We thank the two reviewers for their constructive comments. We have incorporated the review comments and revised the manuscript thoroughly. The review comments and the revision have resulted in a much more complete presentation of the work. While the changes made to the manuscript can be seen in the revised manuscript, we also present here our detailed responses to the review comments (reviewer comments in black, our response in blue).

Responses to the comments from Reviewer #1

GENERAL COMMENTS

The paper describes a very simple approach to attribute model biases in the simulated states and fluxes of the latest version of the Community Land Model (CLM4.5). This is an important and interesting research area, as biases in modelled soil moisture or discharge can for instance substantially affect the prediction and analysis of hydro-climatic extremes such as droughts and/or floods. The approach introduced in the paper is not really innovative as it was first published by Parr et al. in 2015; but it is tested here for a larger study area and a different land-surface model. In general, the method and the results in this paper are well-described, but–to my opinion–not really surprising and rather straightforward. Substantial parts of the results and discussions are dedicated to the differences in bias between the GLEAM-derived datasets and the CLM-runs with and without the bias correction. These results are very straightforward and predictable, as the bias-correction factors were first calibrated against GLEAM. Furthermore, most of the validations/comparisons are performed at aggregated variables (both in space and time), which might mask some of the potential issues. Summarized, I think the topic of this study is interesting, but I have the feeling that the paper (especially the results section) needs some improvements before final publication. Below I list some more specific comments.

SPECIFIC COMMENTS

1. In Section 4.2.1 it is claimed several times that the performance of CLMET is substantially better as compared to the original CLM. To my opinion, these statements need to be revised as they are not necessarily correct; especially not when the reference data is the GLEAM dataset itself. As the bias-correction factors are calculated using the GLEAM data as a reference, it makes perfect sense that applying these correction factors in the model brings the model closer to GLEAM (unless the assumption of time-invariance would not be fulfilled). Therefore, the results discussed from P11-L243 to P14-L305 (i.e. comparison of the bias-corrected CLM evaporation to the GLEAM dataset) only show the robustness of the correction factors. They do not show an improvement of CLMET in reference to CLM. To me, the evaluation of the runoff coefficients and the comparison against alternative datasets of evaporation (FLUXNETMTE, MODIS) is a step in the right direction, but only a small portion of the discussion is dedicated to these results. Therefore, I would suggest to improve the evaluation of the results to really show the impact of applying the method. I would strongly recommend to (1) validate the modelled evaporation against in situ

measurements (for instance data from single eddy-covariance towers) and, (2) extend the evaluation of the model against the alternative datasets of evaporation.

Response: We have followed the reviewer's suggestions in revising the manuscript:

1) Validate the modelled evaporation against in situ measurements:

We selected 16 eddy flux tower stations from the AmeriFlux network to validate model performance (as shown in Figure 1b of the revised manuscript). These stations were previously used to validate the NLDAS-2 surface models by Xia et al. (2015). The 16 stations are located in different sub regions of CONUS with different vegetation cover (i.e., grassland, cropland, needleleaf forest, broadleaf forest, and mixed forest). Considering both consistency in validation period and data availability, we use the year of 2005 for validation at most sites except for three sites: Sylvania Wilderness (2002), Donaldson (2004) and Walnut River (2004).

The model validations are based comparing each station with the model grid cell that encompasses the station. The station-based ET (or latent heat flux, in W/m2) are measured every 30 minutes and aggregated to daily and monthly values. Except for Port Peck and Wind River Crane stations in the northwest CONUS, for all other stations the monthly mean ET from CLMET agrees better with the observed ET than that from CLM (Figure 8 of the revised manuscript). The same statement holds for daily mean ET (Figures 9, and 10 of the revised manuscript). Generally, CLM overestimates ET as compared with station observations, and CLMET alleviates this overestimation, which is consistent with comparisons between modelled ET and satellite-based ET products.

"*In addition, the ET validation is also conducted on the site scale (Figures 8, 9, and 10).* Except for Port Peck and Wind River Crane stations in the northwest CONUS, for all other stations the monthly mean ET from CLMET agrees better with the observed ET than that from CLM *(Figure 8). The same statement holds for daily mean ET (Figures 9 and 10). Generally, CLM overestimates ET as compared with station observations, and CLMET alleviates this overestimation, which is consistent with comparisons between modelled ET and satellite-based ET products.*" (last paragraph of Section 4.2.1 in the revised manuscript)

Xia, Y., Hobbins, M. T., Mu, Q., & Ek, M. B. (2015). Evaluation of NLDAS-2 evapotranspiration against tower flux site observations. Hydrological Processes, 29(7), 1757-1771.

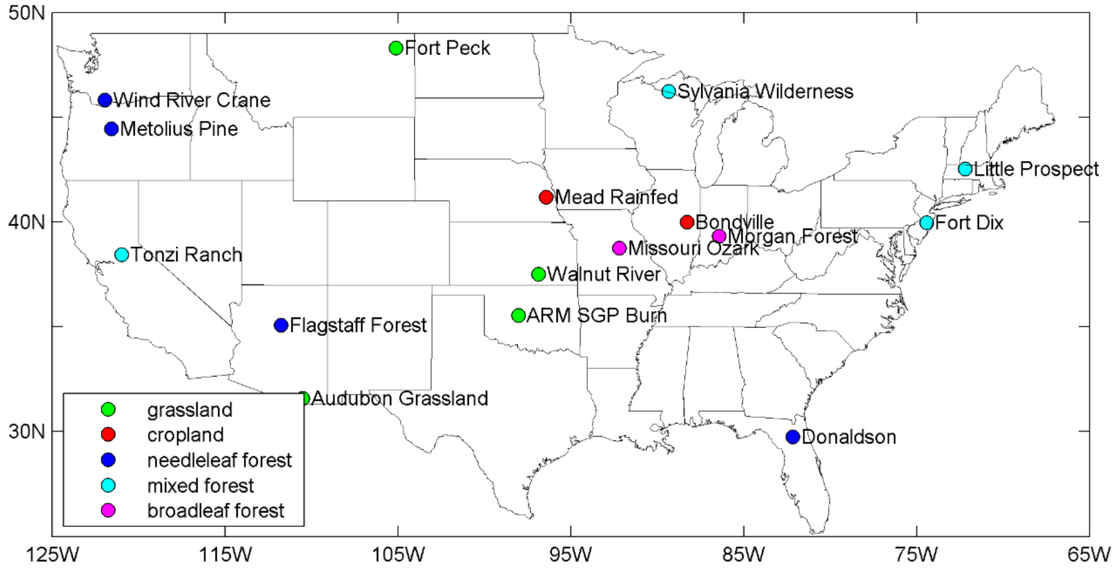


Figure 1b Locations of the 16 AmeriFlux stations with vegetation types.

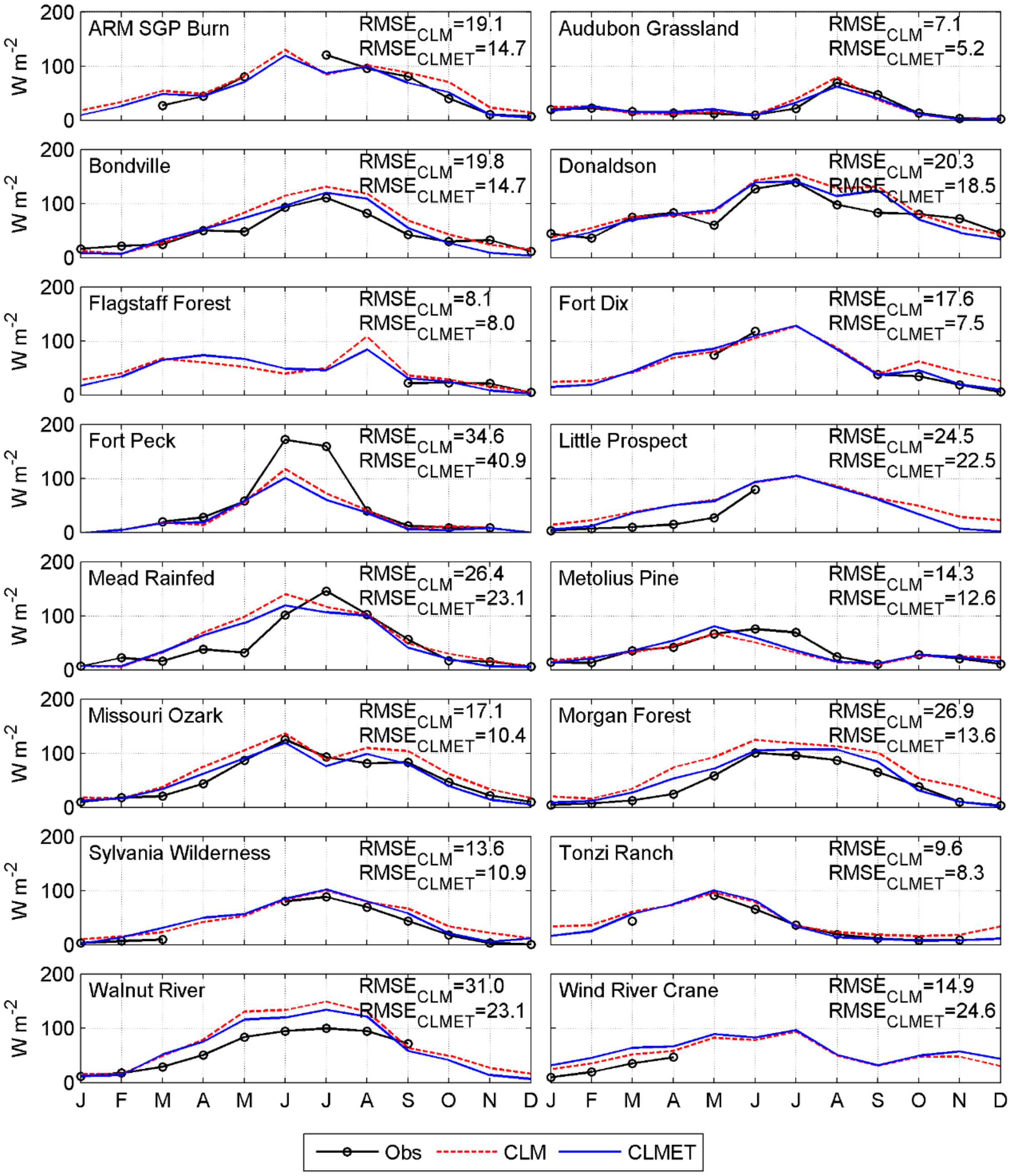


Figure 8 Monthly mean latent heat fluxes from CLM, CLMET and observations at 16 flux tower sites. RMSE$_{CLM}$ and RMSE$_{CLMET}$ represent the root mean square error against observations for CLM and CLMET, respectively. Note that the CLM and CLMET simulations are driven with meteorological forcing at the grid cell level (as opposed to site-specific forcing).

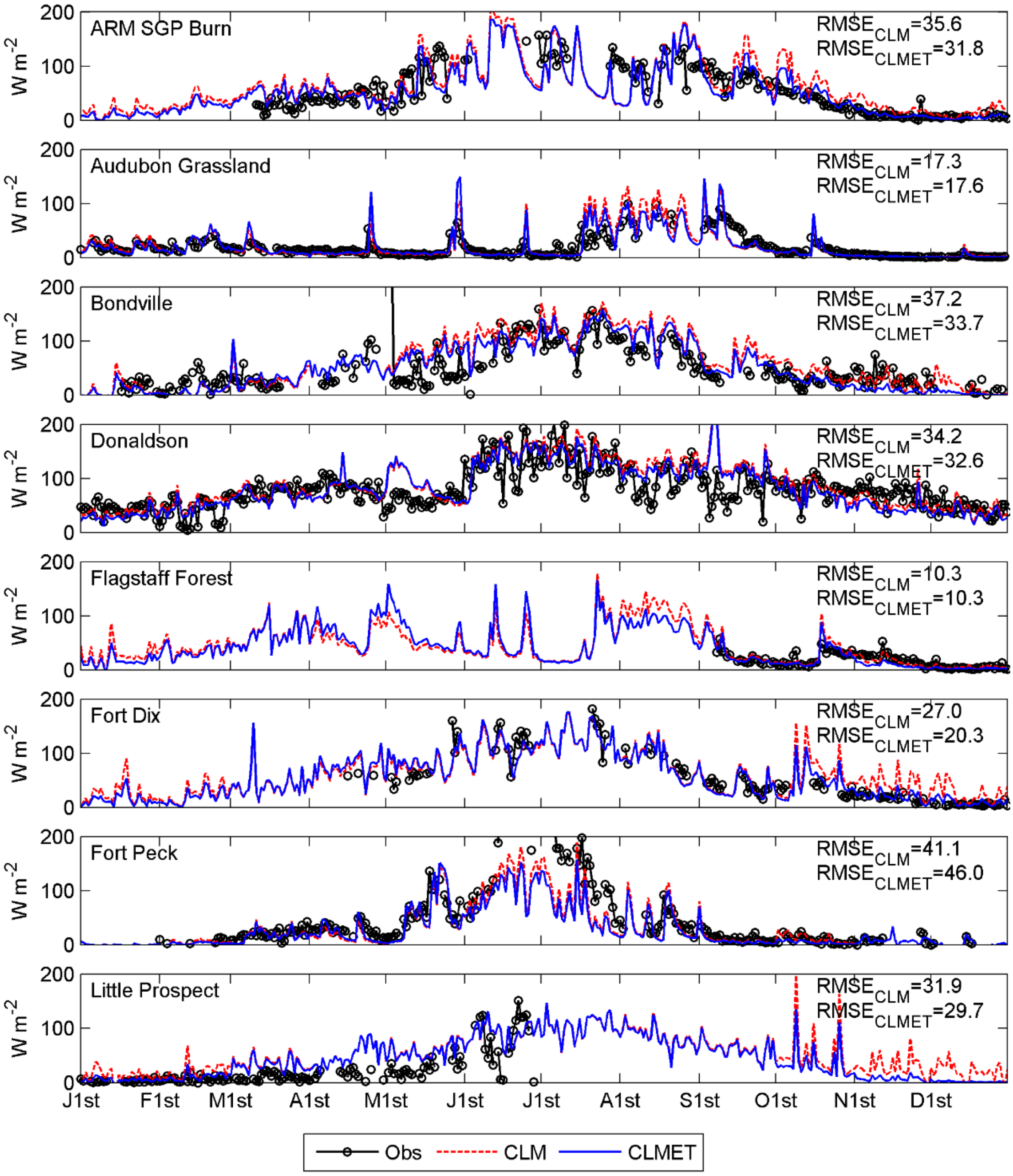


Figure 9 Daily mean latent heat fluxes from CLM and CLMET grids and station observations at ARM SGP Burn, Audubon Grassland, Bondville, Donaldson, Flagstaff Forest, Fort Dix, Fort Peck, and Little Prospect. $RMSE_{CLM}$ and $RMSE_{CLMET}$ represent the root mean square error against observations for CLM and CLMET, respectively.

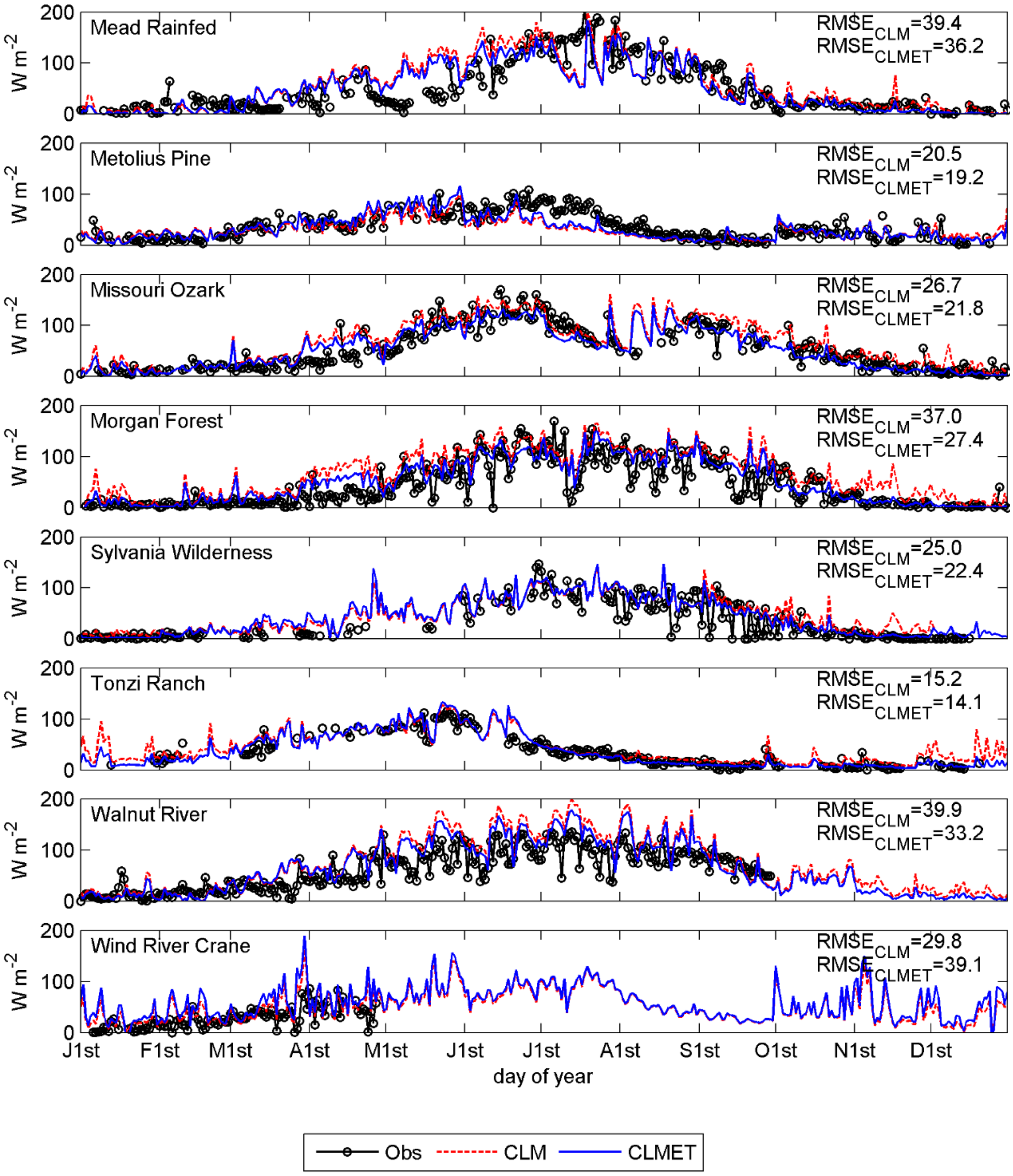


Figure 10 Daily mean latent heat fluxes from CLM and CLMET grids and station observations at Mead Rainfed, Metolius Pine, Missouri Ozark, Morgan Forest, Sylvania Wilderness, Tonzi Ranch, Walnut River, and Wind River Crane. RMSE$_{CLM}$ and RMSE$_{CLMET}$ represent the root mean square error against observations for CLM and CLMET, respectively.

2) extend the evaluation of the model against the alternative datasets of evaporation:

We have deleted the evaluations of ET seasonal cycle and monthly value using the GLEAM dataset, and added the evaluations using the MODIS and FLUXNET-MTE dataset. Therefore GLEAM is used for algorithm calibration while the other two ET products are used for validation. Using MODIS or FLUXNET-MTE ET as a reference, modeled ET from CLMET is the similar to that from CLM over western CONUS, whereas CLMET substantially improves ET simulations over eastern CONUS as compared with CLM. The improvement in CLMET is more evident during September-October-November. We have added the figures (Figures 6 and 7 in the revised manuscript) and revised the relevant texts in the revised manuscript.

