Responses to the comments from Reviewer #1

HESS-2016-696
Wang et al. (2017): Incorporating remote sensing ET into Community Land Model version 4.5

GENERAL COMMENTS

In general, the authors implemented most of my initial suggestions and replied properly to the comments raised by the two reviewers. Below, I list some final comments.

SPECIFIC COMMENTS

1.At P12-Sect. 4.1, the authors now clearly mention that no scaling is applied during periods where the sign of evaporation does not agree between the estimate from GLEAM, and the original CLM run. As can be deduced from Figure 2, this mainly occurs during winter months. I think it should be clearly mentioned in the paper what the impact of this decision is/may be. My guess is that this will introduce a seasonal bias in the "corrected" evaporation, or will even result in discontinuities (especially during the transition of a month with scaling factor to a month without scaling factor).
Response: Point well taken. For example, in the eastern part of the North-West region, as pointed out in comment 2 by the reviewer, because no scaling factors are applied in November and December, the ET bias for CLMET are still substantial during the winter months (which partially explains the lower bias effect on the annual simulated ET, as shown in Figures 3e-3g). In contrast, the ET bias during the summer months is substantially corrected by applying the scaling factors. We have added a sentence to point out the potential effect of this treatment.

"*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.* ***This treatment (scaling in some months and no scaling in other months) may introduce a seasonal bias correction effect in these areas…***" (page 12, lines 250-251)

2. It is obvious from the results in Figures 3 and 5, that the impact of the bias correction algorithm is very limited in the western part of the CONUS. This is also acknowledged several times in the paper, but no clear reason is given why this is the case. As CLM tends to underestimate the GLEAM-derived evaporation in some parts of the West (mainly near the coast), and evaporation in these regions is mainly limited by water, I guess the low impact there could be related to the reasons mentioned at the end of P15. However, the reason explained along these lines is not valid in case of an initial overestimation of CLM, which is for instance happening in the eastern part of the

North-West region (see Figure 3d). Also in that region, the impact of the bias correction algorithm is very low and I think the reason for this should be figured out and explained in the text.

Response: Several factors influence effect of the bias correction algorithm: the difference in the scaling factor between the calibration period and the validation period (as shown in Figure 4 and discussed in page 14), whether ET is overestimated or underestimated (as shown in Figure 7 and discussed in page 15), and whether scaling is applied or not (as shown in Figure 2 and discussed in page 12). The latter two factors are the primary reasons for the smaller improvement in the eastern part of the North-West region. We have added a paragraph in Section 5, together with a previously existing paragraph in the revised format, to explain why the bias correction method performs differently for different areas.

"Qualitatively, whether the Parr et al. (2015) ET bias correction approach improves the quantification of the hydrological cycle depends on whether ET is limited by water or energy and whether ET is underestimated or overestimated. The approach works well when/where ET is not limited by water availability; in water-limited regimes, the approach is effective in correcting the positive ET biases but does not work well if ET is underestimated. Quantitatively, the degree of the model improvement derived from this bias correction algorithm 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.

Another factor affecting the degree of the model improvement is whether the ET scaling is applied at all. As shown in Figure 2, we do not scale ET in some areas of Northwest CONUS during the winter months due to the inconsistence in the ET sign (positive or negative) between GLEAM and CLM. In these areas and season(s), ET in CLMET is not corrected at all. All these three factors (i.e., whether the scaling factor differs significantly between calibration and validation periods, whether ET is underestimated in water-limited regimes, and whether ET scaling is applied at all) influence the effectiveness of the bias correction approach, but one or two of them may dominate for a given region/season. ***For example, regardless of which product is used as the reference for comparison (Figures 3g, 5a4, 5b4), the approach reduces ET biases over the eastern CONUS where the ET scaling is applied in most places/seasons and the scaling factor shows little difference between the calibration and validation periods. In contrast, in the north part of the Midwest, some positive biases still remain in CLMET, as the ET scaling is not applied in winter months and***

the scaling factor differs quite much between these two periods. Over a portion of western CONUS, the bias correction approach is less effective due to the underestimation of ET under a water-limited condition and large differences between calibration and validation periods in the scaling factor. *" (page 19, lines 413-443)*

3. Related to the previous comment, I do not agree with the statement at P14: "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, 2071) underestimate ET in the MODIS product (Michel et al. 2016, Miralles et al. 2016)." I fully agree with the fact that the MODIS algorithm generally produces lower estimates of evaporation, probably explaining the severe positive bias that can be seen in Figure 5a2. However, the reason why this bias in the West is not reduced in CLMET is because the bias correction algorithm has simply no impact in that region (see also the previous comment). As a result, the bias is still present in CLMET. In addition, the negative bias between CLM and the FLUXNET-MTE product in the West is also not alleviated, while this is not discussed in the manuscript.
Response: We have deleted the sentence "partially because the algorithm developed by Mu et al. (2007, 2071) underestimate ET in the MODIS product (Michel et al. 2016, Miralles et al. 2016).", and added a paragraph in Section 5 to explain why the bias correction method performs differently for different areas (including west CONUS). Please see the response to comment 2.

4. I think the fact that the method can hardly deal with underestimations of evaporation in water-limited regimes (see discussion at the end of P15) is an important drawback of the method and somehow summarizes the oddness of technique. This shows that this is a pure post-processing method and that it is not able to fix the real problem, which lies somewhere in the model physics. Therefore, I think that this should be highlighted in the conclusion as well.
Response: We have added a sentence on the drawback of the method in Section 5.

"*Qualitatively, whether the Parr et al. (2015) ET bias correction approach improves the quantification of the hydrological cycle depends on whether ET is limited by water or energy and whether ET is underestimated or overestimated.* ***The approach works well when/where ET is not limited by water availability; in water-limited regimes, the approach is effective in correcting the positive ET biases but does not work well if ET is underestimated*** ……"(page 19, lines 413-417)

TECHNICAL CORRECTIONS

1. Please add a reference to the numbers of global evaporation and runoff listed near the end of P3.
Response: We have added a reference for this in the revised manuscript.

"*Haddeland et al. (2011) compared eleven models in simulating evapotranspiration*

(ET), and found that 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." (page 3, lines 63-66)

2. The paper should be carefully checked for typos and the proper use of English grammar. I am only listing here a few examples, but the list is not limited to the ones below:

Response: We have thoroughly checked manuscript, and made corrections accordingly, including correcting the errors listed below.

a. P6: "We follows → follow"
b. P6: "Although land surface models are cable → capable"
c. P13: "Allthe statistics in CLMET is → are"
d. P15: "CNOUS → CONUS"
e. P15: "... in contrast, when an underestimate→ underestimated"
f. P17: "... scaling factor within each gird→ grid"
g. P20: "... in this study can effectively improves → improve"

3. The paper has currently 15 figures and 4 tables [which could be summarized in figures as well to strengthen the message, as the tables are not really giving a nice overview). I do not know the regulations of Copernicus Publications, but 15 figures seems a lot. Maybe, the authors should think about reducing the number of figures and only keeping the key figures.

Response: We have deleted previous Figures 12, 14, and 15, which results in 12 figures in total in the revised manuscript.

Responses to the comments from Reviewer #2

General comments

Although I was not part of the first review round, I generally agree with the comments from the previous reviewers. That is, the paper describes a simple approach to correct for biases in CLM simulations of evaporation over CONUS by rescaling the model ET to GLEAM. Although the approach is not really innovative and does not result in improved physical understanding, it is effective in reducing biases and improving state and flux estimates, and therefore, can be of interest to the community. I also believe the authors have substantially improved the manuscript after revision, particularly by including additional independent evaporation datasets (other than GLEAM) for validation, also at the daily time scale. Below, I still formulated some minor comments, which I believe could help further improving the manuscript. For future reference, please include line numbers in the manuscript to help tracking comments.

Response: Thank the reviewer for the positive evaluation. We have included line numbers in the revised manuscript.

Specific comments

1. Title: given that GLEAM evaporation is more of a model product (forced by remote sensing observations), than an actual remote sensing product, I would recommend to change the title accordingly.
Response: Following the reviewer's suggestion, we have changed the tile to "Incorporating remote sensing-based ET estimates into Community Land Model version 4.5".

2. Introduction/methods: I think it is important to add a caveat somewhere in the introduction or method section, in that the approach assumes full trust in the GLEAM evaporation, and no trust at all in the CLM evaporation. From a data merging/assimilation perspective, it could be more desirable if information from both CLM and GLEAM would be exploited, based on their relative uncertainties/performances. In its current form, the approach can degrade the performance of ET estimates in places/periods where GLEAM evaporation may perform less well. For instance, GLEAM has a very simple snow module, so it might have issues over snow-covered areas, like the Sierra Nevada (an area that actually stands out in Figure 2).
Response: We have added a sentence to show full trust in the GLEAM evaporation in Section 3 when the GLEAM ET is first described in the manuscript.

"*As such we assume full trust in the GLEAM evaporation data with the bias correction method.*" (Page 9, lines 179-180)

3. Page 4, 3rd paragraph: The authors mention data assimilation as another approach for reducing biases. I would rephrase this part, as the general purpose of data assimilation is not to remove biases, but to correct for random errors (e.g. in the forcings). Biases between model forecasts and observations should be corrected prior to assimilation.
Response: we have changed from "reducing model biases" to "improving model simulations or predictions"

"*Another approach to* ***improving model simulations or predictions*** *is through data assimilation, by merging observational data and land surface models to obtain optimal estimates for next time step*" (page 4, line 82)

4. Page 6, 1st paragraph: The authors state that CLM can realistically capture the spatial pattern of ET over CONUS. This seems to be contradictory to what is shown later in the results section, where large biases occur over space (e.g. Figure 2) and time. Maybe this statement needs to be nuanced, by mentioning the scale to which it applies, or mentioning the reference dataset on which it is based.
Response: we have changed to "overall spatial pattern of ET".

