RESPONSES TO THE REVIEW'S COMMENTS

Reviewer #1

This is such a useful study and I would hope that we could eventually expand the analysis to include more models. To make it even better I would like to suggest the following: I am unhappy with the emphasis on the uncorrected flux data (e.g. page 12, lines 264 to 267). Since we know the fluxes are generally (and variably) underestimated by the flux-observation system, I think it is more useful to consider the evaporative fraction (LE/(LE+H)) rather even than the 'corrected' fluxes which depends on yet more uncertain data (Rn and G). In my paper (Blyth et al, 2010, https://doi.org/10.1175/2009JHM1183.1) I scale the observed evaporation with the ratio of observed sum LE+H and modelled sum LE+H.

You state in the introduction that the greatest uncertainties of change between forest and open come from the flux partition rather than the total absorbed radiation. So a focus on that would be helpful - hence the reliance on the evaporative fraction makes sense. Then separately consider Rn and G.

We thank the reviewer for the insightful assessment of the FLUXNET data and our manuscript. We have added one figure about the change in evaporative fraction (*EF* = *LE*/(*LE*+*H*)) in the revised manuscript (Figure 7, R1). *EF* has advantages as noted by the reviewer, but also disadvantages; particularly in a budget assessment as it is not itself a term in the energy or water budgets. During summer, there is little change in observed daytime EF (Figure 7a) because of the observed decreases in both *LE* and *H* (Figure 2,5). However, the models show increased daytime *EF* due to the decreased *H* and slightly increased *LE* after deforestation. Seasonally, the models suggest year-around increased *EF*, however, which is not shown in the observations during summer (P14, L295-307).

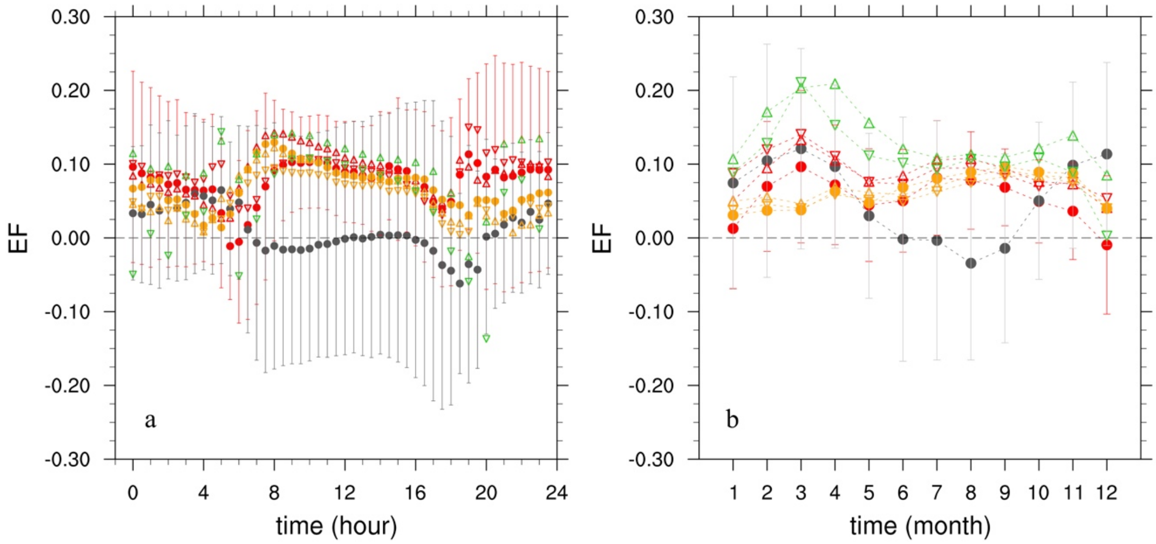

Figure R1. Change in the summer diurnal (a) and seasonal (b) cycle of *EF* (unitless) due to LULCC from forest to open land (open – forest).

Secondly - I wonder if you can do the single-site simulations with the one-soil and two soils options. Give the single-site off-line run a 50% fraction of ’tree’ and ’open’ and compare them - even include a bit of modelled soil moisture to show how it is affecting it. I found that getting the PFT and PFTCOL into the analysis of this paper tends to confuse the issues especially when one is so wayward, while the point about whether or not to have separate soils for each PFT seems essential!

Sorry for the confusion and thank you for the suggestion. The single-site off-line run with 50% “tree” and 50% “open” is actually analogous to the PFT and PFTCOL runs.

No matter what percentage of a grid box (or single site) is covered by a tree (or grass) PFT, or if there are any other PFTs involved in the grid box, the PFT-level open-versus-tree comparisons in a single-site off-line run can be the same as our comparison from the global PFT runs. Within the same grid box, the two PFTs receive the exactly same meteorological forcings. In CLM, biogeophysical and biogeochemical processes are simulated for each PFT independently, and all fluxes to and from the land surface are calculated at the PFT level (Oleson et al. 2013).

The percentage of individual PFTs only matters when aggregating the PFT level properties to the column/grid level, however, which is beyond the scope of this study. We have added more explanation about the PFT-level comparison in the revised manuscript (P8, L168-171).

Oleson, K. W., and co-authors, 2013: Technical Description of version 4.5 of the Community Land Model (CLM), NCAR Technical Note, TN-503+STR, National Center for Atmospheric Research, Boulder, CO, USA, 434pp., [Available at http://www.cesm.ucar.edu/models/cesm1.2/clm/CLM45_Tech_ Note.pdf].

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 R2 (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.

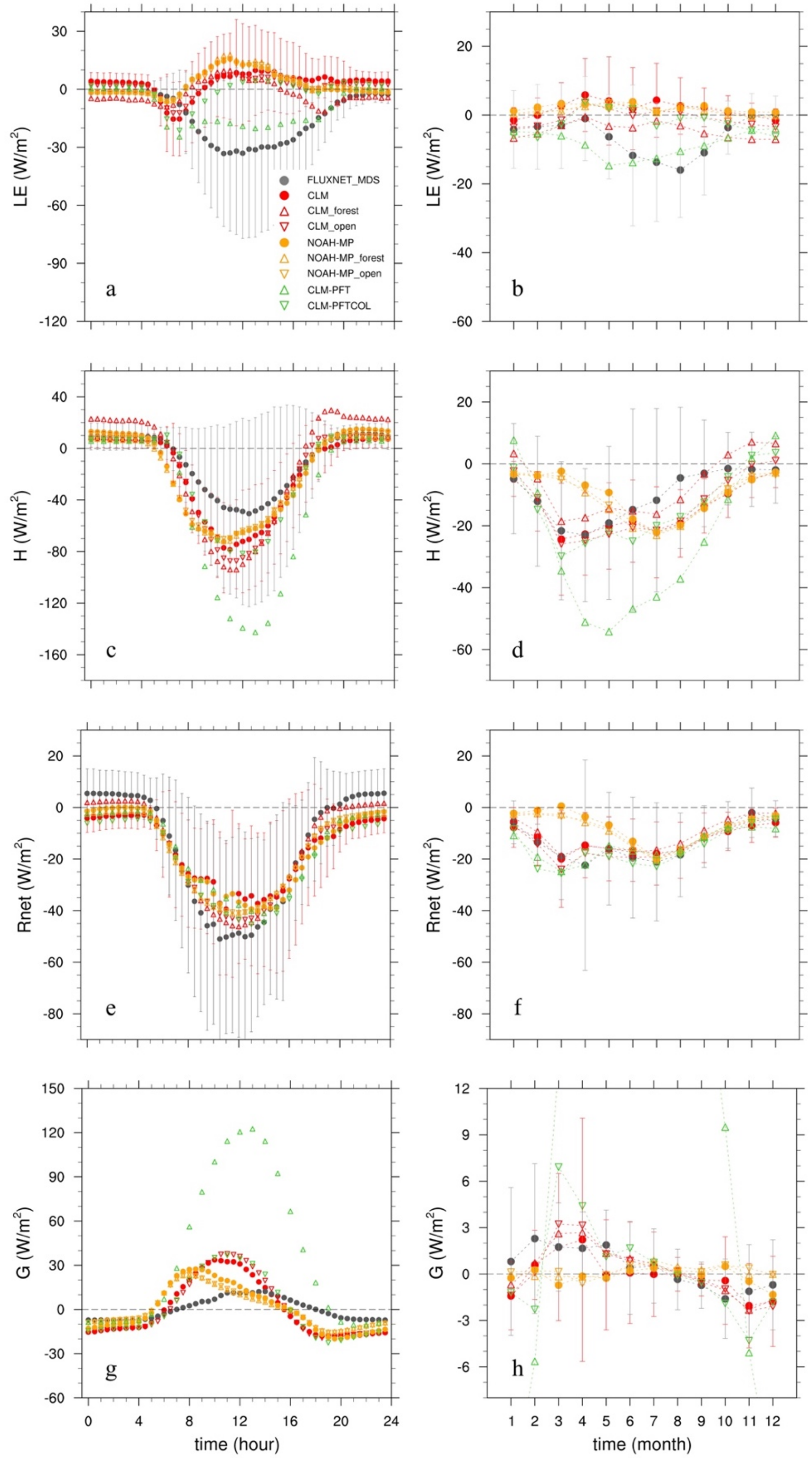

Figure R2. Change in the diurnal (left) and seasonal (right) cycle of LE (W/m^2, a-b), H (W/m^2, c-d), Rnet (W/m^2, e-f), and G (W/m^2, g-h) from forest to open (open – forest) land excluding the pairs 3, 7, 12, and 15.

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.

Pairing FLUXNET Sites to Validate Model Representations of Land Use/Land Cover Change

Liang Chen[1], Paul A. Dirmeyer[1], Zhichang Guo[1], Natalie M. Schultz[2]

1 Center for Ocean-Land-Atmosphere Studies, George Mason University, Fairfax, Virginia, USA

2 School of Forestry and Environmental Studies, Yale University, New Haven, Connecticut, USA

---

* Corresponding author: Liang Chen

Center for Ocean-Land-Atmosphere Studies, George Mason University, Mail Stop 6C5, Fairfax, VA,
22030 USA

Email: lchen15@gmu.edu

Abstract

Land surface energy and water fluxes play an important role in land-atmosphere interactions, especially for the climatic feedback effects driven by land use/land cover change (LULCC). These have long been documented in model-based studies, but the performance of land surface models in representing LULCC-induced responses has not been well investigated. In this study, measurements from proximate paired (open versus forest) flux tower sites are used to represent observed deforestation-induced changes in surface fluxes, which are compared with simulations from the Community Land Model (CLM) and the Noah Multi-Parameterization (Noah-MP) land model. Point-scale simulations suggest CLM can represent the observed diurnal and seasonal changes in net radiation (R_{net}) and ground heat flux (G), but difficulties remain in the energy partitioning between latent (LE) and sensible (H) heat flux. CLM does not capture the observed decreased daytime LE, and overestimates the increased H during summer. These deficiencies are mainly associated with models' greater biases over forest land-cover types and the parameterization of soil evaporation. Global gridded simulations with CLM show uncertainties in the estimation of LE and H at the grid level for regional and global simulations. Noah-MP exhibits a similar ability to simulate the surface flux changes, but with larger biases in H, G, and R_{net} change during late winter and early spring, which are related to a deficiency in estimating albedo. Differences in meteorological conditions between paired sites is not a factor in these results. Attention needs to be devoted to improving the representation of surface heat flux processes in land models to increase confidence in LULCC simulations.