"*The analysis on time series of ET from MODIS, FLUXNET-MTE, and two types of simulations also demonstrates improvement from CLM to CLMET. Climatological seasonal cycles of ET over CONUS and four sub regions for 2000-2011 are shown in Figure 6. CLMET performs better than CLM over CONUS with smaller RMSE (0.31 versus 0.40 against MODIS, 0.19 versus 0.25 against FLUXNET-MTE). The improvement mainly results from reduction of overestimation existing in CLM for SON and DJF. However, the model performance greatly varies with region. As indicated by the ET RMSE values, CLMET and CLM perform similarly over western CONUS, whereas CLMET improves the ET simulation over eastern CONUS no matter which reference data is used. Figure 7 compares the temporal evolution of the simulated ET in CLM and CLMET against MODIS and FLUXNET-MTE ET over CONUS and four sub-regions. It is evident that the bias correction method in CLMET is very effective in reducing overestimation (positive bias), but does not work as well in correcting the underestimation (negative bias). The difference has to do with the specific ET regime, i.e. whether ET is limited by water or energy. When an overestimated ET is overwritten with a lower value, the water on land is sufficient to support the reduced ET; in contrast, when an underestimate ET is overwritten with a higher value, the land surface model checks whether water storage in soil layer and vegetation canopy can sustain the elevated ET and further adjust if necessary to keep with the mass conservation equation. The extent to which ET increases is limited by the availability of water stored in soil layer and vegetation canopy. Therefore, in case of water-limited ET, the actual ET after the water availability check in CLMET can be substantially lower than the corrected ET fed into model.*" (the second paragraph from bottom of Section 4.2.1 in the revised manuscript)

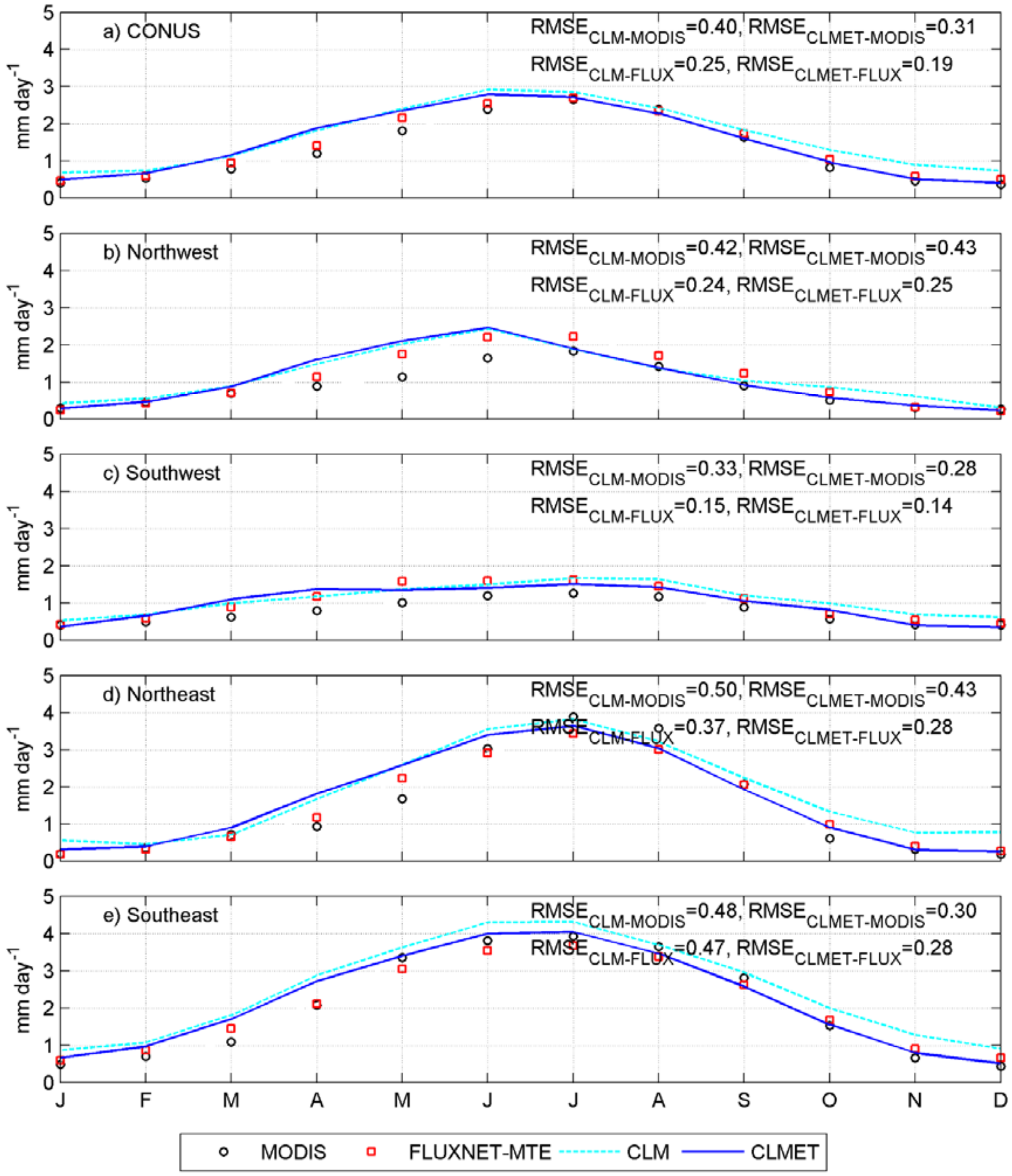


Figure 6 Seasonal cycles of ET from MODIS, FLUXNET-MTE, CLM, and CLMET over CONUS, Northwest, Southwest, Northeast, and Southeast during the period 2000-2011.

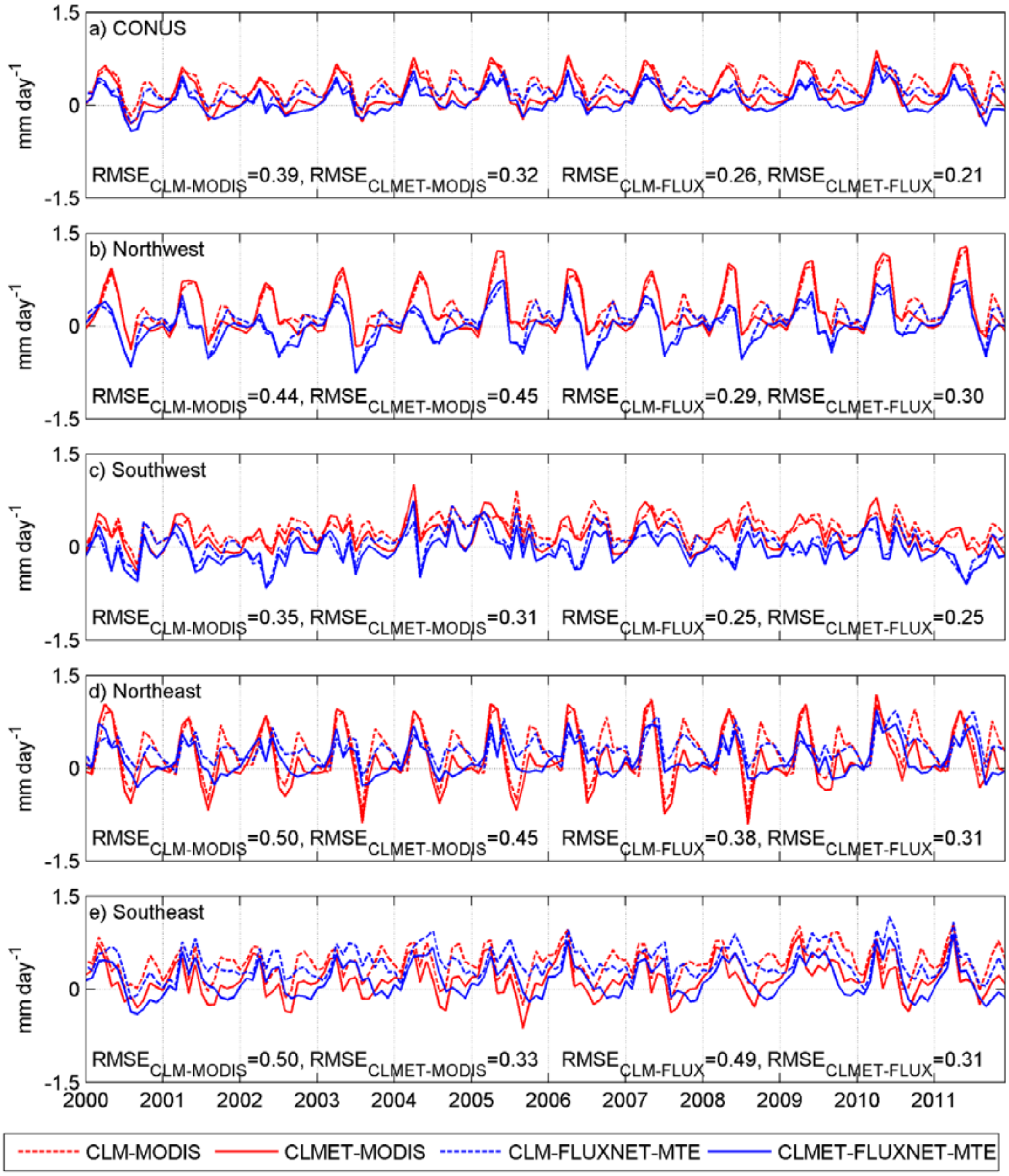


Figure 7 Time series of ET difference between model (CLM or CLMET) and reference data (MODIS or FLUXNET-MTE) over CONUS, Northwest, Southwest, Northeast, and Southeast during the period 2000-2011.

2. It is not clear to me how the statistics in Tables 1 to 4 are exactly calculated. This should be better documented in the manuscript. For instance, the temporal statistics in Table 2: are these calculated per pixel and subsequently averaged over the different study areas (CONUS, NW ...)? Or is the modelled evaporation first aggregated for the study area, and the statistics calculated on the aggregated values? In addition, next to the comparison against the FLUXNET-MTE product, I would also suggest to at least include a validation of the products against actual FLUXNET measurements. Although there are different issues with eddy-covariance measurements as well, a lot of data is

freely available, and these measurements are probably closer to the truth than any of the datasets currently used in the study.

Response:

1) On the calculation of the statistics in Tables 1 to 4:

We have added the following equations to the revised manuscript to show how the statistics is calculated.

$$Bias = \frac{1}{N}\sum_{i=1}^{i=N}\left(\overline{S_i} - \overline{R_i}\right)$$

$$Relative\ bias = \frac{1}{N}\sum_{i=1}^{i=N}\frac{\left(\overline{S_i} - \overline{R_i}\right)}{\overline{R_i}}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{i=N}\left(\overline{S_i} - \overline{R_i}\right)^2}{N}}$$

Where N is the total number of grid cells, and $\overline{S_i}$ ($\overline{R_i}$) are the temporal average of model simulated (reference) value for grid cell i, which is calculated as:

$$\overline{S_i} = \frac{1}{M}\sum_{j=1}^{j=M} S_{i,j}$$

$$\overline{R_i} = \frac{1}{M}\sum_{j=1}^{j=M} R_{i,j}$$

Where Si,j (Ri,j) is model simulated (reference) value on time j and at grid cell i, M is the total number of time series. The statistic RMSE is also used to validate models in reproducing temporal series where M becomes the total number of grid cells, and N becomes the total number of time series.

"In this study, the statistics Bias, Relative bias, and root mean square error (RMSE) are used to validate models in reproducing the spatial pattern against the reference dataset. They are defined as:

$$Bias = \frac{1}{N}\sum_{i=1}^{i=N}\left(\overline{S_i} - \overline{R_i}\right) \quad (1)$$

$$Relative\ bias = \frac{1}{N}\sum_{i=1}^{i=N}\frac{\left(\overline{S_i} - \overline{R_i}\right)}{\overline{R_i}} \quad (2)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{i=N}\left(\overline{S_i} - \overline{R_i}\right)^2}{N}} \quad (3)$$

Where N is the total number of grid cells, and $\overline{S_i}$ ($\overline{R_i}$) *are the temporal average of model simulated (reference) value for grid cell i, which is calculated as:*

$$\overline{S_i} = \frac{1}{M}\sum_{j=1}^{j=M} S_{i,j} \tag{4}$$

$$\overline{R_i} = \frac{1}{M}\sum_{j=1}^{j=M} R_{i,j} \tag{5}$$

Where Si,j (Ri,j) is model simulated (reference) value on time j and at grid cell i, M is the total number of time series. The statistic RMSE is also used to validate models in reproducing temporal series where M becomes the total number of grid cells, and N becomes the total number of time series." (the last two paragraphs of Section 2.2)

2) validation of the products against actual FLUXNET measurements

We have added validations against actual FLUXNET measurements at 16 stations. Please see the response to comment 1 for details.

3. I have the feeling that some issues of the method (e.g. the assumption of time invariant scaling factors or the use of monthly scaling factors) might be masked by the spatiotemporal scales at which the results are analyzed. For instance, why are only time series of the climatological cycle for the entire study area shown in Figure 6? It could be interesting to show some time series from individual pixels as well. Also, an analysis at shorter time scales might show some interesting results. E.g. why do the authors not show a time series of daily evaporation? The same holds for Figure 12: why are these time series not shown at daily time steps and on a pixel basis?

Response:

1) daily series of ET for individual pixels

We have included ET evaluation on daily and monthly scales at 16 pixels, and added figures to compare model simulations with in-situ observations. Please see the response to comment 1 for details.

2) daily series of soil moisture for individual pixels

It is difficult to determine which sites are suitable for validation from total 232 soil moisture observation sites. And the comparison between model simulations and site observations on the daily scale is consistent with the comparison on the monthly scale, as indicated by the comparison for ET. Therefore, we decide to still keep figures on the comparison at the state level (Figures 14 and 15).

4. P6-L116-117: Could the authors be more specific here about what is meant by spatial correlation? Observations from FLUXNET are essentially point measurements.
How are spatial correlations defined here?

Response: Parr et al. (2016) used FLUXNET-MTE (model tree ensemble) ET, which is a gridded ET product, to evaluate CLM4.5. We changed the description as follows:

The spatial correlation coefficients between the simulated annual ET and the FLUXNET-MTE (model tree ensemble) ET are as high as 0.93.

5. P5-L107: I think it should be mentioned here at what temporal resolution the model is applied. From the results in Table 2, I can guess the model is run at a daily resolution. If the latter is the case, I think it should also be justified why the scaling factors are calculated at the monthly time scale. Given that both the simulations and the GLEAM datasets are available at a daily resolution, the scaling factors could as well be calculated at the daily scale. Would this also work? Did the authors test the effect of applying daily scaling factors in the algorithm?

Response:

1) temporal resolution of model: the temporal resolution of model is one hour, which is typical for land surface models. We have added this information into Section 2.3 of the revised manuscript.

2) temporal resolution of scaling factor: This scaling factor characterize the relationship between model biases and ET climatology, and the fundamental assumption is that the nature of the model biases is time-invariant at the inter-annual and longer time scales. The monthly time scale is used here to account for its seasonality. To say that the nature of the model biases varies on a day-to-day time scale does not make physical sense, although technically it can be done. In fact we tested the performance of CLMET based on daily scaling factors. CLMET performance is not improved using daily scaling factors as compared with CLMET using monthly scaling factors.

6. P11-L244-245: Please revise this sentence: GLEAM data is not missing in this period, but is probably masked out in this study as the Northern regions of CONUS are typically covered with snow during these times of the year. GLEAM estimates of sublimation are available for these regions, but I guess they are not considered here (at P7-L140-141, it reads that only interception loss, transpiration and bare-soil evaporation are considered).

Response: we deleted the data records in some parts of west CONUS during the cold seasons by mistake, when GLEAM-derived ET is negative. This led to many missing values in annual ET map in Figure 4 of the original manuscript. We have corrected this mistake and updated Figure 4 (Figure 3 in the revised manuscript). The CONUS-averaged value from CLMET in the update version of annual ET is slightly better than the value in the previous version. We have also updated the table 1 to reflect this change.

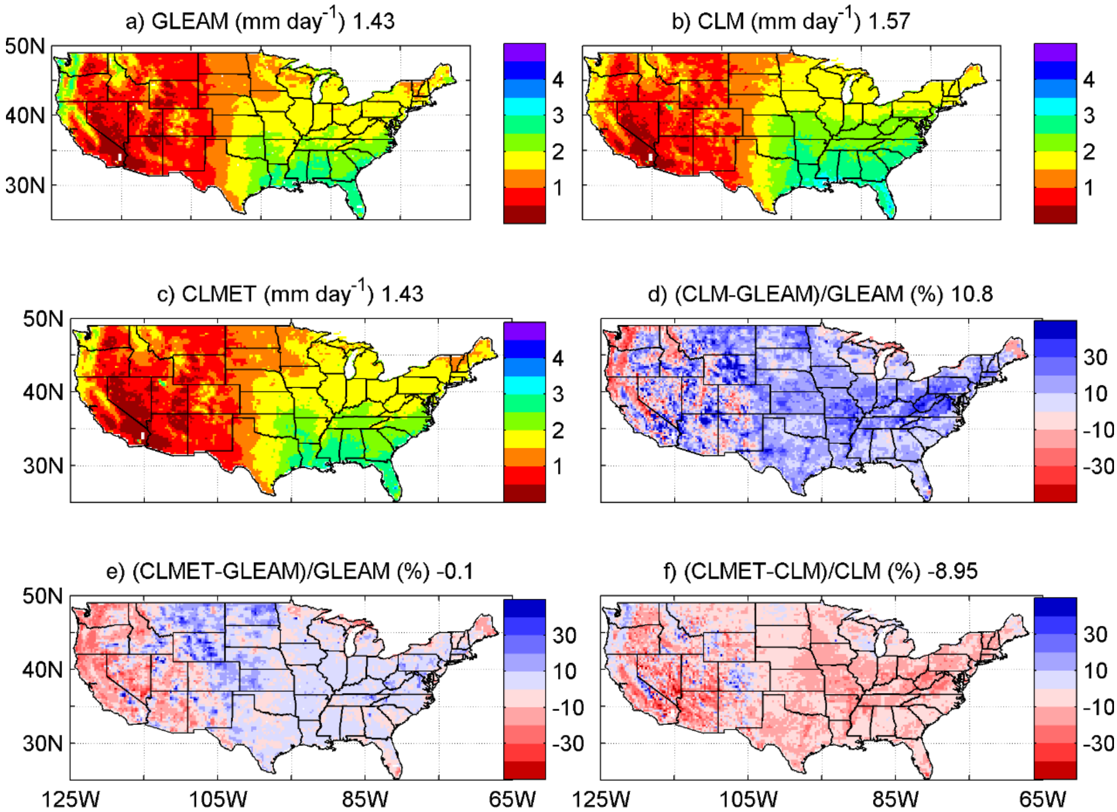


Figure 3 Mean annual ET from a) GLEAM, b) CLM, and c) CLMET, and the relative differences between d) CLM and GLEAM, e) CLMET and GLEAM, and f) CLMET and CLM during 2000-2014. Numbers in titles are CONUS-averaged values.

7. P12-L261-262: If the term "significant" is used, it implies that a statistical test was applied to check this hypothesis. If this is the case, the test should be mentioned here.
Response: we changed to "substantially".

8. Please note that the GLEAM datasets are no "observations" of evaporation. They are estimates of terrestrial evaporation, resulting from applying a simple conceptual model to observation-based datasets of different meteorological variables. GLEAM is kept as simple as possible to minimize the impact of the algorithms and maximize the impact of the meteorological observations on the estimates of evaporation. I would suggest to revise this throughout the manuscript.
Response: we have changed from "observations" to "estimations".

TECHNICAL CORRECTIONS
1. Please use hyphens in "compound adjectives" such as "land-surface models" or "widely-used tools".
Response: the expression of "land surface models" and "widely used tools" are widely used in literature.

2. I would suggest explaining all abbreviations upon their first use. E.g. P3-L68-69: SAC and VIC.

Response: Following reviewer's suggestion, we have spelled out SAC-SMA (Sacramento Soil Moisture Accounting) and VIC (Variable Infiltration Capacity) when they appeared for the first time in the revised manuscript.

"*The Mosaic and Sacramento Soil Moisture Accounting (SAC-SMA) models tend to overestimate ET, whereas the Noah and Variable Infiltration Capacity (VIC) models are likely to underestimate ET.*" (the last sentence of first paragraph, Section 1)

3. P5-L108: Given that no further details are provided in the paper regarding the land surface model used, I would suggest adding a reference here for the CLM model.
Response: Following reviewer's suggestion, we have added a reference about Community Land Model version 4.5 when the model is introduced in the revised manuscript.

Oleson, K. W. et al.: Technical Description of version 4.5 of the Community Land Model (CLM), NCAR Tech. Note, NCAR/TN-503+STR, doi:10.5065/D6RR1W7M, 2013.