"*Parr et al. (2016) found that CLM4.5 can realistically capture the* ***overall*** *spatial* *pattern of ET in CONUS when the model is forced by the NLDAS-2 meteorological variables*" (page 5, lines 115-116)

5. Page 6, 3rd paragraph: I was wondering if, by applying constant monthly scaling factors, there may be jumps in simulated states/fluxes from one month to the other. Alternatively, the authors could consider interpolating scaling factors to get a smooth transition. Please comment on whether such jumps are observed in the simulations (e.g. of daily time scale runoff, soil moisture, etc.)
Response: The scaling factor characterizes 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 finer temporal scale (e.g., 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. Page 7, 2nd paragraph: Please describe which adjustments are made in CLM if soil moisture cannot support the evaporative demand. This is a quite important aspect in view of the approach.
Response: We have specified the adjustments as explained in the following paragraph:

"Since the overwriting process in CLMET may break the water balance, the model checks whether the amount of water stored in vegetation canopy is sufficient to sustain the interception loss and whether the surface soil water storage is sufficient to sustain soil evaporation through the model time step. If not, the interception loss (soil evaporation) rate is set to be equal to the water available in vegetation canopy (soil) divided by the model time step. This adjustment minimizes the imbalance caused by overwriting ET components in CLMET." (page 7, lines 153-159)

7. Equations 1-5: Are absolute values of S and R considered for calculating statistics like bias? If not, and when averaging over grid cells, simulations which have negative bias in one place and positive bias in another place would show up as bias-free, which is incorrect.
Response: It is true that the positive and negative biases are canceled out when they are averaged over grid cells. However, the region-averaged bias can provides us with the systematic error of land surface models, and this statistic has been widely used in model evaluation studies. In fact, we tested evaluating the model performance based on the absolute bias, and the conclusion is consistent with the bias-based conclusion, e.g., larger improvement from CLMET to CLM in eastern CONUS, larger improvement during autumn and winter.

That said, to address the reviewers concern, we have added plots for spatially distributed changes in the absolute value of relative biases (Figures 3g, 5a4, 5b4) to provide a more clear picture of the performance improvement from CLM to CLMET.

8. Results: I believe this section would improve if the authors would at least try to formulate some hypotheses on why biases in CLM simulations of ET occur over some areas and time periods. This could particularly help future studies to improve physical model aspects. Based on your extensive validation analysis of CLM ET, what can be learned? Are ET biases likely originating from biases in soil moisture, biases in radiation/temperature, model physics or parameters (and if so, which parameters)? Is this different for different regions, e.g. water- vs energy-limited regions, etc.?

Responses: We have added one paragraph to specifically talk about the physical conditions (water- or energy-limited ET regimes) that influence the effectiveness of the approach, which serves as the basis for another short paragraph at the end of paper on the implications of our results for physically-based land surface model development:

"*…… 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) in water-limited regimes. 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 underestimated 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 can be increased is limited by the availability of water stored in soil layer and vegetation canopy. Therefore, in water-limited ET regimes, if ET is underestimated in CLM, the actual ET in CLMET after the water availability check can be substantially lower than the corrected ET fed into the model, which diminishes the effect of the bias correction algorithm under such circumstance.*" (page 15, lines 327-340)

"*……However, results from this study can provide useful guidance for physically-based land surface model development. As can be seen from Figure 3g, the bias correction algorithm improves ET estimation over most of the U.S., indicating a strong potential for performance improvement that can be derived from improving the physical parameterization of ET processes in the model. Over regions where the bias correction approach does not improve the ET estimate (which are mostly places where ET is water-limited while the model underestimates ET), parameterizations for other processes that influence soil moisture (e.g., runoff generation, groundwater interactions) are the most likely cause for model biases and should be the focus of physically-based model development effort.*" (page 22, lines 472-480)

9. Page 13, 2nd paragraph, line 12: I believe it is important to stress that improvements in model performance are relative to GLEAM. GLEAM has its own shortcomings, so improving towards GLEAM not necessarily means improving estimates overall.

Response: We have pointed out that the evaluation in the model performance are relative to the GLEAM ET at the beginning of this paragraph in the revised manuscript.

"*Over most of CONUS, CLM overestimates ET and CLMET reduces ET as well as ET biases relative to GLEAM data.*" (page 13, lines 274-276)

10. Page 20, 2nd paragraph: I appreciated the statement that the approach does not replace model improvements through better parameterization of physical processes. This frames the value of your approach, and helps readers to identify shortcomings and advantages of the approach.
Response: We thank the reviewer for the positive evaluation.

Technical corrections

I would strongly encourage the authors to perform an in-depth check of the grammar throughout the manuscript. There are still many mistakes, e.g., on using plural nouns and tenses. The following list is not comprehensive.
Response: We have thoroughly checked manuscript, and made corrections accordingly, including (but not limited to) a long list of corrections suggested by this reviewer.

Figures: Please make sure you refer to panels (a, b, c, …) in all figure captions (for instance figure 4)
Response: Following reviewer's suggestion, we have changed the presentation style in all figure captions.

Incorporating remote sensing-based ET estimates 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 June, 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 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 bias correction 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 with the corrected ET. Using the observations of ET, runoff, and soil moisture content as benchmarks, we demonstrate that CLMET greatly improves the hydrological simulations over most of CONUS, and the improvement is stronger in the eastern CONUS than the west and is the strongest over the southeast CONUS. For any specific region, the degree of the improvement depends on whether the relationship between observational and model ET remains time-invariant (a fundamental hypothesis of the Parr et al. method) and whether water is the limiting factor in places where ET is underestimated. While the bias correction method improves hydrological estimates without improving the physical parameterization of land surface models, results from this study does provide guidance for physically based model development effort.

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 lack spatial and temporal continuity 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 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 tend 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, 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 improving model simulations or predictions 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 models is a typical practice in land data assimilation, and it has been reported that data assimilation of soil moisture helped in reducing model biases (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 ET biases and their effects on other variables 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 historical ET and historical 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 datasets of ET, runoff, soil moisture, are 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 datasets used in CLM4.5 were derived from different sources. The soil texture data were 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 were derived from Moderate Resolution Imaging Spectroradiometer (MODIS) satellite data (Lawrence et al. 2011). Slope and elevation were 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 overall spatial pattern of ET in 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 the large-scale patterns of runoff and ET. In this study CLM4.5 is forced by the NLDAS-2 meteorological forcing (Xia et al., 2012a). The 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, was 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 follow this division in this study, as shown in Figure 1a.

Although land surface models are capable of capturing the large-scale pattern of ET, significant biases were found at finer spatiotemporal scales (Parr et al. 2015, 2016, and Wang et al. 2016), which propagate 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 components 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 the 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 whether the amount of water stored in vegetation canopy is sufficient to sustain the interception loss and whether the surface soil water storage is sufficient to sustain soil evaporation through the model time step. If not, the interception loss (soil evaporation) rate is set to be equal to the water available in vegetation canopy (soil) divided by the model time step. This adjustment 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:

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

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

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

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

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. As such we assume full trust in the GLEAM evaporation data with the bias correction method. GLEAM 3.0a was 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 3.0a was calculated using a Priestley and Taylor equation based on surface net radiation and near-surface air temperature, and was converted to actual evaporation using 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 flux tower site and global scales show 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 of GLEAM dataset is 0.25°, which is consistent with the resolution of CLM4.5 used in this study. The temporal resolution of GLEAM dataset is daily, and the monthly aggregated ET is used to derive the scaling factors.

3.1.2 MODIS and FLUXNET-MTE ET

Two other 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 using a revised 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 was then used to train MTE to produce monthly global ET dataset at the 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 fairly dense, the quality of the FLUXNET-MTE dataset in our study domain is expected to be good. The MODIS dataset is available for 2000-2014, and the FLUXNET-MTE dataset is available for 1982-2011. We chose the overlap period of these two products, 2000-2011, for model validations using MODIS and FLUXNET-MTE dataset.

3.1.3 Flux Tower ET

ET observations (in energy unit) at 16 sites from the AmeriFlux network are used to validate the model on the grid cell scale (Figure 1b). Those sites span 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 these flux tower sites can be found in Xia et al. (2015b). For most sites, the year of 2005 is selected for validation because data for this year has the least amount of missing records; three sites are 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 (Xia et al., 2015; Liao et al., submitted to Journal of Hydrometeorology, 2017). The in-situ observations in Alabama (AL), Illinois (IL), Mississippi (MS), Nebraska (NE), and Oklahoma (OK) from 2006-2010 are selected from NASMD (Figure 1a). A large number of 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. Since the soil layer where measurement was taken varies with stations, we linearly interpolate the volumetric soil water content to the 5 cm and 50 cm depth for all stations to compare with the modeled soil moisture for the 0-10 cm and 0-100 cm layers.