1. Introduction

Earth system models (ESMs) have long been used to investigate the climatic impacts of land use/land cover change (LULCC) (cf. Pielke et al. 2011; Mahmood et al. 2014). Results from sensitivity studies largely depend on the land surface model (LSM) that is coupled to the atmospheric model within ESMs. In the context of the Land-Use and Climate, Identification of Robust Impacts (LUCID) project, Pitman et al. (2009) found disagreement among the LSMs in simulating the LULCC-induced changes in summer latent heat flux over the Northern Hemisphere. de Noblet-Ducoudré et al. (2012) and Boiser et al. (2012) argued that the inter-model spread of LULCC sensitivity (especially regarding the partitioning of available energy between latent and sensible heat fluxes within the different land-cover types) highlights an urgent need for a rigorous evaluation of LSMs. From Phase 5 of the Coupled Model Inter-comparison Project (CMIP5), Brovkin et al. (2013) also found different climatic responses to LULCC among the participating models, and the diverse responses are associated with different parameterizations of land surface processes among ESMs. To deal with the uncertainties in LULCC sensitivity among models, the Land Use Model Inter-comparison Project (LUMIP) has been planned, with a goal to develop metrics and diagnostic protocols that quantify LSM performance and related sensitivities with respect to LULCC (Lawrence et al. 2016).

However, a paucity of useful observations has hindered the assessment of the simulated impacts of LULCC and limited the understanding of the discrepancies among models. In-situ and satellite observations make it possible to quantify the impacts of LULCC on land surface variables. Satellite-derived datasets have been used to explore the albedo, evapotranspiration (ET), and land surface temperature changes due to historical LULCC (Boisier et al. 2013, 2014) and the climatic effects of forest (Li et al. 2015).

Meanwhile, the development of FLUXNET (Baldocchi et al. 2001) enables the study of land surface responses to different land-cover types based on paired field observations from neighboring flux towers over forest and open land (Juang et al. 2007; Lee et al. 2011; Luyssaert et al. 2014; Teuling et al. 2010; Williams et al. 2012). In terms of LSM evaluation, the paired site observations have been mainly used to simulated impacts of LULCC on land surface temperature (Chen and Dirmeyer 2016; Lejeune et al. 2016; Vanden Broucke et al. 2015). However, a more fundamental question, "whether a model can well represent the observed LULCC-induced changes in surface energy fluxes", has not been thoroughly investigated, even though we know that the turbulent fluxes are tightly associated with both energy and water exchange between the land surface and atmosphere.

In this study, we evaluate the performance of the Community Land Model (CLM) version 4.5 and the Noah Multi-Parameterization (Noah-MP) LSM in simulating the impacts of LULCC on surface energy fluxes based on observations from FLUXNET sites. CLM and Noah-MP represent perhaps the two most readily available and widely used state-of-the-art community land models developed in the U.S. CLM is chosen because, as the land component for Community Earth System Model (CESM), it prioritizes the simulation of biogeophysical and biogeochemical processes for climate applications (Oleson et al. 2013). Much effort has gone into improving the representation of the land-atmosphere interactions among different biomes (Bonan et al. 2011), and the model itself has been used for many LULCC sensitivity studies (e.g., Chen and Dirmeyer 2016, 2017; Schultz et al. 2016; Lejeune et al. 2017; Lawrence et al. 2012). Noah-MP has found use mainly in shorter time-scale, limited area applications, such as weather and hydrologic forecasting, and as a LSM run at very high resolution coupled to mesoscale models (e.g., WRF-Hydro, Gochis et al. 2015). It is planned to become the LSM used in global weather and seasonal forecasting applications at the National Centers for Environmental Prediction (NCEP). Its performance over varying land cover types has direct consequences for its use in forecast models.

The rest of this paper is structured as follows. Section 2 describes the datasets used in the study and experimental design. Section 3 presents comparison between observations and model simulations in surface latent and sensible heat flux, ground heat flux, and net radiation. Section 4 shows the uncertainties within the FLUXNET pairs and model simulations. Sections 5 and 6 include discussion and conclusions, respectively.

2. Methodology

2.1 Observational data

We use half-hourly observations from 24 selected pairs of flux sites from the FLUXNET2015 Tier 1 dataset (http://fluxnet.fluxdata.org/data/fluxnet2015-dataset) and 4 pairs from the AmeriFlux dataset (Baldocchi et al. 2001). These observations include meteorological forcings for the LSM, and surface flux measurements for model validation, which include latent heat flux (LE), sensible heat flux (H), ground heat flux (G), and net radiation (R_{net}). All of these variables have been gap-filled (Reichstein et al. 2005; Vuichard and Papale 2015). Table 1 shows the variable names and gap-filling algorithms used in FLUXNET2015. Because there is no directly measured humidity variable reported, which is needed as a meteorological forcing for the LSMs, relative humidity is calculated based on the reported vapor pressure deficit and surface air temperature (Equation 1-2).

$$e_s = 6.11\, exp\left(17.26938818 \frac{T_a}{237.3 + T_a}\right) \quad (1)$$

$$RH = \left(1 - \frac{VPD}{e_s}\right) \times 100 \quad (2)$$

in which T_a is air temperature (°C), e_s is saturation vapor pressure (hPa), VPD is vapor pressure deficit (hPa), and RH is relative humidity (%). Additionally, for the turbulent flux measurements over 18 pairs, FLUXNET2015 provides “corrected” fluxes based on an energy balance closure correction factor, which is calculated for each half-hour as $(R_{net} - G) / (H + LE)$. More details about the data processing can be found on the FLUXNET2015 website (http://fluxnet.fluxdata.org/data/fluxnet2015-dataset/data-processing/).

To simulate local land cover change for each pair, one flux tower is located in forest (deciduous, evergreen or mixed; broadleaf or needleleaf) and the other is in a nearby open land cover type (grassland, cropland or open shrub). Figure 1 shows the locations of the paired sites. Their general characteristics are listed in Table S1. The median linear distance between the paired sites is 21.6 km, and the median elevation difference is 20.0 m. Because of their proximities, the paired sites share similar atmospheric background conditions, however they are not identical (Chen and Dirmeyer 2016). 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. We consider the differences (open minus forest) in observed surface fluxes to be representative of the effects of LULCC (deforestation in this case).

2.2 Model simulations

We have run the offline version of CLM 4.5 and Noah-MP at the point-scale for individual sites. The forcing data, described below, includes downwelling long-wave radiation (W/m^2), downwelling short-wave radiation (W/m^2), air temperature (K), precipitation (mm/s), relative humidity (%), surface pressure (Pa), and wind speed (m/s) at half-hourly time steps. The plant functional type (PFT) in CLM for each site is identified based on its reported land cover type (Table S1) with prescribed climatological satellite phenology (Lawrence and Chase, 2010). Because of the focus on biogeophysical impacts of LULCC in this study, the biogeochemistry Carbon-Nitrogen module has been disabled in our simulations. The initial conditions for each site are generated by cycling through available atmospheric forcings for about 40 years until soil moisture and temperature reach quasi-equilibrium.

The differences in simulated surface fluxes between the paired sites are compared against the observations, so that the performance of CLM in representing LULCC-induced surface flux changes can be evaluated. In the single-point simulations, two types of forcing data are used for each site: 1) measurements at this site; 2) measurements at the neighboring paired site. Consequently, three types of differences in simulated surface fluxes can be calculated: 1) the difference derived from individual forcings; 2) the difference from identical “forest forcings” (both of the paired sites use the same forcings measured at the forest site); 3) the difference from identical “open forcings” (both of the paired sites use the same forcings measured at the open sites). Such an experimental design can well eliminate the influence from the uncertainties of forcing data and the difference in atmospheric background of the paired sites.

The ultimate goal of evaluating CLM’s performance at single-point scale is to assess its capability to be used in global LULCC sensitivity simulations in both offline and coupled modes. The paired sites are close enough that they are typically located within a single grid cell of CESM. Moreover, the sub-grid heterogeneity of CLM allows the biogeophysical processes to be calculated at the individual PFT level (15 PFTs available), and makes it possible to output surface fluxes for individual land cover types. The paired sites can be presented as paired PFTs within a single grid of CESM. They then share the same atmospheric forcings, and their differences can be considered as the impacts of LULCC. It should be noted that the PFT-level calculation is independent of the percentage of individual PFTs in the grid cell. Therefore, the coverage of the PFTs in the shared grid cell does not influence the flux difference between the paired PFTs in the global simulations.

We run CLM offline, globally driven by the CRUNCEP forcings from 1991 to 2010 (Viovy 2011) and present land cover conditions (Lawrence et al. 2012) at a horizontal resolution of 0.9°×1.25°. 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 are comparable. Schultz et al. (2016) found the shared-soil-column configuration for vegetated land units in CLM caused issues with PFT-level ground heat fluxes. They propose an individual-soil-column scheme (PFTCOL) to better represent the PFT-level energy fluxes, so we also extract and examine the output for the paired PFTs from the PFTCOL model configuration. Details about the PFTCOL simulations can be found in Schultz et al. (2016). Additionally, a coupled simulation with Community Atmosphere Model (CAM) has also been conducted. It shows very similar results to the offline simulations, because the paired PFTs in a single model grid box always share the same atmospheric forcings no matter if CLM is run offline or coupled with CAM. Therefore, results from the coupled simulation are not included in this study.

Furthermore, we compare the performance of CLM with Noah-MP (Niu et al. 2011), which serves as a participant model in Land Data Assimilation Systems (LDAS, Cai et al. 2014). Single-point Noah-MP simulations are conducted in the same way as CLM simulations to ensure their comparability. The monthly leaf area index (LAI) of each site is identical to the prescribed satellite-based LAI in the corresponding CLM simulation. Table S2 shows selected options for various physical processes in Noah-MP. Information about all model simulations is summarized in Table 2.