4. P5-L111: Please define "PFT".
Response: Defined

5. P6-L124: I guess this should be section 2.2 instead of 2.3.
Response: Yes, it is 2.2. We have corrected it.

6. P7-L161: The fact that the GLEAM database has three subsets is not relevant here if you only use one.
Response: Following reviewer's suggestion, we have deleted the description of three subsets of GLEAM.

7. P28-Table1: Please correct "COUNS" in the caption. Please also check this at other places in the manuscript: e.g. P14-L315
Response: We have corrected them to "CONUS".

8. P34-Figure3: Please explain in the caption which areas are masked. I guess these are regions covered with snow?
Response: The GLEAM-derived dew and the CLM simulated dew is not consistent in some areas of northwest CONUS. If that happens, the scaling factors became negative, because ET is negative for one and positive for the other. We did not scale ET when the scaling factor is negative, and those areas are masked out in Figure 2. We have added an explanation about it.

Responses to the comments from Reviewer #2

Manuscript Number: hess-2016-696 Title: Incorporating remote sensing ET into Community
Land Model version 4.5 Authors: Dagang Wang, Guiling Wang, Dana T. Parr, Weilin Liao, Youlong Xia, Congsheng
Summary
This paper follows the ET bias correction scheme proposed in Parr et al. 2015 and carries out a regional scale (CONUS) study in order to evaluate the effectiveness/performance of this approach over a large domain in terms of estimating ET, runoff, and soil moisture. The main idea I see is to reduce the ET overestimation in CLM 4.5 by rescaling it down and push the reduced ET back into the model to raise the runoff and soil moisture content – this goal is obviously achieved. The data, experiments and analysis in this study are all carefully chosen and the descriptions are very clear too. The overall quality of the research is good though most of the major conclusions are more or less well expected even without these experiments.
I think the paper can be published in HESS with minor revisions.

Major Comments
Unlike true "state" variables like moisture content or temperature, whose current value directly influences the future state of the underlying dynamic system, ET is not a state variable but a flux variable. Therefore, any effort to incorporate ET information effectively into the land surface model needs a way to propagate the change to ET flux across other parts of the dynamic system (e.g., soil moisture, canopy storage, runoff fluxes, etc.). The approach taken in this paper (following Parr et al. 2015) is to re-run the model (CLMET) and force the ET flux to be a value rescaled relative to the initial run (CLM), where the rescaling factor is pre-calibrated for every location and month. This approach is simple and effective, I think. On the other hand, this approach is also awkward as it looks like an enhanced post-processing" for bias correction instead of tackling the ET overestimation from its root cause, e.g., an underestimated surface resistance. The awkwardness comes in also because the "forced" ET in the CLIMET run will considerably disrupt the model physics itself, e.g. breaking the water balance and sustaining wetter soil without letting the plants transpire more. If we adjust the resistance (or some other related process like to make the water easier/faster to drain from the soil), then most of such physical inconsistency would be gone.
Response: the model bias in ET simulations results from inaccurate information of meteorological conditions (Mueller and Seneviratne, 2014), surface-type data (Hwang and Choi, 2013), model parameters (Ma et al. 2015), and soil water (Decker 2015). Adjusting surface resistance is essentially one of many methods of model parameters calibration, which can reduces model bias as well. However, only making parameter adjustment may results in nonphysical parameter subsets when other inaccurate information is the main cause of the model bias for some regions/seasons (Ray et al. 2015). In this study, we take a different approach to correct simulated ET as a whole instead of adjusting each separate factors, which provide a simple and efficient way to

improve model performance in hydrological estimation without improving the model physics itself. We have added a short discussion in the Section 5.
"Model parameter calibration (e.g., tuning surface resistance) is another way to reduce model bias (Ren et al. 2016). However, the parameter space may contain nonphysical parameter subsets (Ray et al. 2015), which is especially an issue when model parameter tuning is used to offset unrelated model deficits. The method used in this study attempts to avoid such issues through improving the model performance without dealing with calibration of model physical parameters." (the last pargragrah of Section 5 in the revised manuscript)

Mueller, B., and S. I. Seneviratne (2014), Systematic land climate and evapotranspiration biases in CMIP5 simulations, Geophysical Research Letter, 41, 128–134, doi:10.1002/2013GL058055.

Hwang, K., and Choi, M. (2013). Seasonal trends of satellite-based evapotranspiration algorithms over a complex ecosystem in East Asia. Remote Sensing of Environment, 244-263.

Ma, N., Y. Zhang, C.-Y. Xu, and J. Szilagyi (2015), Modeling actual evapotranspiration with routine meteorological variables in the data-scarce region of the Tibetan Plateau: Comparisons and implications, Journal Geophysical Research: Biogeosciences, 120, doi:10.1002/2015JG003006.

Decker, M. (2015). Development and evaluation of a new soil moisture and runoff parameterization for the CABLE LSM including subgrid-scale processes. Journal of Advances in Modeling Earth Systems, 7(4), 1788-1809.

Ray, J., Z. Hou, M. Huang, K. Sargsyan, and L. Swiler (2015), Bayesian calibration of the Community Land Model using surrogates, SIAM/ASA Journal on Uncertainty Quantification, 199 – 233, doi:10.1137/140957998.

The authors have a major assumption that the ET biases won't change from year to year (with seasonal variability, though) so that such static errors can be corrected with static correction factors. So, the entire long ET validation section (4.2.1) is really validating the performance of the new estimation system but this stationarity assumption. It'll be interesting if the results can be compared to a pure "post-processing" approach, i.e., to rescale ET then rebalance the water budget between precipitation, ET, soil moisture, and runoff.
Response: It is hard to rebalance water and energy budgets though post processing without model runs after ET is rescaled. The rescaled ET influences simulations of many components of land surface processes, such as infiltration, soil water/energy transport, which cause changes in land surface states. The land surface states at the current time step is the bases of flux variable simulations for the next time step. All

these processes and connections between adjacent time steps cannot be tackled in the post processing. To obtain the consistency between different components of land surface processes and connect land surface states between adjacent time steps, we really need to re-run CLM and let model resolve all these issues. That is the reason why Parr et al. (2015) proposed the method and we applied this method in CLM on the regional scale.

Details:
Line 65: model -> models
Response: we have changed to "models".

Line 88: intense -> intensive
Response: we have changed to "intensive".

Line 91: past -> historical
Response: we have changed to "historical".

Line 101: Parr et al. -> Parr et al. (2015); into -> for
Response: we have changed to "Parr et al. (2015)" and "for".

Line 111: spell out PFT
Response: we have spelled out PFT (plant functional type).

Line 122: "CONUS" was first mentioned in line 115
Response: we have define "CONUS" (Conterminous United States) in line 115.

Line 155: unbalance -> imbalance
Response: we have changed to "imbalance".

Line 322-334: where does the runoff data come from? GSCD or GRDC? What is GRDS in line 328? And Line 379?
Response: all these should be GSCD (Global Streamflow Characteristics Dataset). We have corrected them.

Line 413: replace -> to replace
Response: we have changed to "to replace".

Incorporating remote sensing ET into Community Land Model version 4.5

Dagang Wang[1, 2, 3, 4]*, Guiling Wang[4*], Dana T. Parr[4], Weilin Liao[1, 2], Youlong Xia[5], Congsheng Fu[4]

[1] School of Geography and Planning, Sun Yat-sen University, Guangzhou, China

[2] Guangdong Key Laboratory for Urbanization and Geo-simulation, Sun Yat-sen University, Guangzhou, China

[3] Key Laboratory of Water Cycle and Water Security in Southern China of Guangdong High Education Institute, Sun Yat-sen University, Guangzhou, P.R. China

[4] Department of Civil and Environmental Engineering, University of Connecticut, Storrs, USA

[5] National Centers for Environmental Prediction/Environmental Modeling Center, and I. M. System Group at NCEP/EMC, College Park, Maryland, USA

Submitted to special issue of Hydrology and Earth System Sciences: Observations and modeling of land surface water and energy exchanges across scales, in Honor of Eric F. Wood

Revised April, 2017

*Corresponding authors: Dr. Dagang Wang, School of Geography Science and Planning, Sun Yat-sen University, Guangzhou, P. R. China 510275, wangdag@mail.sysu.edu.cn, (86) 2084114575. Dr. Guiling Wang, Department of Civil and Environmental Engineering, University of Connecticut, Storrs, USA, guiling.wang@uconn.edu

Abstract

Land surface models bear substantial biases in simulating surface water and energy budgets despite of the continuous development and improvement of model parameterizations. To reduce model biases, Parr et al. (2015) proposed a method incorporating satellite-based evapotranspiration (ET) products into land surface models. Here we apply this method to the Community Land Model version 4.5 (CLM4.5) and test its performance over the conterminous US (CONUS). We first calibrate a relationship between the observational ET from the Global Land Evaporation Amsterdam Model (GLEAM) product and the model ET from CLM4.5, and assume that this relationship holds beyond the calibration period. During the validation or application period, a simulation using the default CLM4.5 ("CLM") is conducted first, and its output is combined with the calibrated observational-vs-model ET relationship to derive a corrected ET; an experiment ("CLMET") is then conducted in which the model-generated ET is overwritten using the corrected ET. Using the observations of ET, runoff, and soil moisture content as benchmarks, we demonstrate that CLMET greatly reduces the biases existing in CLM. The improvement differs with region, being more significant in eastern CONUS than western CONUS, with the most striking improvement over the southeast CONUS. This regional dependence reflects primarily the regional dependence in the degree to which the relationship between observational and model ET remains time-invariant (a fundamental hypothesis of the Parr et al. method). The bias correction method provides an alternative way to improve the performance of land surface models, which could lead to more realistic drought evaluations with improved ET and soil moisture estimates.

Key words: evapotranspiration; land surface model; bias correction; CLM

1. Introduction

Land surface models are widely used tools in simulating and predicting the Earth's water and energy budgets over a wide range of spatiotemporal scales (Rodell et al., 2004, Haddeland et al. 2011, Getirana, 2014, Xia et al. 2012a, b, Xia et al. 2016a, b). For example, the Global Land Data Assimilation System (GLDAS) was designed to simulate the terrestrial water and energy budgets over the globe using multiple land surface models (Rodell et al., 2004); and its regional counterpart, the North America Land Data Assimilation System (NLDAS), utilizes four land surface models and focuses on the conterminous United States at a much higher resolution (Rodell et al., 2004, Xia et al. 2012a, b). Products from these two operational systems have been widely used in estimating terrestrial water storage changes (Syed et al. 2008), investigating land-atmosphere coupling strength (Spennemann and Saulo, 2015), analyzing soil moisture variability (Cheng et al. 2015), studying the impact of soil moisture on dust outbreaks (Kim and Choi 2015), and improving data quality of in-situ soil moisture observations (Dorigo et al. 2013, Xia et al. 2015). These model-based estimates of land surface fluxes and state variables are considered important surrogate for observations, as observational data for some components of the global water and energy cycles are scarce in many regions of the world, and are not spatially and temporally continuous where they do exist. However, land surface models are subject to large uncertainties. Haddeland et al. (2011) compared eleven models in simulating evapotranspiration (ET) and found that the range across models is very wide. The global ET on land surface ranges from 415 to 586 mm year^{-1}, and the runoff ranges from 290 to 457 mm year^{-1}. Xia et al. (2012a-b, 2016a-b) documented large disparity among the four models in NLDAS phase 2 (NLDAS-2) at both the continental and basin scales. The Mosaic and Sacramento Soil Moisture Accounting (SAC-SMA) models tend to overestimate ET, whereas the Noah and Variable Infiltration Capacity (VIC) models are likely to underestimate ET.

Great efforts have been made to improve model performance over the years, through enhancing both the model parameterization of land surface processes and the model input data. For instance, during the past ten years, the Community Land Model (CLM) has been upgraded from version 2 to version 4.5 (Bonan et al. 2002, Oleson et al. 2008, Oleson et al. 2013), accompanied by increasingly accurate and high resolution surface datasets (Lawrence et al. 2011). Comparison with observations of runoff, evapotranspiration, and total water storage demonstrated continuous improvement of the model performance (Lawrence et al. 2011). The Noah model is another example of continuous upgrade from its original version since 1980s (Mahrt et al. 1984). Recent model developments were on vegetation canopy energy balance, the layered snowpack, frozen soil and infiltration, soil moisture-groundwater interaction and related runoff production, and vegetation phenology (Niu et al. 2011). Despite the improved understanding and parameterization of physical processes and better input data, substantial model biases remain (e.g., Parr et al. 2016, Wang et al. 2016).

Another approach to reducing model biases is through data assimilation, by merging observational data and land surface models to obtain optimal estimates for next time step. Fusing soil moisture observations into land surface model is a typical practice in land data assimilation, and it has been reported that data assimilation of soil moisture helped in reducing model bias (Reichle and Koster 2005, Kumar et al. 2008, Yin et al. 2015). However, data assimilation is a computationally intensive task, especially when implementing a multi-model ensemble approach. Moreover, data assimilation approach is not applicable to future prediction. Parr et al. (2015) proposed an alternative approach to reducing model biases, and applied it to the Variable

Infiltration Capacity (VIC) model over the Connecticut River Basin for both historical simulations and future projections. The Parr et al. (2015) approach assumes that the relationship between the model evapotranspiration (ET) and observational ET remain unchanged from one period to another, and hence the relationship estimated from the calibration period can be used to correct the ET biases and their effects for any period, historically or in the future. When applied to VIC over the Connecticut River Basin, Parr et al. (2015) found that the ET bias correction approach significantly reduces systematic biases in the estimates of both past ET and past river flow, and qualitatively influences the projected future changes in drought and flood risks.

To establish the robustness of the Parr et al. (2015) method, it needs to be evaluated over different regions and different climate regimes based on different models. In this study, we implement the Parr et al. (2015) approach for CLM4.5 and evaluate its performance over the whole Conterminous United States (CONUS). The land surface model, study area, and the bias correction method are introduced in Section 2. The data for model calibration and validation, including dataset of ET, runoff, soil moisture, is described in Section 3. Section 4 presents the calibration and validation results. Finally, the main findings are summarized and discussed in Section 5.

2 Model and Methodology

2.1 Model and Forcing Data

CLM4.5 (Oleson et al. 2013) in its offline mode with the prescribed vegetation phenology is used in this study. The land surface dataset used in CLM4.5 is derived from different sources. The soil texture data are taken from Bonan et al. (2012), which was generated using the International Geopshere-Biosphere Programme soil data (Global Soil Data Task, 2000). Both the percentage of plant functional types (PFTs) and the leaf area index within each grid cell are derived from Moderate Resolution Imaging Spectroradiometer (MODIS) satellite data (Lawrence et al. 2011). Slope and elevation are obtained from the U.S. Geological Survey HYDRO1K 1 km data set (Verdin and Greenlee, 1996). Parr et al. (2016) found that CLM4.5 can realistically capture the spatial pattern of ET over CONUS when the model is forced by the NLDAS-2 meteorological variables. The spatial correlation coefficients between the simulated annual ET and the FLUXNET-MTE (model tree ensemble) ET are as high as 0.93. Wang et al. (2016), using multiple atmospheric forcing datasets, also reported that CLM4.5 can reasonably reproduce large-scale pattern of runoff and ET. In this study CLM4.5 is forced by the NLDAS-2 meteorological forcing (Xia et al., 2012a). NLDAS-2 forcing is available during 1979-present at hourly resolution on a 0.125° grid system, but is aggregated to a 0.25° resolution in this study as the driving forcing for CLM4.5. CONUS is chosen as the study domain over the globe for the high quality of atmospheric forcing data in this region.

2.2 Methodology

The division of CONUS into Northwest, Southwest, Northeast, and Southeast, which is based on the 40°N latitude line and the 98°W longitude line, is defined by Lohmann et al. (2004). This division was later adopted by Xia et al. (2012a) and Tian et al. (2014) when land surface models were evaluated over CONUS. We follows this division in this study, as shown in Figure 1a.

Although land surface models are cable of capturing large-scale pattern of ET, significant biases were found at finer spatiotemporal scales (Parr et al. 2015, Parr et al. 2016, and Wang et al. 2016), which propagates to influence other components of the hydrological cycle including runoff and soil moisture (Parr et al. 2015). Following Parr et al. (2015), we derived the climatology of modeled ET for each model grid cell and for each month based on a simulation during the calibration period and climatology of observational ET from satellite-based ET data at the same spatiotemporal resolution during the same period, and estimate the scaling factor between observational ET and the model ET. This scaling factor, which has its unique spatial variability and seasonal cycle, is assumed to be time-invariant at the inter-annual and longer time scales. To correct the ET biases in model simulations during any period, two types of simulations are conducted sequentially. In the first type of simulation, named as CLM, we run the default CLM4.5 and save the output for three component of ET, i.e., interception loss, plant transpiration, and soil evaporation, at the PFT level for every time step. The corrected interception loss, plant transpiration, and soil evaporation are then derived by multiplying the simulated values with the ET scaling factor, and will be used as input for the second type of simulation, named as CLMET. In CLMET, we re-run CLM4.5 for the same period as in the first type, but overwrite the three ET components simulated by the model with the corrected values. Since ET simulations affect the partitioning of precipitation between ET and runoff, the bias correction in ET is expected to have direct positive impact on runoff generation and therefore soil moisture.

In this study, we use 1986-1995 as the calibration period and 2000-2014 as the validation period. The simulations during the calibration period are obtained from a 16-year (1980-1995) CLM run with the first 6-year run disregarded as the spinup. Both CLM and CLMET runs during the validation period starts with the initial condition of January 1st 1996 obtained from the calibration period. The time step for both CLM and CLMET runs is one hour. Since the overwriting process in CLMET may break the water balance, the model checks if the interception loss exceeds the water stored in vegetation canopy and if the surface soil water is sufficient to support soil evaporation, and makes adjustment if needed. This minimizes the imbalance caused by overwriting ET components in CLMET.

In this study, the statistics Bias, Relative bias, and root mean square error (RMSE) are used to validate models in reproducing the spatial pattern against the reference dataset. They are defined as:

$$Bias = \frac{1}{N}\sum_{i=1}^{i=N}\left(\overline{S_i} - \overline{R_i}\right) \quad (1)$$

$$Relative\ bias = \frac{1}{N}\sum_{i=1}^{i=N}\frac{\left(\overline{S_i} - \overline{R_i}\right)}{\overline{R_i}} \quad (2)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{i=N}\left(\overline{S_i} - \overline{R_i}\right)^2}{N}} \quad (3)$$

Where N is the total number of grid cells, and $\overline{S_i}$ ($\overline{R_i}$) are the temporal average of model simulated (reference) value for grid cell i, which is calculated as:

$$\overline{S_i} = \frac{1}{M}\sum_{j=1}^{j=M} S_{i,j} \quad (4)$$

$$\overline{R_i} = \frac{1}{M}\sum_{j=1}^{j=M} R_{i,j} \quad (5)$$

Where $S_{i,j}$ ($R_{i,j}$) is model simulated (reference) value on time j and at grid cell i, M is the total number of time series. The statistic RMSE is also used to validate models in reproducing temporal series where M becomes the total number of grid cells, and N becomes the total number of time series.

3 Data

3.1 ET

3.1.1 GLEAM ET

GLEAM (The Global Land Evaporation Amsterdam Model) version 3.0a (Miralles et al. 2011, Martens et al. 2016) is used to calibrate the ET scaling factors and to validate CLM and CLMET. GLEAM 3.0a is derived based on reanalysis net radiation and air temperature, a combination of gauge-based, reanalysis and satellite-based precipitation and satellite-based vegetation optical depth, spanning the 35-year period 1980–2014 (http://www.gleam.eu/). Potential evaporation in GLEAM 3.0 is calculated using a Priestley and Taylor equation based surface net radiation and near-surface air temperature, and is converted to actual evaporation based on the multiplicative evaporative stress factor. The dataset has been used in studying soil moisture-temperature coupling (Miralles et al. 2012), the impact of land surface on precipitation (Guillod et al. 2015), and the climate control on land surface evaporation (Miralles et al., 2014). Recent evaluations conducted at both tower and global scales shows that GLEAM-based ET is superior to MODIS-based and the Surface Energy Balance System (SEBS) based ET products (Michel et al. 2016, Miralles et al. 2016). The spatial resolution for GLEAM dataset is 0.25°, which is consistent with the resolution of CLM4.5 used in this study. The temporal resolution of GELAM dataset is daily, and the monthly aggregated ET is used to derive the scaling factors.

