Title

A Systematic Evaluation of Noah-MP in Simulating Land-Atmosphere Energy, Water, and Carbon Exchanges Over the Continental United States

Journal

Journal of Geophysical Research: Atmospheres, 122(22)

ISSN

2169-897X

Authors

Ma, Ning
Niu, Guo-Yue
Xia, Youlong
et al.

Publication Date

2017-11-27

DOI

10.1002/2017jd027597

Peer reviewed

A Systematic Evaluation of Noah-MP in Simulating Land-Atmosphere Energy, Water, and Carbon Exchanges Over the Continental United States

Ning Ma[1,2,3], Guo-Yue Niu[2,4], Youlong Xia[5], Xitian Cai[6], Yinsheng Zhang[1], Yaoming Ma[1,3], and Yuanhao Fang[7]

[1] Key Laboratory of Tibetan Environment Changes and Land Surface Processes, Institute of Tibetan Plateau Research, and Center for Excellence in Tibetan Plateau Earth Sciences, Chinese Academy of Sciences, Beijing, China, [2] Department of Hydrology and Atmospheric Sciences, University of Arizona, Tucson, AZ, USA, [3] University of Chinese Academy of Sciences, Beijing, China, [4] Biosphere 2, University of Arizona, Tucson, AZ, USA, [5] Environmental Modeling Center, National Centers for Environmental Prediction, College Park, MD, USA, [6] Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ, USA, [7] Department of Hydrology and Water Resources, Hohai University, Nanjing, China

Abstract

Accurate simulation of energy, water, and carbon fluxes exchanging between the land surface and the atmosphere is beneficial for improving terrestrial ecohydrological and climate predictions. We systematically assessed the Noah land surface model (LSM) with mutiparameterization options (Noah-MP) in simulating these fluxes and associated variations in terrestrial water storage (TWS) and snow cover fraction (SCF) against various reference products over 18 United States Geological Survey two-digital hydrological unit code regions of the continental United States (CONUS). In general, Noah-MP captures better the observed seasonal and interregional variability of net radiation, SCF, and runoff than other variables. With a dynamic vegetation model, it overestimates gross primary productivity by 40% and evapotranspiration (ET) by 22% over the whole CONUS domain; however, with a prescribed climatology of leaf area index, it greatly improves ET simulation with relative bias dropping to 4%. It accurately simulates regional TWS dynamics in most regions except those with large lakes or severely affected by irrigation and/or impoundments. Incorporating the lake water storage variations into the modeled TWS variations largely reduces the TWS simulation bias more obviously over the Great Lakes with model efficiency increasing from 0.18 to 0.76. Noah-MP simulates runoff well in most regions except an obvious overestimation (underestimation) in the Rio Grande and Lower Colorado (New England). Compared with North American Land Data Assimilation System Phase 2 (NLDAS-2) LSMs, Noah-MP shows a better ability to simulate runoff and a comparable skill in simulating R_n but a worse skill in simulating ET over most regions. This study suggests that future model developments should focus on improving the representations of vegetation dynamics, lake water storage dynamics, and human activities including irrigation and impoundments.

1 Introduction

Land surface models (LSMs) provide an interface between the atmosphere and the terrestrial ecosystems within the context of Earth system models (ESMs) as well as ecohydrological models by representing the biogeophysical and biogeochemical processes from the top of plants to the bedrocks (Sellers et al., 1997). On the one hand, LSMs feed the atmosphere of an ESM with energy, water, and carbon fluxes. On the other hand, LSMs have been widely used to investigate the impacts of climate change and anthropogenic activities on terrestrial water and ecosystems at multiple spatial and temporal scales (Barlage et al., 2015; Piao et al., 2015; Yan & Dickinson, 2014; Zeng et al., 2016). In the past few decades, LSMs have evolved from simple “bucket” schemes (Manabe, 1969) to the third generation LSMs (Pitman, 2003; Sellers et al., 1997), which aim to incorporate more feedback mechanisms between climate, water, and terrestrial ecosystems with more detailed representations of water and ecosystem dynamics (Niu & Zeng, 2012). Nevertheless, with more complex interactions of processes and uncertain parameters, such third generation LSMs need to be tested against various available observation data products to identify their strengths as well as deficiencies for developing next generation models (Best et al., 2015; Clark et al., 2015; Luo et al., 2012).

LSM evaluations have been primarily made at local scales against in situ flux observations with a fetch of only hundreds of meters for the sake of more accurate measurements (Chen et al., 1996; Schaefer et al., 2012; Stöckli et al., 2008). However, for use in large-scale water, weather, and climate models, it is necessary for a given LSM to be assessed off-line at much larger scales (e.g., basin, regional, continental, and global scales) (Pitman, 2003; Sellers et al., 1997). Across a wide range of climates and ecosystems, large-scale testing can help address models' weaknesses and strengths that a local-scale testing cannot. Also, given some spatially constant model parameters (e.g., runoff parameters) across different river basins, the LSMs' performance becomes less dependent on the optimization of model parameters at large scales. The main challenge in evaluating LSMs at a large scale has been primarily the deficiency of spatial representativeness of the local measurements for a relatively larger LSM grid (Niu & Zeng, 2012; Roads et al., 2003). Fortunately, the fast-developing satellite observations from space and the upscaled gridded data from ground-based observational networks have greatly facilitated LSMs evaluations at the large scales (Anav et al., 2015; Bonan et al., 2011; Cai, Yang, David, et al., 2014; Cai, Yang, Xia, et al., 2014; Demaria et al., 2016; Mao et al., 2012; Toure et al., 2016; Xia, Cosgrove, Mitchell, Peters-Lidard, Ek, Brewer, et al., 2016; Xia, Cosgrove, Mitchell, Peters-Lidard, Ek, Kumar, et al., 2016).

By improving the realism of Noah in a variety of representations of terrestrial biophysical and hydrological processes, the Noah LSM with multiparameterization options (Noah-MP) (Niu et al., 2011) was designed for a full spectrum of environmental conditions worldwide. Noah-MP has been

extensively tested at local scales, focusing on the effects of new representations of a specific physical process on the improvements of the model's performance (e.g., Gao et al., 2015; Gayler et al., 2014; Zhang et al., 2016; Zheng et al., 2015). Noah-MP was also evaluated at a global scale focusing primarily on runoff, soil moisture, snow depth, and snow water equivalent (SWE) over 50 global river basins (Yang et al., 2011). Afterward, Cai, Yang, David, et al. (2014) assessed the Noah-MP's performance in simulating some major hydrological variables including terrestrial water storage (TWS), evapotranspiration (ET), runoff, and soil moisture in the Mississippi River Basin. Further, Cai, Yang, Xia, et al. (2014) compared the skill of four LSMs including Noah-MP, Noah, Community Land Model (CLM), and Variable Infiltration Capacity (VIC) model in simulating the water balance over the continental United States (CONUS), demonstrating that Noah-MP performs best in reproducing soil moisture and TWS. It should be noted that, however, the climates and ecosystems of the numerous river basins over the CONUS are noticeably different (Guttman & Quayle, 1996). Rice et al. (2016) found that the watershed characteristics due to the differences between precipitation and evaporative demand have a strong influence on the long-term variability of streamflow across CONUS, thereby impacting the magnitude of changes in the hydrologic cycle. In the context of large-sample hydrological studies, Gupta et al. (2014) argued that the flexibility and transferability of a model cannot be clarified without assessments over different basins encompassing a wide range of land surface and climatic conditions. This supports the need of actively evaluating a land surface model's performance over many regions with hydroclimatic regimes that vary considerably within the model domain. Recently, Xia, Cosgrove, Mitchell, Peters-Lidard, Ek, Brewer, et al. (2016) evaluated Noah-MP and Noah-I (the research version of Noah in the North American Land Data Assimilation System Phase 2 (NLDAS-2)) in runoff simulation over 12 National Weather Service River Forecast Centers within CONUS. However, such an evaluation of Noah-MP at a river basin scale with a wide range of climates is limited only for runoff. With regard to other variables (e.g., those associated with vegetation dynamics), it is not yet well investigated how Noah-MP would perform in different regions with different hydroclimatic conditions.

Noah-MP has been widely used by the atmospheric and hydrological communities for weather forecast and short-term climate predictions by coupling to the Weather Research and Forecast Model (Barlage et al., 2015) as well as hydrological predictions by coupling to the National Water Model (Cosgrove et al., 2017). As such, it is critically important to accurately document the model's performance not only in hydrological simulations (e.g., soil moisture, runoff, ET, and TWS) as done by previous studies but also other closely related variables, for example, gross primary productivity (GPP), net radiation (R_n), sensible heat flux (H), and snow cover fraction (SCF). Most advanced third generation LSMs link transpiration to plant photosynthesis through the same stomatal control, and the allocation of GPP

into leaf and root carbon storages may directly affect the simulated leaf area that further feeds back to the simulation of GPP and transpiration. In this regard, errors in the GPP simulation could propagate to the estimation of transpiration and ET (Bonan et al., 2011; Schaefer et al., 2012). Therefore, identifying the source of ET modeling biases requires a simultaneous analysis of the simulated GPP. Additionally, LSMs need to conserve energy for the partitioning of R_n into turbulent sensible and latent heat fluxes (Best et al., 2015; Pitman, 2003), the simulated R_n may therefore impact the model's ability of predicting both sensible and latent heat fluxes. To our knowledge, it is not yet known the effectiveness in modeling R_n for Noah-MP at the regional/continental scales. Furthermore, as a key component in both surface water and energy balances, snow cover is extremely critical for estimating surface $R_n$and snowmelt runoff (Chen et al., 2014). Snow cover fraction (SCF), that is, the fraction of a model grid cell covered by snow, is one of the most common representations of subgrid snow distributions in ESMs (Niu & Yang, 2007). Evaluation of SCF can be also regarded as an indirect assessment of snow mass and snow depth since they are highly correlated (Niu & Yang, 2007). As done for CLM (Toure et al., 2016) and VIC (Demaria et al., 2016), evaluating SCF would benefit a more comprehensive understanding of snow processes modeling for Noah-MP.

In recognition of the above needs, the present study aims (1) to systematically assess the Noah-MP's performances in simulating GPP, R_n, H, ET (in lieu of latent heat flux, LE), runoff, terrestrial water storage anomaly (TWSA), and SCF over 18 United States Geological Survey (USGS) two-digital hydrological unit code (HUC2) regions of CONUS; (2) to identify possible causes to poor performances in modeling any key variables; and (3) to compare Noah-MP's ability with the current NLDAS-2 LSMs. This study differs with the above mentioned large-scale evaluations of Noah-MP in the following aspects: (i) evaluating additional variables including R_n and its four components, GPP, H, and SCF across diverse hydroclimatic regimes; (ii) additional bias source analyses of the simulated R_n, GPP, ET, and TWSA for future model improvements; and (iii) a comparison of Noah-MP to the NLDAS-2 LSMs for potential uses in the next generation NLDAS system (e.g., NLDAS-3).

The paper is structured as follows: Section 2 provides a brief description of Noah-MP and model configuration, reference data sets to validate the model outputs, and evaluation methods. Section 3 shows the spatial and temporal analyses of above mentioned variables in the 18 HUC2 regions.
Section 4 presents the evaluation conclusions and suggestions for future model improvements.