3. Surface energy fluxes and their changes

First, we analyze the diurnal and seasonal cycles of surface energy fluxes and the LULCC-induced changes. The diurnal cycle analysis is primarily focused on summer (DJF for the two austral sites and JJA for the other sites). The seasonal cycle for the austral sites is shifted by 6 months to keep summer in the middle of the time series when comparing or compositing with the Northern Hemisphere sites. The results shown below are composites averaged over all open (or forest) sites or open-forest pairs. Not all sites have energy-balance corrected fluxes available;

exclusion of those sites shows very similar results for uncorrected fluxes to the average over all sites (or pairs, not shown). There are also some pairs with relatively large changes in surface fluxes. Exclusion of those pairs shows very consistent patterns with the results including all sites, even though there is a slight influence on the magnitude of the changes (Figure S1). Therefore, all sites are included in our analyses for each variable.

3.1 Latent heat flux (LE)

Figure 2a-b shows the diurnal cycle of *LE* averaged over all the open sites and forest sites during summer. Compared with the observations without energy-balance correction, single-point CLM simulations overestimate *LE* for the open sites with both their actual meteorological forcings and the nearby forest forcings, but underestimate *LE* over the forest sites. The extracted PFT-level output from the global simulations also exhibit similar biases. Relative to CLM, Noah-MP simulations show better agreement with observations over the open sites, but a greater underestimation over forest. The energy-balance correction tends to increase the values of *LE*. Therefore, both CLM and Noah-MP have negative biases compared to the corrected fluxes (except LE_CORR_25 over the open sites).

Figure 2c shows the difference in the diurnal cycle of *LE* due to LULCC (deforestation). It should be noted that there is a substantial spread among the pairs in model simulations and especially observations, indicating the diverse geographical backgrounds and specific vegetation changes of these paired sites. The observations suggest an overall lower summer daytime *LE* over the open land compared to forest. In spite of the considerable spread among the energy-balance corrected *LE* observations (Figure 2ab), the differences between the forest and open lands show consistent signals. However, both CLM and Noah-MP single-point simulations fail to represent the observed decreased daytime *LE* as a result of deforestation. The simulated *LE* over the open land is usually slightly greater than the forest from 10:00 to 16:00 at local time. Such a discrepancy may be attributed to the large underestimation of daytime forest *LE* in the models. Meanwhile, simulations by different forcings of the paired sites show robust signals, implying that the bias of the simulated *LE* sensitivity should not be attributed to the uncertainties of the forcing data. For the CLM global simulations, the PFTCOL case exhibits a similar diurnal pattern to the single-point simulations, while decreased daytime *LE* is found consistently only in the PFT simulations. As CLM-PFT is less physically realistic than CLM-PFTCOL from a soil hydrologic perspective, its superior performance is curious.

To explore the mechanism of the *LE* changes within CLM, we examine the changes in the three components of evapotranspiration; namely canopy evaporation, canopy transpiration, and ground evaporation (Figure 3). Unfortunately, these separate components are not measured and cannot be directly validated. The CLM, PFT and PFTCOL simulations show an agreement in decreased canopy evaporation after deforestation with the greatest decrease during the early morning. There also is an agreement in an overall decreased canopy transpiration, but CLM simulations do not exhibit an obvious change during the morning when greatly decreased canopy transpiration can be found in the PFT and PFTCOL simulations. The main discrepancy among model versions is found in ground evaporation, which increases after deforestation in the CLM and PFTCOL simulations. The increased ground evaporation has exceeded the decreased canopy evaporation and transpiration, resulting in slightly increased *LE* (Figure 2c). Interestingly, the PFT simulations, which have known issues with PFT-level ground heat flux (Schultz et al. 2016), show decreased daytime ground evaporation. Along with decreased canopy evaporation, transpiration, and ground evaporation, the total *LE* decreases sharply after deforestation in the PFT simulations, which agrees better with the observations than other simulations (Figure 2c). However, the decreased ground evaporation may be associated with a problematic soil-column scheme at sub-grid scale, which undermines the credibility of the agreement between the observations and PFT simulations.

Figure 4 shows the changes in monthly *LE* after deforestation across the annual cycle. There is clear and consistent seasonality in the *LE* changes from the observations. The four types of observations show decreased *LE* (up to -24.0 W/m^2) during local summer. There is little change in *LE* in the uncorrected observations during the winter season. However, there is significantly increased *LE* (up to +17.9 W/m^2) in the energy-balance corrected observations in late winter and early spring. Neither CLM nor Noah-MP capture the observed seasonality of *LE* change. As found in the change in the diurnal cycle of the *LE*, the PFTCOL simulations exhibit a similar pattern to the single-point simulations, while the PFT simulations show decreased *LE* throughout the year with the maximum from May to August, and the best correlation ($R = 0.81$, $P < 0.01$) with observations.

3.2 Sensible heat flux (H)

Figure 5a-b shows the diurnal cycle of *H* averaged over all open and forest sites during local summer. Generally, the models overestimate *H* throughout the day, with the largest positive bias during midday. Compared with the observations without energy-balance correction, the overestimation can be up to 86.5 W/m^2 from CLM over the forest during noon and 46.4 W/m^2 over the open sites. The difference in *H* between the forest and open sites is shown in Figure 5c. Robust signals are found among the four types of observations, so results from the energy-balance corrected observations are not included hereafter, but are shown in Figure S2. Both observations and models exhibit a clear diurnal pattern of change in *H* after deforestation – a small nighttime increase and a large daytime decrease. Observations show a large spread among the 28 pairs, which is much greater than that from the CLM simulations, indicating uncertainties and variability among the observed fluxes and the robustness of simulated *H* sensitivity to LULCC in the LSM. Compared with the observations, CLM shows a greater *H* decrease, which is twice as much as in the observations. The overestimated *H* decrease may be related to the large positive bias in *H* over the forest sites (Figure 5b). Additionally, the PFT simulations show the largest *H* decrease, which may be associated with the ground heat issues in the shared-soil-column scheme.

Seasonally, decreased *H* is found throughout the year after deforestation in both observations and models (except for the same-forest-forcing CLM simulations in winter, Figure 6). The greatest decrease is observed during spring, when both of the single-point CLM and PFTCOL simulations show good agreement. However, CLM and Noah-MP simulations also show a large decrease during summer, which has not been observed in the FLUXNET dataset. Again, the PFT simulations show the greatest *H* decrease among the simulations and the largest bias compared with the observations during the warm season.

Additionally, evaporative fraction (*EF*), which is defined as the ratio of *LE* to the available energy (*LE*+*H*), is a useful diagnostic of the surface energy balance (Gentine et al. 2011). Meanwhile, most of the correction methods to solve the imbalance issue of surface energy budget assume the Bowen ratio for small- and large-scale eddies are similar or even equal (Wilson et al. 2002; Foken 2008; Zhou and Wang 2016). Under such an assumption, *EF* can be independent of energy closure issue, because EF is related to the Bowen ratio (B) as:

$$EF = (1 + B)^{-1} \tag{3}$$

Figure 7 shows the change in the diurnal (summer only) and seasonal cycle of *EF* due to LULCC from forest to open land. During summer, there are small changes in observed daytime *EF* (Figure 7a) because of the decreases in both *LE* and *H*. However, both CLM and Noah-MP show increased daytime *EF* due to the decreased *H* and slightly increased *LE* after deforestation. Seasonally, the models show year-around increased *EF*, however, which is not observed in FLUXNET from June to September, further demonstrating the models' deficiencies in representing energy partitioning during summer.

3.3 Diurnal and seasonal cycle of ground heat flux (G) and net radiation (R_{net})

Figure 8a shows the change in the diurnal cycle of *G* after deforestation. Both the observations and models exhibit increased *G* during the day and decreased *G* during the night. However, models overestimate the magnitude of the *G* change, and discrepancies also exist in the timing of maximum change. The greatest increase in *G* is observed during early afternoon, while the greatest increase in simulated *G* occurs at noon in CLM (single-point and PFTCOL) and during morning in Noah-MP. Because *G* is strongly correlated with R_{net} (Santanello and Friedl 2003), we examine the timing of maximum observed G and R_{net} during summer. There are some sites showing about a 1-hour lag between maximum R_{net} and G (not shown). Therefore, the lag between simulated and observed peaks in G change can be partially attributed to the uncertainties in G measurements that are commonly estimated with heat flux plates installed at some depth (e.g., 5~10 cm) below the surface (Wang and Bou-Zeid 2012), while the LSM simulated G is calculated at the surface. Meanwhile, the G changes (in both the diurnal and seasonal cycle) in the PFT simulations are further from the observations than the other simulations. Such disagreement further confirms the issues with the sub-grid soil column scheme in CLM, which is discussed in the following section. The changes in observed G also have a clear seasonal pattern – an increase during the warm season and a decrease during the cold season (Figure 8b). This seasonality is well captured by the CLM simulations (especially the simulations with identical forcings for the paired sites) in both magnitude and timing, but not evident in Noah-MP simulations.

After exploring the three flux components of the surface energy balance, it is worthwhile to examine the change in R_{net} after deforestation. During summer, the observations show that R_{net} slightly increases during the night, and decreases considerably (up to -65.7 W/m^2) during the day, which can be attributed to the increased albedo after deforestation (Figure 9a). Decreased daytime R_{net} is also found in the CLM simulations, but with a slightly smaller magnitude. Seasonally, there is a good agreement between the observations and CLM simulations, showing a large R_{net} decrease during spring and summer but a relatively small decrease during autumn and winter (Figure 9b). The Noah-MP simulations are comparable to CLM, but with a notable deficiency in simulating the R_{net} change during late winter and early spring.

4. Uncertainty Analysis

4.1 Uncertainties among the FLUXNET pairs

The results discussed above are based on composites averaged over all forest and open sites. It is worthwhile to examine the uncertainties in surface flux changes among different paired sites. Figure 10a shows the changes in summer daytime (8:00 ~ 16:00) *LE* from the observations and model simulations across the 28 pairs. This time period is chosen because it is the time of greatest differences in surface energy fluxes (Figure 2c, 5c, 7a, 8a). The observations show decreased *LE* associated with deforestation over 23 pairs, among which the pairs of evergreen needleleaf forest and open shrub (No. 16~25) exhibit consistent decreases and the pairs of deciduous broadleaf forest and crops (No. 1~4) show the overall greatest decrease. However, both CLM and Noah-MP show relatively weak increases over most of the pairs, which further demonstrate their deficiency in simulating *LE* change. Additionally, for both CLM or Noah, the choice of forcings does not exert much influence on the simulated change in summer daytime *LE*.