4 Results

4.1 Calibration of ET Scaling Factor

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. This treatment (scaling in some months and no scaling in other months) may introduce a seasonal bias correction effect in these areas. The model simulations generally agree better with GLEAM estimations during the warm seasons, whereas the difference between simulations and GLEAM estimations remains large during the cold seasons. The scaling factors greatly vary with region. For instance, the area-averaged scaling factors for November are 0.34, 0.58, 0.28, and 0.52 for Northwest, Southwest, Northeast, and Southeast, 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 especially severe over the Northeast CONUS where simulated ET is almost five times of GLEAM estimate 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 average ET from the four 0.25° model grid cell within each 0.5° observational 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 technically at the point scale, we spatially average observed soil moisture in each state and compare the averaged observations with the model simulations averaged across 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 and CLMET reduces ET as well as ET biases relative to GLEAM data. 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 are 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 average, 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 representing each of the four 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 appears 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. Not surprisingly, 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 datasets 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. 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 are greater than in CLM for the whole CONUS, Northwest, Southwest, and Northeast in MAM. The most considerable improvement occurs in SON compared with the other three seasons. CLMET deteriorates the ET estimate for MAM by enhancing the 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 the two types of simulations also demonstrates improvement from CLM to CLMET. Climatological seasonal cycles of ET over CONUS and four sub regions for the period 2000-2011 are shown in Figure 6. CLMET outperforms CLM over CONUS with a smaller RMSE (0.31 versus 0.40 against MODIS, 0.19 versus 0.25 against FLUXNET-MTE). The improvement mainly results from reduction of the 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) in water-limited regimes. 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 underestimated 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 can be increased is limited by the availability of water stored in soil layer and vegetation canopy. Therefore, in water-limited ET regimes, if ET is underestimated in CLM, the actual ET in CLMET after the water availability check can be substantially lower than the corrected ET fed into the model, which diminishes the effect of the bias correction algorithm under such circumstance.

In addition, the ET validation is also conducted at 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 the 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 coefficients in CLM and CLMET are 0.18 and 0.21, which are comparable to the GSCD-based runoff coefficient (0.22). However, CLM underestimates runoff in most areas of CONUS due to an overestimation of ET. CLMET alleviates the underestimation by reducing 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) reduced 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 (not shown). The same value of the ET scaling factor within each grid 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 are 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, reductions in all three components of ET interception loss, plant transpiration, and soil evaporation from CLM to CLMET slow down moisture depletion from the soil. As a result, the water content in different soil layers increases with reduced ET. Figure 12 shows soil water at the surface and root-zone layers simulated by CLM and CLMET, and their differences in August. From CLM to CLMET, the changes over CONUS show an overwhelmingly increasing 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 estimates of ET and soil moisture resulting from the bias correction method in this study may prove useful for improving drought monitoring and assessment.

Due to the strong spatial heterogeneity of soil moisture and the lack of large-scale distributed data, the comparisons between observed soil moisture and modeled soil moisture from CLM and CLMET are done based on the spatial averages across stations within each state and at 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. The soil in CLM shows dry biases over most of the examined states with the exception of soil moisture at the top 10 cm layer in Alabama and Illinois, and CLMET generally alleviate these dry biases. The RMSE values against the NASMD observations in CLMET is smaller than or at least the same as the RMSE values in CLM. An exception exists for the top 0-10 cm layer in Alabama and Illinois 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. The improvements in Alabama, Mississippi, Nebraska, and Oklahoma are summarized in Table 5.

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 improved estimation of the terrestrial hydrological cycle across most of the model domain, with different degrees of improvement among the CONUS sub-regions.

Qualitatively, whether the Parr et al. (2015) ET bias correction approach improves the quantification of the hydrological cycle depends on whether ET is limited by water or energy *and* whether ET is underestimated or overestimated. The approach works well when/where ET is not limited by water availability; in water-limited regimes, the approach is effective in correcting the positive ET biases but does not work well if ET is underestimated. Quantitatively, the degree of the model improvement derived from this bias correction algorithm 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.

Another factor affecting the degree of the model improvement is whether the ET scaling is applied at all. As shown in Figure 2, we do not scale ET in some areas of Northwest CONUS during the winter months due to the inconsistence in the ET sign (positive or negative) between GLEAM and CLM. In these areas and season(s), ET in CLMET is not corrected at all. All these three factors (i.e., whether the scaling factor differs significantly between calibration and validation periods, whether ET is underestimated in water-limited regimes, and whether ET scaling is applied at all) influence the effectiveness of the bias correction approach, but one or two of them may dominate for a given region/season. For example, regardless of which product is used as the reference for comparison (Figures 3g, 5a4, 5b4), the approach reduces ET biases over the eastern CONUS where the ET scaling is applied in most places/seasons and the scaling factor shows little difference between the calibration and validation periods. In contrast, in the north part of the Midwest, some positive biases still remain in CLMET, as the ET scaling is not applied in winter months and the scaling factor differs quite much between these two periods. Over a portion of western CONUS, the bias correction approach is less effective due to the underestimation of ET under a water-limited condition and large differences between calibration and validation periods in the scaling factor.

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 the total ET, we conducted additional experiments in which the scaling factor 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 do not show further improvement, which is likely due to 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, they found 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 estimates 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 improve 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 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. However, results from this study can provide useful guidance for physically-based land surface model development. As can be seen from Figure 3g, the bias correction algorithm improves ET estimation over most of the U.S., indicating a strong potential for performance improvement that can be derived from improving the physical parameterization of ET processes in the model. Over regions where the bias correction approach does not improve the ET estimate (which are mostly places where ET is water-limited while the model underestimates ET), parameterizations for other processes that influence soil moisture (e.g., runoff generation, groundwater interactions) are the most likely cause for model biases and should be the focus of physically-based model development effort.

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. The authors thank Mr. Brecht Martens and two anonymous reviewers for their constructive comments.

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, 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, 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, 2015.

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 1, 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

Table 5 Root mean square error (RMSE) values of monthly volumetric soil moisture ($m^{-3}m^{-3}$) simulated by CLM and CLMET relative to the quality-controlled NASMD for the top 0-10 cm soil layer and for the top 0-100 cm soil layer over Alabama, Illinois, Mississippi, Nebraska, and Oklahoma.

	top 0-10 cm soil water content		top 0-10 cm soil water content	
	CLM	CLMET	CLM	CLMET
Alabama	0.044	0.048	0.027	0.020
Illinois	0.019	0.021	0.038	0.034
Mississippi	0.048	0.042	0.070	0.060
Nebraska	0.014	0.014	0.032	0.025
Oklahoma	0.050	0.045	0.039	0.032

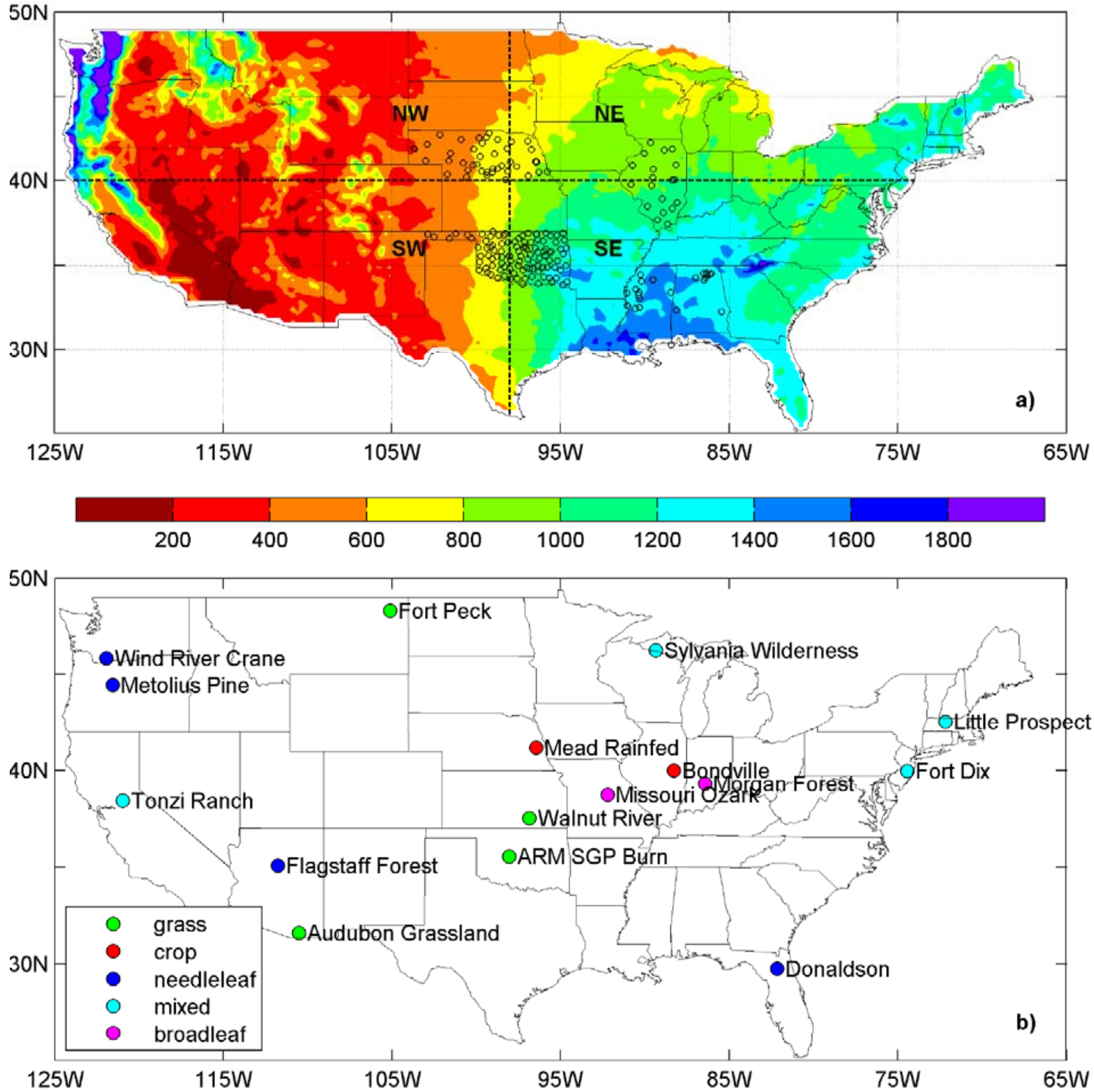

Figure 1 a) Mean annual (1980-2015) precipitation in mm over conterminous USA (CONUS). NW, SW, NE, and SE represent Northwest, Southwest, Northeast, and Southeast CONUS, respectively. The black circles represent sites of in-situ soil moisture observations in Alabama, Illinois, Mississippi, Nebraska,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