2 Materials and Methods

2.1 A Brief Description of Noah-MP

Noah-MP numerically describes the states of terrestrial energy, water, carbon, and associated flux exchanges between the land surface and the

atmosphere that are controlled by terrestrial hydrometeorological and ecohydrological processes (Niu et al., 2011). Based on Noah (Chen & Dudhia, 2001; Ek et al., 2003), Noah-MP was first augmented with vegetation and groundwater dynamics (Niu et al., 2007) and then equipped with multiple schemes for each of ecohydrological process. Noah-MP has a structure of one canopy layer, three snow layers, and four soil layers. It computes surface temperature by iteratively solving the surface energy balance of solar radiation, longwave radiation, sensible heat, latent heat, and ground heat fluxes. It employs a “semitile” method to deal with surface heterogeneity. The advantage of the semitile method over the traditional “tile” method is to avoid overlapping of shadows on the ground or snow surface. Noah-MP explicitly represents evaporation from the soil surface, canopy interception loss, and transpiration through formulations in analogy to the Ohm's law considering aerodynamic and stomatal resistances to the water vapor and carbon fluxes within and over the plant canopies. Plant transpiration is limited by the stomatal resistance, which is linked to photosynthesis that is further controlled by the root zone soil moisture. Noah-MP assumes uniformly distributed roots in the vertical direction and varying root depth depending on vegetation types.

Noah-MP solves the one-dimensional Richards' equation to compute vertical soil moisture distribution using the Clapp-Hornberber water retention relationship (Clapp & Hornberger, 1978). The upper boundary condition for solving soil moisture is the infiltration rate that is computed as the residual of precipitation minus surface runoff, which is parameterized through a simple TOPMODEL-based runoff scheme (Niu et al., 2005). Noah-MP treats an unconfined aquifer as a reservoir or bucket underlying the soil column to account for water exchanges between the soil and the bucket through gravity and capillary forces (Niu et al., 2007).

Noah-MP represents photosynthesis of C3 plants following the model of Farquhar et al. (1980) modified by Collatz et al. (1991), whereas that of C4 plants following Collatz et al. (1992). The rate of gross photosynthesis is computed as the minimum of three limiting factors: Rubisco limitation, light limitation, and that associated with transport of photosynthate for C3 plants and PEP-carboxylase limitation for C4 plants. It also includes a short-term vegetation phenology model that describes allocation of the assimilated carbon to carbon storages in various parts of the plant (e.g., leaf, stem, wood, and root), death due to cold and drought stresses, and turnover due to senescence, herbivory, or mechanical loss (Dickinson et al., 1998). Leaf area index (LAI) is converted from the leaf carbon storage through specific leaf area, which is dependent on vegetation types.

2.2 Atmospheric Forcing, Vegetation and Soil Parameters, Model Initialization and Simulation

We used the NLDAS-2 hourly, 0.125° × 0.125° atmospheric forcing fields (Xia, Mitchell, Ek, Sheffield, et al., 2012) of air temperature, specific

humidity, wind speed, surface pressure, downward shortwave radiation, downward longwave radiation, and precipitation (available online http://ldas.gsfc.nasa.gov/nldas/NLDAS2forcing_download.php) to drive Noah-MP. We employed the global 1 km hybrid State Soil Geographic Database and the USGS 24-category vegetation data (available online http://www.ral.ucar.edu/research/land/technology/lsm.php) to determine the soil and vegetation parameters, respectively. Both 1 km data sets are first aggregated to 0.125° with the dominant soil and vegetation types to match the spatial resolution of the NLDAS-2 atmospheric forcing. The soil and vegetation parameters are then determined for each soil and vegetation types through the look-up tables in Noah-MP, respectively. The soil column is divided into four layers with thicknesses of 0.1 m, 0.3 m, 0.6 m, and 1 m from the surface to the bottom.

We chose the same model physical options as the EXP6 in Yang et al. (2011), which includes all the augmentations in Niu et al. (2011) (Table 1). For simulations over a continental scale, calibration of model parameters may be impracticable because of the huge computational demand (Clark et al., 2017) and the difficulties in the explicit representations of subgrid heterogeneity (Clark et al., 2015; Niu & Zeng, 2012), although limited calibration did improve the model performance for partial outputs over specific regions (Cai, Yang, David, et al., 2014; Cuntz et al., 2016). We used the default values for all the parameters, which were from the manually optimized parameters used in the EXP6 of Yang et al. (2011). Most key hydrological parameters are still spatially constant for all regions, for example, the maximum groundwater discharge rate, the saturated hydraulic conductivity decay factor, and the micropore fraction controlling upward capillary flow of groundwater (see more details in Niu et al., 2011).

Table 1
Options Used in the Present Noah-MP Simulation

Schemes of physical processes	Option used in this study
Dynamic vegetation	Dynamically predict LAI and GVF
Canopy stomatal resistance	Ball-Berry
Soil moisture factor for stomatal resistance	Noah
Runoff and groundwater	TOPMODEL with groundwater
Surface layer drag coefficient	Monin-Obukhov similarity theory
Super cooled liquid water	Non-iterative
Frozen soil permeability	Linear effects
Radiation transfer	Modified two stream
Ground snow surface albedo	CLASS
Precipitation partition	Jordan91
Snow/soil temperature time scheme	Semi-implicit

Note. For detailed explanations, refer to Niu et al. (2011).

The model is integrated for 36 years at a time step of 1 h from 1 January 1980 to 31 December 2015. Starting from arbitrary, spatially homogeneous initial conditions, we ran Noah-MP for a total of 216 years by repeating the run from 1980 to 2015 for six times and saved the model prognostic

variables at the end of the last cycle. As demonstrated by Cai, Yang, David, et al. (2014), this long spin-up run is undoubtedly enough for the soil water and groundwater to reach an equilibrium state. We then used these saved prognostic variables as initial conditions for another 36 year run from 1980 to 2015. We aggregated the hourly outputs to monthly and yearly for further analyses.

2.3 USGS Two-Digital Hydrological Unit Code Regions

To provide a uniform geographical framework for hydrological research and water resource management, USGS divides the United States into successively smaller hydrological units at four levels: regions, subregions, accounting units, and cataloging units (Seaber et al., 1987). Given a code and named as “hydrological unit code” (HUC), a *hydrologic unit* refers to the upstream area of a specific point on the stream that contributes surface water runoff directly to the outlet. The HUC system has been widely used by not only the scientific community but also federal government as the basis for hydrological monitoring and water resources management over the United States (Sabo et al., 2010; Smith et al., 2002). The first level of hydrological units, *region*, comprises either the drainage area of a major river (e.g., Missouri region) or the combined drainage areas of a series of rivers (e.g., Texas-Gulf region) (Seaber et al., 1987). Every region is named by a unique two-digital number by USGS, that is, HUC2 (Figure 1). Throughout the paper, we conducted our analyses over 18 USGS HUC2 regions of CONUS. The mean area of a HUC2 region is approximately 500,000 km^2 (Velpuri et al., 2013).

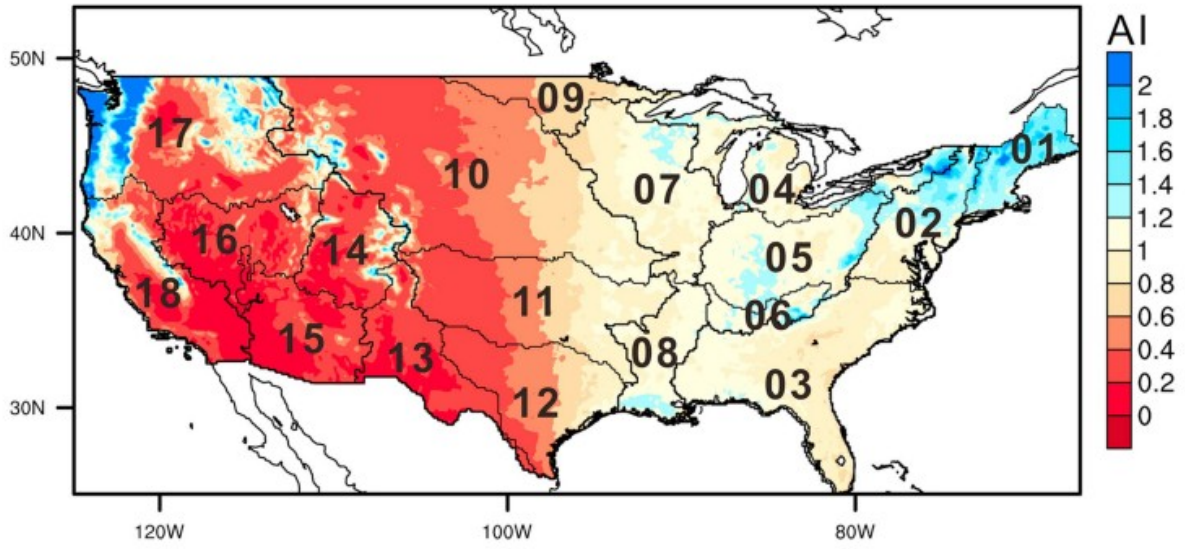

Figure 1

The multiyear mean (1980–2015) annual aridity index (AI) over CONUS. The polygons with thicker black boundaries represent the USGS two-digital hydrologic unit code (HUC2) regions with their corresponding two-digit codes: 01: New England, 02: Mid-Atlantic, 03: South Atlantic-Gulf, 04: Great Lakes, 05: Ohio, 06: Tennessee, 07: Upper Mississippi, 08: Lower Mississippi, 09: Souris-Red-Rainy, 10:

Missouri, 11: Arkansas-White-Red, 12: Texas-Gulf, 13: Rio Grande, 14: Upper Colorado, 15: Lower Colorado, 16: Great Basin, 17: Pacific Northwest, and 18: California.

To characterize the climate background of each region, we calculated the aridity index (AI) (Fu & Feng, 2014), defined as the ratio of annual precipitation to annual potential evapotranspiration, which is calculated with the Penman approach (Penman, 1948), using the NLDAS-2 forcing data and the Noah-MP simulated R_n and ground heat flux. Figure 1illustrates the 1980–2015 climatological mean AI over CONUS. The HUC2 regions in the eastern CONUS (e.g., the New England, Mid-Atlantic, South Atlantic-Gulf, Great Lakes, Ohio, Tennessee, Upper Mississippi, and Lower Mississippi) with AI ranging from 1.2 to 2.0 are characterized as humid regions where precipitation are usually larger than the evaporative demand, while the regions in the western CONUS (e.g., the Rio Grande, Great Basin, part of Pacific Northwest, California, Upper Colorado, and Lower Colorado) with AI lower than 0.6 are characterized as arid regions where precipitation are much lower than the evaporative demand. For the regions in the central CONUS (e.g., the Souris-Red-Rainy, Missouri, Arkansas-White-Red, and Texas-Gulf), AI ranges from 1.2 to 0.6 because of the dramatic gradients in annual precipitation that decreases from the east to the west, indicating a transition regime from humid to arid.

2.4 Reference Data Products

2.4.1 NASA/GEWEX SRB Radiation

We used the global monthly radiation data product at 1° resolution from the National Aeronautics and Space Administration/Global Energy and Water cycle EXperiment (NASA/GEWEX) Surface Radiation Budget (SRB) project at Langley Research Center (Release 3.0 version, available online https://eosweb.larc.nasa.gov/project/srb/srb_table). The NASA/GEWEX SRB radiation products have been widely used for correcting the radiative components of reanalysis forcing (Mathis et al., 2015) and evaluating ESM simulations (Luo et al., 2016; Xia, Cosgrove, Mitchell, Peters-Lidard, Ek, Kumar, et al., 2016). Compared to multiple ground stations of the Baseline Surface Radiation Network, the root-mean-square error are 23.3 and 11.1 W m^{-2} for the monthly shortwave and longwave radiation fluxes, respectively (Zhang et al., 2015, 2013), suggesting an overall satisfactory accuracy of NASA/GEWEX SRB radiation products. We used its monthly downward and upward shortwave and longwave radiation data during January 1984 to December 2007. We resampled the NASA/GEWEX SRB products from 1° to 0.125° using a nearest-neighbor interpolation method, to match the model resolution and thus calculate the spatial pattern of the modeled R_n′s relative bias.