The changes in R_{net} over individual pairs are shown in Figure 10b. There are 27 pairs (all except number 21) showing decreased R_{net} after deforestation, with the greatest decreases over the pairs of evergreen needleleaf forest and grassland. Both CLM and Noah-MP well captures the observed decreases in R_{net} over most of the pairs.

It should be noted that pair 15 shows large *LE* and R_{net} changes in Figure 10. This pair consists of a site over valley grassland and the other site over mountain evergreen needleleaf forest with 60.29 km separation and 1186 m elevation difference. There are significantly different air temperature and downwelling longwave radiation measurements between the sites (Figure S3). Such large differences in *LE* and R_{net} here are likely associated with the distinct although proximate geographical sites. Even though the exclusion of this site does not make a significant change to the composite analysis in section 3 (not shown), it may raise another question if the simulated sensitivity of surface energy fluxes is associated with the inconsistencies of atmospheric forcings of LSMs at the single pair level.

4.2 Uncertainties within the forcings for LSMs

Based on the composite analysis in section 3, we have found that the simulated changes in surface energy fluxes with identical forcings (either from forest or open sites) are consistent with the simulations with individual forcings, demonstrating that the overall sensitivities of surface energy fluxes are robust among the choices of different forcings. In this sub-section, we explore the uncertainties of the simulated surface flux changes due to the different forcings for individual pairs, especially with the focus on the roles of separation and elevation difference in the simulated sensitivity of surface energy fluxes.

Since we have simulations with identical forcings for the paired sites, the difference in surface flux changes between “forest forcings” and “open forcings” can be considered as the simulated sensitivity of surface energy fluxes to variation in the atmospheric forcings. Figure 11 shows the relationship with separation and elevation difference for individual pairs. Overall, the flux changes are not associated with the separation and elevation difference between the paired sites, further confirming the robustness of simulated signals from paired-site simulations. Nevertheless, some "outliers" are identified. In the CLM simulations, only pair 15 shows large differences in *LE* and *H* change. However, pairs 3, 7, and 12 also exhibit large differences in Noah-MP simulations. The uncertainties in pairs 12 and 15 may be attributed to their large elevation differences. For pair 7 in Australia, Noah-MP shows greater sensitivity of *H* and R_{net} to atmospheric forcings over the evergreen broadleaf forest than grassland (not shown), leading to large differences in the surface flux changes. However, this is the only pair with evergreen broadleaf forest, and its behavior in Noah-MP needs further investigation. Even though the pair 3 sites are close with small elevation difference, we found considerably different downwelling shortwave and longwave radiation between the two sites (not shown), which may explain the uncertainties in the Noah-MP simulations.

5. Discussion

This study has examined simulated changes in the surface energy budget in response to local land cover change based on paired proximate FLUXNET sites with differing land cover. Our results suggest that CLM well represents the observed changes in R_{net} and *G*; but there remain issues in simulating the energy partitioning between *LE* and *H*, which also further confirms the large uncertainties in simulated ET responses to LULCC revealed in several recent studies (e.g., Pitman et al. 2009; Boisier et al. 2012, 2014; de Noblet-Ducoudré et al. 2012; Vanden Broucke et al. 2015). Based on the observations, deforestation generally leads to a decrease in summer daytime R_{net}, accompanied by decreased *LE* and *H*. On one hand, CLM captures the observed signal of *H* change, but overestimates the decrease due to its large overestimation of *H* over the forest. On the other hand, the model underestimates the *LE* over the forest, leading to an opposite signal (a slight increase) of *LE* change comparing to the observations. Simulations in Noah-MP show similar biases. Therefore, uncertainties in current LULCC sensitivity studies may persist specifically in the representation of turbulent fluxes over forest land-cover types.

Scrutinizing the three components of ET suggests that the simulated increase in summer daytime *LE* is mainly attributable to a large increase in ground evaporation, which counteracts the decreased canopy evaporation and transpiration. This may raise another issue about the soil resistance parameterization in CLM4.5. Previous studies indicate that the model generates excessive ground evaporation when the canopy is sparse or absent (Swenson and Lawrence 2014; Tang et al. 2015). If there is overestimated ground evaporation over the open land, such a bias can also contribute to the disagreement in the LULCC-induced ET changes. Swenson and Lawrence (2014) have implemented a dry surface layer for the soil resistance parameterization to solve this issue for the upcoming CLM5. An extension of the evaluation with CLM5 would be useful to examine if the issue within the soil resistance parameterization is responsible for the uncertainties in ET changes.

Besides the uncertainties in estimating turbulent fluxes over different land cover types, the simulations show that differences in the meteorological forcings between nearby paired sites seem to have little impact on the simulation of surface flux changes due to LULCC. Many LSMs besides CLM employ a sub-grid tiling parameterization where multiple land surface types exist within a single grid box, each maintaining a separate set of surface balances and returning a weighted average set of fluxes to the atmosphere based on areal coverage of each surface type. In this arrangement, each land surface type within a grid box receives the same meteorological forcing from the overlying atmospheric model. It appears from our forcing-sensitivity studies that this arrangement does not significantly impact the simulation of surface flux changes associated with LULCC on the grid scale.

That said, the sub-grid comparison between different land cover types may yet be problematic due to the shared soil column issue for vegetated land units in CLM (Schultz et al. 2016). Both the single-point observations and simulations show significant differences in surface soil moisture between most of the paired sites, even though no clear drying or wetting pattern is found (Figure S4). The differences between the paired sites suggests that the shared soil column for vegetated land in CLM may not well represent soil moisture and temperature at the sub-grid scale, which may influence the simulations of land surface energy and water fluxes. We find an unreasonably large change in PFT-level G between forest and open land especially for the seasonal cycle in PFT simulations, while both observations, single-point and PFTCOL simulations show a seasonal change with a very small range (within ± $3 W/m^2$). As G is the calculated as the residual of the surface energy budget in CLM (Oleson et al. 2013), this sub-grid G issue may cast even more uncertainties on the calculation of LE and H at the PFT level, as well as their aggregated values at the grid level for regional or global simulations. Therefore, caution should be taken when examining the LULCC sensitivity which involves sub-grid PFT changes.

Compared with CLM, Noah-MP exhibits a similar ability to simulate surface flux changes, except for a deficiency in simulating H and R_{net} changes during late winter and early spring. We have examined the daytime albedo change after deforestation, calculated from available shortwave radiation terms, from observations and model simulations during local late winter/early spring (February ~ April, FMA) and summer (Figure 12). Both CLM and Noah-MP agree with the observations during summer. However, Noah-MP does not capture the observed albedo increase over nearly half of the pairs during late winter/early spring. Greater disagreement is also found during the local winter season (DJF, not shown), suggesting a deficiency in snowmelt timing or snow albedo sensitivity to LULCC, despite improvement in the snow surface albedo simulations by implementation of the Canadian Land Surface Scheme (CLASS; Verseghy, 1991) in Noah-MP (Niu et al. 2011).

Finally, it should be recognized that the observational data are not perfect. In particular, there may be systematic biases or even trends in specific instruments that contribute to the perceived differences between paired sites (e.g., site 3). Ideally, redundant instrumentation at sites, or in this case the rotation of an extra set of instruments among nearby paired sites, could be used to identify, quantify and account for significant systematic biases in measurements for suspicious variables. Furthermore, footprints of the flux towers may bias the comparison of surface fluxes between the open and forest sites (Baker et al. 2003; Griebel et al. 2016). In other words, the observed differences between sites can only be partially attributed to LULCC because their environmental conditions may also be different. 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 observed differences in surface climate conditions can be attributed to LULCC (e.g., Lejeune et al. 2017; Luyssaert et al. 2014; Teuling et al. 2010; Vanden Broucke et al 2015). Meanwhile, model simulations with the different forcings can effectively examine the effects of the local environment of individual sites, because their footprints can also be taken by the meteorological measurements. Our results show robust signals of LULCC-induced changes in surface fluxes, implying that impacts of footprints at individual sites are probably trivial.

6. Conclusions

This study has evaluated the performance of two state-of-the-art LSMs in simulating the LULCC-induced changes in surface energy fluxes. Observations from 28 FLUXNET pairs (open versus forest) are used to represent the observed flux changes following deforestation, which are compared with the LSM simulations forced with meteorological data from the observation sites. Diurnal and seasonal cycles of the flux changes have been investigated.

The single-point simulations in CLM and Noah-MP show the greatest bias in simulating *LE* change. Significantly decreased daytime *LE* is observed during local summer, but not captured by the models. The observed *LE* changes also exhibit an evident seasonality, which is not represented in the model. The energy partitioning between *LE* and *H* might be a common issue within the LSMs. Other studies have noted problems in the simulation of surface fluxes by LSMs, including poor performance relative to non-physical statistical models (Best et al. 2015, Haughton et al. 2016).

The sub-grid comparison from the global simulations in CLM yields unrealistic changes in *G* and *H* when the soil column is shared among vegetated land units, even though there is a better agreement in *LE* change with the observations. The individual-soil-column scheme improves the representation of the PFT-level energy flux changes, but uncertainties still remain as with the point-scale simulations. Therefore, these uncertainties must be considered when interpreting global experiments of LULCC sensitivity studies with current LSMs.

Consistent aggregate performance across many paired sites suggests the problems in these LSMs may not lie primarily with parameter selection at individual sites, but with more fundamental issues of the representation of physical processes in LSMs. The simulation of LULCC may or may not have become more consistent among models since LUCID (de Noblet-Ducoudré et al. 2012), but consistency with observed biophysical responses appears to be lacking. LUMIP (Lawrence et al. 2016) will be a step toward better LSM simulation of LULCC responses, and ultimately better simulations of the response of climate to LULCC.

Acknowledgements:

This study was supported by the National Science Foundation (AGS-1419445). This work used eddy covariance data acquired and shared by the FLUXNET community, including these networks: AmeriFlux, CarboEuropeIP, CarboItaly, CarboMont, Fluxnet-Canada, GreenGrass, ICOS, and OzFlux-TERN. The ERA-Interim reanalysis data are provided by ECMWF and processed by LSCE. The FLUXNET eddy covariance data processing and harmonization was carried out by the European Fluxes Database Cluster, AmeriFlux Management Project, and Fluxdata project of FLUXNET, with the support of CDIAC and ICOS Ecosystem Thematic Center, and the OzFlux, ChinaFlux and AsiaFlux offices. We thank all site investigators and flux networks for their work to make our model evaluation possible. The authors wish to thank Ahmed Tawfik at the National Center for Atmospheric Research for his assistance of preparing forcing datasets for the AmeriFlux sites. Computing resources for the CLM and Noah-MP experiments were provided by the NSF/CISL/Yellowstone supercomputing facility.