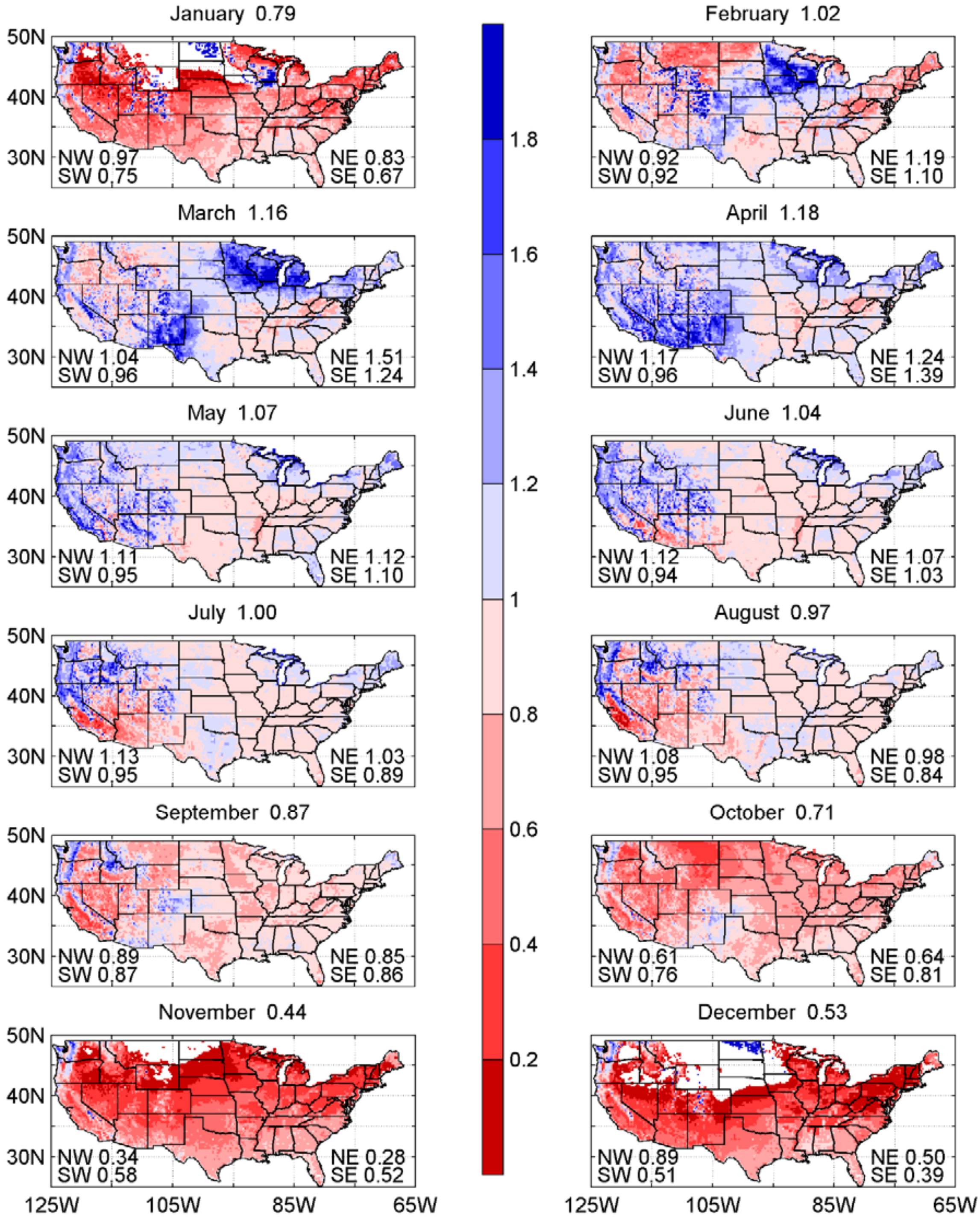

Figure 2 Scaling factor as the ratio of the CLM simulated ET to the GLEAM ET for each month during 1986-1995. The numbers in titles are CONUS-averaged values, and the numbers of within figures are area-averaged values for each of four sub regions (NW, SW, NE, and SE). The areas with negative scaling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

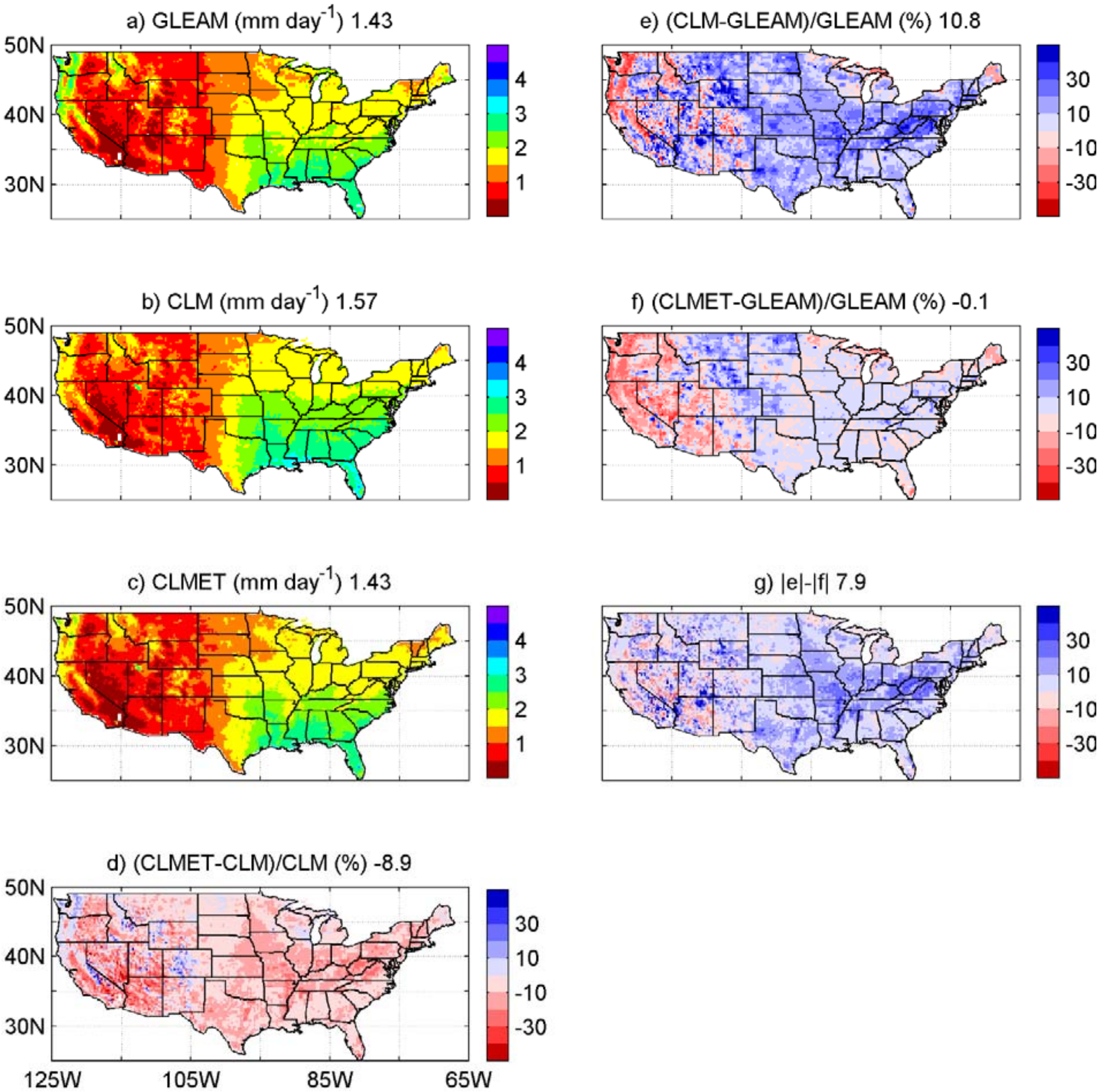

Figure 3 Mean annual ET from a) GLEAM, b) CLM, and c) CLMET, the relative difference between d) CLMET and CLM, e) CLM and GLEAM, f) CLMET and GLEAM, and g) the difference between absolute value of e) and absolute value of f) during the period 2000-2014. Numbers in titles are CONUS-averaged values.