2.4.2 FLUXNET MTE GPP, *H*, and LE

We used the FLUXNET model tree ensembles (MTE) GPP, *H*, and LE data from Max Planck Institute for Biogeochemistry (available online https://www.bgc-jena.mpg.de/geodb/projects/Home.php). The MTE was first trained with

measured flux data of 198 FLUXNET towers across a wide range of biomes worldwide. With inputs of the fraction of absorbed photosynthetic active radiation derived from remote sensing, climate, and land cover data, this approach generates the monthly, 0.5° × 0.5° gridded GPP, *H*, and LE data sets during 1982–2008 over the global continents (Jung et al., 2010, 2011). These data sets have been widely employed to evaluate LSMs-simulated (Anav et al., 2015; Bonan et al., 2011; Xia, Cosgrove, Mitchell, Peters-Lidard, Ek, Brewer, et al., 2016; Xia, Cosgrove, Mitchell, Peters-Lidard, Ek, Kumar, et al., 2016) and satellite-derived (Frankenberg et al., 2011) land-atmosphere carbon, water, and energy exchanges. While the uncertainty in the FLUXNET MTE products is not negligible owning to the uneven spatial distribution of flux towers (e.g., few towers in Asia and Africa) selected for training the model tree, we have the greatest confidence in it over the “data-rich” areas such as CONUS where most FLUXNET sites were incorporated (see the supporting information Figure S1 in Jung et al., 2010). In this study we used the monthly FLUXNET MTE GPP, *H*, and LE data during 1982–2008 to assess the corresponding Noah-MP simulation outputs. To calculate the relative difference between model and products, we resampled the 0.5° FLUXNET MET products to 0.125° using a nearest-neighbor interpolation method.

2.4.3 GRACE Terrestrial Water Storage Anomaly

We used the global 1°, monthly Gravity Recovery and Climate Experiment (GRACE)-observed TWSA products (Release 5.0 version, available online http://grace.jpl.nasa.gov) released by three different processing centers: GeoforschungsZentrum Potsdam, Center for Space Research at University of Texas, Austin, and Jet Propulsion Laboratory based on the same Earth's gravity field. While these products are slightly different, the noise of different solutions can be effectively reduced by arithmetic average of the three products (Sakumura et al., 2014). For this reason, we used the average of all the three products during 2003–2015 in this study. Because of the postprocessing of original GRACE observation data, there tends to be signal loss of surface mass variations (Landerer & Swenson, 2012). Therefore, Landerer and Swenson (2012) developed a gain factor data set to reduce the differences between the original and the reprocessed GRACE data. Thus, we multiplied the TWSA products this gain factor. According to Kumar et al. (2016), the total measurement error of the GRACE TWSA ranges from 0 to 40 mm over most parts of CONUS, but higher errors may occur on the West Coast, the lower Mississippi River basin, and Florida due to the spatial leakage error from a limited range of spherical harmonic (Seo et al., 2006).

2.4.4 Water Level Data of the Great Salt Lake and Michigan Lake

To quantify the impacts of lakes on the modeled TWS, we selected two regions with large surface areas of water bodies: the Great Lakes (HUC2 code: 04) and the Great Basin (HUC2 code: 16). In the Great Lakes, we used the monthly water level data of the Michigan Lake (available

online https://tidesandcurrents.noaa.gov/) observed at five stations around the lake including Ludington, Holland, Calumet Harbor, Milwaukee, and Sturgeon Bay Canal during 2003–2015. All these data have been verified by NOAA Center for Operational Oceanographic Products and Services. Although the water level data of the five stations are almost same, we used their arithmetic average values to reduce the uncertainties in the water level measurements for such a large lake. With regard to the Great Basin region, we used the monthly water level data of the Great Salt Lake (available online https://waterdata.usgs.gov/nwis/nwismap/?site_no=10010000) observed at the Saltair Boat Harbor by USGS during 2003–2015. Since the ratio of the lake area to the entire area of a region is essentially small, the changes in the lake area may have very little influence on the results. We therefore assumed that the lake area of the Michigan Lake (57,756 km^2) and the Great Salt Lake (4,402 km^2) hold static during this 13 year period.

2.4.5 MODIS Snow Cover Fraction

We used the Moderate Resolution Imaging Spectroradiometer (MODIS) SCF product from the NASA Terra Satellite (Riggs et al., 2006). MODIS snow-covered area is determined by the normalized snow difference index and other threshold tests (Hall & Riggs, 2007). The MOD10CM version 5 of SCF data (available online https://nsidc.org/data/) is based on the 0.05° × 0.05° climate-modeling grid (CMG) over a global view. With a temporal resolution of monthly, this product is generated by assembling only high-quality CMG level-3 daily snow cover product (MOD10C1) whose confidence index is higher than 70% during the month, thereby basically avoiding the contamination of cloud (Riggs et al., 2006). An evaluation by Hall and Riggs (2007) suggests that the accuracy of MODIS SCF product reaches 93% on average compared to in situ snow measurements. While the MOD10CM data extend from March 2000 to the present, we evaluated the simulated SCF based only on the period of January 2004 to December 2015 since there are some data gaps before 2004.

2.4.6 USGS Water Watch Hydrological Unit Runoff

We used the USGS Water Watch hydrological unit runoff data (available online https://waterwatch.usgs.gov/). The USGS runoff of each hydrological unit was generated by combining the historical flow data collected at stream gauges, the drainage basins of the stream gauging stations, and the boundaries of the hydrological units (Brakebill et al., 2011). This data set contains the monthly runoff from 1901 to 2016 at various hydrological unit scales (e.g., from HUC2 regional scale to HUC8 cataloging unit scale) and has been regarded as a close surrogate of the natural runoff in hydroclimate studies (Ashfaq et al., 2013; Oubeidillah et al., 2014; Velpuri et al., 2013). In the present study, we used its monthly runoff data at a HUC2 regional scale for the period spanning January 1982 through December 2008.

2.5 NLDAS-2 LSMs Simulated Outputs

The rapid advances of land surface modeling techniques enable the community to provide more accurate initial land surface states for climate modeling and weather forecasting. Under this circumstance, the NLDAS team has developed a standard framework that consists of atmospheric forcing, soil/vegetation data sets, LSMs, and model validations to gain a reliable description of land surface processes over CONUS (Mitchell et al., 2004). The second phase of this project, NLDAS-2 (Xia, Mitchell, Ek, Cosgrove, et al., 2012; Xia, Mitchell, Ek, Sheffield, et al., 2012), has produced country-wide water and energy cycle results. Since Noah-MP is driven by same NLDAS-2 atmospheric forcings as other NLDAS-2 models, a direct comparison between Noah-MP and NLDAS-2 LSMs may be allowed. To this end, we used the simulated R_n, ET, and runoff results of four operational NLDAS-2 LSMs (available online http://www.emc.ncep.noaa.gov/mmb/nldas/) including Mosaic, Noah (version 2.8), Sacramento Soil Moisture Accounting (SAC) and VIC (version 4.0.3) (Xia, Cosgrove, Mitchell, Peters-Lidard, Ek, Brewer, et al., 2016; Xia, Cosgrove, Mitchell, Peters-Lidard, Ek, Kumar, et al., 2016). Note that since SAC is intended for lumped hydrological modeling without simulation of net radiation, the present comparison of R_n is only based on the other three LSMs. In-line with the length of the reference data products, we used the monthly outputs of ET and runoff during 1982–2008 as well as the monthly outputs of R_n during 1984–2007 from the NLDAS-2 LSMs in the comparison. For a given variable, the multimodel ensembles mean (MME) is calculated as the arithmetic average of the NLDAS-2 LSMs outputs.

2.6 Regional Upscaling and Evaluation Methods

There are considerable differences in the spatial resolution between the Noah-MP simulated outputs (i.e., 0.125°) and the above reference data sets (e.g., 0.05° for MODIS SCF; 0.5° for FLUXNET MTE GPP, H, and LE; and 1° for NASA/GEWEX SRB radiation, and GRACE TWSA). While a great number of methods have been used in previous studies to either downscale or upscale the reference data to the same resolution as the model grids, Xia, Cosgrove, Mitchell, Peters-Lidard, Ek, Brewer, et al. (2016) argued that evaluating LSMs performances at a same larger spatial extent is more appropriate. Considering each of the HUC2 regions may be, to some extent, hydrologically self-similar, upscaling both the reference data and the model outputs into a same HUC2 region would reduce the error induced by spatial interpolation of the reference data. Hence, we used regional masks at different spatial resolutions that match the grids of the various reference data, NLDAS-2 LSMs and Noah-MP model outputs to derive regional averages of the monthly (or annual) results over all the HUC2 regions.

To quantify the performance of Noah-MP, we calculated the Pearson correlation coefficient (R), root-mean-square error (RMSE), relative bias (RB), and Nash-Sutcliffe efficiency (NSE) between the Noah-MP outputs and the reference data at various time scales following Ma et al. (2015) and Cai, Yang, Xia, et al. (2014).

Note that most of the data products used in the present study for validation are based on either satellite retrievals or spatiotemporal upscaling of measurements at local sites, involving a variety of uncertainties that may lead to systematic errors. Taking the FLUXNET MTE products as an example, some explanatory variables that used to train the MTE were assumed static over years, thereby leading to an underestimated interannual variance in FLUXNET MTE GPP, ET, and H (Jung et al., 2011). Furthermore, the energy components are not always balanced for NASA/GEWEX R_n, FLUXNET MTE H, and LE (Xia, Cosgrove, Mitchell, Peters-Lidard, Ek, Kumar, et al., 2016). And this may be also true for water components including NLDAS-2 precipitation, USGS HUC2 runoff, GRACE TWSA, and FLUXNET MET ET (Cai, Yang, Xia, et al., 2014). Although quantifying the uncertainties in the reference data products is well beyond the scope of the current study, the potential errors in the "observations" cannot be completely ignored.

3 Results and Discussions

3.1 Net Radiation

Both NASA/GEWEX SRB-derived and Noah-MP-modeled R_n are computed as the sum of four radiative components, that is, downward shortwave radiation (DSR), downward longwave radiation (DLR), upward shortwave radiation (USR), and upward longwave radiation (ULR). On the whole, the spatial pattern of the multiyear (1984–2007) average of the modeled annual R_n follows that of the NASA/GEWEX SRB product (Figures 2a and 2b), both showing large values in the southeastern CONUS and the West Coast. The regional averages of annual R_n in Figure 2d suggest that Noah-MP is able to capture the interregional variability with R, NSE, RB, and RMSE being 0.905, 0.59, −0.03, and 7.5 W m^{-2}, respectively. In most parts of southeastern and western CONUS, the relative biases are within ±10% (Figure 2c). However, in the Rocky Mountains and northeastern CONUS, Noah-MP produces much smaller R_n than does the NASA/GEWEX SRB with a relative bias of approximately −40% (Figure 2c).

