data, at least for some sites.

True, we have revised the sentence: "*Below we show that the differences in meteorology are usually small and not likely a dominant factor in simulated surface flux differences in most of the pairs*" (P6, L133).

L165-166 do PFT in CLM are the same as the land cover reported for FLUXNET sites?

Yes. For a grid cell in CLM, the sub-grid heterogeneity is described as the percentage of each PFT (totally 15 PFTs are potentially available). The two paired sites are close enough spatially that we can consider them as two different PFTs within a single climate model grid cell. Because the surface fluxes are calculated at the individual PFT level, we can extract the output of the corresponding PFT based on the reported land cover type of each flux site. We have added more explanation in the revised manuscript (P8, L165, L168-171).

Figure 1: source of land cover?

The land cover type of each site is based on the reported land cover in FLUXNET database. We have added this information in the revised manuscript (P35, L696-697).

Figure 2: label each panel with "a", "b", and "c". Also, in caption Table reference is missing. Note that the difference is calculated as closed-open canopy?

Thank you for the suggestion. We have added the panel labels for these figures. Also, we have added the note that the difference is calculated as "*open–forest*" in the figure captions (Figure 2).

Figures 5-10: DO NOT USE SAME AS. It is very difficult to flip pages back and forth in order to understand the figure.

Agree. We have changes the captions for those figures.

Figures 11 and 13 are very difficult to follow. Not sure what you mean observations or model also the arrows showing LC conversion. Also, instead of 1-7, why not directly label using actual simulation type?

Thank you for the suggestion. The source "observations or models" just means how each column is calculated (based on observations or model simulations). Yes, the arrows just show the land cover change (from a forest type to an open type). We have replaced the numbers with actual observation or simulations types (Figure 10 and 12).

Reference:

Lejeune, Q., S. I. Seneviratne, and E. L. Davin, 2017: Historical Land-Cover Change Impacts on Climate: Comparative Assessment of LUCID and CMIP5 Multimodel Experiments. J. Climate, 30, 1439-1459, doi:10.1175/JCLI-D-16-0213.1.

Luyssaert, S. and Coauthors, 2014: Land management and land-cover change have impacts of similar magnitude on surface temperature. Nature Climate Change, 4, 389-393, doi:10.1038/nclimate2196.

Teuling, A. J. and Coauthors, 2010: Contrasting response of European forest and grassland energy exchange to heatwaves. Nature Geoscience, 3, 722-727, doi:10.1038/ngeo950.

Vanden Broucke, S., S. Luyssaert, E. L. Davin, I. Janssens, and N. van Lipzig, 2015: New insights in the capability of climate models to simulate the impact of LUC based on temperature decomposition of paired site observations. Journal of Geophysical Research: Atmospheres, 120, 5417-5436, doi:10.1002/2015JD023095.