REFERENCES

Baker, I., A. S. Denning, N. Hanan, L. Prihodko, M. Uliasz, P. Vidale, K. Davis, and P. Bakwin, 2003: Simulated and observed fluxes of sensible and latent heat and CO2 at the WLEF-TV tower using SiB2.5. Global Change Biol., **9**, 1262-1277, doi:10.1046/j.1365-2486.2003.00671.x.

Baldocchi, D. and Coauthors, 2001: FLUXNET: A New Tool to Study the Temporal and Spatial Variability of Ecosystem–Scale Carbon Dioxide, Water Vapor, and Energy Flux Densities. *Bull. Amer. Meteor. Soc.,* **82**, 2415-2434, doi:10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2.

Betts, A. K., 2004: Understanding Hydrometeorology Using Global Models. *Bull. Amer. Meteor. Soc.,* **85**, 1673-1688, doi:10.1175/BAMS-85-11-1673.

Boisier, J., N. de Noblet-Ducoudré, and P. Ciais, 2014: Historical land-use-induced evapotranspiration changes estimated from present-day observations and reconstructed land-cover maps. *Hydrology and Earth System Sciences,* **18**, 3571-3590,

Boisier, J., N. de Noblet-Ducoudré, and P. Ciais, 2013: Inferring past land use-induced changes in surface albedo from satellite observations: a useful tool to evaluate model simulations. *Biogeosciences,* **10**, 1501-1516,

Boisier, J. P. and Coauthors, 2012: Attributing the impacts of land-cover changes in temperate regions on surface temperature and heat fluxes to specific causes: Results from the first

LUCID set of simulations. *Journal of Geophysical Research: Atmospheres,* **117**, n/a-n/a, doi:10.1029/2011JD017106.

Bright, R. M., K. Zhao, R. B. Jackson, and F. Cherubini, 2015: Quantifying surface albedo and other direct biogeophysical climate forcings of forestry activities. *Global Change Biol.,* **21**, 3246-3266, doi:10.1111/gcb.12951.

Brovkin, V. and Coauthors, 2013: Effect of Anthropogenic Land-Use and Land-Cover Changes on Climate and Land Carbon Storage in CMIP5 Projections for the Twenty-First Century. *J. Climate,* **26**, 6859-6881, doi:10.1175/JCLI-D-12-00623.1.

Cai, X., Z. Yang, Y. Xia, M. Huang, H. Wei, L. R. Leung, and M. B. Ek, 2014: Assessment of simulated water balance from Noah, Noah-MP, CLM, and VIC over CONUS using the NLDAS test bed. *Journal of Geophysical Research: Atmospheres,* **119**, 13,751-13,770, doi:10.1002/2014JD022113.

Chen, L., and P. A. Dirmeyer, 2016: Adapting observationally based metrics of biogeophysical feedbacks from land cover/land use change to climate modeling. *Environmental Research Letters,* **11**, 034002, doi:10.1088/1748-9326/11/3/034002.

Chen, L., and P. A. Dirmeyer, 2016: Impacts of Land Use/Land Cover Change on Afternoon Precipitation over North America. *J. Climate,* doi:10.1175/JCLI-D-16-0589.1.

de Noblet-Ducoudré, N. and Coauthors, 2012: Determining Robust Impacts of Land-Use-Induced Land Cover Changes on Surface Climate over North America and Eurasia:

Results from the First Set of LUCID Experiments. *J. Climate,* **25**, 3261-3281, doi:10.1175/JCLI-D-11-00338.1.

Dirmeyer, P. A., R. D. Koster, and Z. Guo, 2006: Do Global Models Properly Represent the Feedback between Land and Atmosphere? *J. Hydrometeor.,* **7**, 1177-1198, doi:10.1175/JHM532.1.

Foken, T., 2008: THE ENERGY BALANCE CLOSURE PROBLEM: AN OVERVIEW. *Ecol. Appl.,* **18**, 1351-1367, doi:10.1890/06-0922.1.

Gochis, D.J., W. Yu, and D.N. Yates, 2015. The WRF-Hydro Model Technical Description and User's Guide, Version 3.0. NCAR Technical Document, 120 pp., [Available at https://www.ral.ucar.edu/sites/default/files/public/images/project/WRF_Hydro_User_Guide_v3.0.pdf].

Griebel, A., L. T. Bennett, D. Metzen, J. Cleverly, G. Burba and S. K. Arndt, 2016: Effects of inhomogeneities within the flux footprint on the interpretation of seasonal, annual, and interannual ecosystem carbon exchange. Agricultural and Forest Meteorology, **221**, 50-60, doi:10.1016/j.agrformet.2016.02.002.

Janowiak, J. E., V. E. Kousky, and R. J. Joyce, 2005: Diurnal cycle of precipitation determined from the CMORPH high spatial and temporal resolution global precipitation analyses. *Journal of Geophysical Research: Atmospheres,* **110**, n/a-n/a, doi:10.1029/2005JD006156.

Juang, J., G. Katul, M. Siqueira, P. Stoy, and K. Novick, 2007: Separating the effects of albedo from eco-physiological changes on surface temperature along a successional chronosequence in the southeastern United States. *Geophys. Res. Lett.,* **34**, n/a-n/a, doi:10.1029/2007GL031296.

Lawrence, D. M. and Coauthors, 2016: The Land Use Model Intercomparison Project (LUMIP) contribution to CMIP6: rationale and experimental design. *Geoscientific Model Development,* **9**, 2973-2998,

Lawrence, D., and K. Vandecar, 2015: Effects of tropical deforestation on climate and agriculture. *Nature climate change,* **5**, 27-36,

Lawrence, P. J., and T. N. Chase, 2010: Investigating the climate impacts of global land cover change in the community climate system model. *Int. J. Climatol.,* **30**, 2066-2087, doi:10.1002/joc.2061.

Lawrence, P. J. and Coauthors, 2012: Simulating the Biogeochemical and Biogeophysical Impacts of Transient Land Cover Change and Wood Harvest in the Community Climate System Model (CCSM4) from 1850 to 2100. *J. Climate,* **25**, 3071-3095, doi:10.1175/JCLI-D-11-00256.1.

Lee, X. and Coauthors, 2011: Observed increase in local cooling effect of deforestation at higher latitudes. *Nature,* **479**, 384-387, doi: 10.1038/nature10588.

Lejeune, Q., S. I. Seneviratne, and E. L. Davin, 2016: Historical land-cover change impacts on climate: comparative assessment of LUCID and CMIP5 multi-model experiments. *J. Climate,* doi:10.1175/JCLI-D-16-0213.1.

Li, Y., M. Zhao, S. Motesharrei, Q. Mu, E. Kalnay, and S. Li, 2015: Local cooling and warming effects of forests based on satellite observations. *Nature communications,* **6**, doi: 10.1038/ncomms7603.

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.

Mahmood, R. and Coauthors, 2014: Land cover changes and their biogeophysical effects on climate. *Int. J. Climatol.,* **34**, 929-953, doi:10.1002/joc.3736.

Niu, G. and Coauthors, , 2011: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research: Atmospheres,* **116**, D12109, doi:10.1029/2010JD015139.

Oleson, K. W., and co-authors, 2013: Technical Description of version 4.5 of the Community Land Model (CLM), NCAR Technical Note, TN-503+STR, National Center for Atmospheric Research, Boulder, CO, USA, 434pp., [Available at http://www.cesm.ucar.edu/models/cesm1.2/clm/CLM45_Tech_Note.pdf].

Pielke, R. A. and Coauthors, , 2011: Land use/land cover changes and climate: modeling analysis and observational evidence. *Wiley Interdisciplinary Reviews: Climate Change,* **2**, 828-850, doi:10.1002/wcc.144.

Pitman, A. J. and Coauthors, , 2009: Uncertainties in climate responses to past land cover change: First results from the LUCID intercomparison study. *Geophys. Res. Lett.,* **36**, n/a-n/a, doi:10.1029/2009GL039076.

Portmann, F. T., S. Siebert, and P. Döll, 2010: MIRCA2000—Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling. *Global Biogeochem. Cycles,* **24**, L14814, doi:10.1029/2008GB003435.

Reichstein, M. and Coauthors, , 2005: On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. *Global Change Biol.,* **11**, 1424-1439, doi:10.1111/j.1365-2486.2005.001002.x.

Santanello, J. A., and M. A. Friedl, 2003: Diurnal Covariation in Soil Heat Flux and Net Radiation. *J. Appl. Meteor.,* **42**, 851-862, doi:10.1175/1520-0450(2003)042<0851:DCISHF>2.0.CO;2.

Schultz, N. M., X. Lee, P. J. Lawrence, D. M. Lawrence, and L. Zhao, 2016: Assessing the use of subgrid land model output to study impacts of land cover change. *Journal of Geophysical Research: Atmospheres,* **121**, 6133-6147, doi:10.1002/2016JD025094.

Swenson, S. C., and D. M. Lawrence, 2014: Assessing a dry surface layer-based soil resistance parameterization for the Community Land Model using GRACE and FLUXNET-MTE data. *Journal of Geophysical Research: Atmospheres,* **119**, 10,299-10,312, doi:10.1002/2014JD022314.

Tang, J., W. J. Riley, and J. Niu, 2015: Incorporating root hydraulic redistribution in CLM4.5: Effects on predicted site and global evapotranspiration, soil moisture, and water storage. *Journal of Advances in Modeling Earth Systems,* **7**, 1828-1848, doi:10.1002/2015MS000484.

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.

Verseghy, D. L., 1991: Class—A Canadian land surface scheme for GCMS. I. Soil model. *Int. J. Climatol.,* **11**, 111-133, doi:10.1002/joc.3370110202.

Viovy, N., 2011. CRUNCEP data set for 1901–2008, [Available at https://www.earthsystemgrid.org/dataset/ucar.cgd.ccsm4.CRUNCEP.v4.html].

Vuichard, N., and D. Papale, 2015: Filling the gaps in meteorological continuous data measured at FLUXNET sites with ERA-Interim reanalysis. *Earth System Science Data,* **7**, 157-171, doi:10.5194/essd-7-157-2015.

Williams, C. A. and Coauthors, 2012: Climate and vegetation controls on the surface water balance: Synthesis of evapotranspiration measured across a global network of flux towers. *Water Resour. Res.,* **48**, W06523, doi:10.1029/2011WR011586.

Wilson, K. and Coauthors, 2002: Energy balance closure at FLUXNET sites. *Agric. For. Meteorol.,* **113**, 223-243, doi:10.1016/S0168-1923(02)00109-0.