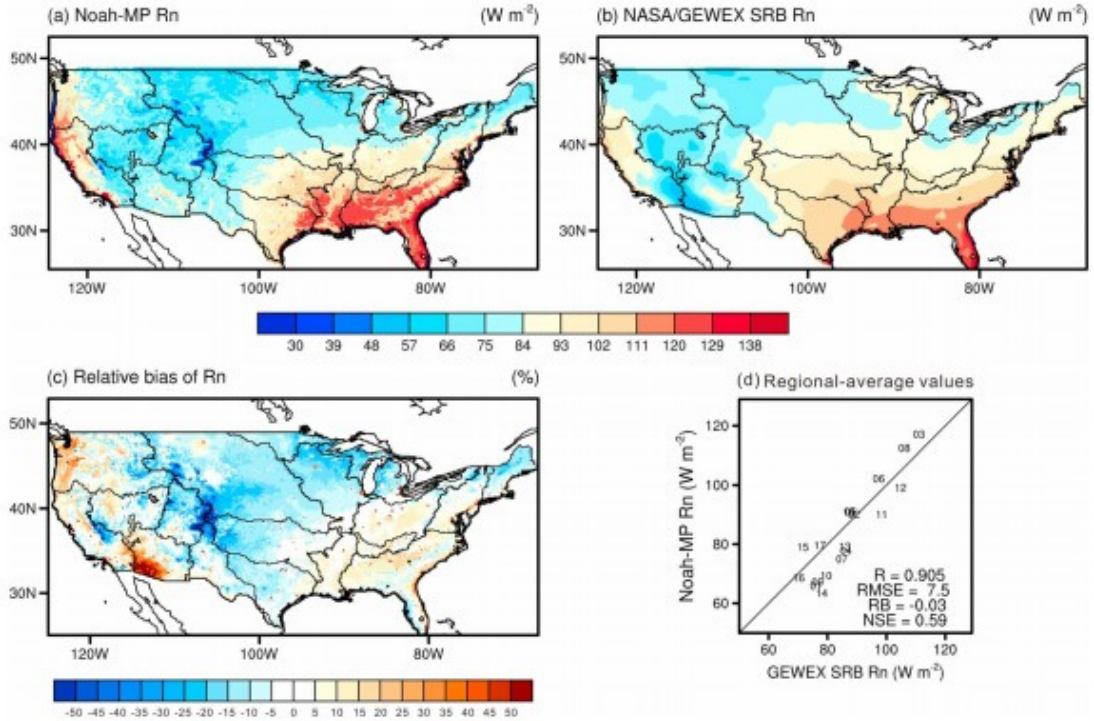

Figure 2

Spatial pattern of the multiyear (1984–2007) mean net radiation (R_n) from (a) Noah-MP, (b) NASA/GEWEX SRB, (c) the relative bias [=(Noah-MP − NASA/GEWEX SRB)/NASA/GEWEX SRB × 100%] over CONUS, and (d) comparison of the regional averages over the 18 HUC2 regions shown as their corresponding HUC2 codes.

To explore the possible reasons for the significant discrepancy of R_n in the Rocky Mountains and northeastern CONUS, we further investigated the four components that used to derive R_n. DSR and DLR are actually the NLDAS-2 forcing to drive Noah-MP, while USR and ULR are the modeling results from Noah-MP. The NLDAS-2 DSR and DLR agree very well with the NASA/GEWEX SRB product as witnessed from the spatial pattern of the relative biases (Figures S1c and S1g in the supporting information) with NSE values of 0.83 and 0.92 for DSR and DLR, respectively, over the 18 HUC2 regions (Figures S1d and S1h). The relative biases of the simulated ULR are within ±5% at nearly all grids, indicating a very accurate simulation of the land surface temperature (Figures S1m–S1p). However, Noah-MP yields remarkably larger USR values (Figures S1i–S1l) caused by the higher modeled surface albedo over the western Rocky Mountains and the northeastern boreal areas, especially during winter season (December, January, and February) (Figure S2). Naturally, these areas are usually covered by snow in winter due to its high elevation and latitude (see section 3.6 for spatial pattern of SCF), the albedo is therefore expected to be higher during this period. As seen from Figure S2, the modeled surface albedo reaches up to 0.6, about twice that of the NASA/GEWEX SRB product, suggesting that the modeled R_n tends to be more plausible because of the

more reasonable modeled cold season albedo over such high-elevation and high-latitude regions. Further comparisons to MODIS albedo and LAI products for different vegetation types over these snow-covered regions would be helpful to more fully evaluating the two-stream radiation scheme with consideration of within- and between-canopy gap probabilities (Niu & Yang, 2004) in Noah-MP, which already showed advantages over other schemes at local scales (Chen et al., 2014). Though not directly compared to the MODIS albedo products, the comparison of the modeled SCF to the MODIS SCF product (see section 3.6) implies a more accurate estimation of surface albedo than the NASA/GEWEX SRB data, because the MODIS SCF reflects not only the high reflectivity of the snow on the ground but also the shading effects of the overlying vegetation.

Noah-MP is also capable of simulating R_n in terms of seasonal variability over all HUC2 regions with NSE ranging from 0.65 to 0.94 (Figure 3). The modeled R_n is generally comparable to the NASA/GEWEX SRB R_n (Zhang et al., 2015, 2013; see section 2.4.1) with RMSE values ranging from 10.8 W m^{-2} to 28.0 W m^{-2} (Figure 3) over the 18 HUC2 regions. However, it appears that the interannual variability of the modeled monthly R_n is smaller than that of the NASA/GEWEX SRB data during 1984–2007 over the majority of the regions. Owning to the larger modeled albedo (Figure S2), Noah-MP yields smaller R_n values during winter and early spring over the snow-covered northeastern boreal regions (e.g., the New England, Great Lakes, Souris-Red-Rainy, Upper Mississippi, and Missouri) and the western Rocky Mountains (e.g., the Upper Colorado) in cold seasons. In warm seasons, however, the modeled R_n generally agrees well with that of NASA/GEWEX SRB over all the regions (Figure 3).

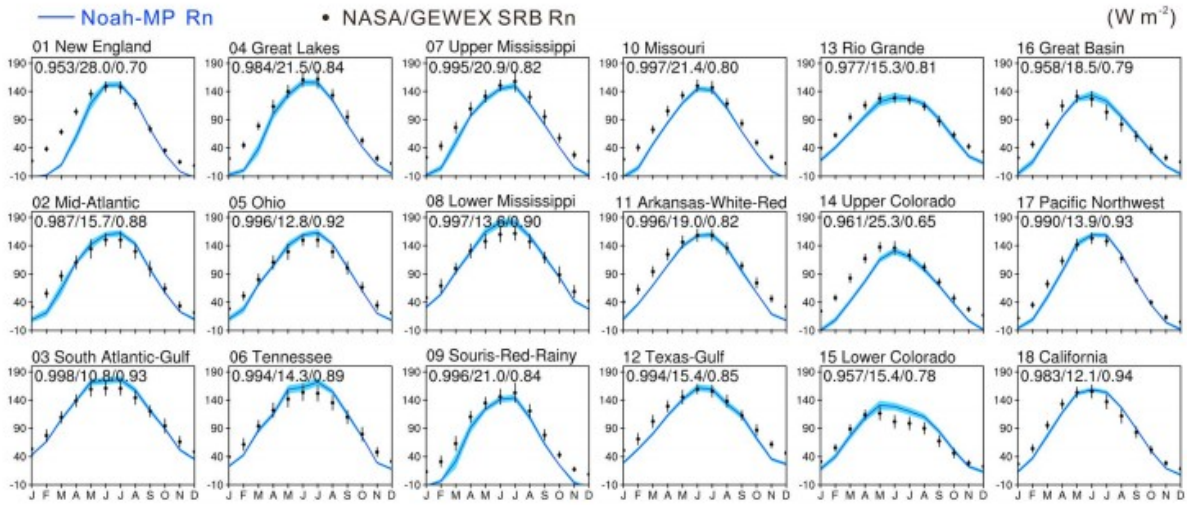

Figure 3

The multiyear (1984–2007) mean annual cycle of net radiation (R_n) from Noah-MP (blue line) and NASA/GEWEX SRB (black dot) over the 18 HUC2 regions. The light blue shaded area and the black error bar represent the standard derivation of the Noah-MP and NASA/GEWEX SRB data, respectively,

to indicate interannual variability for each month of a year. Also shown on the top of each panel are modeling metrics in the order of *R*, RMSE, and NSE.

3.2 Gross Primary Production

Noah-MP produces a spatial pattern of annual GPP similar to that of the FLUXNET MTE product over CONUS (Figures 4a and 4b), both showing less carbon uptake over the arid regions in the western CONUS (except the coastal areas) and more carbon uptake in the humid regions of the eastern CONUS. However, the relative biases are roughly 50% in the highly vegetated area of the central and eastern CONUS (Figure 4c). In particular, the simulated GPP in the Ohio, Tennessee, Upper Mississippi, and Lower Mississippi (i.e., regions “05”, “06”, “07”, and “08” in Figure 4d) are higher than 2,400 g C m^{-2} yr^{-1}, while the FLUXNET MTE product only has approximately 1,600 g C m^{-2} yr^{-1}. Regarding the comparison of multiyear mean values over the 18 HUC2 regions, Noah-MP overestimates GPP with RMSE and RB values of 542.5 g C yr^{-1} and 0.4, respectively (Figure 4d). The performance of Noah-MP in modeling GPP is generally comparable to that of CLM, JULES (Joint UK Land Environment Simulator) and ORICHIDEE (ORganizing Carbon and Hydrology in Dynamic EcosystEms) (Anav et al., 2015), which also remarkably overestimates GPP in the eastern CONUS. Similarly, Mao et al. (2012) also documented that CLM produces much more GPP than does MODIS in the evergreen forest ecosystems over the world. In the arid regions of the western CONUS with most grasslands and shrublands, the bias in the modeled annual GPP is much smaller.