3.1.2 MODIS and FLUXNET-MTE ET

Another two gridded ET products are used for independent evaluations: MODIS ET and FLUXNET-MTE (model tree ensemble) ET. Mu et al. (2007, 2011) produced a MODIS-based global ET dataset by revising the Penman–Monteith (PM) equation. The dataset is arguably the most widely used remote-sensing-based global ET product (Miralles et al. 2016). Monthly version of the MODIS-based product at the 0.5° spatial resolution are used to validate the model with the bias correction method. The FLUXNET-MTE global ET dataset was derived from 253 FLUXNET eddy covariance towers distributed over the globe using the model tree ensemble (MTE) approach

(Jung et al., 2009, 2010). The record gaps of half hourly eddy covariance fluxes were filled first, and the complete tower-based dataset is used to train MTE to produce monthly global ET dataset at 0.5° spatial resolution. The data have been used to study the ET trend (Jung et al., 2010) and to improve canopy processes in a land surface model (Bonan et al., 2011). As FLUXNET sites over the CONUS are very dense, the quality of the FLUXNET-MTE dataset in our study domain is expected to be high. The MODIS dataset is available from 2000-2014, and the FLUXNET-MTE dataset is available from 1982-2011. We chose the overlap period of those two products, 2000-2011, for model validations using MODIS and FLUXNET-MTE dataset.

3.1.3 Tower flux ET

ET observations (in energy unit) at 16 site from the AmeriFlux network are used validate model on the grid cell scale (Figure 1b). Those sites spans four sub-regions (i.e., NW, SW, NE, and SW) of CONUS with five different vegetation types (i.e., grass, crop, evergreen needleleaf forest, mixed forest, and deciduous broadleaf forest). More details about the tower flux sites can be seen in Xia et al. (2015b). The year of 2005 is selected for validation, as data at most sites in this year is available and the missing records are minimum. However, there are three exceptions due to data availability: 2002 for the site of Sylvania Wilderness, 2004 for the sites of Donaldson and Walnut River. Both daily and monthly ET observations at these 16 sites are compared with model simulations.

3.2 Observation-based Runoff Coefficient

The runoff coefficient (the ratio of runoff to precipitation) of Global Streamflow Characteristics Dataset (GSCD) version 1.9 (Beck et al., 2013, Beck et al., 2015) is used to evaluate the model performance in simulating runoff. The GSCD dataset was produced based on streamflow observations from approximately 7500 catchments over the globe. A data-driven approach was adopted to derive the gridded streamflow characteristics at the 0.125° resolution on a global scale. This dataset is relatively reliable for the grid cells within which a large number of catchments data is used. The uncertainty is low in North America, Europe, and southeastern Australia where a large number of observations are available.

3.3 In-situ soil moisture observations

The North American Soil Moisture Database (NASMD) is used to evaluate the model performance in simulating soil moisture in both the surface (0-10cm) and root-zone (0-100cm) layers. NASMD was initiated in 2011 to provide support for developing climate forecasting tools, calibrating land surface models, and validating satellite-derived soil moisture algorithms. A homogenized procedure has been implemented, as the measurement stations are across a variety of in-situ networks. In addition, a quality control (QC) algorithm was applied to the measurement records. Liao et al. (submitted to Journal of Hydrometeorology, 2016) developed an additional QC algorithm to further improve data quality of the NASMD soil moisture based on the approach of Xia et al. (2015). The soil moisture after QC agree more closely with a manual-checked benchmark. More details on the QC algorithm and the comparison with the benchmark can be found in Liao et al. (2016). The in-situ observations in the states of Alabama (AL), Illinois (IL), Mississippi (MS), Nebraska (NE), and Oklahoma (OK) from 2006-2010 are selected from NASMD (Figure 1a), as a large number stations is evenly distributed over these states and observation records during this period are relatively complete after QC. The numbers of stations in AL, IL, MS, NE, and OK are 10, 19, 14, 45, 105, and 39, respectively, as shown in Figure 1a. Since the soil layer in which measurement is conducted varies with stations, we interpolate the volumetric soil water content to the 5 cm and 50 cm depth for all stations using the liner interpolation method to compare with the modeled soil moisture in the 0-10 cm and 0-100 cm layers.

4 Results

4.1 ET scaling factor calibration

Figure 2 shows the climatological scaling factors for each month over CONUS based on the 1986-1995 period. The GLEAM-derived dew and the CLM simulated dew is not consistent in some areas of northwest CONUS. If that happens, the scaling factors became negative, because ET is negative for one and positive for the other. We did not scale ET when the scaling factor is negative, and those areas are masked out in Figure 2. The model simulations generally agree better with GLEAM estimations during the warm seasons, whereas the difference between simulations and GLEAM estimations are large during the cold seasons. The scaling factors greatly vary with region, as indicated by area-averaged values for four sub regions. For instance, the area-averaged values are 0.34, 0.58, 0.28, and 0.52 for Northwest, Southwest, Northeast, and Southeast in November, respectively. The overestimation is overwhelming during October, November, December, and January, whereas underestimation occurs in many areas during March, April, and May. The overestimation is very severe with simulations being almost 5 times of GLEAM estimations for Northeast CONUS in December.

4.2 Evaluation

We evaluate the effectiveness of the ET bias correction approach in CLM4.5 by comparing results from CLM and CLMET with the reference dataset. The evaluation metrics examined include bias, relative bias, and root mean square error (RMSE) as described in Section 2.2. Since the spatial resolution of some gridded reference data is not consistent with the model resolution, we upscale the finer resolution data to match the coarser resolution data using simple arithmetic averages. For example, when the MODIS and FLUXNET-MTE ET are used for validation, we aggregate the four 0.25° modeled ET within each 0.5° grid cell; for the GSCD runoff data, we aggregate observations from 0.125° to 0.25° to match the model resolution. As in-situ soil moisture observations are essentially on the point scale, we spatially average observed soil moisture in each state and compare the averaged observations with the averaged model simulations over grid cells within the same state.

4.2.1 ET

Figure 3 shows the multi-year averages (2000-2014) of ET derived from GLEAM, simulated by CLM and CLMET, and the relative bias of simulations against GLEAM. Over most of CONUS, CLM overestimates ET relative to GLEAM data, and CLMET reduces ET as well as ET biases. The averaged relative bias in CLM over CONUS is 10.8%, with relative bias exceeding 10% in a substantial portion of CONUS; and in CLMET, the CONUS-averaged relative bias is reduced to -0.1%, and it is within 10% over most of CONUS. This improvement is more significant over eastern CONUS than western CONUS. Table 1 shows the statistics on the model performance with these two schemes during different seasons and in four sub regions. CLM overestimates the CONUS-averaged ET in all other seasons except for March-April-May (MAM), and the largest overestimation occurs in Northeast CONUS during December-January-February (DJF) with a relative bias as large as 146.4%. The underestimation in MAM is largest over Southwest CONUS with a relative bias of -17.9%. CLMET substantially improves the model performance as indicated by the various metrics. All the statistics in CLMET is superior to those in CLM with a few exceptions in bias or relative bias. The improvement from CLM to CLMET is more substantial for September-October-November (SON) and DJF than MAM and June-July-August (JJA). The relative bias of 51% (77.7%) in CLM is reduced to 7.8% (18.9%) in CLMET over CONUS during SON (DJF). For the regional comparison, the improvement is greatest over Southeast CONUS. All the positive biases in all seasons over Southeast CONUS are substantially reduced.

To understand the differences between CLM and CLMET, we select four months from each of seasons, January, April, July, and November, to examine the relationship between the relative bias of model simulations and the scaling factor changes from calibration period (1986-1995) to validation period (2000-2014) in Figure 4. The improvement from CLM to CLMET is evident, especially in January and November (Figure 4a-4b). Although the bias is dramatically reduced in CLMET, it remains large in Northeast CONUS in January (Figure 4b1). In addition, the bias in CLMET seems larger in western CONUS than eastern CONUS (Figure 4b). The spatial patterns of the relative biases in CLMET and the scaling factor differences between the two periods demonstrate a great degree of similarity (Figure 4b-4c), and the scatter plots between the two quantities (Figure 4d) reflect a strong correlation. This suggests that the degree to which CLMET can improve model performance in simulating ET greatly depends on how stable the scaling factor is from the calibration period to the validation period, i.e., how well the assumption of a time-invariant scaling relationship holds. Over most of CONUS, changes in the scaling factor are within 10% (Figure 4d). This temporal stability of the relationship between observed ET and simulations guarantees improvements from CLM to CLMET.

CLM and CLMET performances are also evaluated using two independent observation dataset of ET, MODIS-based and FLUXNET-MTE-based ET (Figure 5, Tables 2 and 3). For the multi-year averaged ET, the relative bias in CLMET is smaller than that in CLM, and the improvement is greater in eastern CONUS than western CONUS as compared with either MODIS- or FLUXNET-MTE-based ET. Note that there is still a substantial overestimation in western CONUS in CLMET compared with the MODIS ET, partially because the algorithm developed by Mu et al. (2007, 2011) underestimate ET in the MODIS product (Michel et al. 2016, Miralles et al. 2016). With the reference of the MODIS or FLUXNET-MTE ET, CLMET corrects bias for all other three seasons except for MAM (Tables 2 and 3). Bias, relative bias and RMSE in CLMET is greater than CLM for the whole CNOUS, Northwest, Southwest, and Northeast in MAM. Among all other three seasons, SON is the reason when model performance is improved most from CLM to CLMET. The performance in CLMET against MODIS or FLUXNET-MTE is similar to the model performance against GLEAM for annual mean, JJA, SON, and DJF but with smaller magnitudes. CLMET deteriorates the ET simulation for MAM by intensifying overestimation already occurring in CLM, which is different from the validation against the GLEAM-based ET.

The analysis on time series of ET from MODIS, FLUXNET-MTE, and two types of simulations also demonstrates improvement from CLM to CLMET. Climatological seasonal cycles of ET over CONUS and four sub regions for 2000-2011 are shown in Figure 6. CLMET performs better than CLM over CONUS with smaller RMSE (0.31 versus 0.40 against MODIS, 0.19 versus 0.25 against FLUXNET-MTE). The improvement mainly results from reduction of overestimation existing in CLM for SON and DJF. However, the model performance greatly varies with region. As indicated by the ET RMSE values, CLMET and CLM perform similarly over western CONUS, whereas CLMET improves the ET simulation over eastern CONUS no matter which reference data is used. Figure 7 compares the temporal evolution of the simulated ET in CLM and CLMET against MODIS and FLUXNET-MTE ET over CONUS and four sub-regions. It is evident that the bias correction method in CLMET is very effective in reducing overestimation (positive bias), but does not work as well in correcting the underestimation (negative bias). The difference has to do with the specific ET regime, i.e. whether ET is limited by water or energy. When an overestimated ET is overwritten with a lower value, the water on land is sufficient to support the reduced ET; in contrast, when an underestimate ET is overwritten with a higher value, the land surface model checks whether water storage in soil layer and vegetation canopy can sustain the elevated ET and further adjust if necessary to keep with the mass conservation equation. The extent to which ET increases is limited by the availability of water stored in soil layer and vegetation canopy. Therefore, in case of water-limited ET, the actual ET after the water availability check in CLMET can be substantially lower than the corrected ET fed into model.

In addition, the ET validation is also conducted on the site scale (Figures 8, 9, and 10). Except for Port Peck and Wind River Crane stations in the northwest CONUS, for all other stations the monthly mean ET from CLMET agrees better with the observed ET than that from CLM (Figure 8). The same statement holds for daily mean ET (Figures 9 and 10). Generally, CLM overestimates ET as compared with station observations, and CLMET alleviates this overestimation, which is consistent with comparisons between modelled ET and satellite-based ET products.

4.2.2 Runoff

Using the runoff coefficient (the ratio of runoff to total precipitation) derived from GSCD as the benchmark, we evaluate the model performance in CLM and CLMET in simulating runoff (Figure 11). The CONUS averaged runoff coefficient in CLM and CLMET are 0.18 and 0.21, which is comparable with the GSCD-based runoff coefficient (0.22). However, CLM underestimate the runoff in most areas of CONUS due to overestimate of ET. CLMET alleviates the underestimation by decreasing ET therefore increasing the runoff, especially over eastern CONUS. The relative bias of CLMET against GSCD is 1.1%, which is much smaller than the value in CLM (-9.2%). Table 4 shows the regional difference in runoff simulations in CLM and CLMET. The improvement is greater over Eastern CONUS than Western CONUS, which is consistent with the improvement of ET simulations. The most striking improvement occurs in Southeast CONUS, with the relative bias (RMSE) decreased from -24.7% (0.091) to -8.2% (0.06).

Because only the multi-year mean annual runoff coefficient is available for GSCD, we cannot examine the seasonal dependency of the model performance improvement.

The increase in runoff from CLM to CLMET is mainly due to the increase in subsurface runoff (Figure 12). The same value of the ET scaling factor within each gird cell are applied to three components of ET (interception loss, plant transpiration and soil evaporation) in this study. Because interception loss accounts for a small portion of total ET, the absolute change of interception loss (decrease from CLM to CLMET over most areas) is much smaller compared with plant transpiration and soil evaporation (not shown). As a result, the increase in throughfall does not change much from CLM to CLMET, which leads to smaller increases in surface runoff. By contrast, plant transpiration and soil evaporation is more significantly reduced by CLMET, inducing wetter soil and therefore more subsurface runoff.

4.2.3 Soil moisture

As analyzed in Section 4.2.2, reduction in all three components of ET interception loss, plant transpiration, and soil evaporation from CLM to CLMET slows down moisture depletion from the soil. As a result, the water content at different soil layers increases with the reduced ET. Figure 13 shows soil water at the surface and root-zone layers simulated from CLM and CLMET, and their differences during the summer season (JJA). From CLM to CLMET, the changes over CONUS show an overwhelmingly increase signal for both surface and root-zone soil moisture. The moisture increase in the top 0-100 cm soil layer from CLM to CLMET in central CONUS is very evident, which may have significant implications in drought monitoring and assessment. For example, Central Great Plains experienced a severe drought in summer of 2012, and soil moisture derived from land surface models was used to evaluate the intensity of the drought event (Hoerling et al. 2014, Livneh and Hoerling 2016). Unfortunately, land surface models tend to systematically overestimate drought (Milly and Dunne 2016, Ukkol et al. 2016). The more accurate simulations of ET and soil moisture resulting from the bias correction method used in this study may prove useful in better drought monitoring and assessment.

Figures 14 and 15 show the comparisons between observed soil moisture and modeled soil moisture from CLM and CLMET on the monthly scale during 2006-2010 for the top 0-10 cm and top 0-100 cm soil, respectively. The soil water increase from CLM to CLMET is more evident during SON and DJF, which is consistent with changes in ET that also features more decreases during SON and DJF. Because the soil in CLM shows dry bias over most states with the exception of soil moisture at the top 10 cm layer in Alabama, CLMET generally alleviate the dry bias in CLM. Therefore, the RMSE values against the NASMD observations in CLMET is smaller or at least the same to RMSE values in CLM. An exception exists for the top 0-10 cm layer in Alabama where a wet bias is found in CLM. The soil water content difference between CLM and CLMET is larger for the 0-100 cm layer than the 0-10 cm layer, because plant transpiration, to which a large fraction of ET and therefore a large fraction of ET bias correction are associated, primarily depletes moisture from the rooting zone which is deeper than 10 cm. As such, the improvement is more evident for the top 0-100 cm layer. For example, in Mississippi, the RMSE is reduced from 0.048 m^3 m^{-3} in CLM to 0.042 in CLMET at the top 0-10 cm layer, and from 0.07 to 0.06 m^3 m^{-3} at the top 0-100 cm layer.

5 Summary and discussions

In this study, we implemented the on-line bias correction approach proposed by Parr et al. (2015) to CLM4.5, and evaluated the effectiveness of the approach in reducing model biases over CONUS. The bias correction algorithm was calibrated using the GLEAM ET product combined with the default CLM4.5 output over the period of 1986-1995, and was validated over the period of 2000-2014 using both gridded and site-based ET datasets, the GSCD runoff product, and the NASMD soil moisture data. Results from all evaluation metrics indicate substantial improvement in the estimation of the terrestrial hydrological cycle.

The degree to which the Parr et al. (2015) approach improves the quantification of the hydrological cycle differs among the CONUS sub-regions, and is highly related to whether the fundamental assumption of Parr et al. (2015) (on a time-invariant relationship characterizing the default model biases) holds or not. Although the scaling factors between observations and simulations do not change much from the calibration period to the validation period over most regions in most seasons, dramatic changes do exist in some areas. Differences in the scaling factors between the calibration and verification/application periods greatly influence the effectiveness of the bias correction method, with large differences causing the approach to be less effective leaving substantial biases in CLMET. Northeast CONUS during winter is an example of having a large bias in CLMET due to greater changes in the ET scaling factor from the calibration period to the verification period. Overall, the approach reduces land surface dry biases over eastern CONUS in CLM4.5.

For a given grid cell and given month, the scaling factors for all three ET components, i.e., interception loss, plan transpiration, soil evaporation, are the same in this study, set to be the ratio of the remote sensing ET to the modeled ET. Since the GLEAM dataset contains values of three components besides total ET, we conducted additional experiments in which the scaling factors for each ET component was estimated separately, using the ratio of each ET component from the GLEAM product to the corresponding ET component from CLM during the same calibration period. However, results based on the component-specific scaling factors do not show any improvement, which is likely caused by the inaccurate partitioning of ET into interception loss, plan transpiration, soil evaporation. Miralles et al. (2016) compared the ET partitioning for three widely used remote sensing based ET products, and found that the contribution of each component to ET is dramatically different among these three products. For instance, the percentage of global ET accounted for by soil evaporation ranges from 14% to 52%, and the ranges are even larger at the regional and local scales. Because the in-situ measurements of separate components of ET is very scarce, it is particularly challenging to validate the accuracy of the remote sensing based estimated of the three ET components. These challenges led Miralles et al. (2016) to recommend against the use of any single product in partitioning ET.

The bias correction method evaluated in this study can effectively improves the estimates of surface fluxes and state variables in the absence of improved physical parameterizations in land surface models. It is applicable to not only historical simulations but also future predictions (Parr et al. 2015). It provides an alternative approach to, but would in no way to replace, model improvement through better parameterization of physical processes. Development of better physical parameterizations has to be based on improved understanding of physical processes, more effective mathematical formulations, and higher quality surface type dataset, which requires a long-term commitment from the land surface modeling community. Model parameter calibration (e.g., tuning surface resistance) is another way to reduce model bias (Ren et al. 2016). However, the parameter space may contain nonphysical parameter subsets (Ray et al. 2015), which is especially an issue when model parameter tuning is used to offset unrelated model deficits. The method used in this study attempts to avoid such issues through improving the model performance without dealing with calibration of model physical parameters.

6. Data availability

The GLEAM ET data was provided by the GLEAM team at the website www.GLEAM.eu. The MODIS ET data by NTSG, University of Montana at the website http://www.ntsg.umt.edu/project/mod16. The FLUXNET-MTE ET data was provided by Max Planck Institute for Biogeochemistry at the website https://www.bgc-jena.mpg.de/geodb/projects/Data.php. The GSCD runoff data was provided by the Amsterdam Critical Zone Hydrology Group at the website http://hydrology-amsterdam.nl/valorisation/GSCD.html. The original NASMD soil moisture data is available at the website http://soilmoisture.tamu.edu/. The quality-controlled NASMD soil moisture data can be obtained from the authors upon request. Latent fluxes measurements at tower sites is available Flux http://ameriflux.lbl.gov/

Author contributions

D. Wang and G. Wang designed the study. D. Wang conducted model simulations and data analysis with input from G. Wang, D. Parr and C. Fu, D. Wang and G. Wang wrote the paper with input from Y. Xia. W. Liao and Y. Xia contributed to data processing.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

This study is supported by National Natural Science Foundation of China (Grant No. 51379224), and the Fundamental Research Funds for the Central Universities.

References

Ahmed, M., Sultan, M., Yan, E., and Wahr, J.: Assessing and Improving Land Surface Model Outputs Over Africa Using GRACE, Field, and Remote Sensing Data, Surv. Geophys., 37, 1-28, 2016.

Beck, H. E., Dijk, A. I. J. M., Miralles, D. G., Jeu, R. A. M. D., Bruijnzeel, L. A., Mcvicar, T. R., and Schellekens, J.: Global patterns in base flow index and recession based on streamflow observations from 3394 catchments, Water Resour. Res., 49, 7843-7863, 2013.

Beck, H., De Roo, A., and Van Dijk, A.: Global Maps of Streamflow Characteristics Based on Observations from Several Thousand Catchments, J. Hydrometeorol., 2015.