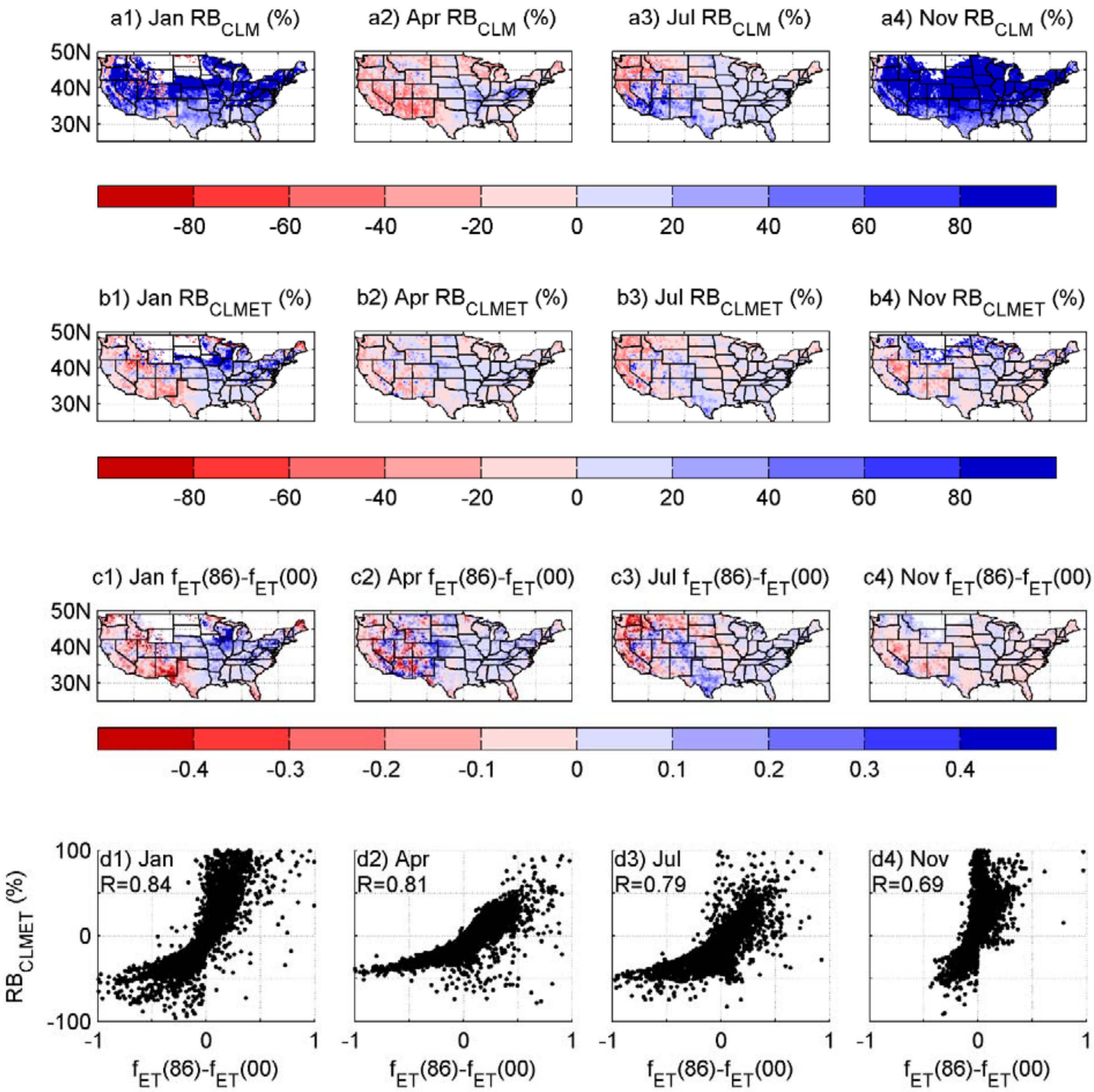

Figure 4 a) Relative bias (RB) for CLM (RB_{CLM}), b) RB for CLMET (RB_{CLMET}) during the period 2000-2014, c) difference in scaling factor f_{ET} between the period 1986-1995 and the period 2000-2014 ($f_{ET}(86)$- $f_{ET}(00)$), and d) scatter plots of $f_{ET}(86)$- $f_{ET}(00)$ versus RB_{CLMET} in 1) January (Jan), 2) April (Apr), 3) July (Jul), and 4) November (Nov).

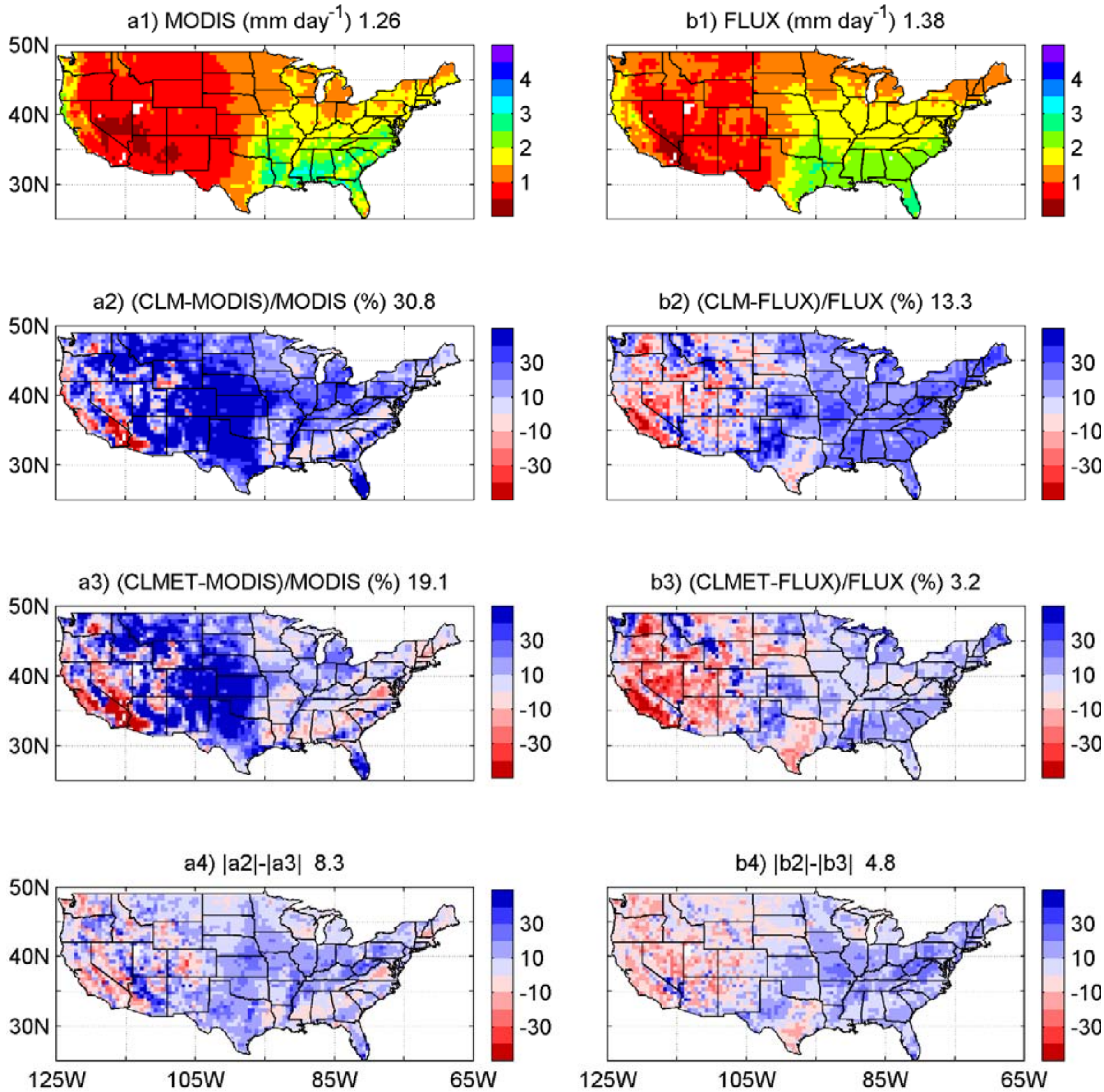

Figure 5 Mean annual ET from a1) MODIS, b1) FLUXNET-MTE, the relative differences between a2) CLM and MODIS, b2) CLM and FLUXNET-MTE, a3) CLMET and MODIS, and b3) CLMET and FLUXNET-MTE, and the differences between a4) absolute value of a2 and absolute value of a3, and b4) absolute value of b2 and absolute value of b3 during the period 2000-2011. Numbers in titles are CONUS-averaged values.