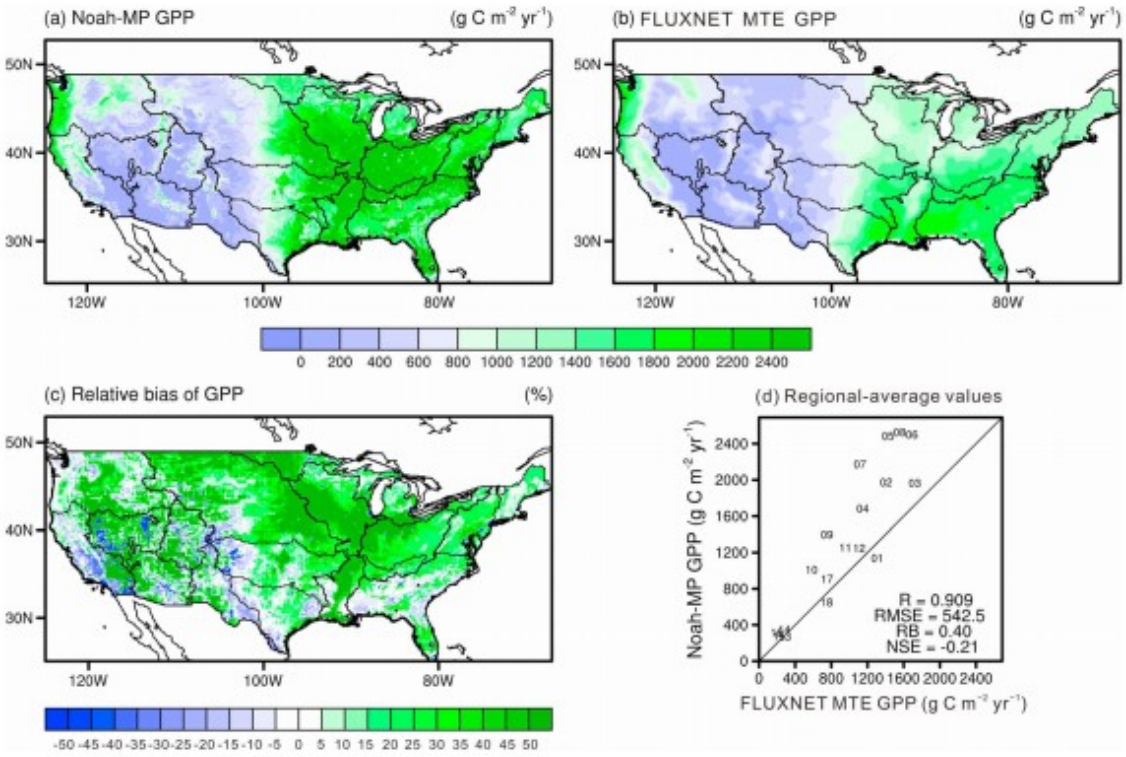

Figure 4

Spatial pattern of the multiyear (1982–2008) mean gross primary production (GPP) from (a) Noah-MP, (b) FLUXNET MTE, (c) the relative bias [=Noah-MP − FLUXNET MET)/FLUXNET MET × 100%] over CONUS, and (d) comparison of the regional averages over the 18 HUC2 regions shown as their corresponding HUC2 codes.

Similar to the modeling results of CLM, JULES, and ORICHIDEE (Anav et al., 2015), Noah-MP overestimates GPP in both magnitude and interannual variability compared to that of FLUXNET MTE mainly over the humid and the transitional regions. With regard to the annual cycle (Figure 5), the NSE values over the highly vegetated regions are primarily lower than 0.6 (e.g., “05”, “06”, “07”, “08”, and “09” in Figure 5). In particular, Noah-MP simulates earlier leafing-out in the spring and greater carbon uptake in the later spring over most central and eastern areas of CONUS, as reflected by the large positive bias during this period (Figure 5). For example, the simulated GPP is at least twice that of FLUXNET MTE over the Lower and Upper Mississippi regions in spring. This may be caused by the Noah-MP's scheme of carbon allocation to shoot and root, which probably allocates too much assimilated carbon into shoots in spring, accelerating the plants' carbon uptake through photosynthesis (Niu et al., 2011). Such biases may be alleviated by modifying the carbon allocation scheme (Gim et al., 2017), refining the temperature limitation (or heat stress) (Schaefer et al., 2012) as well as introduction of nitrogen limitation (Cai et al., 2016; Stöckli et al., 2008).

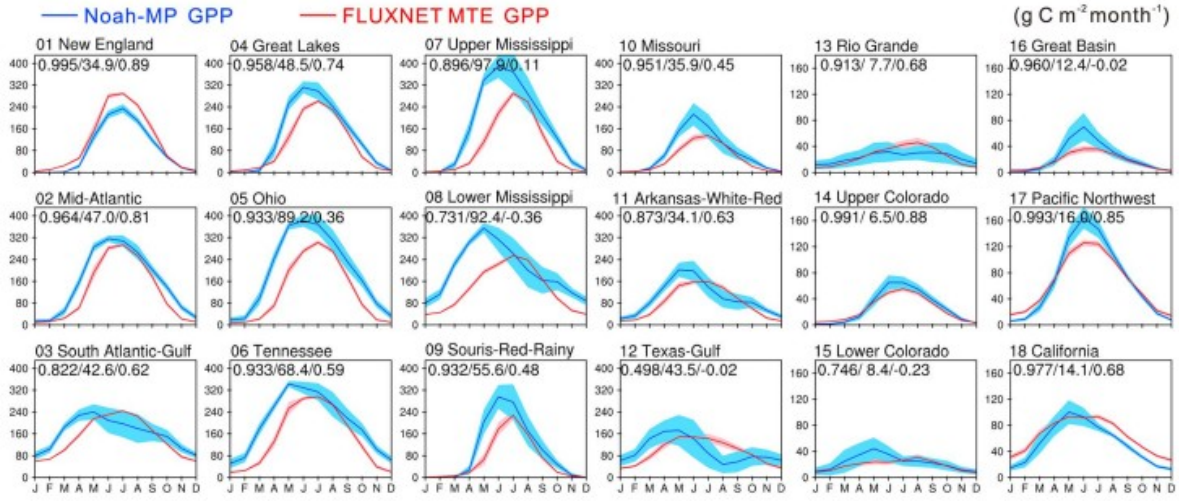

Figure 5

The multiyear (1982–2008) mean annual cycle of gross primary production (GPP) from Noah-MP (blue line) and FLUXNET MTE (red line) over the 18 HUC2 regions. The light blue and red shaded areas represent the standard derivation of the Noah-MP and FLUXNET MTE GPP data, respectively, to indicate the interannual variability for each month of a year. Also shown on the top of each panel are modeling metrics in the order of *R*, RMSE, and NSE.

3.3 Evapotranspiration

In general, the spatial pattern of modeled ET resembles that of FLUXNET MTE, both declining from the southeastern to the northern and western

CONUS (Figures 6a and 6b). Interestingly, the spatial pattern of the relative bias is very similar to that of GPP (comparing Figure 6c with Figure 4c). Noah-MP underestimates ET over the Central Valley in California by approximately −50% but overestimates ET by 40%–50% in most areas of the eastern and central CONUS except parts of South Atlantic-Gulf and Texas-Gulf (Figure 6c). At a regional scale, the largest discrepancy appears in the Lower Mississippi (region “08” in Figure 6d), where Noah-MP yields 1047.8 mm yr^{-1}, while FLUXNET MTE has only 792.2 mm yr^{-1}. Over the whole CONUS domain, the model overestimates ET by 22% relative to FLUXNET MTE (see the RB value in Figure 6d). In general, the simulated surface hydrological variables are often highly dependent on the quality of NLDAS-2 precipitation and other forcing variables (Xia, Mitchell, Ek, Sheffield, et al., 2012). For this reason, the modeled ET biases may be also partially attributed to the uncertainties in the forcings.

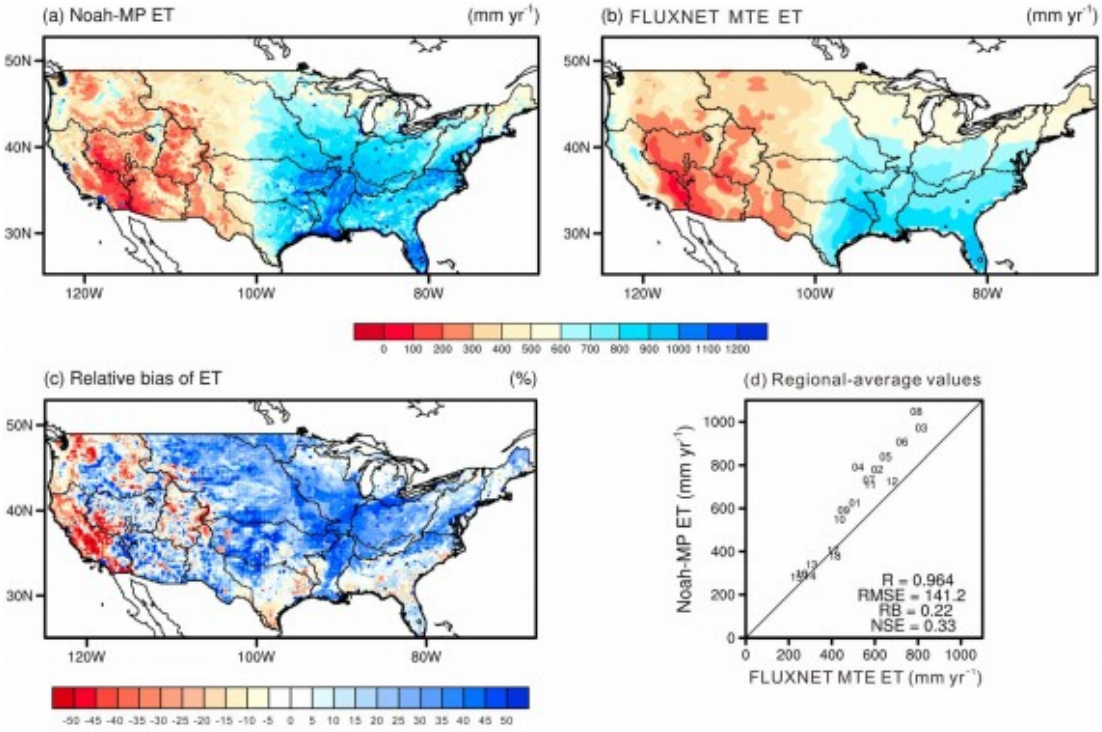

Figure 6

Spatial pattern of the multiyear (1982–2008) mean evapotranspiration (ET) from (a) Noah-MP, (b) FLUXNET MET, (c) the relative bias [=(Noah-MP − FLUXNET MET)/FLUXNET MTE × 100%] over CONUS, and (d) comparison of the regional averages over the 18 HUC2 regions shown as their corresponding HUC2 codes.

With regard to the seasonal variability, Noah-MP basically captures the annual cycle of the FLUXNET MTE ET product with NSE values of 15 regions being larger than 0.8 (Figure 7). In the Lower Colorado and California (i.e., regions “15” and “18” in Figure 7), however, the simulated ET does not agree well with that of FLUXNET MTE, leading to much smaller NSE values. In response to the earlier leaf emergence in spring and later greater carbon

uptake (Figure 5), Noah-MP produces higher ET in spring and early summer in Ohio, Tennessee, Upper Mississippi, and Lower Mississippi (regions “05”, “06”, “07”, and “08” in Figure 7) under the same stomata control on transpiration and photosynthesis. As ET is the sum of transpiration, canopy interception loss, and soil evaporation, the magnitude of the biases in ET is less notable than those in GPP (see Figures 4c and 6c). In the central and eastern regions with greater vegetation coverage, transpiration is dominant (Figure S3), and thus contributing the most to the biases in the total ET, while in the western arid regions where are less vegetated, the biases in modeled ET are relatively smaller (Figure 6d).