Bonan, G. B., Lawrence, P. J., Oleson, K. W., Samuel, L., Martin, J., Markus, R., Lawrence, D. M., and Swenson, S. C.: Improving canopy processes in the Community Land Model version 4 (CLM4) using global flux fields empirically inferred from FLUXNET data, Journal of Geophysical Research Biogeosciences, 116, G2014, 2011.

Bonan, G. B., Oleson, K., Vertenstein, M., Levis, S., Zeng, X., Dai, Y., Dickinson, R., and Yang, Z.: The land surface climatology of the Community Land Model coupled to the NCAR Community Climate Model, J. Climate, 15, 3123-3149, 2002.

Cai, X., Yang, Z. L., Xia, Y., Huang, M., Wei, H., Leung, L. R., and Ek, M. B.: 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, 713-751, 770, 2014.

Cheng, S., Guan, X., Huang, J., Ji, F., and Guo, R.: Long - term trend and variability of soil moisture over East Asia, J. Geophys. Res., 120, 8658-8670, 2015.

Dickinson, R. E., Oleson, K., Bonan, G., Hoffman, F. M., Thornton, P., Vertenstein, M., Yang, Z., and Zeng, X.: The Community Land Model and Its Climate Statistics as a Component of the Community Climate System Model, J. Climate, 19, 2302-2324, 2010.

Dorigo, W. A., Xaver, A., Vreugdenhil, M., Gruber, A., Hegyiová, A., Sanchisdufau, A. D., Zamojski, D., Cordes, C., Wagner, W., and Drusch, M.: Global automated quality control of in situ soil moisture data from the International Soil Moisture Network, Vadose Zone J., 12, 918-924, 2013.

Getirana, A. C. V., Dutra, E., Guimberteau, M., Kam, J., Li, H. Y., Decharme, B., Zhang, Z., Ducharne, A., Boone, A., Balsamo, G., Rodell, M., Toure, A. M., Xue, Y., Peterslidard, C. D., Kumar, S., Arsenault, K. R., Drapeau, G., Leung, L. R., Ronchail, J., and Sheffield, J.: Water balance in the Amazon Basin from a land surface model ensemble, J. Hydrometeorol., 15, 2586-2614, 2014.

Guillod, B. P., Orlowsky, B., Miralles, D., Teuling, A. J., Blanken, P. D., Buchmann, N., Ciais, P., Ek, M., Findell, K. L., Gentine, P., Lintner, B. R., Scott, R. L., Van Den Hurk, B. J. J. M., and Seneviratne, S. I.: Land-surface controls on afternoon precipitation diagnosed from observational data: uncertainties and confounding factors, Atmos. Chem. Phys., 14, 8343-8367, 2014.

Haddeland, I., Clark, D. B., Franssen, W., Ludwig, F., Voß, F., Arnell, N. W., Bertrand, N., Best, M. J., Folwell, S. S., Gerten, D., Gomes, S., Gosling, S. N., Hagemann, S., Hanasaki, N., Harding, R. J., Heinke, J., Kabat, P., Koirala, S., Oki, T., Polcher, J., Stacke, T., Viterbo, P., Weedon, G. P., and Yeh, P. J. F.: Multimodel estimate of the global terrestrial water balance: setup and first results, J. Hydrometeorol., 12, 869-884, 2011.

Hoerling, M., Eischeid, J., Kumar, A., Leung, R., Mariotti, A., Mo, K., Schubert, S., and Seager, R.: Causes and Predictability of the 2012 Great Plains Drought, B. Am. Meteorol. Soc., 95, 269-282, 2014.

Jung, M., Reichstein, M., Ciais, P., Seneviratne, S. I., Sheffield, J., Goulden, M. L., Bonan, G., Cescatti, A., Chen, J., Jeu, R. D., Dolman, A. J., Eugster, W., Gerten, D., Gianelle, D., Gobron, N., Heinke, J., Kimball, J., Law, B. E., Montagnani, L., Mu, Q., Mueller, B., Oleson, K., Papale, D., Richardson, A. D., Roupsard, O., Running, S., Tomelleri, E., Viovy, N., Weber, U., Williams, C., Wood, E., Zaehle, S., and Zhang, K.: Recent decline in the global land evapotranspiration trend due to limited moisture supply., Nature, 467, 951-954, 2010.

Jung, M., Reichstein, M., and Bondeau, A.: Towards global empirical upscaling of FLUXNET eddy covariance observations: Validation of a model tree ensemble approach using a biosphere model, Biogeosciences, 6, 2001-2013, 2009.

Kim, H. and Choi, M.: Impact of soil moisture on dust outbreaks in East Asia: Using satellite and assimilation data, Geophys. Res. Lett., 42, 2789-2796, 2015.

Kumar, S. V., Reichle, R. H., Peters-Lidard, C. D., Koster, R. D., Zhan, X., Crow, W. T., Eylander, J. B., and Houser, P. R.: A land surface data assimilation framework using the land information system: Description and applications, Adv. Water Resour., 31, 1419-1432, 2008.

Lawrence, D. M., Oleson, K. W., Flanner, M. G., Thornton, P. E., Swenson, S. C., Lawrence, P. J., Zeng, X., Yang, Z., Levis, S., Sakaguchi, K., Bonan, G. B., and Slater, A. G.: Parameterization improvements and functional and structural advances in Version 4 of the Community Land Model, J. Adv. Model. Earth Sy., 3, 365-375, 2011.

Livneh, B. and Hoerling, M. P.: The Physics of Drought in the U.S. Central Great Plains, J. Climate, 29, 6783-6804, 2016.

Lohmann, D., Mitchell, K. E., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., Cosgrove, B. A., Sheffield, J., Duan, Q., Luo, L., Higgins, R. W., Pinker, R. T., and Tarpley, J. D.: Streamflow and water balance intercomparisons of four land surface models in the North American Land Data Assimilation System project, Journal of Geophysical Research Atmospheres, 109, 585-587, 2004.

Mahrt, L. and Pan, H.: A two-layer model of soil hydrology, Bound.-Lay. Meteorol., 29, 1-20, 1984.

Martens, B., Miralles, D. G., Lievens, H., Schalie, R. V. D., Jeu, R. A. M. D., Férnandez-Prieto, D., Beck, H. E., Dorigo, W. A., and Verhoest, N. E. C.: GLEAM v3: satellite-based land evaporation and root-zone soil moisture, Geoscientific Model Development Discussions, 1-36, 2016.

Michel, D., Jiménez, C., Miralles, D. G., Jung, M., Hirschi, M., Ershadi, A., Martens, B., Mccabe, M. F., Fisher, J. B., Mu, Q., Seneviratne, S. I., Wood, E. F., and Fernández-Prieto, D.: The WACMOS-ET project - Part 1: Tower-scale evaluation of four remote sensing-based evapotranspiration algorithm, Hydrol. Earth Syst. Sc., 20, 803-822, 2016.

Milly, P. C. D. and Dunne, K. A.: Potential evapotranspiration and continental drying, Nat. Clim. Change, 6, 946-949, 2016.

Miralles, D. G., Berg, M. J. V. D., Gash, J. H., Parinussa, R. M., Jeu, R. A. M. D., Beck, H. E., Holmes, T. R. H., Jiménez, C., Verhoest, N. E. C., Dorigo, W. A., Teuling, A. J., and Dolman, A. J.: El Niño – La Niña cycle and recent trends in continental evaporation, Nat. Clim. Change, 4, 122-126, 2014.

Miralles, D. G., Berg, M. J. V. D., Teuling, A. J., and Jeu, R. A. M. D.: Soil moisture – temperature coupling: A multiscale observational analysis, Geophys. Res. Lett., 39, L21707, 2012.

Miralles, D. G., Holmes, T. R. H., Jeu, R. A. M. D., and Gash, J. H.: Global land-surface evaporation estimated from satellite-based observations, Hydrology and Earth System Sciences Discussions, 7, 453-469, 2011.

Miralles, D. G., Jiménez, C., Jung, M., Michel, D., Ershadi, A., Mccabe, M. F., Hirschi, M., Martens, B., Dolman, A. J., Fisher, J. B., Mu, Q., Seneviratne, S. I., Wood, E. F., and Fernández-Prieto, D.: The WACMOS-ET project – Part 2: Evaluation of global terrestrial evaporation data sets, Hydrol. Earth Syst. Sc., 20, 823-842, 2016.

Mu, Q., Heinsch, F. A., Zhao, M., and Running, S. W.: Development of a global evapotranspiration algorithm based on MODIS and global meteorology data, Remote Sens. Environ., 111, 519-536, 2007.

Mu, Q., Zhao, M., and Running, S. W.: Improvements to a MODIS global terrestrial evapotranspiration algorithm, Remote Sens. Environ., 115, 1781-1800, 2011.

Niu, G., Yang, Z., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M., and Xia, Y.: 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, 2011.

Oleson, K. W., Niu, G. Y., Yang, Z. L., Lawrence, D. M., Thornton, P. E., Lawrence, P. J., Stöckli, R., Dickinson, R. E., Bonan, G. B., Levis, S., Dai, A., and Qian, T.: Improvements to the Community Land Model and their impact on the hydrological cycle, Journal of Geophysical Research Atmospheres, 113, 811-827, 2008.

Oleson, K. W., Lawrence, D. M., Bonan, G. B., Drewniak, B., Huang, M., Koven, C. D., Levis, S., Li, F., Riley, W. J., Subin, Z. M., Swenson, S. C., Thornton, P. E., Bozbiyik, A., Fisher,

R. A., Kluzek, E., Lamarque, J.-F., Lawrence, P. J., Leung, L. R., Lipscomb, W., Muszala, S., Ricciuto, D. M., Sacks, W. J., Sun, Y., Tang, J. Y., and Yang, Z.-L.: Technical Description of version 4.5 of the Community Land Model (CLM), NCAR Tech. Note, NCAR/TN-503+STR, doi:10.5065/D6RR1W7M, 2013.

Parr, D., Wang, G., and Bjerklie, D.: Integrating Remote Sensing Data on Evapotranspiration and Leaf Area Index with Hydrological Modeling: Impacts on Model Performance and Future Predictions, J. Hydrometeorol., 16, 2086-2100, 2015.

Parr, D., Wang, G., and Fu, C.: Understanding Evapotranspiration Trends and their Driving Mechanisms over the NLDAS Domain Based on Numerical Experiments Using CLM4.5, Journal of Geophysical Research Atmospheres, 121, 7729-7745, 2016.

Ray, J., Z. Hou, M. Huang, K. Sargsyan, and L. Swiler: Bayesian calibration of the Community Land Model using surrogates, SIAM/ASA Journal on Uncertainty Quantification, 199 – 233, 2015.

Reichle, R. H. and Koster, R. D.: Global assimilation of satellite surface soil moisture retrievals into the NASA Catchment land surface model, Geophys. Res. Lett., 32, 177-202, 2005.

Ren, H., Z. Hou, M. Huang, J. Bao, Y. Sun, T. Tesfa, and R. Leung: Classification of hydrological parameter sensitivity and evaluation of parameter transferability across 431 US MOPEX basins, J. Hydrol., 536, 92–108, 2016.

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J. C., Mitchell, K., Meng, C. J., Arsenault, K. R., Cosgrove, B. A., Radakovich, J., Bosilovich, M. G., Entin, J. K., Walker, J. P., Lohmann, D., and Toll, D. L.: The Global Land Data Assimilation System, B. Am. Meteorol. Soc., 85, 381-394, 2004.

Sheffield, J. and Wood, E. F.: Characteristics of global and regional drought, 1950-2000: Analysis of soil moisture data from off-line simulation of the terrestrial hydrologic cycle, Journal of Geophysical Research Atmospheres, 112, D17115, 2007.

Spennemann, P. C. and Saulo, A. C.: An estimation of the land-atmosphere coupling strength in South America using the Global Land Data Assimilation System, Int. J. Climatol., 35, 4151-4166, 2015.

Swenson, S. C. and Lawrence, D. M.: A GRACE‐based assessment of interannual groundwater dynamics in the Community Land Model, Water Resour. Res., 51, 8817-8833, 2015.

Syed, T. H., Famiglietti, J. S., Rodell, M., Chen, J., and Wilson, C. R.: Analysis of terrestrial water storage changes from GRACE and GLDAS, Water Resour. Res., 44, 339-356, 2008.

Ukkola, A. M., Kauwe, M. G. D., Pitman, A. J., Best, M. J., Abramowitz, G., Haverd, V., Decker, M., and Haughton, N.: Land surface models systematically overestimate the intensity, duration and magnitude of seasonal-scale evaporative droughts, Environ. Res. Lett., 11, 2016.

Wang, A., Zeng, X., and Guo, D.: Estimates of global surface hydrology and heat fluxes from the Community Land Model (CLM4.5) with four atmospheric forcing datasets, J. Hydrometeorol., 17, 2493-2510, 2016.

Xia, Y., Cosgrove, B. A., Mitchell, K. E., Peters Lidard, C. D., Ek, M. B., Kumar, S., Mocko, D., and Wei, H.: Basin‐scale assessment of the land surface energy budget in the National Centers for Environmental Prediction operational and research NLDAS-2 systems, J. Geophys. Res., 121, 196-220, 2016a.

Xia, Y., Ford, T. W., Wu, Y., Quiring, S. M., and Ek, M. B.: Automated Quality Control of In Situ Soil Moisture from the North American Soil Moisture Database Using NLDAS-2 Products, Journal of Applied Meteorology & Climatology, 54, 2015a.

Xia, Y., Hobbins, M. T., Mu, Q., & Ek, M. B. Evaluation of NLDAS-2 evapotranspiration against tower flux site observations. Hydrological Processes, 29(7), 1757-1771, 2015b.

Xia, Y., Mitchell, K. E., Ek, M. B., Cosgrove, B., Sheffield, J., Luo, L., Alonge, C., Wei, H., Meng, J., Livneh, B., Duan, Q., and Lohmann, D.: Continental-scale water and energy flux analysis and validation for North American Land Data Assimilation System project phase 2 (NLDAS‐2): 2. Validation of model‐simulated streamflow, J. Geophys. Res., 117, D3110, 2012a.

Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., Luo, L., Alonge, C., Wei, H., Meng, J., Livneh, B., Lettenmaier, D., Koren, V., Duan, Q., Mo, K., Fan, Y., and Mocko, D.: Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products, Journal of Geophysical Research Atmospheres, 117, D3109, 2012b.

Xia, Y., Peters-Lidard, C. D., and Luo, L.: Basin-scale assessment of the land surface water budget in the National Centers for Environmental Prediction operational and research NLDAS-2 systems, J. Geophys. Res., 121, 196-220, 2016b.

Yin, J., Zhan, X., Zheng, Y., Liu, J., Fang, L., and Hain, C. R.: Enhancing Model Skill by Assimilating SMOPS Blended Soil Moisture Product into Noah Land Surface Model, J. Hydrometeorol., 16, 917-931, 2015c.

Table 1 Spatial evaluations of simulated ET from two different types of runs (CLM and CLMET) against GLEAM-derived ET over CONUS, Northwest (NW), Southwest (SW), Northeast (NW), and Southeast (SW) annually and seasonally during the period 2000-2014. March-April-May: MAM, June-July-August: JJA, September-October-November: SON, December-January-February: DJF

Season	Region	Bias (mm day^{-1})		Relative bias (%)		RMSE (mm day^{-1})	
		CLM	CLMET	CLM	CLMET	CLM	CLMET
Annual	CONUS	0.137	-0.006	10.8	-0.1	0.266	0.144
	NW	0.029	-0.03	7.9	0.3	0.25	0.199
	SW	0.074	-0.025	10.2	-3.1	0.181	0.118
	NE	0.138	-0.012	9.6	-0.1	0.243	0.132
	SE	0.315	0.041	15.6	2.1	0.355	0.099
MAM	CONUS	-0.081	-0.062	-5.8	-3.3	0.351	0.227
	NW	-0.138	-0.074	-6.7	-2.7	0.326	0.244
	SW	-0.211	-0.122	-17.9	-9.3	0.318	0.206
	NE	-0.191	-0.078	-8.3	-2.8	0.429	0.293
	SE	0.19	0.023	8.9	1.5	0.346	0.165
JJA	CONUS	0.094	-0.041	6.4	-1.3	0.451	0.331
	NW	-0.137	-0.121	-3.9	-4.0	0.487	0.408
	SW	0.147	-0.006	18.3	-0.9	0.352	0.232
	NE	0.045	-0.124	2.5	-2.7	0.55	0.452
	SE	0.332	0.075	9.1	2.1	0.414	0.181
SON	CONUS	0.360	0.055	51	7.8	0.428	0.155
	NW	0.271	0.044	76.4	14.0	0.346	0.147
	SW	0.228	0.044	39.5	5.0	0.282	0.117
	NE	0.481	0.077	50.4	7.3	0.527	0.242
	SE	0.499	0.061	34.5	4.1	0.531	0.11
DJF	CONUS	0.182	0.009	77.7	18.9	0.265	0.115
	NW	0.114	-0.013	104.2	28.8	0.252	0.122
	SW	0.132	-0.014	42.3	-1.9	0.182	0.056
	NE	0.239	0.077	146.4	65.3	0.334	0.199
	SE	0.24	0.004	49.5	2.7	0.292	0.072

Table 2. Similar to Table 1, but based on comparison with MODIS-derived ET during the period 2000-2011.

Season	Region	Bias (mm day^{-1})		Relative bias (%)		RMSE (mm day^{-1})	
		CLM	CLMET	CLM	CLMET	CLM	CLMET
Annual	CONUS	0.321	0.177	30.8	19.1	0.427	0.321
	NW	0.28	0.232	35.8	27.9	0.367	0.326
	SW	0.282	0.183	39.7	25.6	0.428	0.36
	NE	0.278	0.125	19.6	9.1	0.316	0.193
	SE	0.431	0.159	24.9	10.6	0.538	0.348
MAM	CONUS	0.514	0.533	50.1	55.8	0.631	0.635
	NW	0.564	0.628	67.2	74.5	0.636	0.687
	SW	0.345	0.438	45.9	61.8	0.538	0.599
	NE	0.547	0.655	51.7	61.9	0.58	0.675
	SE	0.596	0.436	34.6	25.8	0.735	0.578
JJA	CONUS	0.251	0.116	18.2	12.1	0.759	0.691
	NW	0.263	0.281	23.8	25.6	0.704	0.71
	SW	0.344	0.192	28.8	14.5	0.806	0.724
	NE	0.028	-0.144	2.9	-2.4	0.662	0.564
	SE	0.31	0.052	13.2	5.8	0.829	0.72
SON	CONUS	0.345	0.039	48.2	9.8	0.459	0.284
	NW	0.261	0.038	56.8	9.4	0.369	0.261
	SW	0.284	0.096	55.9	20.8	0.43	0.306
	NE	0.448	0.043	47.4	5.6	0.483	0.207
	SE	0.417	-0.019	32.1	2.7	0.547	0.329
DJF	CONUS	0.181	0.025	82.2	28	0.383	0.276
	NW	0.043	-0.049	77.6	40.4	0.385	0.365
	SW	0.156	0.007	70.5	19.4	0.292	0.191
	NE	0.091	-0.051	96.7	14.8	0.344	0.214
	SE	0.403	0.169	87.5	33.6	0.474	0.281

Table 3. Similar to Table 3, but based on comparison with FLUXNET-MTE ET during the period 2000-2011.