Xu, Z., R. Mahmood, Z. Yang, C. Fu, and H. Su, 2015: Investigating diurnal and seasonal climatic response to land use and land cover change over monsoon Asia with the Community Earth System Model. *Journal of Geophysical Research: Atmospheres,* **120**, 1137-1152, doi:10.1002/2014JD022479.

Zhang, L. and Coauthors, , 2016: Evaluation of the Community Land Model simulated carbon and water fluxes against observations over ChinaFLUX sites. *Agric. For. Meteorol.,* **226–227**, 174-185, doi:10.1016/j.agrformet.2016.05.018.

Zhou, C., and K. Wang, 2016: Biological and Environmental Controls on Evaporative Fractions at AmeriFlux Sites. J. Appl. Meteor. Climatol., **55**, 145-161, doi:10.1175/JAMC-D-15-0126.1.

Table 1. Information about the variables used from FLUXNET2015. The marginal distribution sampling (MDS) filling method is based on Reichstein et al. (2005); and the ERA-interim filling method can be found in Vuichard and Papale (2015).

Name	Gap-filling	Description
SW_IN_F	MDS and ERA-interim	downwelling shortwave radiation
LW_IN_F	MDS and ERA-interim	downwelling longwave radiation
PA_F	MDS and ERA-interim	atmospheric pressure
TA_F	MDS and ERA-interim	air temperature
VPD_F	MDS and ERA-interim	vapor pressure deficit
P_F	ERA-interim	precipitation
WS_F	ERA-interim	wind speed
LE_F_MDS	MDS	latent heat flux
H_F_MDS	MDS	sensible heat flux
G_F_MDS	MDS	ground heat flux
NETRAD	n/a	net radiation
LE_CORR	n/a	corrected LE_F_MDS by energy balance closure correction factors. LE_CORR_25, LE_CORR, and LE_CORR_75 are calculated based on 25, 50, and 75th percentiles of the factors, respectively.
H_CORR	n/a	corrected H_F_MDS by energy balance closure correction factors. H_CORR_25, H_CORR, and H_CORR_75 are calculated based on 25, 50, and 75th percentiles of the factors, respectively.

Table 2. Information about model simulations. "Nearby" observations indicate that the paired sites have the identical forcings either from the companion forest or open sites.

Name	Forcings	Description
CLM	observations from individual sites	single-point CLM simulations with its own observations
CLM_forest	observations only from forest sites	single-point CLM simulations with the (nearby) forest observations
CLM_open	observations only from open sites	single-point CLM simulations with the (nearby) open land observations
CLM-PFT	CRUNCEP	global CLM simulations with default soil-column scheme with PFT-level output
CLM-PFTCOL	CRUNCEP	global CLM simulations with default individual-soil-column scheme scheme with PFT-level output
NOAH-MP	observations from individual sites	single-point NOAH-MP simulations with its own observations
NOAH-MP_forest	observations only from forest sites	single-point NOAH-MP simulations with the (nearby) forest observations
NOAH-MP_open	observations only from open sites	single-point NOAH-MP simulations with the (nearby) open land observations

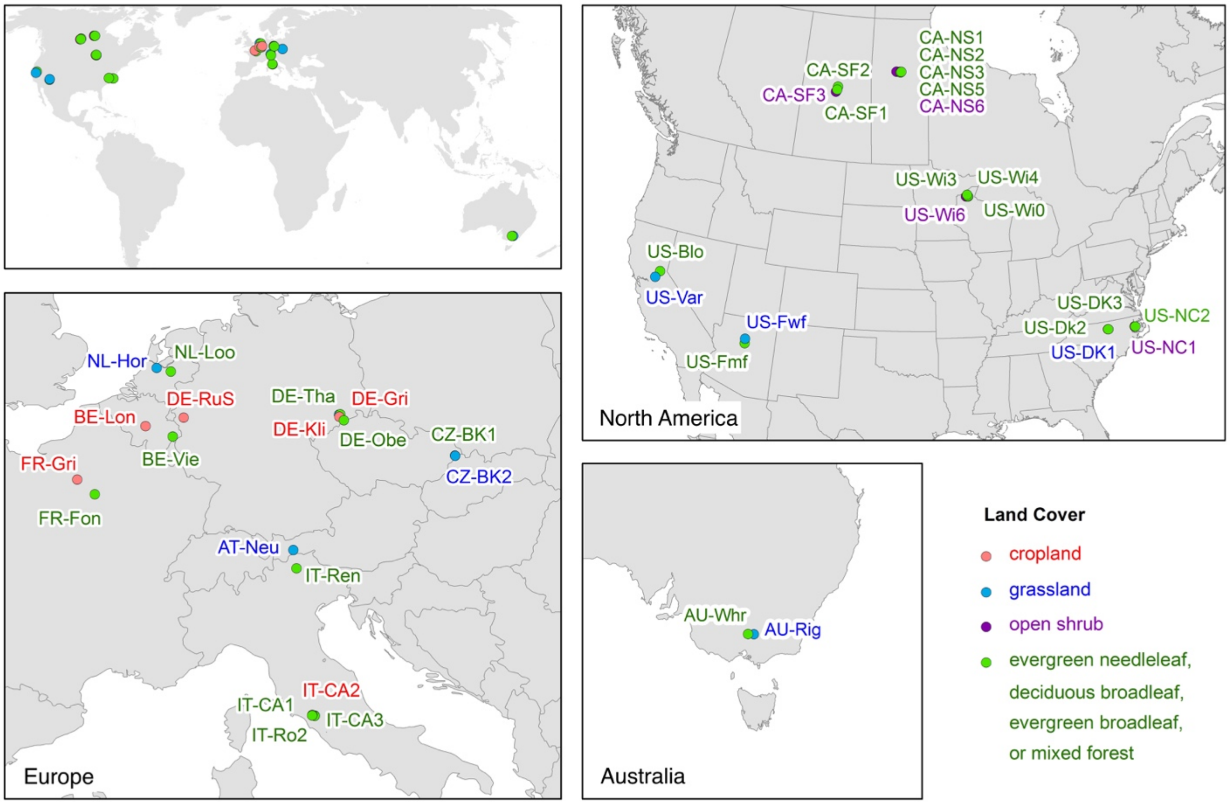

Figure 1. Location and land cover type of the paired sites. The land cover type of each site is based on the reported land cover in FLUXNET database.

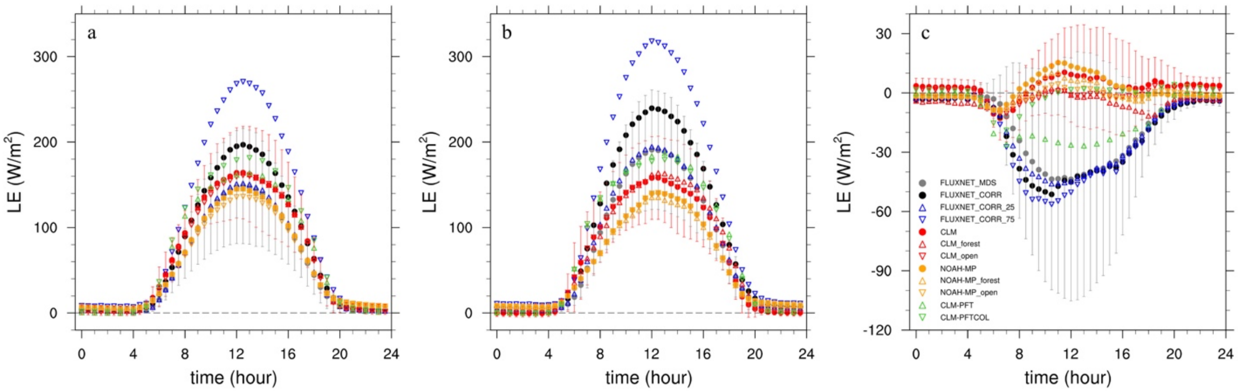

Figure 2. The diurnal cycle of *LE* (W/m^2) averaged over all the open sites (a) and forest sites (b) and their difference (open – forest, c) during the summer. The gray error bars indicate the standard deviation of the observed LE (MDS) among the sites; the red error bars are for the simulated *LE* in the CLM case. Details about the four types of FLUXNET observations can be found in Table 1. Information about model simulations in CLM and Noah is described in Table 2.

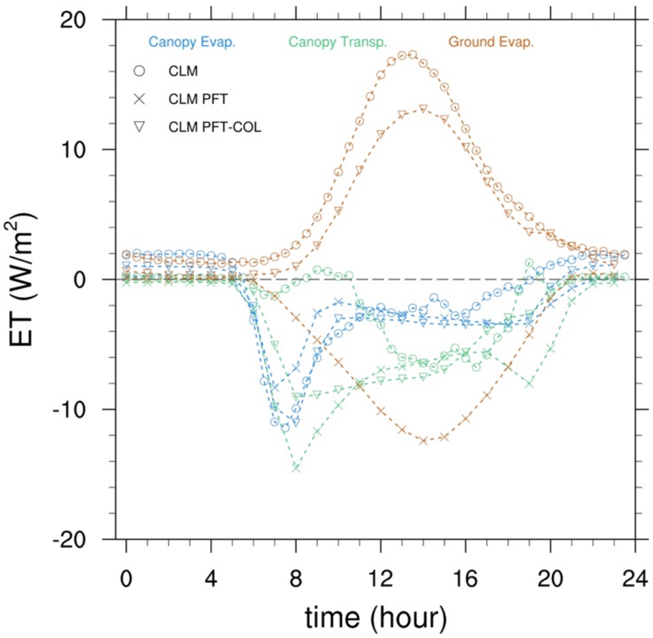

Figure 3. Change in the diurnal cycle of components (colors) of evapotranspiration (canopy evaporation, canopy transpiration, and ground evaporation) due to LULCC from forest to open land (open – forest).

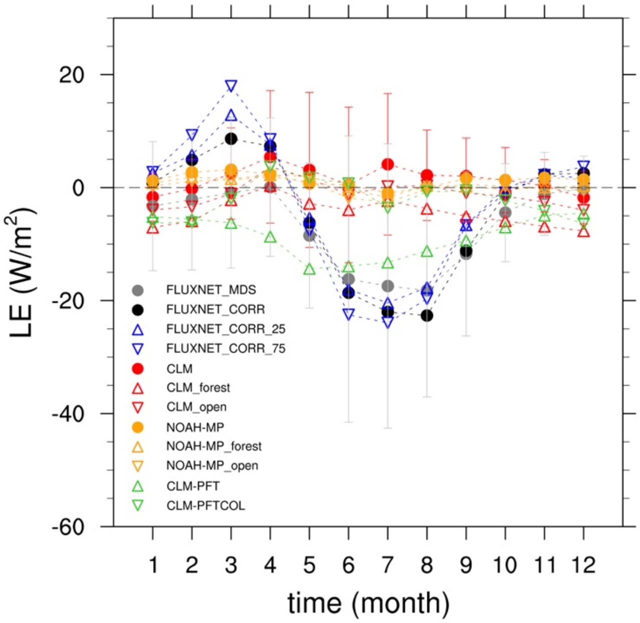

Figure 4. Change in the seasonal cycle of *LE* (W/m^2) due to LULCC from forest to open land (open – forest).