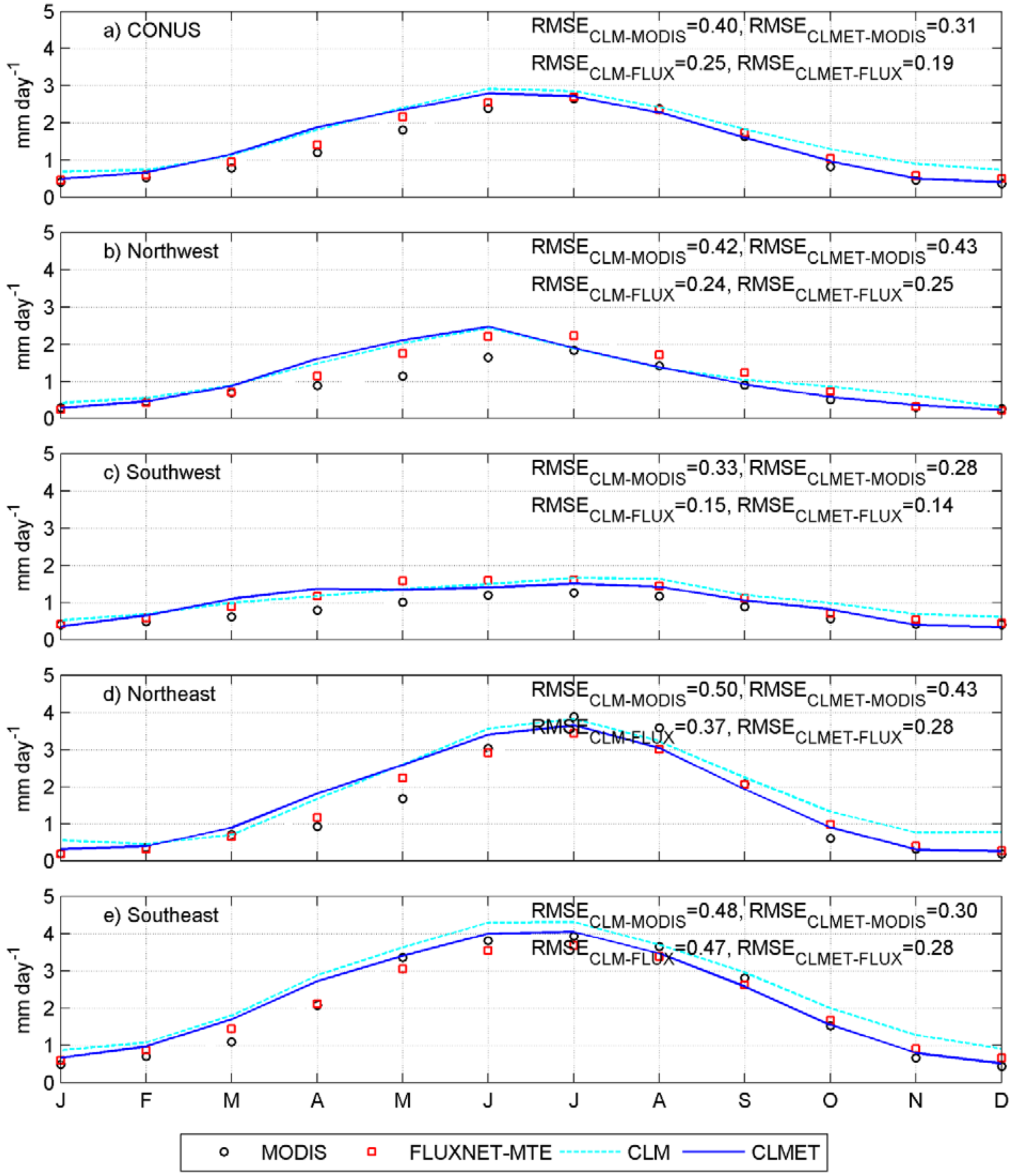

Figure 6 Seasonal cycles of ET from MODIS, FLUXNET-MTE, CLM, and CLMET over a) CONUS, b) Northwest, c) Southwest, d) Northeast, and e) Southeast during the period 2000-2011.

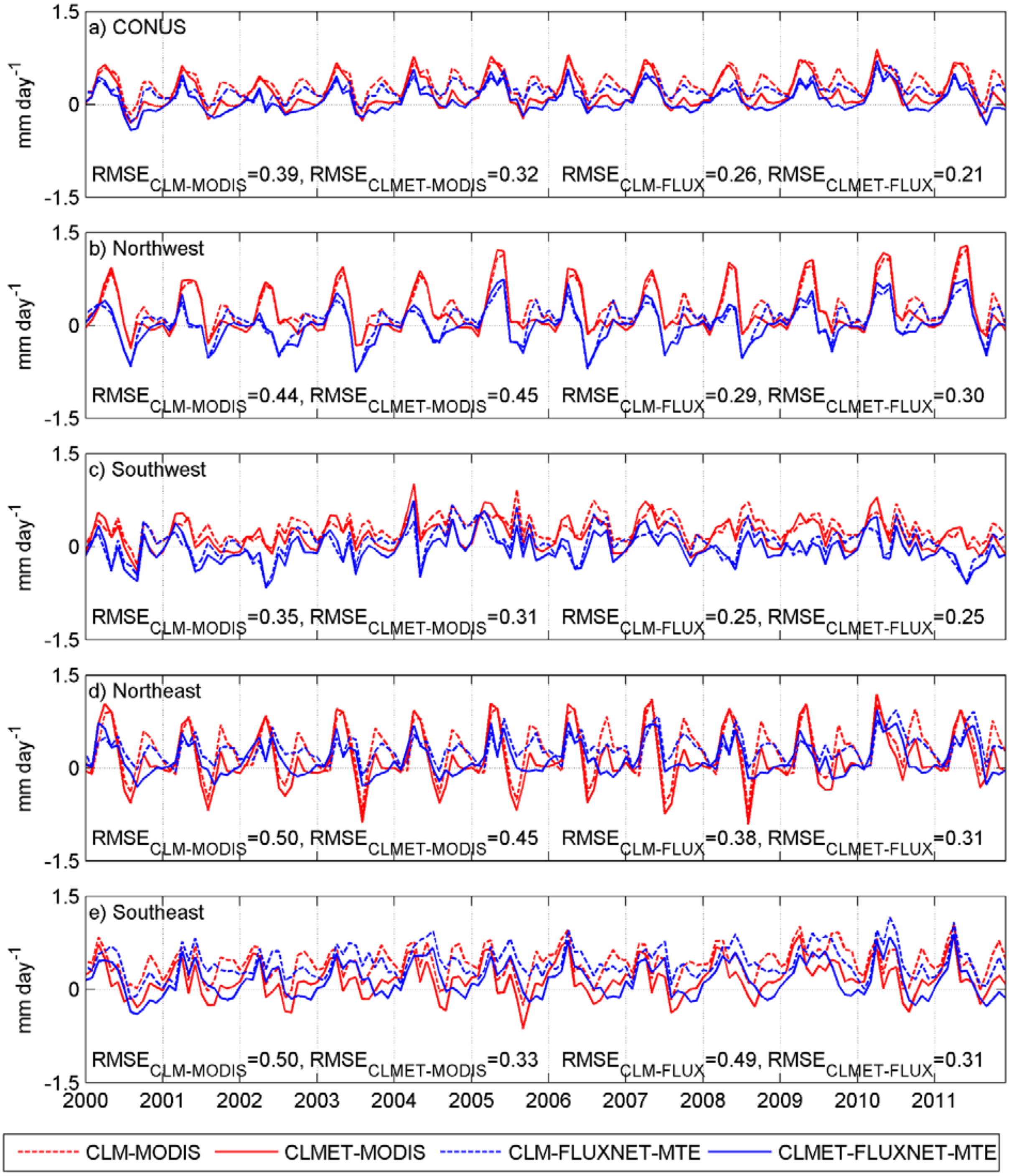

Figure 7 Time series of ET difference between model (CLM or CLMET) and reference data (MODIS or FLUXNET-MTE) over a) CONUS, b) Northwest, c) Southwest, d) Northeast, and e) Southeast during the period 2000-2011.