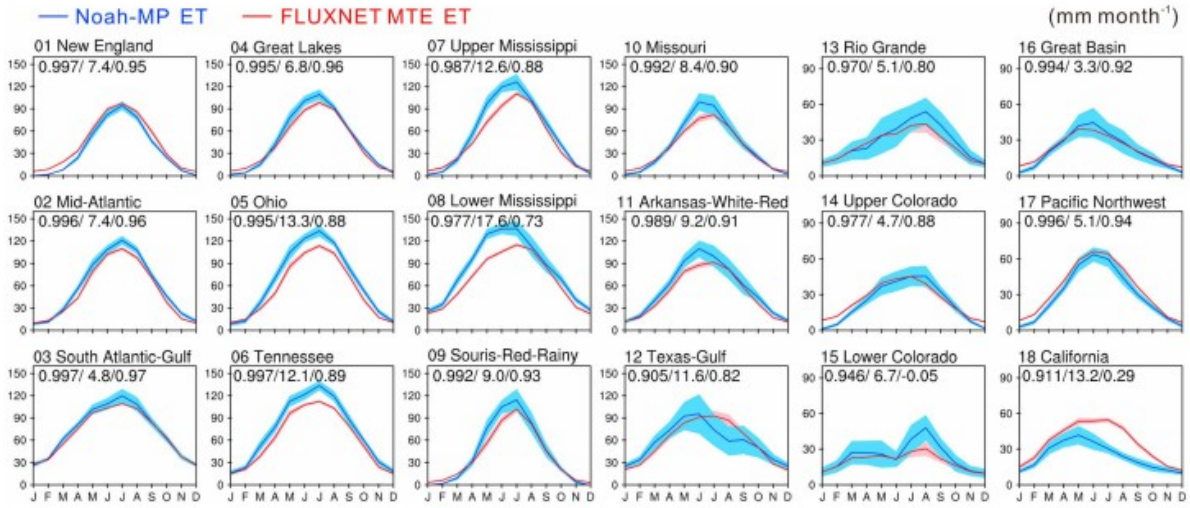

Figure 7

The multiyear (1982–2008) mean annual cycle of evapotranspiration (ET) from Noah-MP (blue line) and FLUXNET MTE (red line) over the 18 HUC2 regions. The light blue and red shaded areas represent the standard derivation of the Noah-MP and FLUXNET MTE ET data, respectively, to indicate the interannual variability for each month of a year. Also shown on the top of each panel are modeling metrics in the order of *R*, RMSE, and NSE.

To confirm that the overestimated ET during spring is mainly caused by the overestimated GPP, we ran an additional experiment (denote as EXP_LAI) without the dynamic vegetation model but with an LAI climatology (with seasonal cycles) prescribed for each land use type. In EXP_LAI, the green vegetation fraction is calculated from the prescribed LAI using the method of Niu et al. (2011). As seen, the model biases are reduced to a large extent when the dynamic vegetation model is not used (Figure 8). In particular, the relative biases drop to <20% over most parts of the central and eastern CONUS (see Figures 6c and 8b). In terms of the comparison of 18 HUC2 regions at a multiyear mean annual scale, the NSE increases from 0.33 to 0.90, RMSE decreases from 141.2 mm yr^{-1} to 55.6 mm yr^{-1}, and RB reduces from 0.22 to 0.04 (Figure 6d versus Figure 8c).

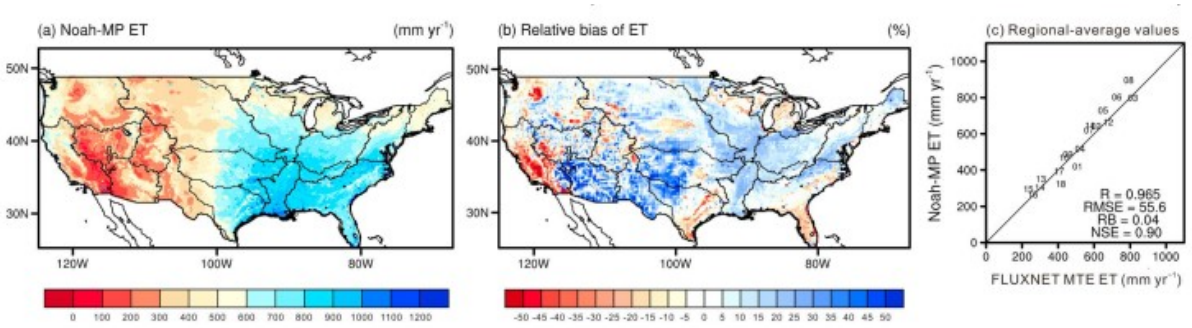

Figure 8

Spatial pattern of multiyear (1982–2008) mean annual evapotranspiration (ET) from (a) the EXP_LAI experiment of Noah-MP simulation with prescribed LAI, (b) the relative bias [=(Noah-MP − FLUXNET MTE)/FLUXNET MTE × 100%] over CONUS, and (c) comparison of regional averages over the 18 HUC2 regions shown as their corresponding HUC2 codes.

3.4 Sensible Heat Flux

Noah-MP produces a spatial pattern of *H* that is comparable with that of FLUXNET, both showing smaller values in the northeast and increasing gradually toward the southwest (Figures 9a and 9b). To balance the overestimated ET, the modeled *H* is underestimated over most area of CONUS except for the West Coast and parts of southeastern CONUS. With negative relative biases above −40%, the most obvious underestimation of *H* occurs in the same areas of the northeastern CONUS where LE is apparently overestimated (Figures 6c and 9c). In the western costal and part of the southeastern CONUS, however, Noah-MP overestimates *H* by more than 50% (Figure 9c), as also reflected by the positive biases over the South Atlantic-Gulf, California, and Northwest (i.e., regions “03”, “17”, and “18” in Figure 9d). Regarding the multiyear mean regional-average values, Noah-MP underestimates *H* by −16% with RMSE and NSE of 10.1 W m^{-2} and 0.5, respectively (Figure 9d).

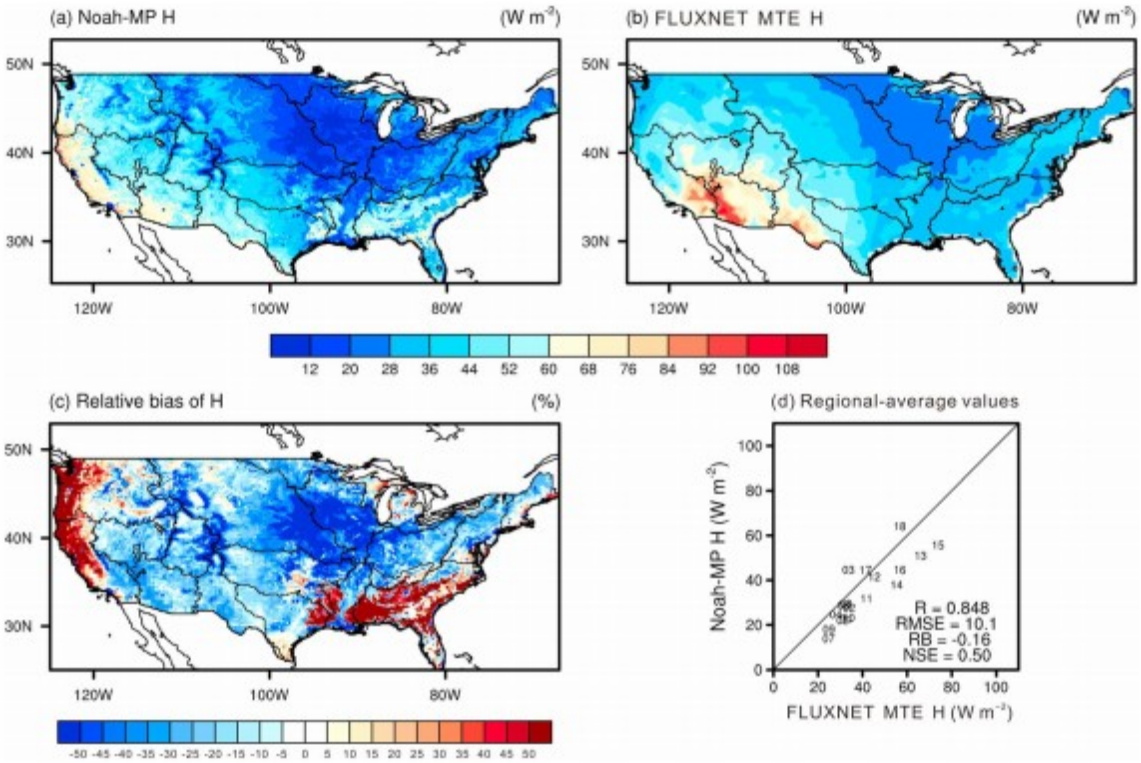

Figure 9

Spatial pattern of the multiyear (1982–2008) mean sensible heat flux (*H*) from (a) Noah-MP, (b) FLUXNET MET, (c) the relative bias [=Noah-MP − FLUXNET MET)/FLUXNET MET × 100%] over CONUS, and (d) the comparison of regional averages over the 18 HUC2 regions shown as their corresponding HUC2 codes.

Similar to GPP and LE, the Noah-MP simulated *H* also shows a larger interannual variability than that of FLUXNET MTE in all the CONUS regions, especially over the central transitional regions (Figure 10). We would, however, draw the attention again that the FLUXNET MTE products have a much lower interannual variability when compared to the FLUXNET tower observations (Jung et al., 2011; Koster et al., 2015). In view of the regional-averaged monthly variation, Noah-MP underestimates *H* throughout the year over most arid regions (e.g., Great Basin, Rio Grande, Upper Colorado, and Lower Colorado) with the NSE values being lower than 0.6. In the majority of the eastern humid and central transitional regions, Noah-MP significantly underestimates *H* in the early growing season to compensate for the overestimated LE (see Figure 7) but shows much less biases in the second half of the year (Figure 10). These model biases are alleviated in EXP_LAI (figures not shown) accordingly compared to the experiment with the dynamic vegetation module.

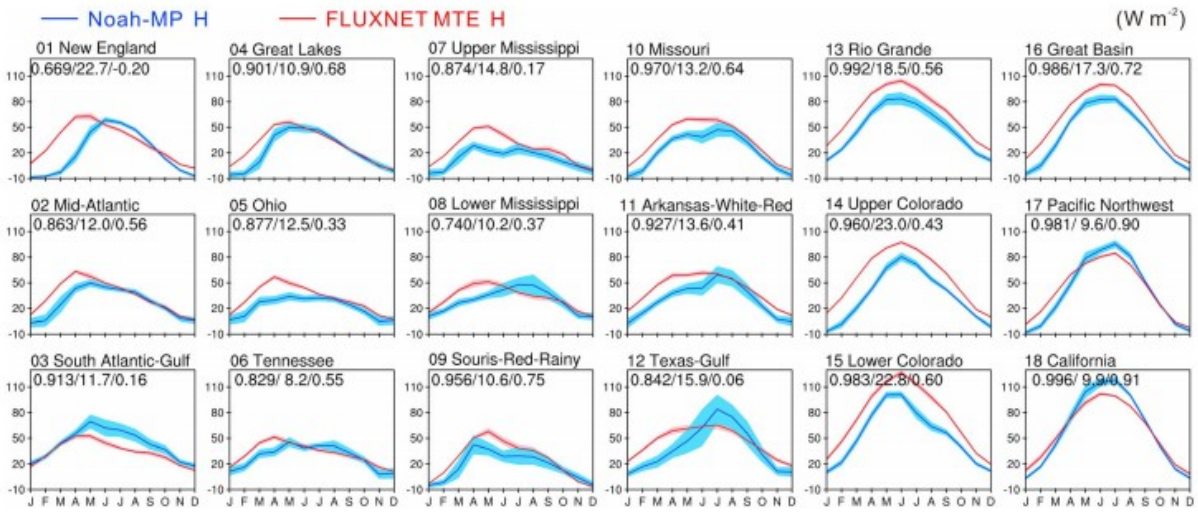

Figure 10

The multiyear (1982–2008) mean annual cycle of sensible heat flux (*H*) from the Noah-MP (blue line) and FLUXNET MTE (red line) over the 18 HUC2 regions. The light blue and red shaded areas represent the standard derivation of the Noah-MP and FLUXNET MTE *H* data, respectively, to indicate the interannual variability for each month of a year. Also shown on the top of each panel are modeling metrics in the order of *R*, RMSE, and NSE.