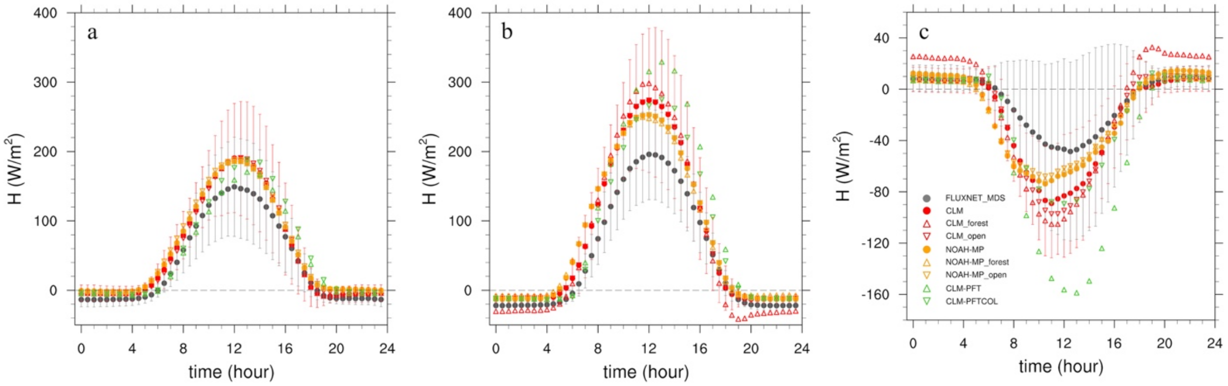

Figure 5. The diurnal cycle of *H* (W/m^2) averaged over all the open sites (a) and forest sites (b) and their difference (open – forest, c) during the summer. The gray error bars indicate the standard deviation of the observed *H* among the sites; the red error bars are for the simulated *H* in the CLM case.

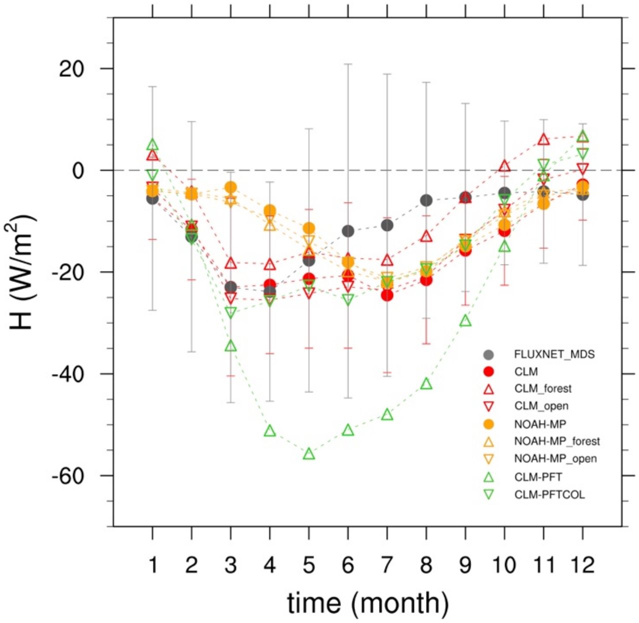

Figure 6. Change in the seasonal cycle of *H* (W/m^2) due to LULCC from forest to open land (open – forest).

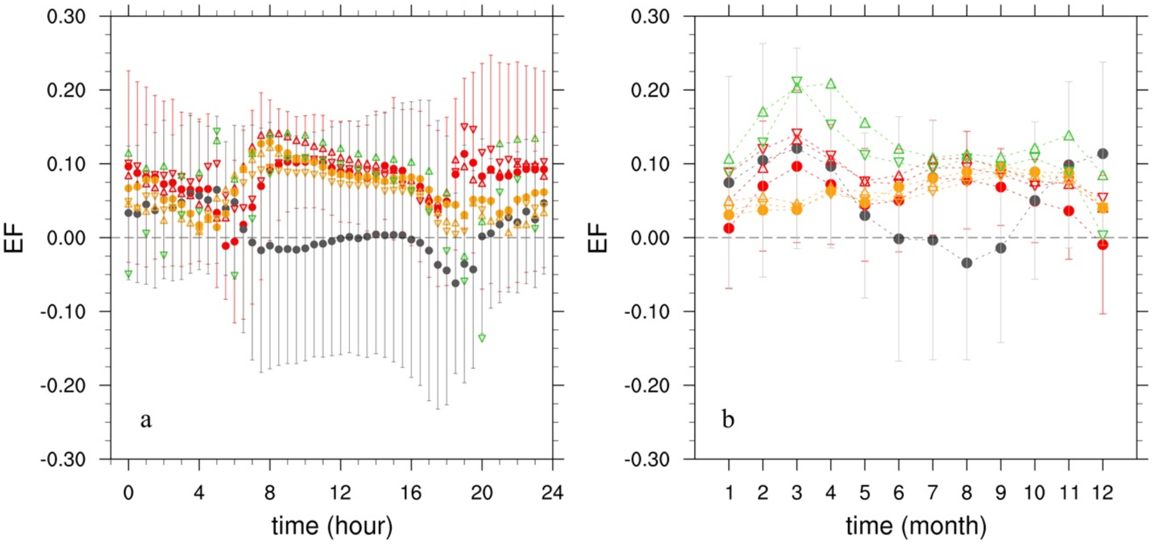

Figure 7. Change in the summer diurnal (a) and seasonal (b) cycle of *EF* (unitless) due to LULCC from forest to open land (open – forest).

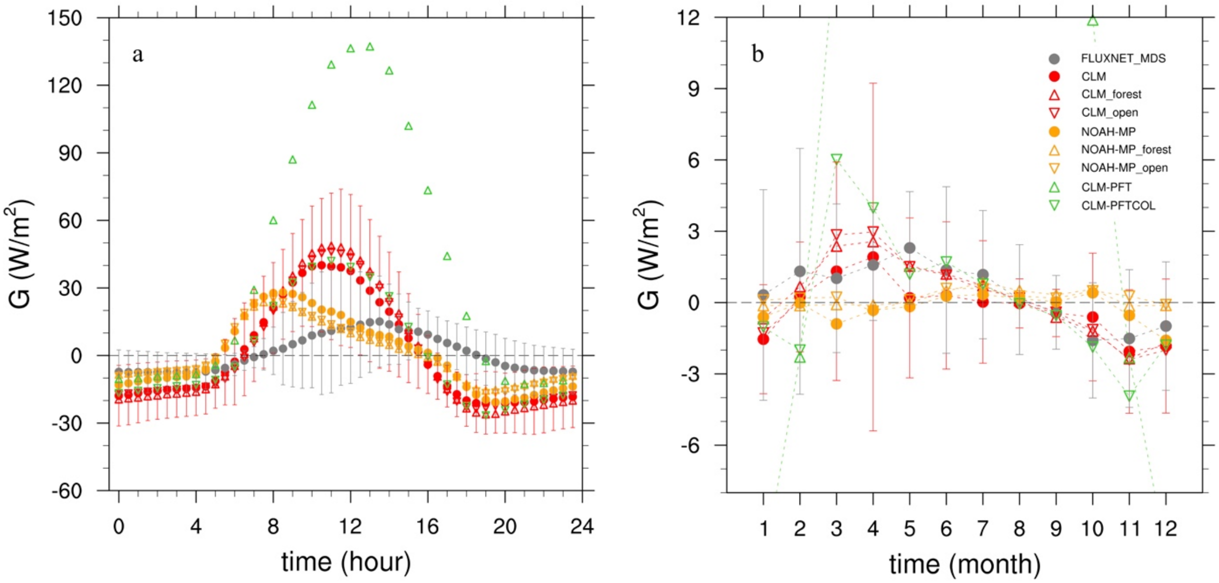

Figure 8. Change in the summer diurnal (a) and seasonal (b) cycle of *G* (W/m^2) due to LULCC from forest to open land (open – forest). It should be noted that the changes in the CLM-PFT simulation are much further from the observations than the other simulations. Some of its values are beyond the limit of the figure (b). The smallest value is -11.2 W/m^2 in January; while the largest value is 52.9 W/m^2 in May.

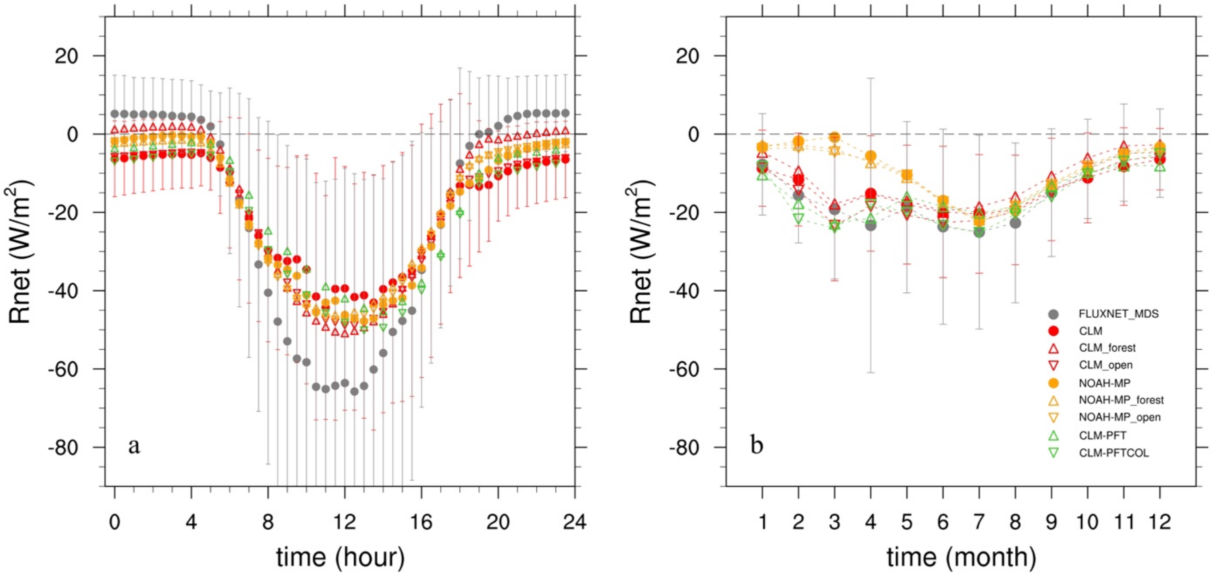

Figure 9. Change in the summer diurnal (a) and seasonal (b) cycle of R_{net} (W/m^2) due to LULCC from forest to open land (open – forest).