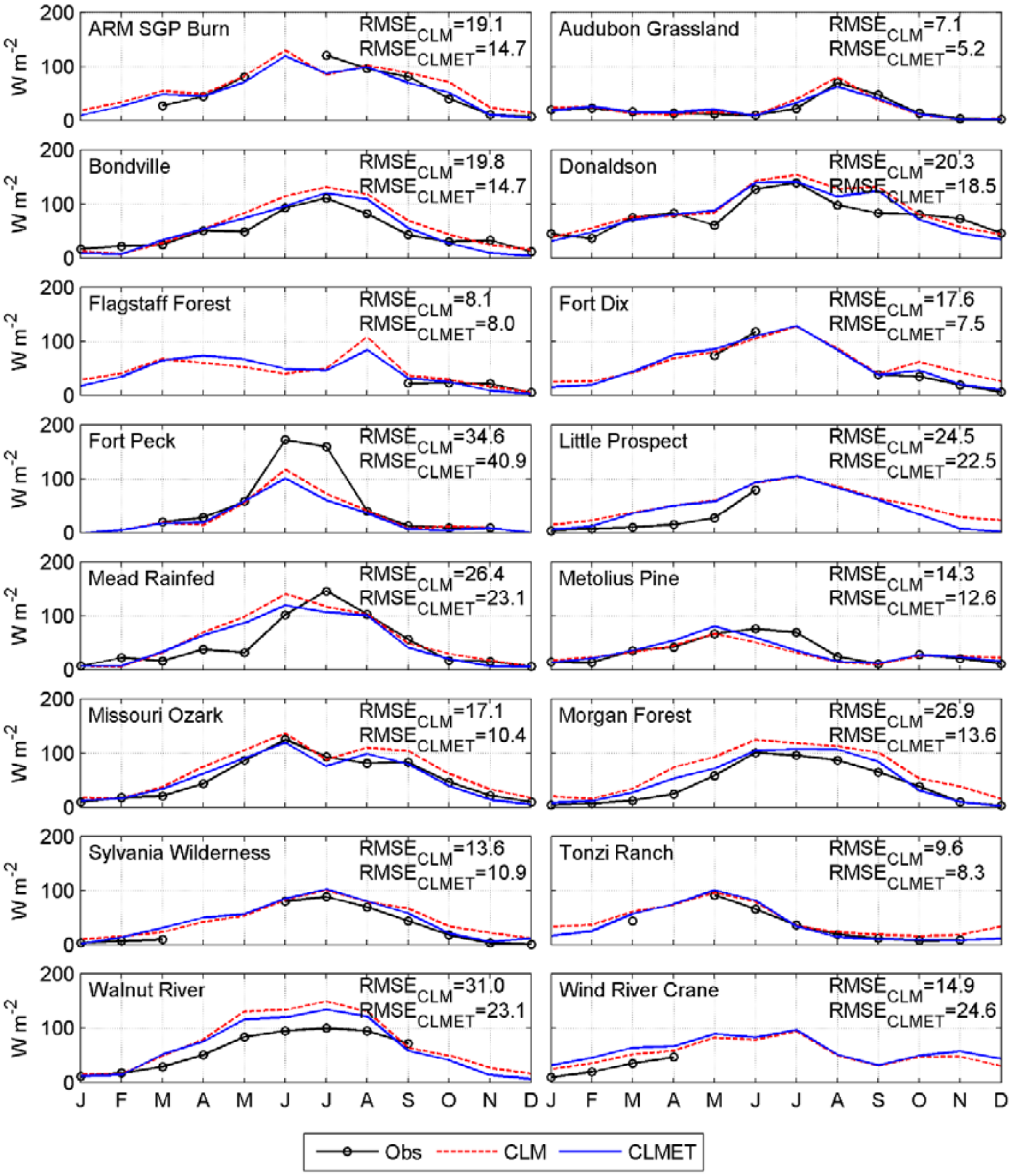

Figure 8 Monthly mean latent heat fluxes from CLM, CLMET and observations at 16 flux tower sites. $RMSE_{CLM}$ and $RMSE_{CLMET}$ represent the root mean square error against observations for CLM and CLMET, respectively. Note that the CLM and CLMET simulations are driven with meteorological forcings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

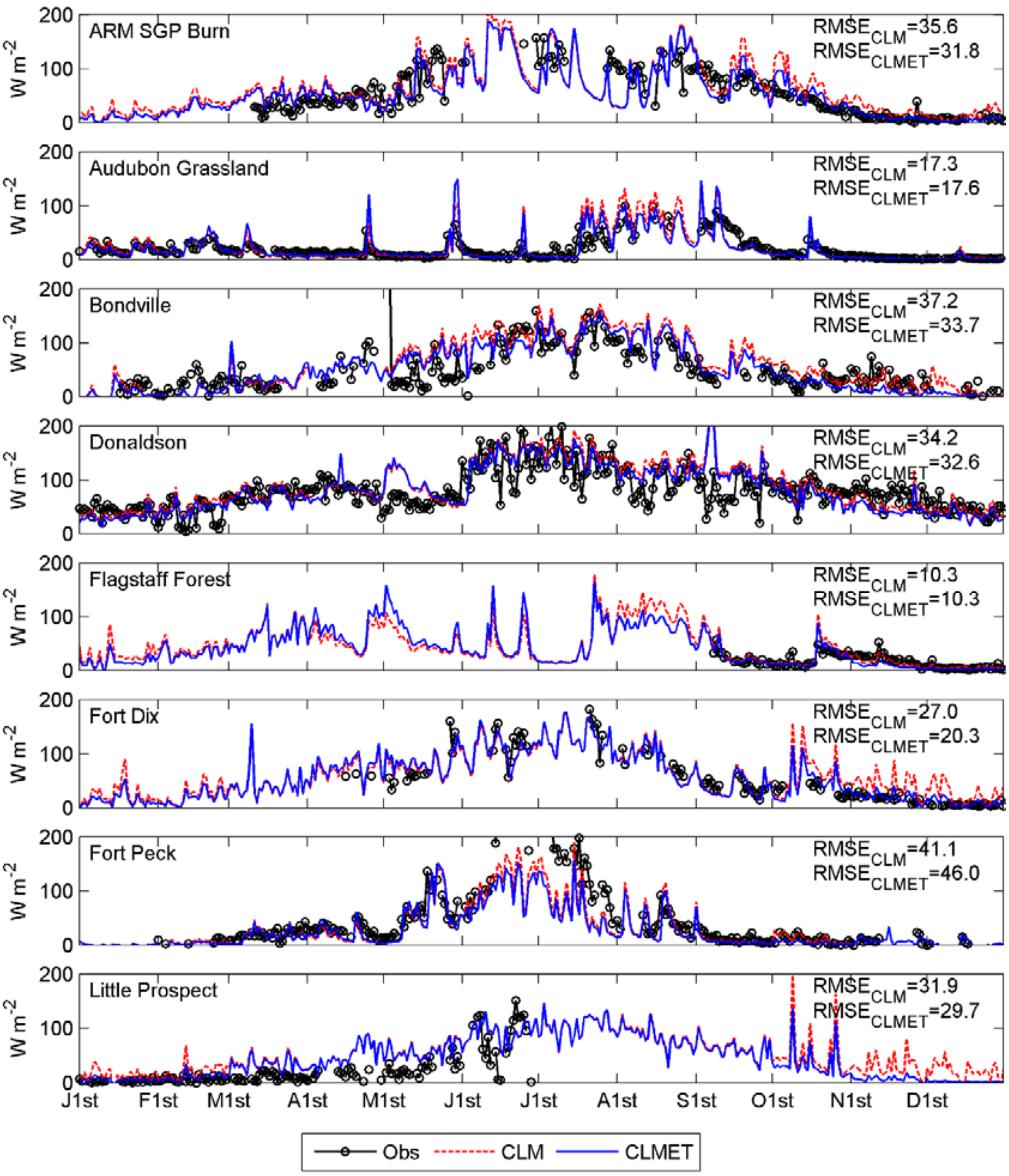

Figure 9 Daily mean latent heat fluxes from CLM and CLMET grids and station observations at ARM SGP Burn, Audubon Grassland, Bondville, Donaldson, Flagstaff Forest, Fort Dix, Fort Peck, and Little Prospect. RMSE$_{CLM}$ and RMSE$_{CLMET}$ represent the root mean square error against observations for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