Season	Region	Bias (mm day^{-1})		Relative bias (%)		RMSE (mm day^{-1})	
		CLM	CLMET	CLM	CLMET	CLM	CLMET
Annual	CONUS	0.207	0.065	13.3	3.2	0.328	0.24
	NW	0.07	0.013	5.8	0.0	0.222	0.234
	SW	0.051	-0.047	6.8	-4.7	0.244	0.241
	NE	0.309	0.165	21.9	12.2	0.334	0.238
	SE	0.427	0.154	21.3	7.6	0.461	0.248
MAM	CONUS	0.27	0.292	15.8	19.5	0.418	0.399
	NW	0.266	0.33	22.4	28.0	0.349	0.401
	SW	-0.042	0.051	-7.3	2.5	0.298	0.301
	NE	0.288	0.401	21.6	30.4	0.338	0.435
	SE	0.561	0.4	26.4	18.5	0.6	0.448
JJA	CONUS	0.197	0.063	7.0	0.5	0.608	0.517
	NW	-0.149	-0.13	-8.7	-7.5	0.506	0.506
	SW	0.029	-0.122	9.2	-6.1	0.594	0.555
	NE	0.415	0.257	13.6	8.8	0.492	0.369
	SE	0.565	0.304	16.9	9.4	0.779	0.585
SON	CONUS	0.216	-0.088	20.3	-9.4	0.353	0.294
	NW	0.072	-0.151	9.2	-22.8	0.224	0.286
	SW	0.132	-0.055	21.1	-5.2	0.311	0.277
	NE	0.356	-0.034	33.7	-1.1	0.473	0.385
	SE	0.346	-0.091	21.2	-5.4	0.396	0.23
DJF	CONUS	0.149	-0.004	40.1	-1	0.268	0.189
	NW	0.104	0.014	27	-4.9	0.279	0.26
	SW	0.086	-0.063	20.9	-14.4	0.17	0.129
	NE	0.176	0.037	78.5	19.2	0.329	0.208
	SE	0.236	0.002	42.8	0.8	0.282	0.129

Table 4 Statistics of simulated annual runoff coefficient (ratio of runoff to total precipitation) against GSCD observations over CONUS, Northwest (NW), Southwest (SW), Northeast (NW), and Southeast (SW) during the period 2000-2014.

	Bias		Relative bias (%)		RMSE	
	CLM	CLMET	CLM	CLMET	CLM	CLMET
CONUS	-0.053	-0.027	-18.5	-6.7	0.198	0.192
Northwest	-0.046	-0.036	-13.5	-5.6	0.146	0.144
Southwest	-0.026	-0.019	-19.9	-11.4	0.373	0.373
Northeast	-0.06	-0.022	-15.7	-1.5	0.108	0.092
Southeast	-0.074	-0.026	-24.7	-8.2	0.091	0.06

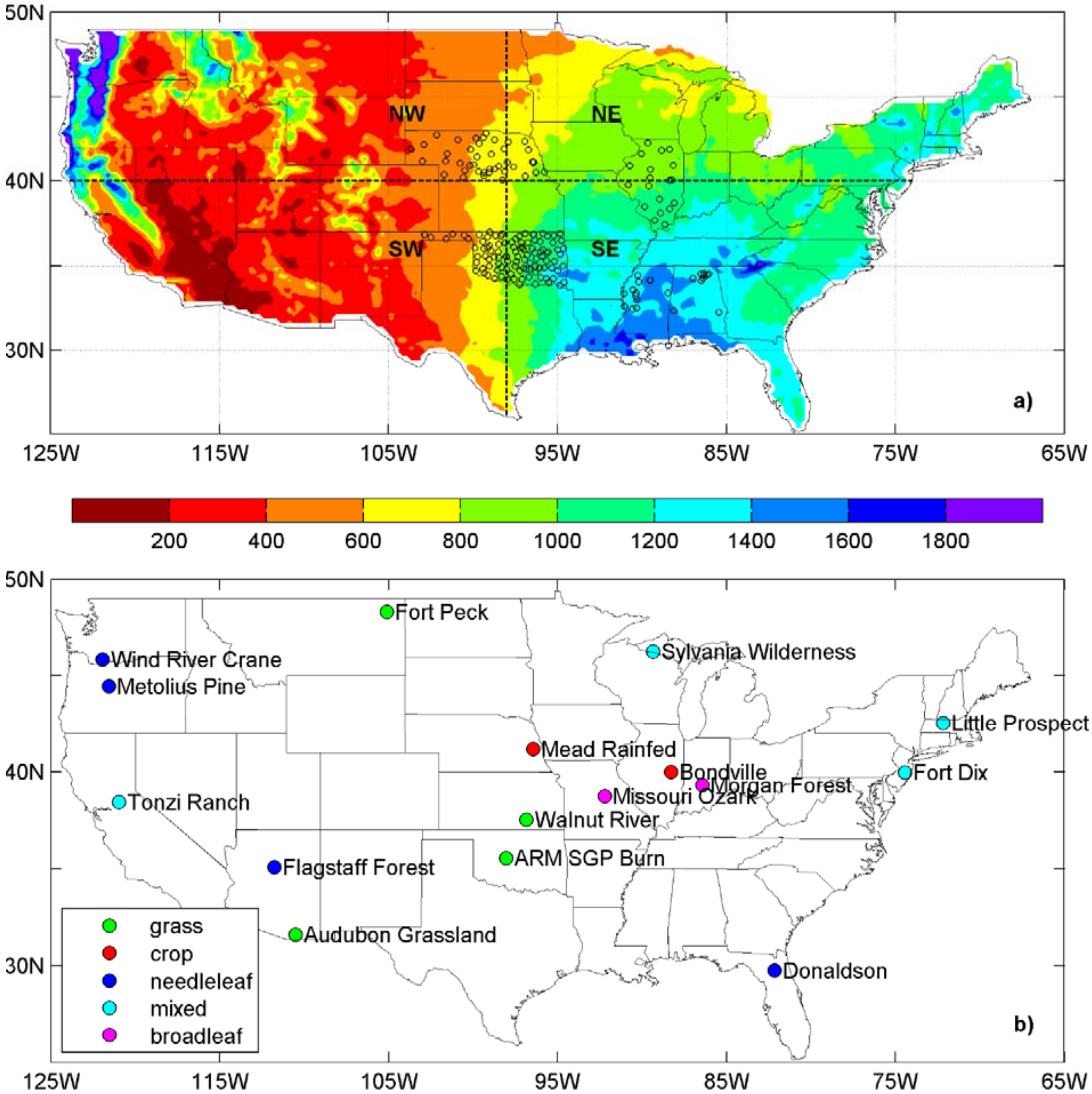


Figure 1 a) Mean annual (1980-2015) precipitation in mm over conterminous USA (CONUS). NW, SW, NE, and SE represent Northwest, Southwest, Northeast, and Southeast, respectively. The black circles represent sites of in-situ soil moisture observations in Alabama, Illinois, Mississippi, Nebraska, and Oklahoma. b) Locations of the 16 AmeriFlux stations with vegetation types.

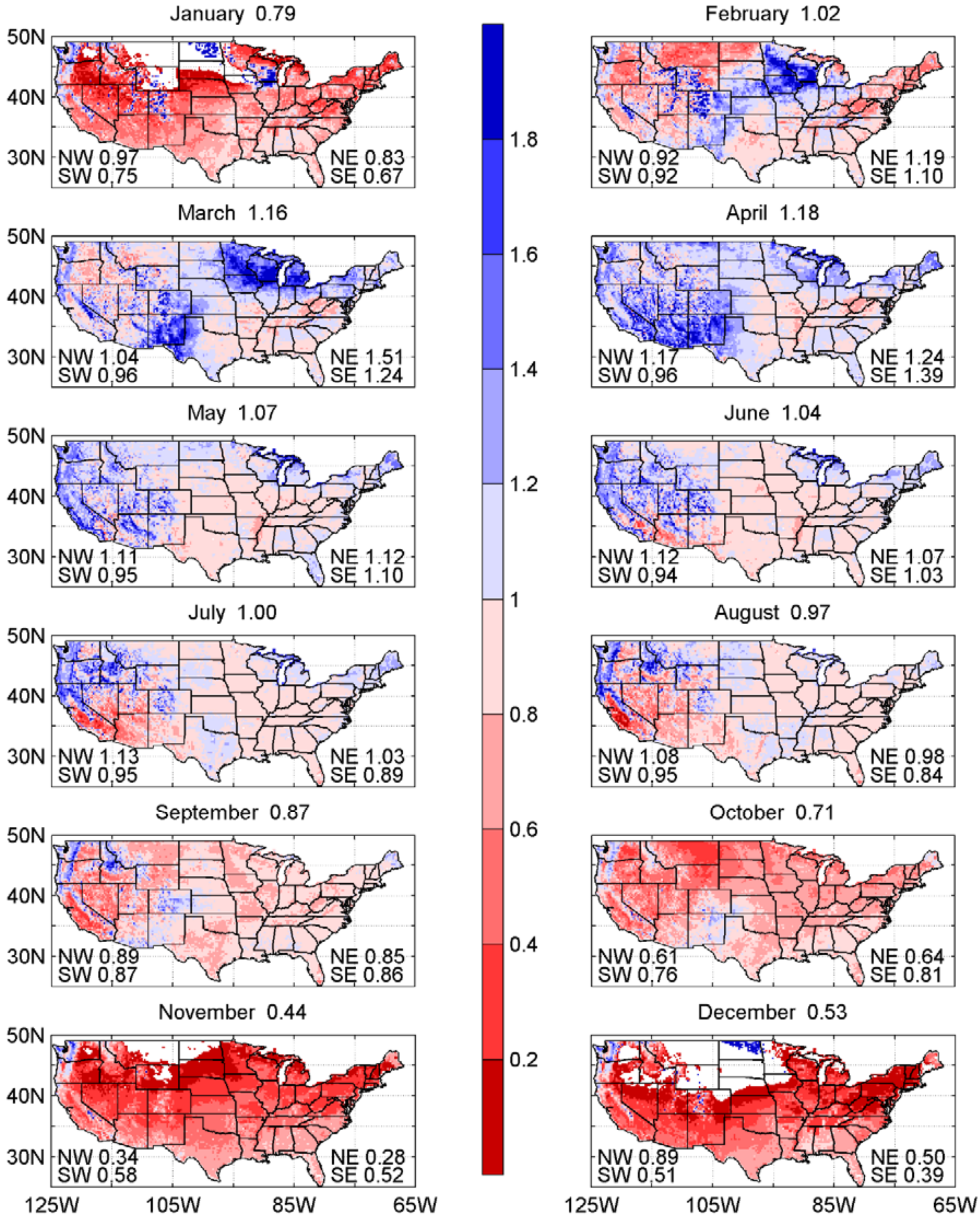


Figure 2 Scaling factors of the CLM simulated ET to the GLEAM ET for each month during 1986-1995. The numbers in titles are CONUS-averaged values, and the number of within figures are area-averaged values for each of four sub regions (NW, SW, NE, and SE). The areas with negative scaling factors are masked out.

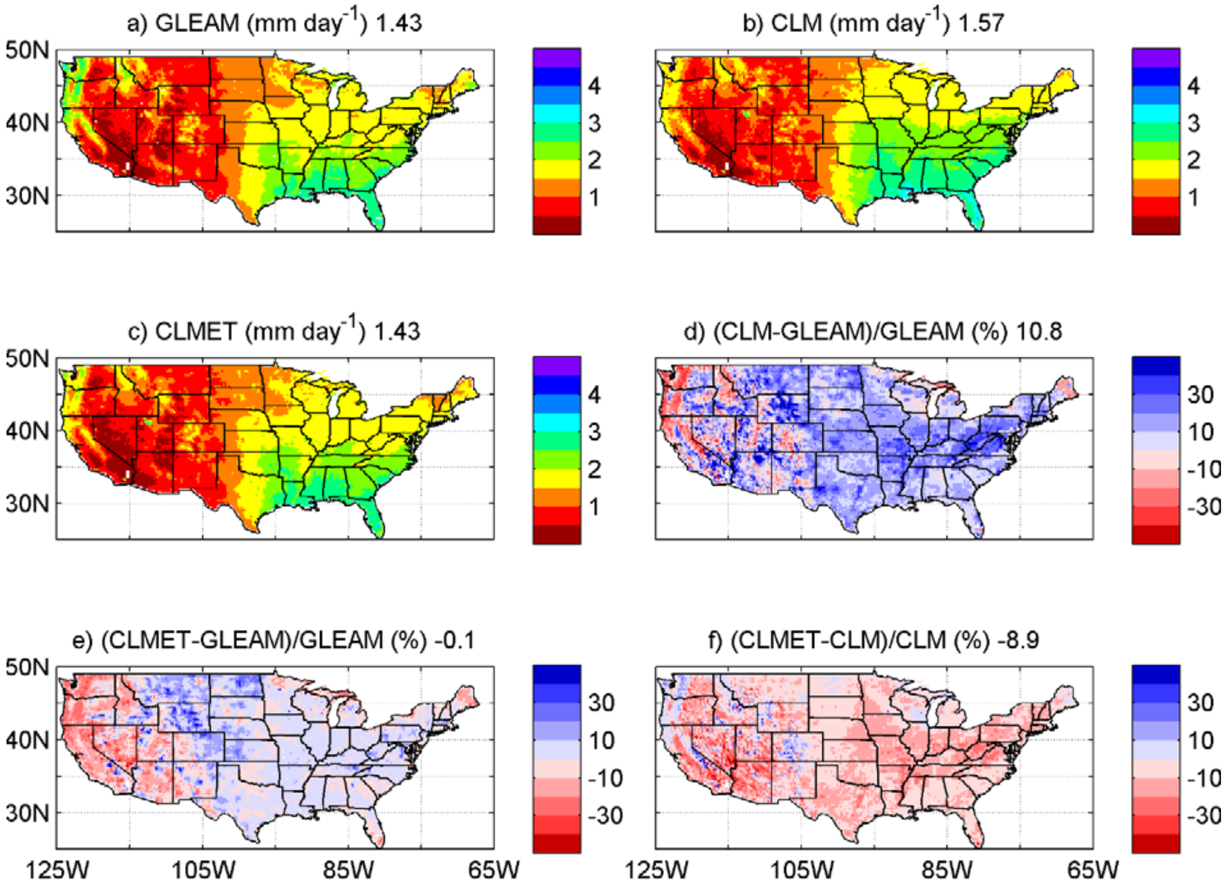
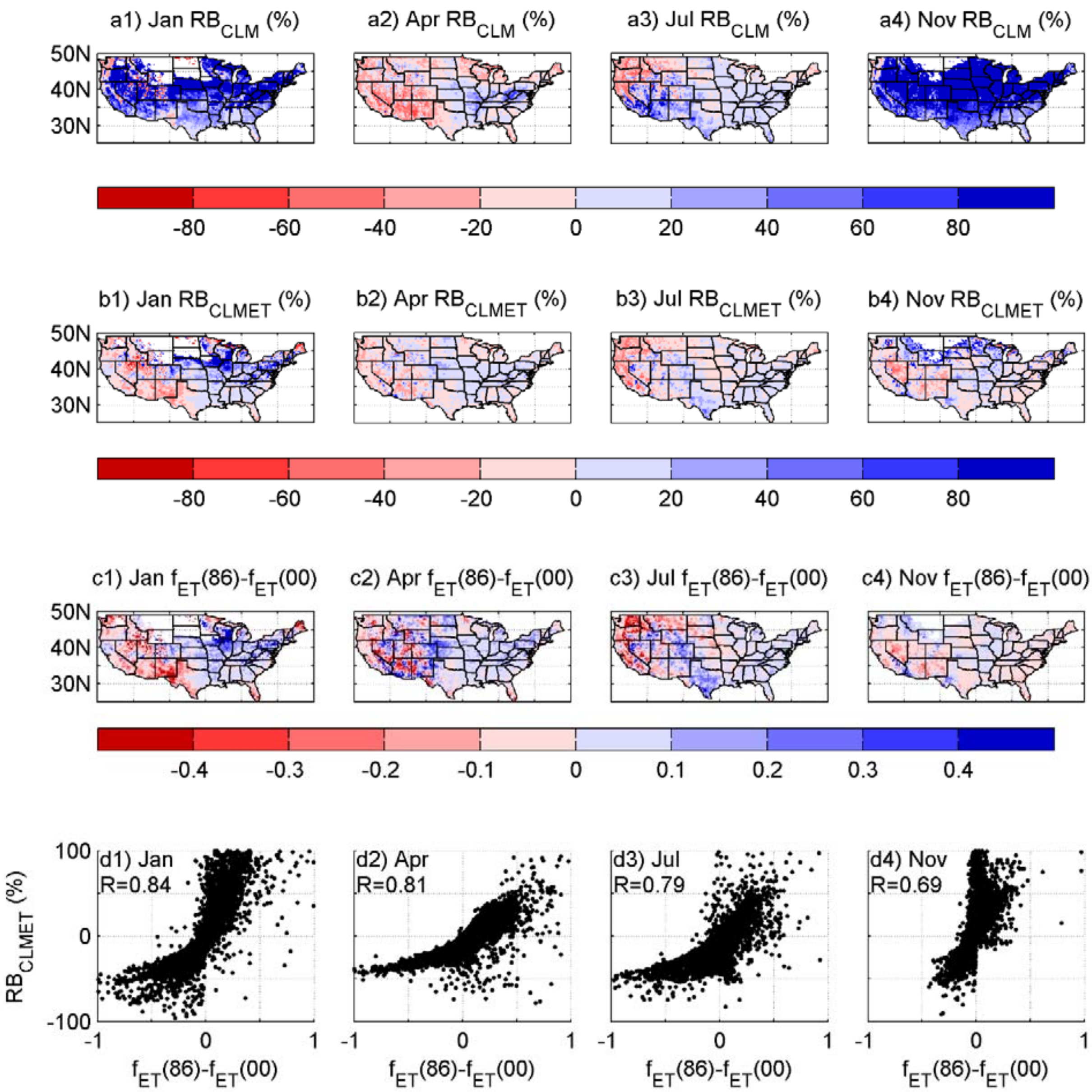


Figure 3 Mean annual ET from a) GLEAM, b) CLM, and c) CLMET, and the relative difference between d) CLM and GLEAM, e) CLMET and GLEAM, and f) CLMET and CLM during 2000-2014. Numbers in titles are CONUS-averaged values.



an

Figure 4 Relative bias (RB) for CLM (RB_{CLM}), RB for CLMET (RB_{CLMET}) during the period 2000-2014, difference in scaling factor f_{ET} between the period 1986-1995 and the period 2000-2014 ($f_{ET}(86)$- $f_{ET}(00)$), and scatter plots of $f_{ET}(86)$- $f_{ET}(00)$ versus RB_{CLMET} in January (Jan), April (Apr), July (Jul), and November (Nov).

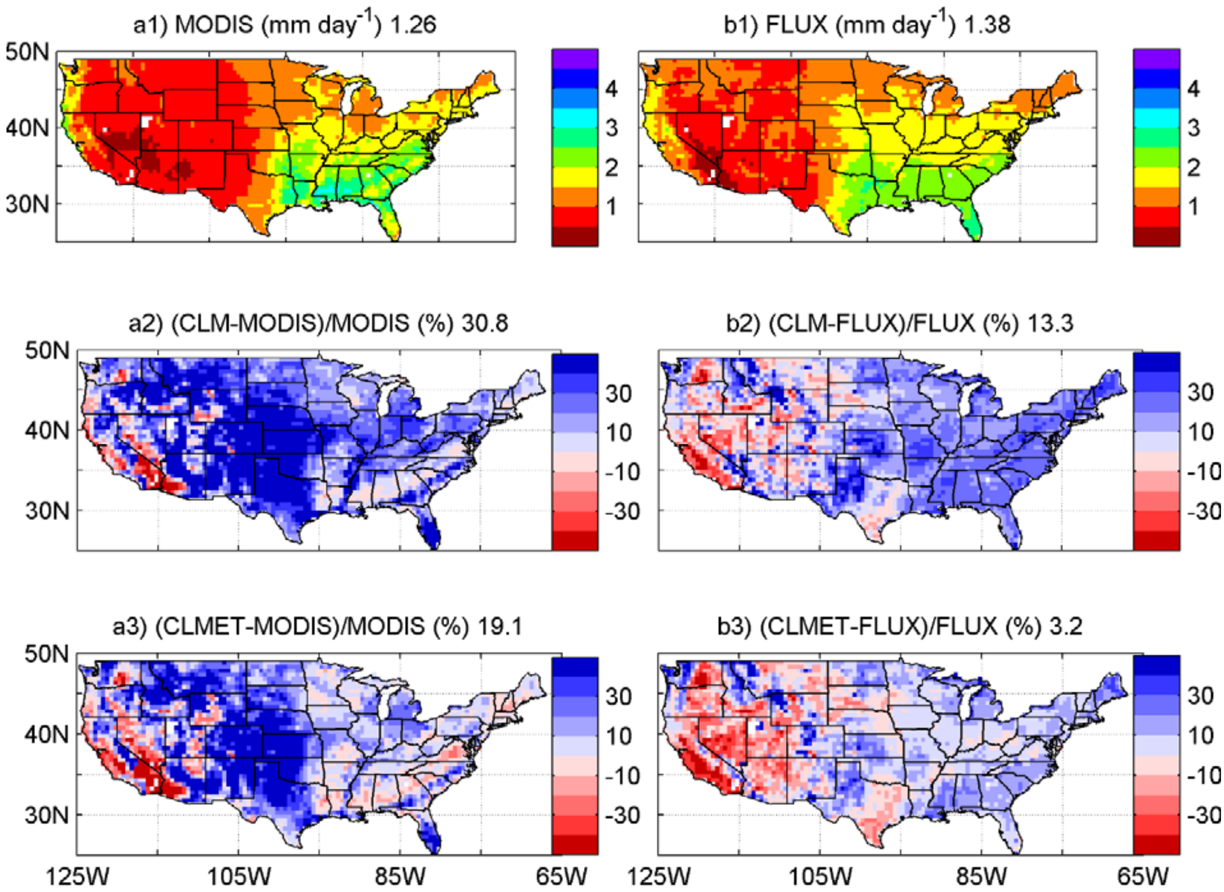


Figure 5 Mean annual ET from a1) MODIS, b1) FLUXNET-MTE, and the relative differences between a2) CLM and MODIS, b2) CLM and FLUXNET-MTE, a3) CLMET and MODIS, and b3) CLMET and FLUXNET-MTE during 2000-2011. Numbers in titles are CONUS-averaged values.