3.5 Terrestrial Water Storage

The Noah-MP simulated terrestrial water storage is the sum of four model outputs including the groundwater storage, total soil water depth, snow water equivalent, and canopy-intercepted water. In the present study, the modeled monthly TWS anomaly (TWSA) and its components from January 2003 to December 2015 are computed as the deviation from the mean value during the same period with that of GRACE (i.e., January 2004 to December 2009). That is,

$$\mathrm{TWSA} = \Delta\mathrm{GW} + \Delta\mathrm{SW} + \Delta\mathrm{SWE} + \Delta\mathrm{CW}, \quad (1)$$

where ΔGW, ΔSW, ΔSWE, and ΔCW are monthly anomalies of ground water storage, total soil moisture, snow water equivalent, and canopy-intercepted water, respectively. Note that similar to Xia, Cosgrove, Mitchell, Peters-Lidard, Ek, Brewer, et al. (2016) and Xia et al. (2017), the water storage in rivers, lakes, and reservoirs is not explicitly represented in Noah-MP and thus excluded in the present analysis. The modeled monthly TWSA agrees generally well with that of GRACE over all the HUC2 regions with R ranging from 0.659 to 0.934 ($P < 0.001$) (Figure 11). Of the total 18 HUC2 regions, 10 regions have NSE values >0.6 at a monthly scale. The simulated TWSA's amplitude is more comparable to that of GRACE in most regions than the multimodel ensembles mean of the four LSMs used in the NLDAS-2 framework (Xia, Cosgrove, Mitchell, Peters-Lidard, Ek, Brewer, et al., 2016) possibly because of the inclusion of a bucket groundwater (or an unconfined aquifer) below the soil column of a traditional LSM (Niu et al., 2007, 2011). The present result is also supported by Cai, Yang, Xia, et al. (2014), which

shows that Noah-MP performs best in simulating TWSA over CONUS compared to Noah, VIC, and CLM.

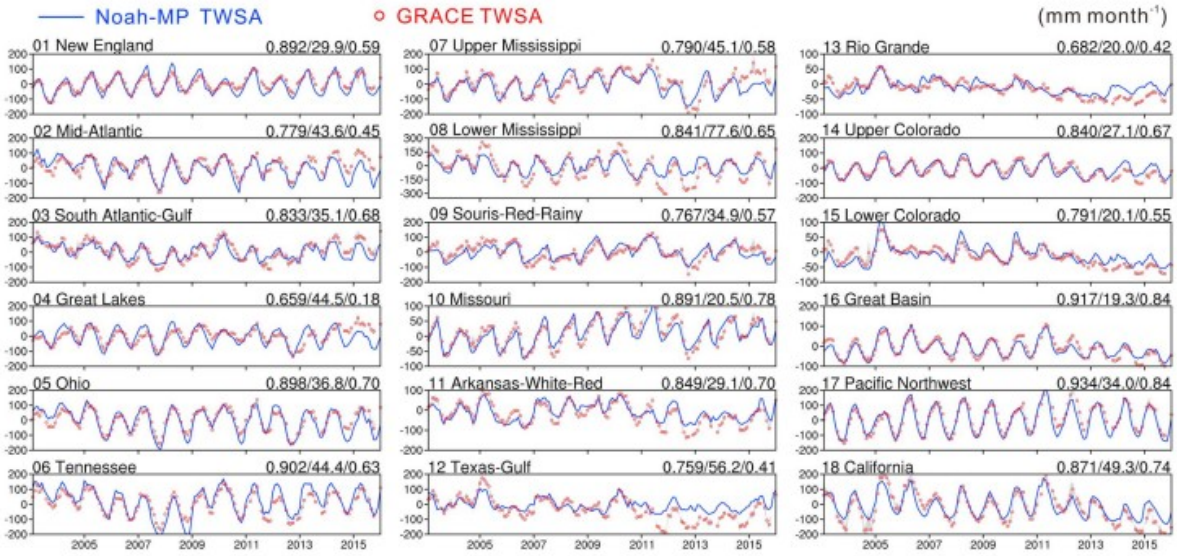

Figure 11

The Noah-MP simulated (blue line) and GRACE-observed (red circle) monthly terrestrial water storage anomaly (TWSA) from 2003 to 2015 over the 18 HUC2 regions. The gray shaded area is the standard deviation of the three GRACE products, representing the uncertainty of the GRACE TWSA values. Also shown on the top of each panel are modeling metrics in the order of *R*, RMSE, and NSE.

Consistent with previously reported droughts in recent years over the western United States (Griffin & Anchukaitis, 2014), Noah-MP is able to reproduce the recent drying trend since 2011 over the Rio Grande, Lower Colorado, Upper Colorado, and California but to a lesser extent (regions “13”, “14”, “15”, and “18” in Figure 11). In the Lower Mississippi, Arkansas-White-Red and Texas-Gulf (regions “08”, “11”, and “12” in Figure 11), Noah-MP also shows shallower summer troughs than those of GRACE in recent years since 2011. Taking the Lower Mississippi region as an example, the mean amplitude of annual TWS variation over the whole 13 year period modeled by Noah-MP (202.0 mm) is 33.9% less than that of GRACE (305.7 mm). Based on the MODIS-derived United States irrigation map (Ozdogan & Gutman, 2008), all the above regions are intensively developed with a large body of groundwater-pumping irrigated agriculture, for example, the Central Valley, High Plains, and Mississippi Embayment. Statistically, the total annual virtual groundwater withdrawal from the aquifers in these three agricultural areas is even close to the capacity of Lake Mead, the largest reservoir of the United States (Marston et al., 2015). In this regard, a primary reason for such discrepancies may be that Noah-MP does not yet incorporate anthropogenic impacts such as agricultural irrigation and impoundments, but these effects have been detected by GRACE. Thus, it is not surprising to witness above remarkable discrepancy in the Lower Mississippi region since it boasts 32,430 km^2 of irrigated farms, accounting for 12% the total regional area, a much higher percentage than any other HUC2 regions (U.S. Department of

Agriculture-National Agricultural Statistics Service, 2013). In fact, recently intensified pumping of “fossil” water from the aquifers has caused abnormally low TWS because of the enhanced rate of water loss to the atmosphere through ET. For example, Castle et al. (2016) pointed out that the groundwater withdrawals have caused at least $22 \pm 5.5\ km^3\ yr^{-1}$ additional water loss via ET in the Colorado River Basin. Additionally, a water budget survey over the Central Valley of California, which is the most productive agricultural area in the United States, documents that annual ET exceeds annual precipitation by ~60% due to irrigation, which leads to an annual average groundwater depletion of $1.85 \times 10^9\ m^3$ (Faunt, 2009). Though challenging, incorporating modules that account for human-induced water use, especially in basins with plenty of heavily managed lands (e.g., irrigation and impoundments), is crucial for more accurate modeling TWS variability across multiple spatial and temporal scales.

Noah-MP also shows a relatively lower ability in simulating TWS over the regions with high percentage of water bodies, for example, the Great Lakes with the lowest NSE of 0.18 (region “04” in Figure 11). The TWSA amplitude is largely impacted by changes in the water level of the water bodies such as rivers, reservoirs and lakes (Cai, Yang, Xia, et al., 2014; Xia, Cosgrove, Mitchell, Peters-Lidard, Ek, Brewer, et al., 2016; Xia et al., 2017). To interpret the discrepancies of the modeled TWSA from the GRACE observed, we selected two regions with relatively larger water bodies, that is, the Great Lakes and the Great Basin. The Michigan Lake is located in the former, while the Great Salt Lake is situated in the latter. Starting from equation 1, we further incorporated the lake water storage variations into the modeled TWSA; that is,

$$TWSA = \Delta GW + \Delta SW + \Delta SWE + \Delta CW + \Delta LW, \quad (2)$$

where ΔLW is the anomaly (relative to the same time period of GRACE) of the lake water storage. We calculated ΔLW using the measured water level and lake area data for the Michigan Lake and the Great Salt Lake, respectively. As seen in Figure 12, the biases of the modeled TWSA are largely reduced with consideration of the lake water storage changes over both regions. This is more apparent for the Great Lakes region, of which the NSE value increases from 0.18 to 0.76 after including the lake water storage of the Michigan Lake. Furthermore, the overall trend of TWSA from 2012 to 2015 is greatly improved in consistence with the observed rise of the water level of the Michigan Lake in recent years (data not shown). Though not as apparent as that in the Great Lakes, the simulated TWSA of the Great Basin also becomes slightly closer to the GRACE measured with a higher NSE value of 0.88 (Figure 12). This suggests that future LSMs should explicitly represent lake, reservoir, and river water storage dynamics. Moreover, assimilating satellite altimetry data (e.g., ICESat and future ICESat-2) of water level

change will be of help for monitoring the changes in water storage of the terrestrial water bodies and total TWS at a regional scale.

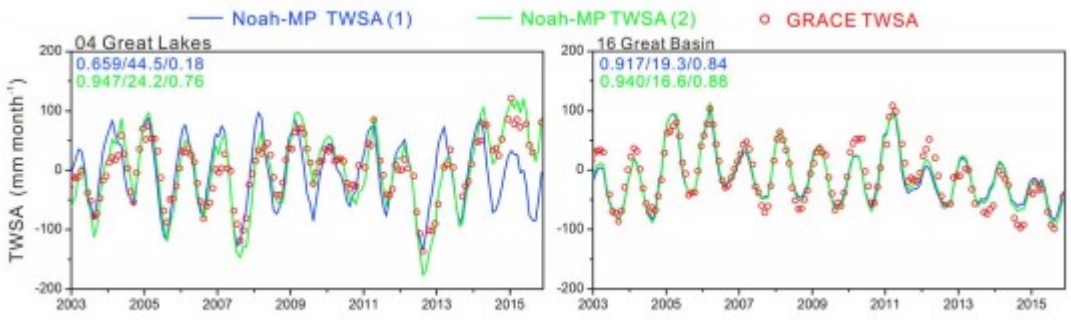

Figure 12

Comparison of the Noah-MP simulated (blue and green lines) and the GRACE-observed (red circles) monthly terrestrial water storage anomaly (TWSA) in the Great Lakes and the Great Basin during 2003–2015. Noah-MP TWSA (1) (blue) are the results from Noah-MP without consideration of lake water storage, while Noah-MP TWSA (2) (green) are the results from Noah-MP with added lake water storage. The numbers in the upper left corner of each panel are *R*, RMSE, and NSE, respectively.

3.6 Snow Cover Fraction

Noah-MP separately simulates SCF on the ground (SCF_g) following Niu and Yang (2007) and that on the vegetation canopy (SCF_v) following Deardorff (1978). Considering the shading effects of the canopy, we calculated the grid SCF as the average of SCF_v and SCF_g weighted by the green vegetation fraction (GVF) and the bare ground fraction (1–GVF), respectively: $SCF = SCF_v \times GVF + SCF_g \times (1 - GVF)$, for comparing to the MODIS SCF product. Similar to Demaria et al. (2016), we specifically divided the snow season into three phases, that is, buildup (November to December), stable accumulation (January to February), and melting (March to April) phases. Noah-MP yields a geographical distribution of snow cover in consistence with that of MODIS during each phase (Figure 13), particularly in the accumulation phase. The spatial pattern of both Noah-MP-simulated and MODIS-derived SCF follows the latitudinal gradient in the eastern and central CONUS but becomes more dependent on altitudinal gradient over the Rocky Mountains in the west. Averaged over the 18 HUC2 regions, the RB is only 0.04 in the stable accumulation phase but increases to 0.14 during the melting phase (Figures 13c, 13f, and 13i). NSE is as high as 0.97 in the stable accumulation phase, while it slightly decreases in the other two phases mainly because of the degraded model performance over the New England region (“01” in Figure 13) where SCF is remarkably overestimated.