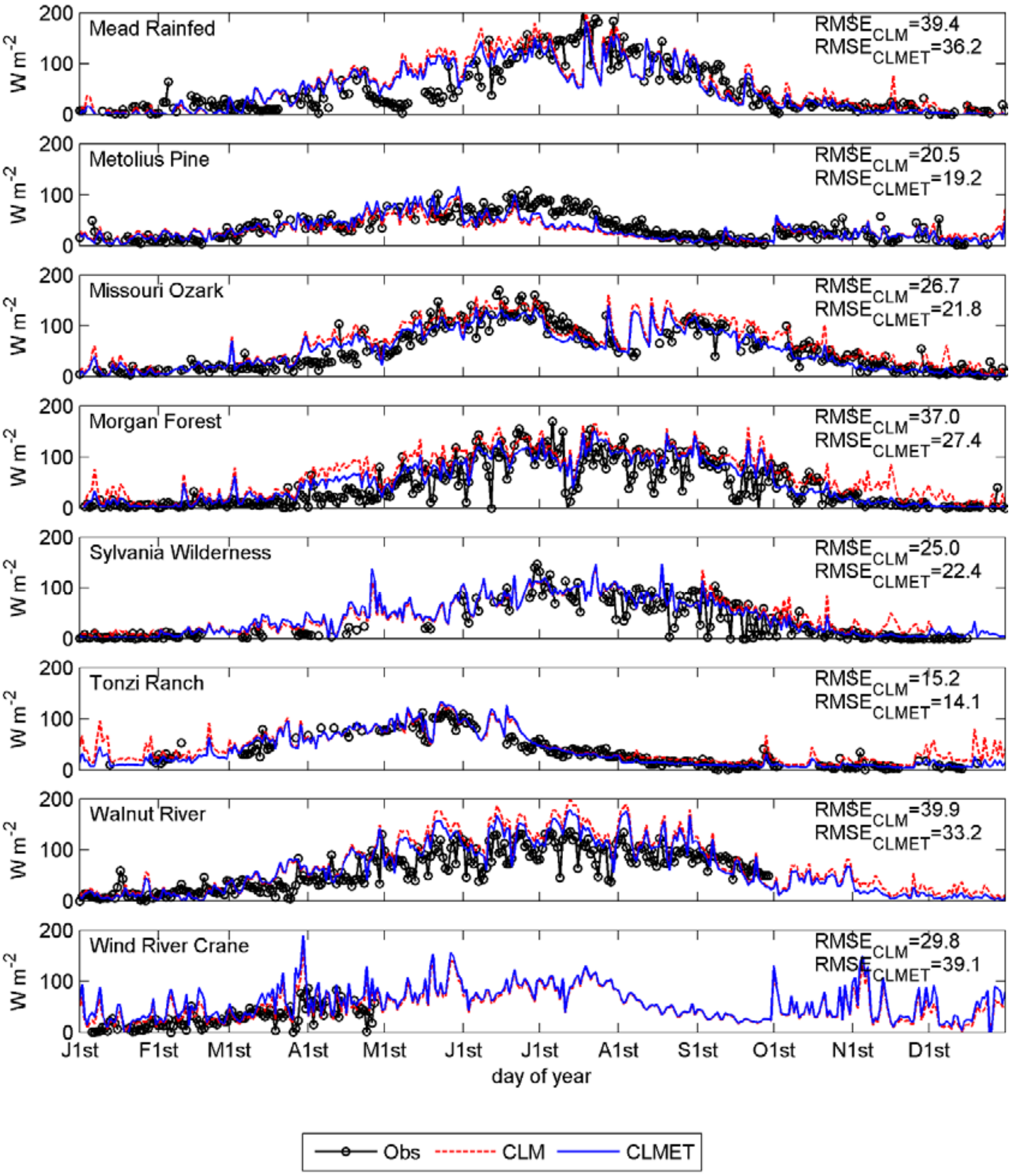

Figure 10 Daily mean latent heat fluxes from CLM and CLMET grids and station observations at Mead Rainfed, Metolius Pine, Missouri Ozark, Morgan Forest, Sylvania Wilderness, Tonzi Ranch, Walnut River, and Wind River Crane. $RMSE_{CLM}$ and $RMSE_{CLMET}$ represent the root mean square error against observations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

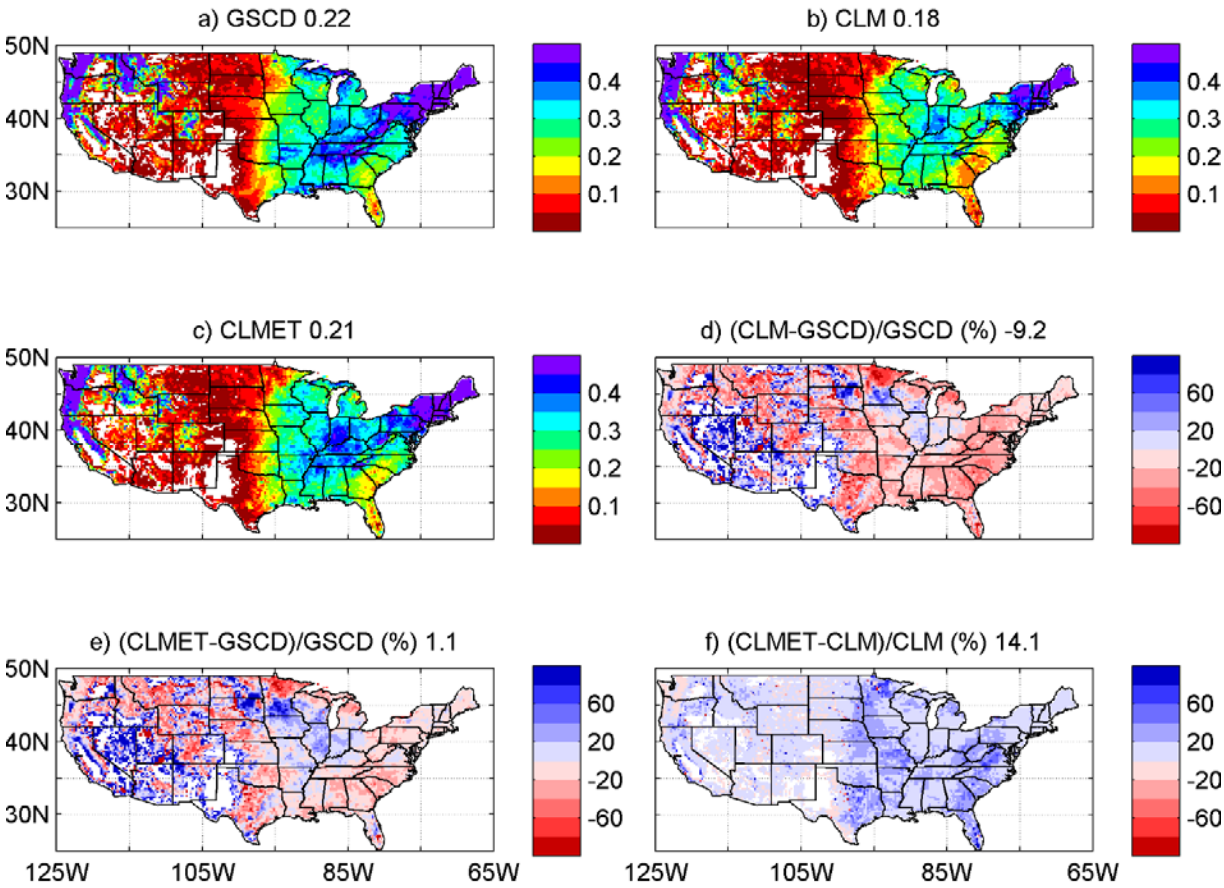

Figure 11 Mean annual runoff coefficient (the ratio runoff to total precipitation) from a) Global Streamflow Characteristics Dataset (GSCD), b) CLM, and c) CLMET, and the relative differences between d) CLM and GSCD, e) CLMET and GSCD, and f) CLMET and CLM during the period 2000-2014. Runoff coefficient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

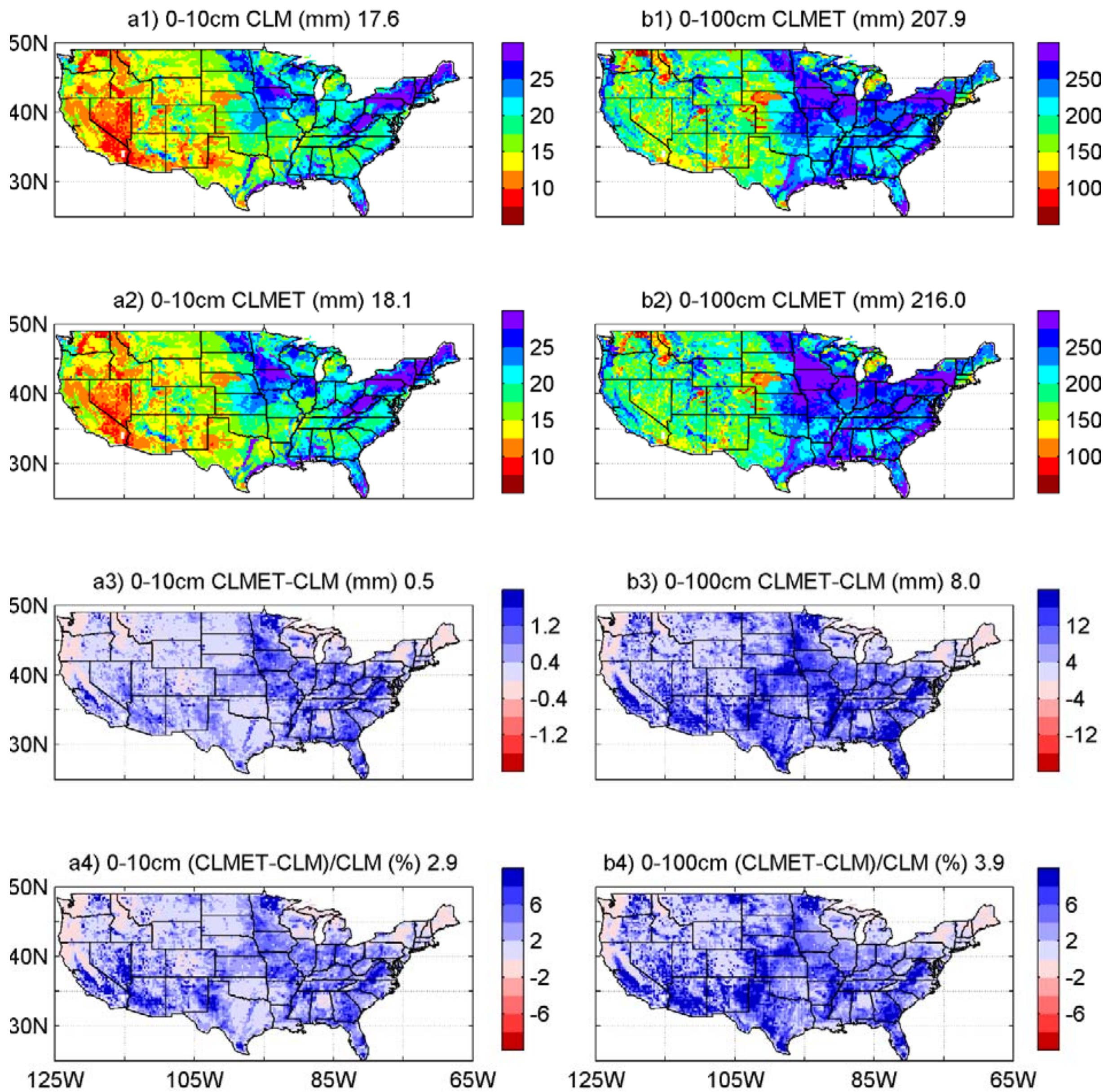

Figure 12 Simulated soil moisture (mm) in the top a) 0-10 cm and b) 0-100 layers in August from 1) CLM and 2) CLMET, 3) their differences, and 4) their relative differences during the period 2000-2014.