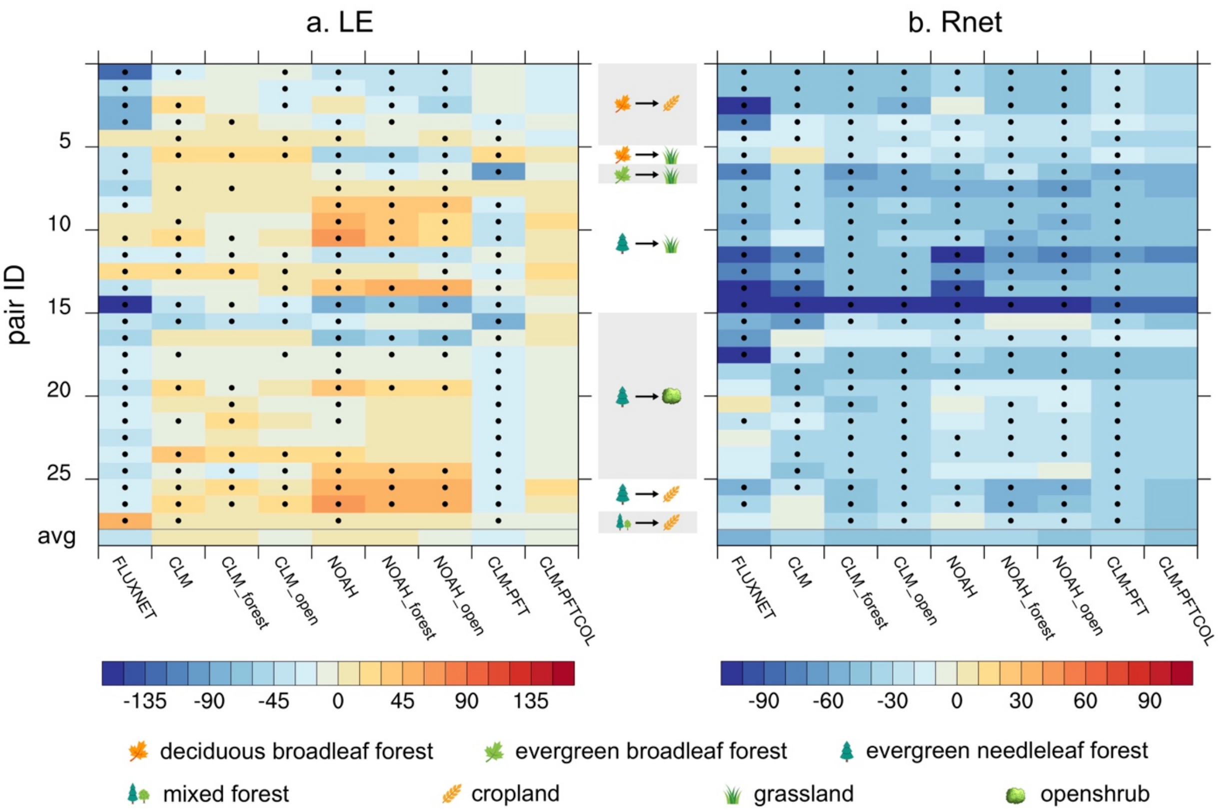

Figure 10. Change (open – forest) in observed and simulated *LE* (a) and R_{net} (b) during summer daytime (averaged during the period 08:00 ~ 16:00) over individual pairs and their averages. The vertical labels show the pair ID from 1 to 28 based on Table S1. The pairs are grouped based on the type of LULCC (shown as the icons in the middle). The bottom row is the average over all pairs. The Student's t-test is performed on the daily (daytime average) time series for each pair. Dots indicate statistically significant changes at the 95% confidence level. No significant test is carried out for the CLM-PFTCOL simulation (the last column), because we only have long-term averaged hourly output for each month.

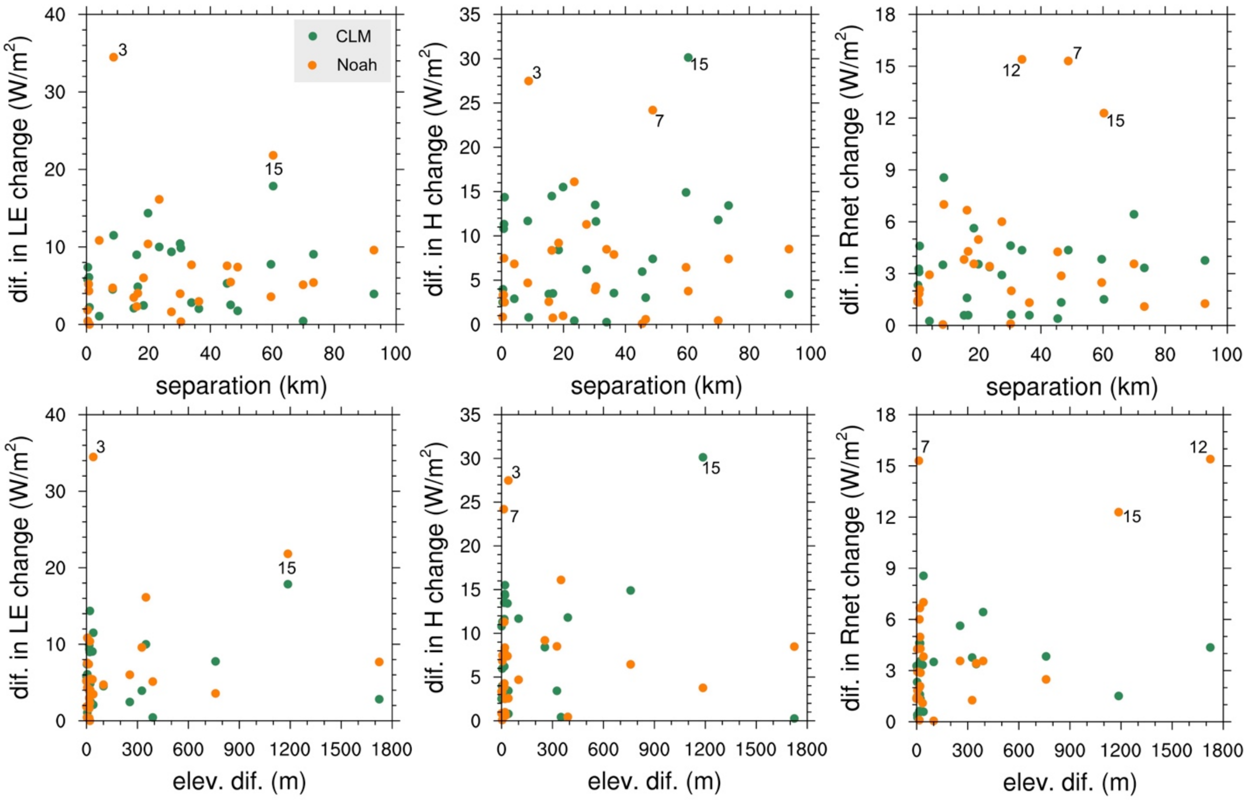

Figure 11. Sensitivity of differences in simulated surface energy flux changes (left column: *LE*, middle: *H*, and right: R_{net}) between "forest forcing" and "open forcing" simulations to site separation (top) and elevation difference (bottom) between the forest and open sites in individual pairs. The pairs No. 3, 7, 12 and 15 are labeled because of the greatest differences in surface fluxes changes.

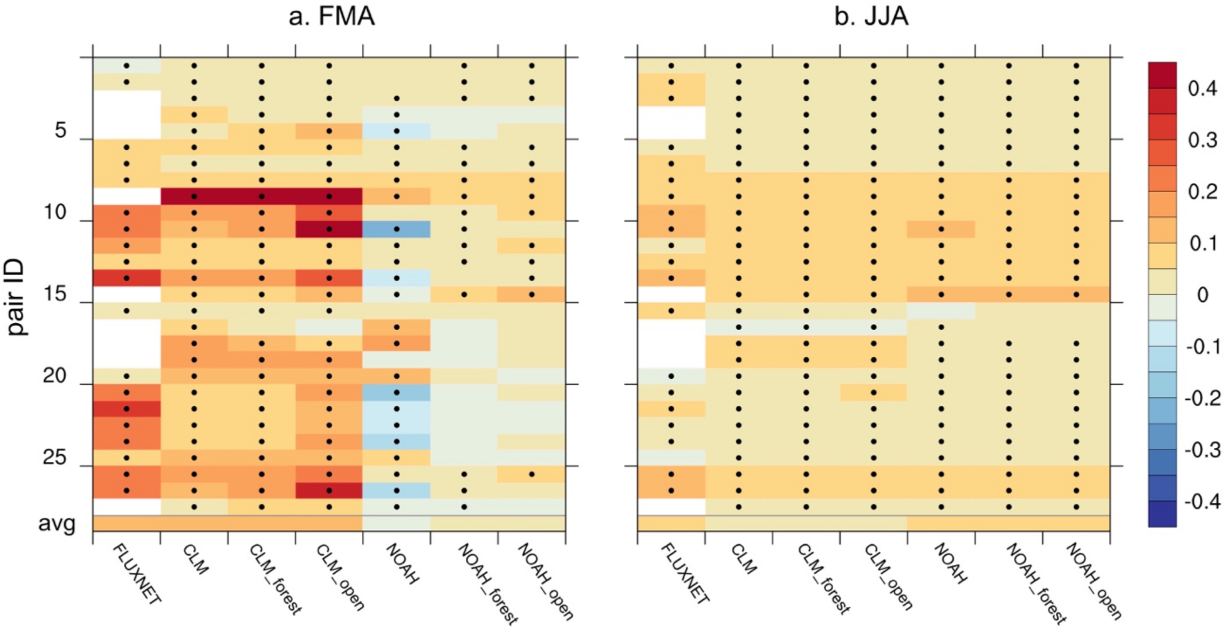

Figure 12. Change (open – forest) in observed and simulated daytime albedo during late winter/early spring (FMA, a) and summer (JJA, b). White areas indicate missing observations.