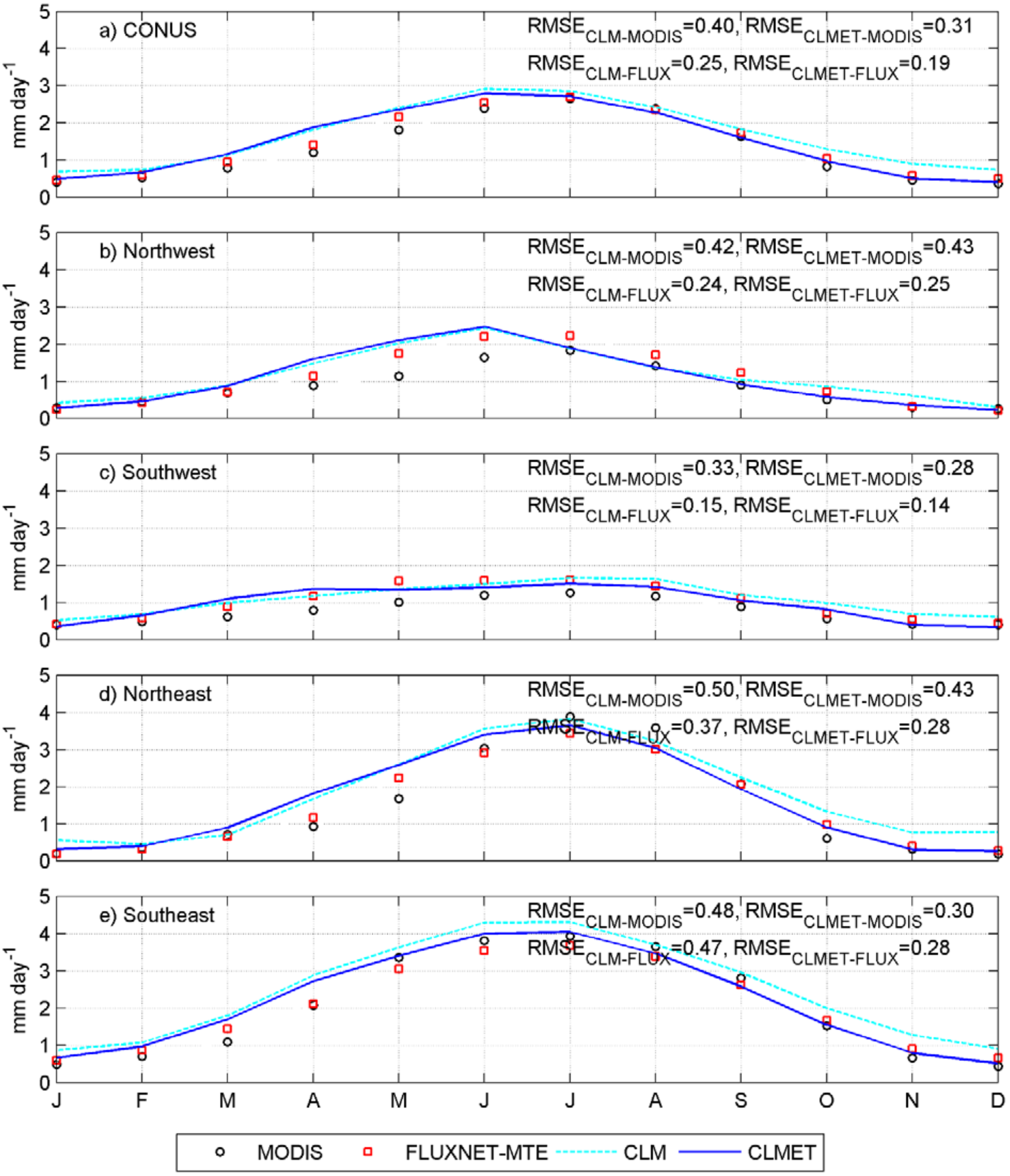


Figure 6 Seasonal cycles of ET from MODIS, FLUXNET-MTE, CLM, and CLMET over CONUS, Northwest, Southwest, Northeast, and Southeast during the period 2000-2011.

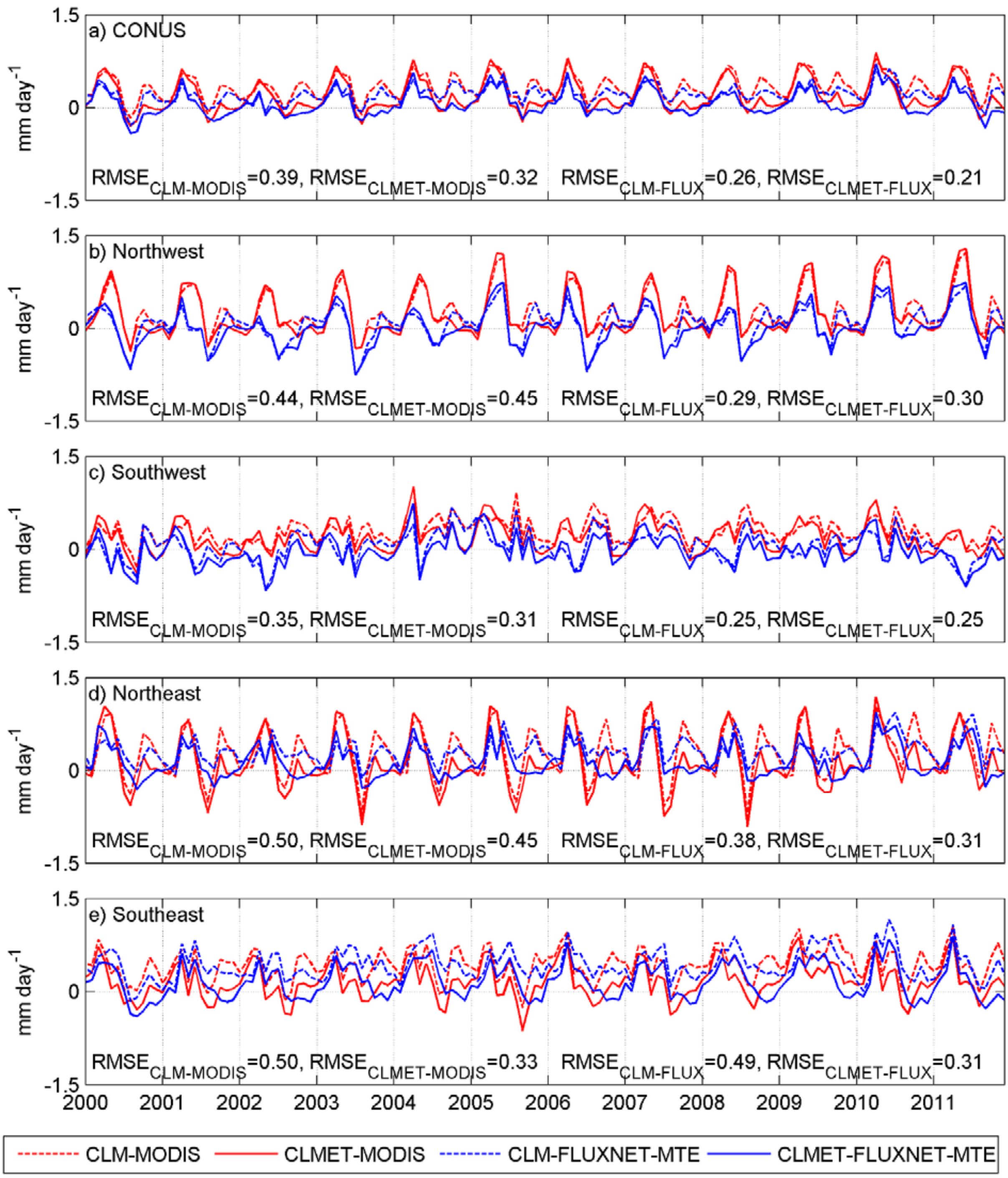


Figure 7 Time series of ET difference between model (CLM or CLMET) and reference data (MODIS or FLUXNET-MTE) over CONUS, Northwest, Southwest, Northeast, and Southeast during the period 2000-2011.

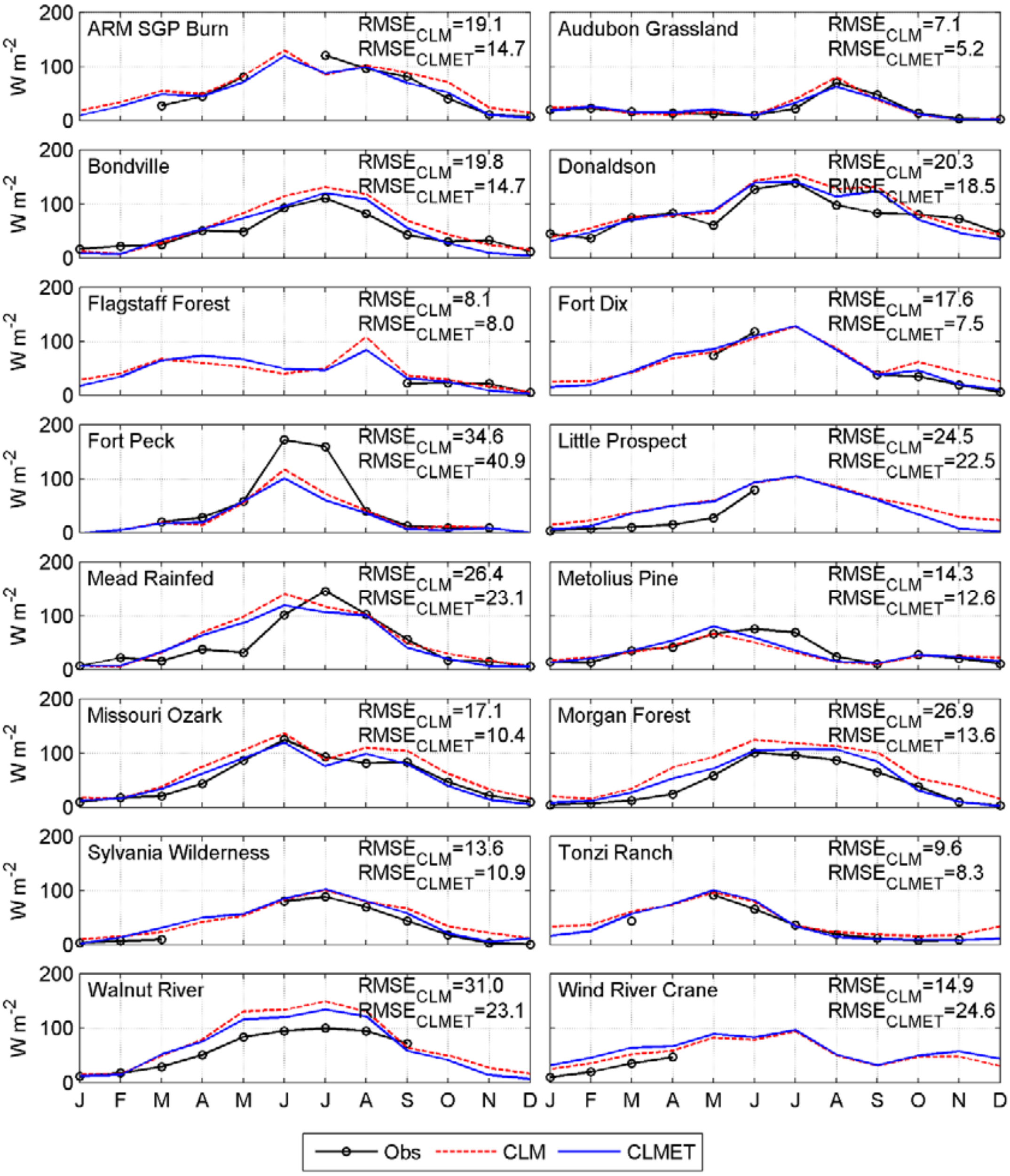


Figure 8 Monthly mean latent heat fluxes from CLM, CLMET and observations at 16 flux tower sites. RMSECLM and RMSECLMET represent the root mean square error against observations for CLM and CLMET, respectively. Note that the CLM and CLMET simulations are driven with meteorological forcing at the grid cell level (as opposed to site-specific forcing).

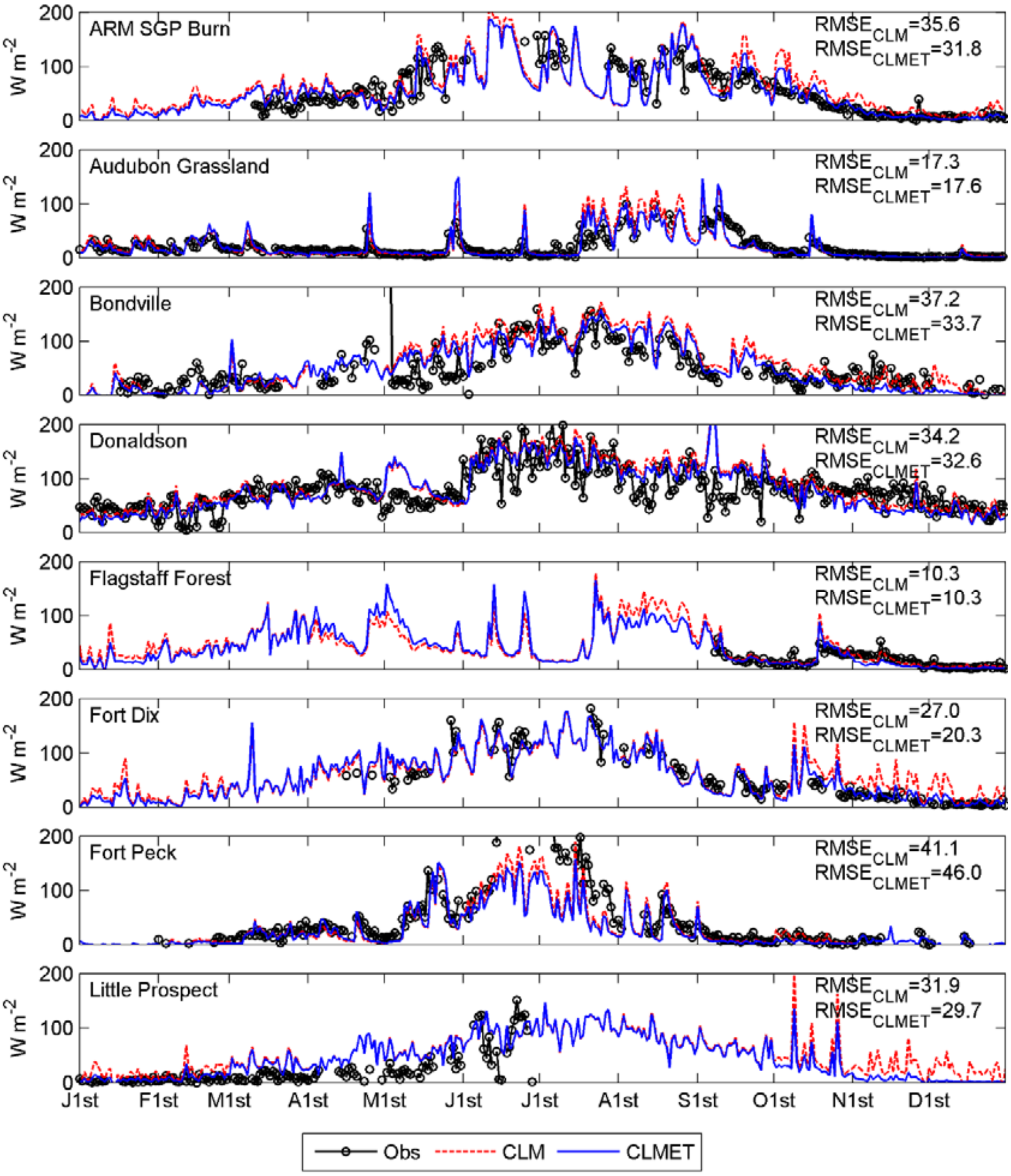


Figure 9 Daily mean latent heat fluxes from CLM and CLMET grids and station observations at ARM SGP Burn, Audubon Grassland, Bondville, Donaldson, Flagstaff Forest, Fort Dix, Fort Peck, and Little Prospect. RMSECLM and RMSECLMET represent the root mean square error against observations for CLM and CLMET, respectively.

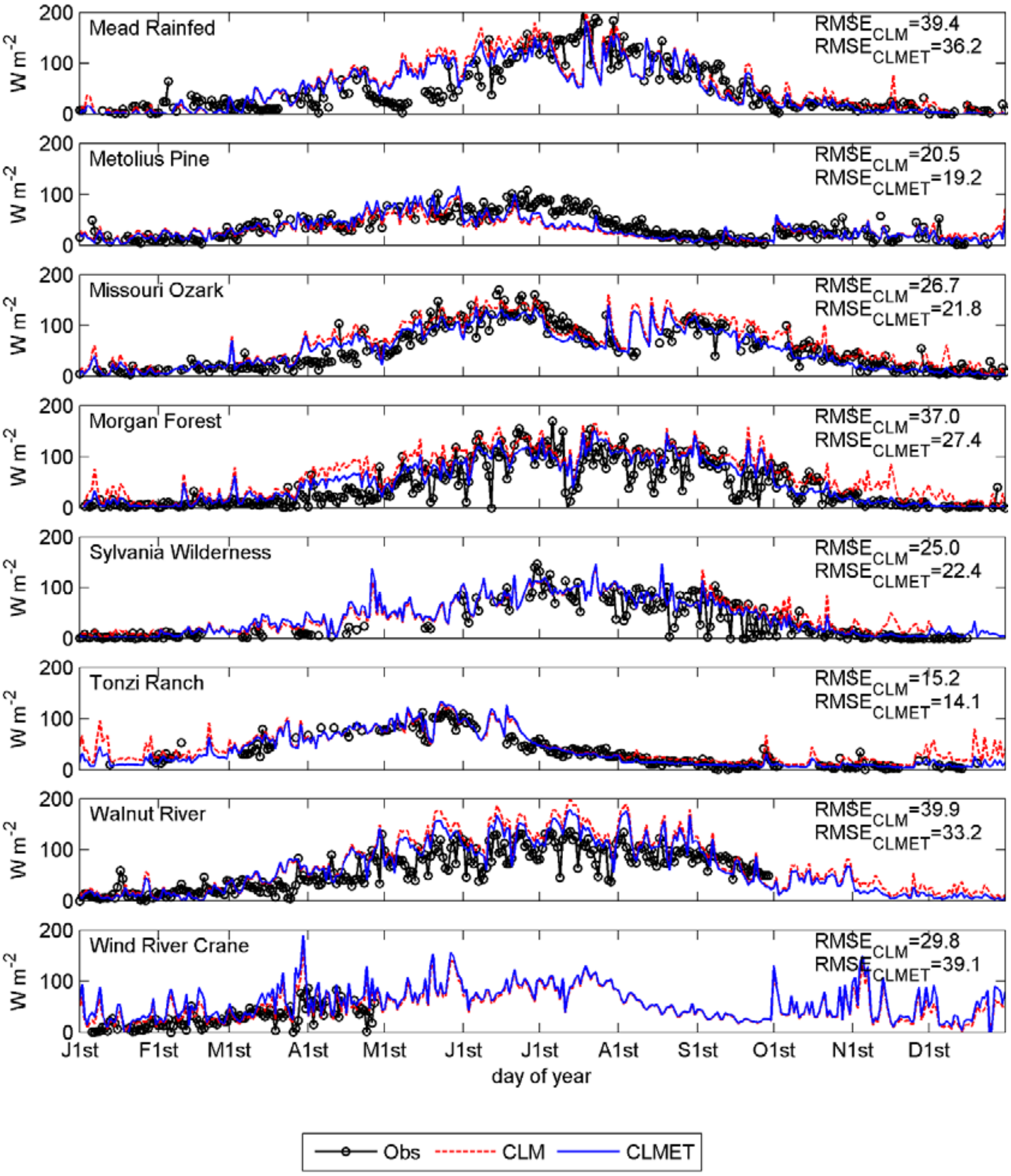


Figure 10 Daily mean latent heat fluxes from CLM and CLMET grids and station observations at Mead Rainfed, Metolius Pine, Missouri Ozark, Morgan Forest, Sylvania Wilderness, Tonzi Ranch, Walnut River, and Wind River Crane. RMSECLM and RMSECLMET represent the root mean square error against observations for CLM and CLMET, respectively.

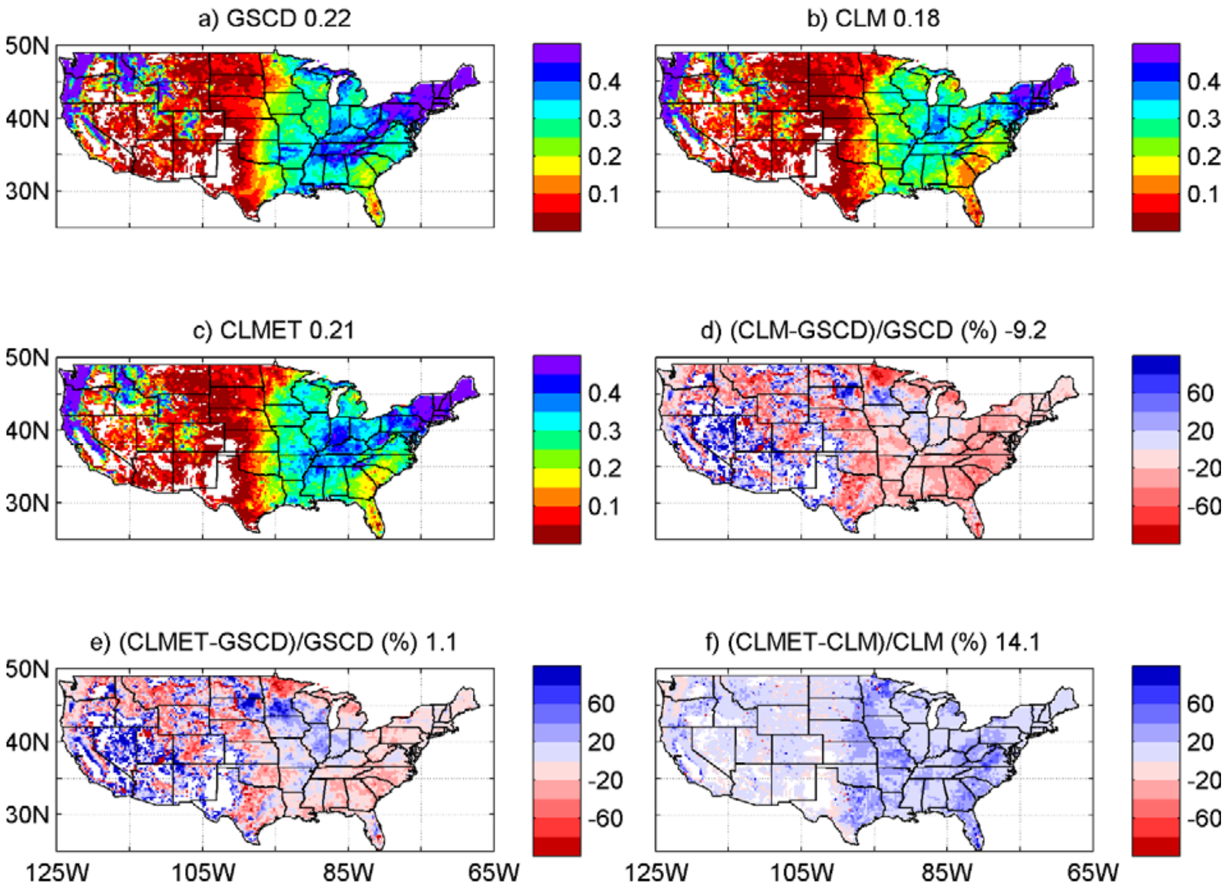


Figure 11 Mean annual runoff coefficient (the ratio runoff to total precipitation) from a) Global Streamflow Characteristics Dataset (GSCD), b) CLM, and c) CLMET, and the relative differences between d) CLM and GSCD, e) CLMET and GSCD, and f) CLMET and CLM during 2000-2014. Runoff coefficient less than 0.02 is blanked out. Numbers in titles are CONUS-averaged values.

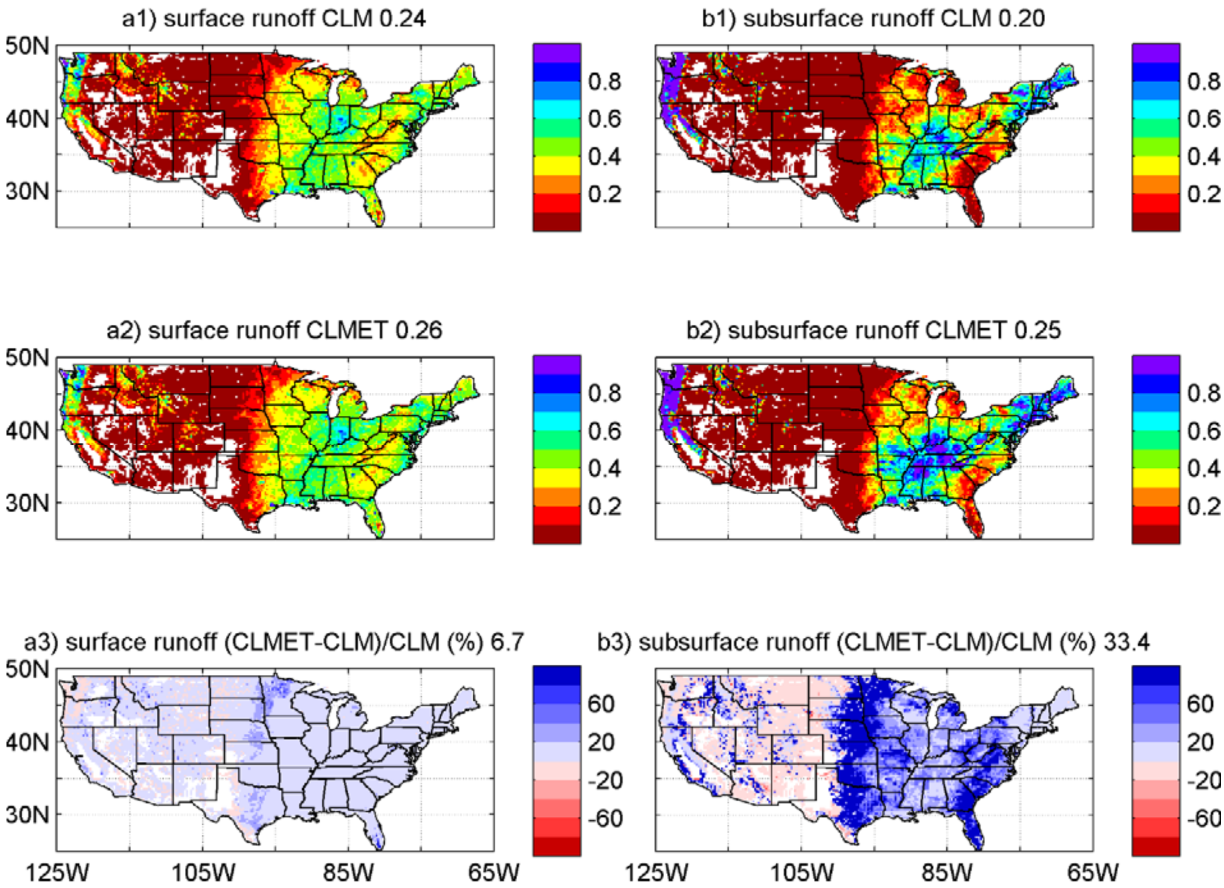


Figure 12 Surface runoff and subsurface runoff simulated in CLM and CLMET and their relative differences during 2000-2014. Numbers in titles are the CONUS-averaged values.

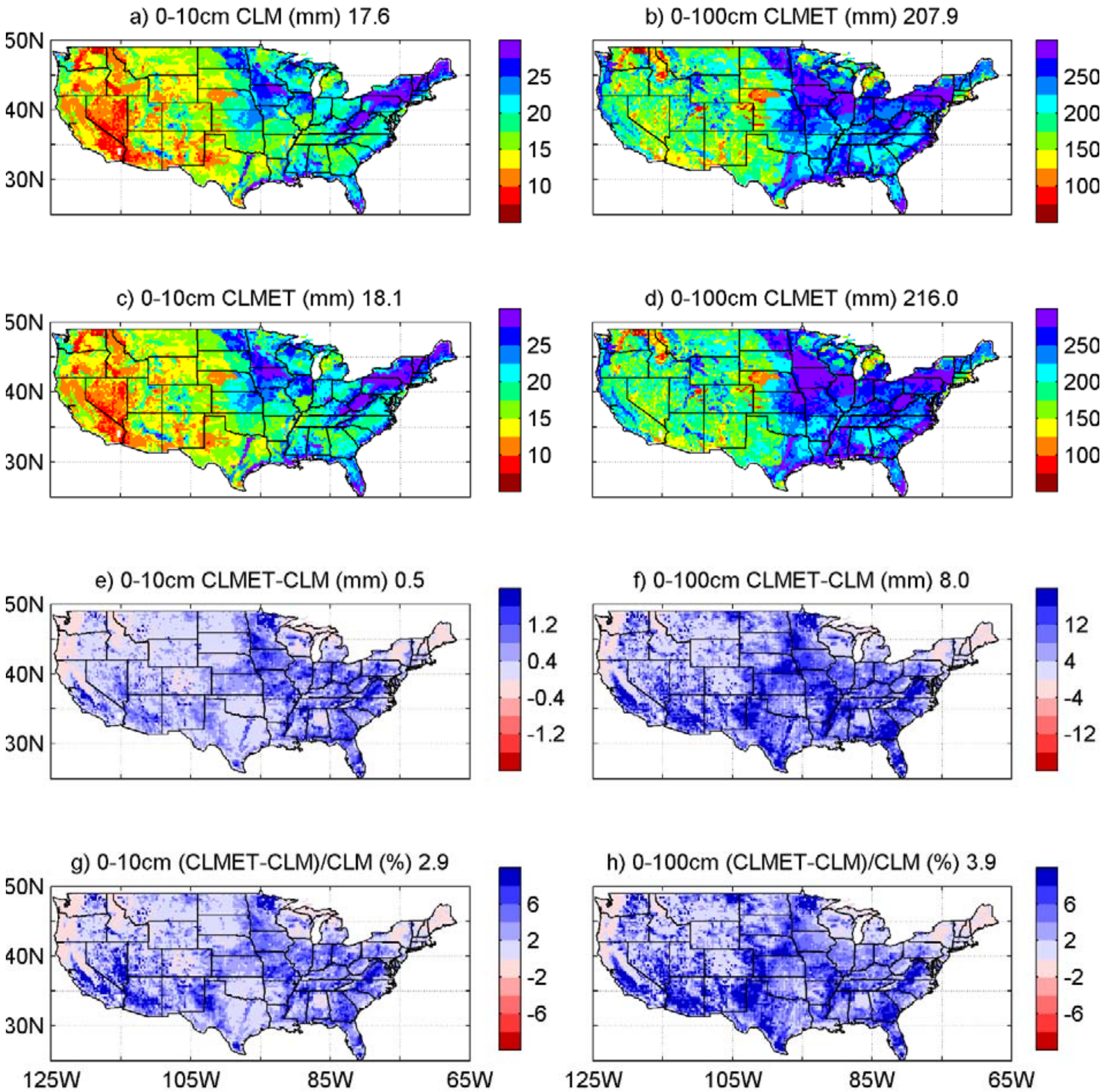


Figure 13 Simulated soil moisture (mm) in the top 0-10 cm and 0-100 layers in August from CLM and CLMET, their differences, and their relative differences during 2000-2014.

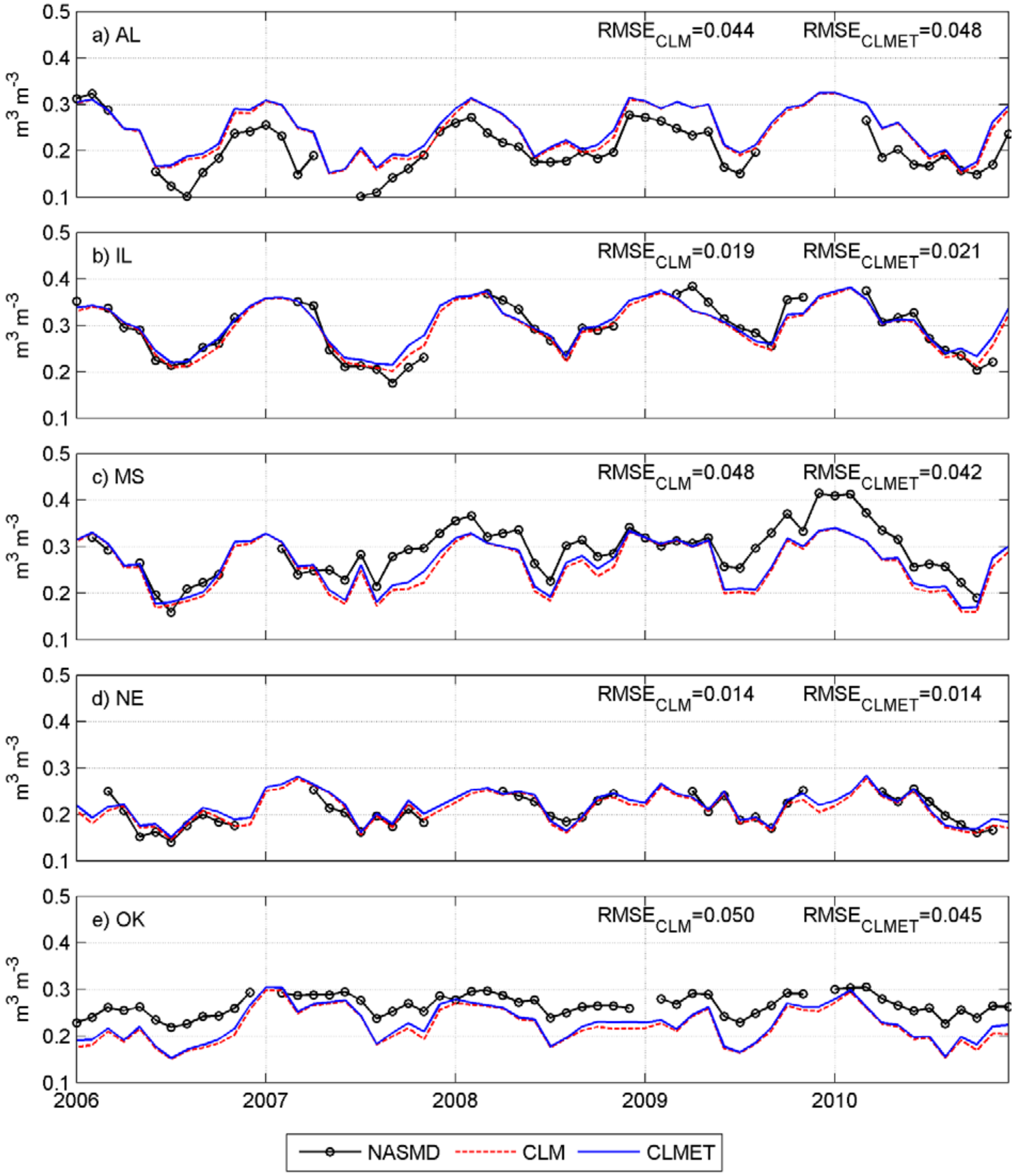


Figure 14 Monthly volumetric soil water content (m^{-3} m^{-3}) in the top 0-10cm soil layer from the quality-controlled NASMD, CLM, and CLMET over the state of Alabama (AL), Illinois (IL), Mississippi (MS), Nebraska (NE), and Oklahoma (OK) for the period of 2006-2010.

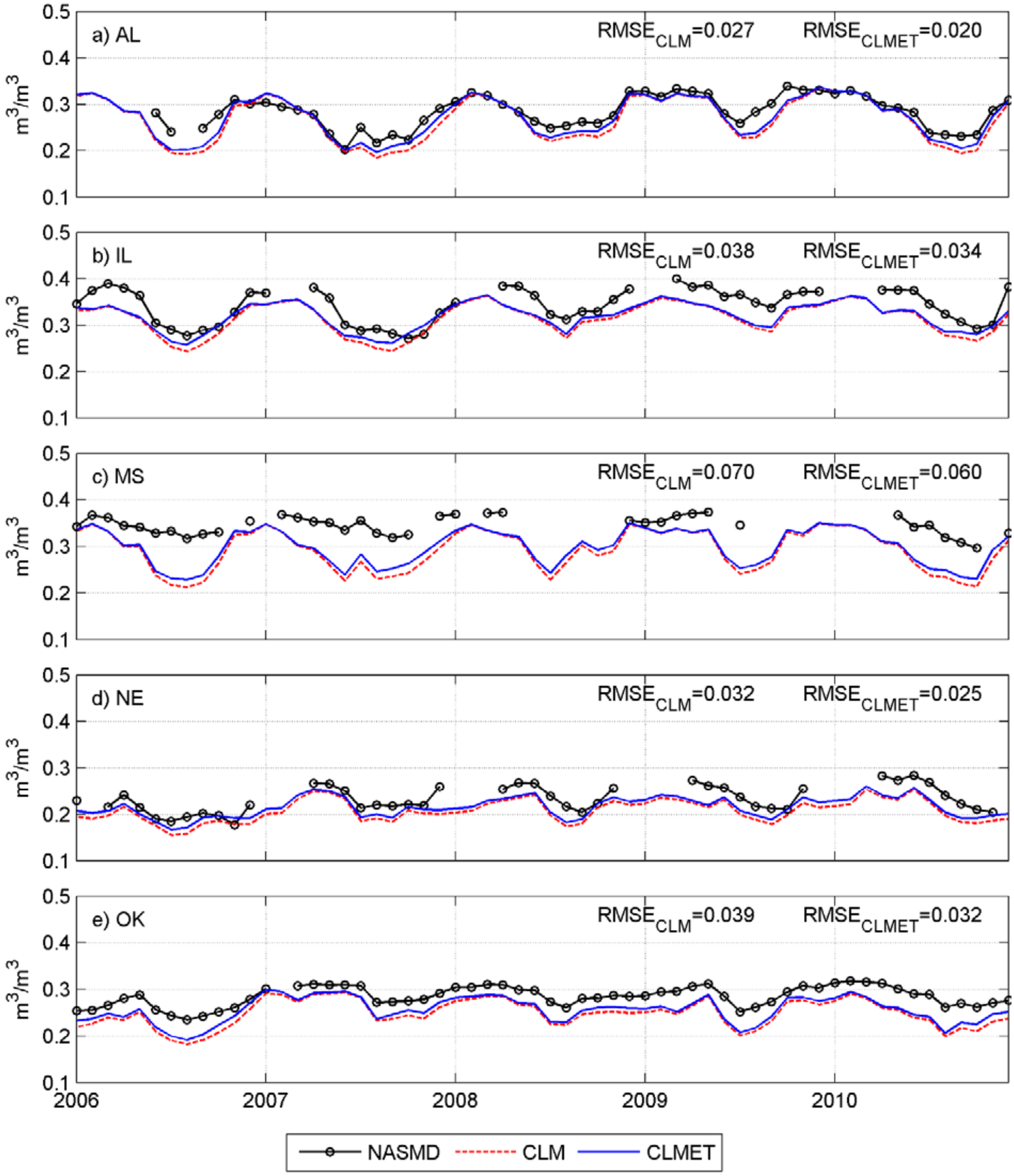


Figure 15 Same as Figure 14, but for the top 0-100cm soil layer.