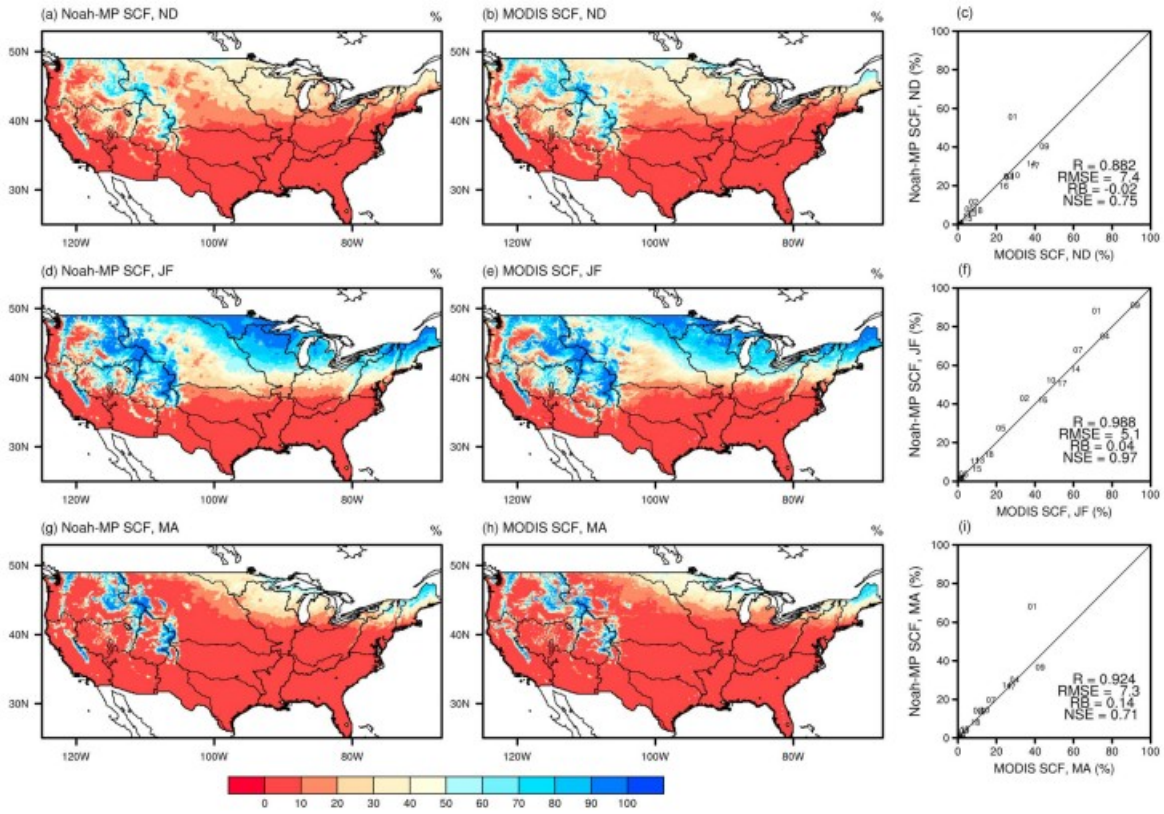

Figure 13

The spatial patterns of the Noah-MP simulated and the MODIS-derived multiyear mean (2004–2015) snow cover fraction (SCF) and the regional averages of the 18 HUC2 regions (shown as their corresponding HUC2 codes) over CONUS during the snow buildup ((a–c) November and December), stable accumulation ((d–f) January and February), and melting ((g–i) March and April) phases.

Regarding the 12 snow-covered regions in the western and northern CONUS, 11 of them (except the New England) have NSE values >0.8 and RMSE <6% with respect to the interannual variability of modeled SCF (Figure 14). The high efficiency in simulating SCF, in-line with other studies of modeling SWE and snow depth using Noah-MP at local (Chen et al., 2014; Niu et al., 2011; Wrzesien et al., 2015) and basin scales (Yang et al., 2011), demonstrates a good skill of Noah-MP in snow process modeling. This may be attributed to the introduction of (i) a SCF scheme that considers the hysteresis of the SCF-snow depth relationship for accumulation and melting phases (Niu & Yang, 2007), (ii) a module of snow interception by the vegetation canopy (Niu & Yang, 2004), and (iii) a multilayer structure for the snowpack on the ground that facilitates an accurate prediction of snow skin temperature, liquid water retention and refreezing within the snowpack, and snow densification (Niu et al., 2011).

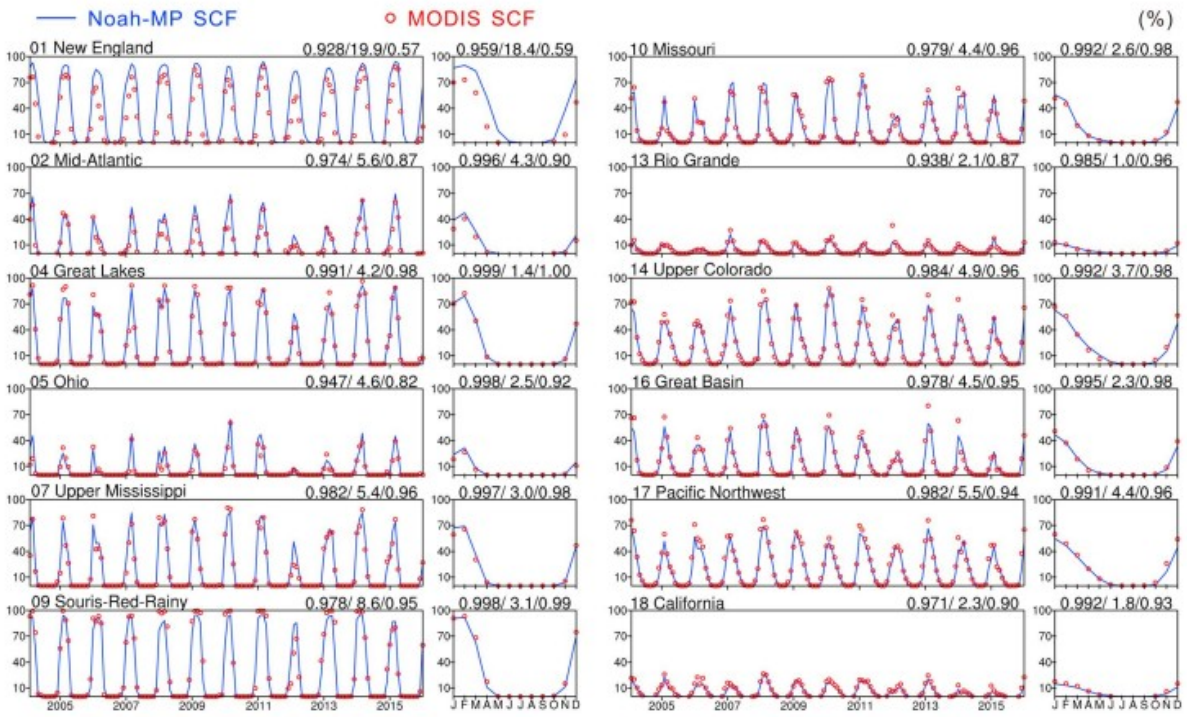

Figure 14

The Noah-MP simulated (blue line) and the MODIS-derived (red circles) monthly snow cover fraction (SCF) and their multiyear mean annual cycle during 2004–2015 over 12 selected HUC2 regions in the northern and western CONUS. Also shown on the top of each panel are modeling metrics in the order of *R*, RMSE, and NSE.

Noah-MP's excessive SCF in the New England during all phases (region “01” in Figures 13 and 14) may be due partially to underestimating sublimation (i.e., lower modeled latent heat flux in cold season as seen in Figure 7). In addition, other causes to the significant SCF bias may be twofold. The first may stem from the possible NLDAS cold bias (Arsenault et al., 2014), leading to excessive partitioning of precipitation into snowfall (controlled by the surface air temperature) and delayed snowmelt in the spring. The second is due likely to the spatial resolution of the present simulation (0.125°), which might be too coarse to resolve the subgrid heterogeneity of snow cover, as has been highlighted by Toure et al. (2016) in a recent evaluation of SCF simulated by CLM.

3.7 Runoff

While previous studies have used a simple river routing model to route the surface and subsurface flows (Cai, Yang, Xia, et al., 2014; Xia, Cosgrove, Mitchell, Peters-Lidard, Ek, Brewer, et al., 2016), one cannot neglect the uncertainties in river routing models due to the model parameters and unrealistically low topographic relief derived from the course-resolution DEM. In fact, the present study is intended to make use of the USGS runoff data at a HUC2 regional scale by evaluating the simulated runoff on a monthly scale, rather than daily or hourly scales for modeling flash floods. The residence

time of water in the world's largest rivers (e.g., the Amazon River and the Mississippi River) is about 2 weeks, and the area of the HUC2 regions ensures the residence time to be much smaller than half a month. As such, we directly calculated the regional runoff by integrating the runoff outputs at each grid over all grids in a region. There are 13 regions of which NSE values are >0.6 at a monthly scale during 1982–2008 (Figure 14), suggesting that Noah-MP is able to capture the seasonality and magnitude of runoff over most of the HUC2 regions. In general, both Noah-MP and USGS show relatively larger (smaller) runoff over the humid (arid) regions. The model performs best in the South Atlantic-Gulf with an NSE value of 0.91, followed by the Lower Mississippi, Great Basin, and Arkansas-White-Red, of which all the NSE values exceed 0.8 at a monthly scale.

However, the modeled runoff does not match well the USGS results over the New England, Rio Grande, and Lower Colorado with even negative NSE values (regions “01”, “13”, and “15” in Figure 14). In the New England, Noah-MP underestimates runoff by roughly −40% with a delayed peak by 1 month. Meanwhile, the modeled ET shows very little bias in this region (see region “01” in Figure 7). Therefore, it appears that the precipitation forcing of the NDLAS-2 is presumably underestimated over this region (climatologically, precipitation is roughly equal to the sum of runoff and ET). The delayed runoff peak is due mainly to the delayed snowmelt (Figure 15), which is likely caused by the potential cold bias in the temperature forcing of NLDAS-2 (Arsenault et al., 2014). In the Rio Grande and Lower Colorado, Noah-MP overestimates runoff by roughly two times throughout the year. Because the modeled ET values are also higher than the observed values (see regions 13 and 15 in Figure 7), the precipitation forcing of NLDAS-2 may be overestimated in these two regions. In addition to the possible error of the forcing, uncertainties in the model structure and parameters may be also responsible for low NSE values in these regions. As mentioned above, the two key runoff parameters in Noah-MP, that is, the maximum groundwater discharge rate (~5 mm s^{-1}) and the saturated hydraulic conductivity decay factor (=6.0), are spatially constant over the whole CONUS domain in this study. A thorough calibration of the model parameters including not only the runoff parameters but also snow, soil, and vegetation properties will certainly improve the runoff simulations (cf. Cai, Yang, David, et al., 2014; Cuntz et al., 2016).

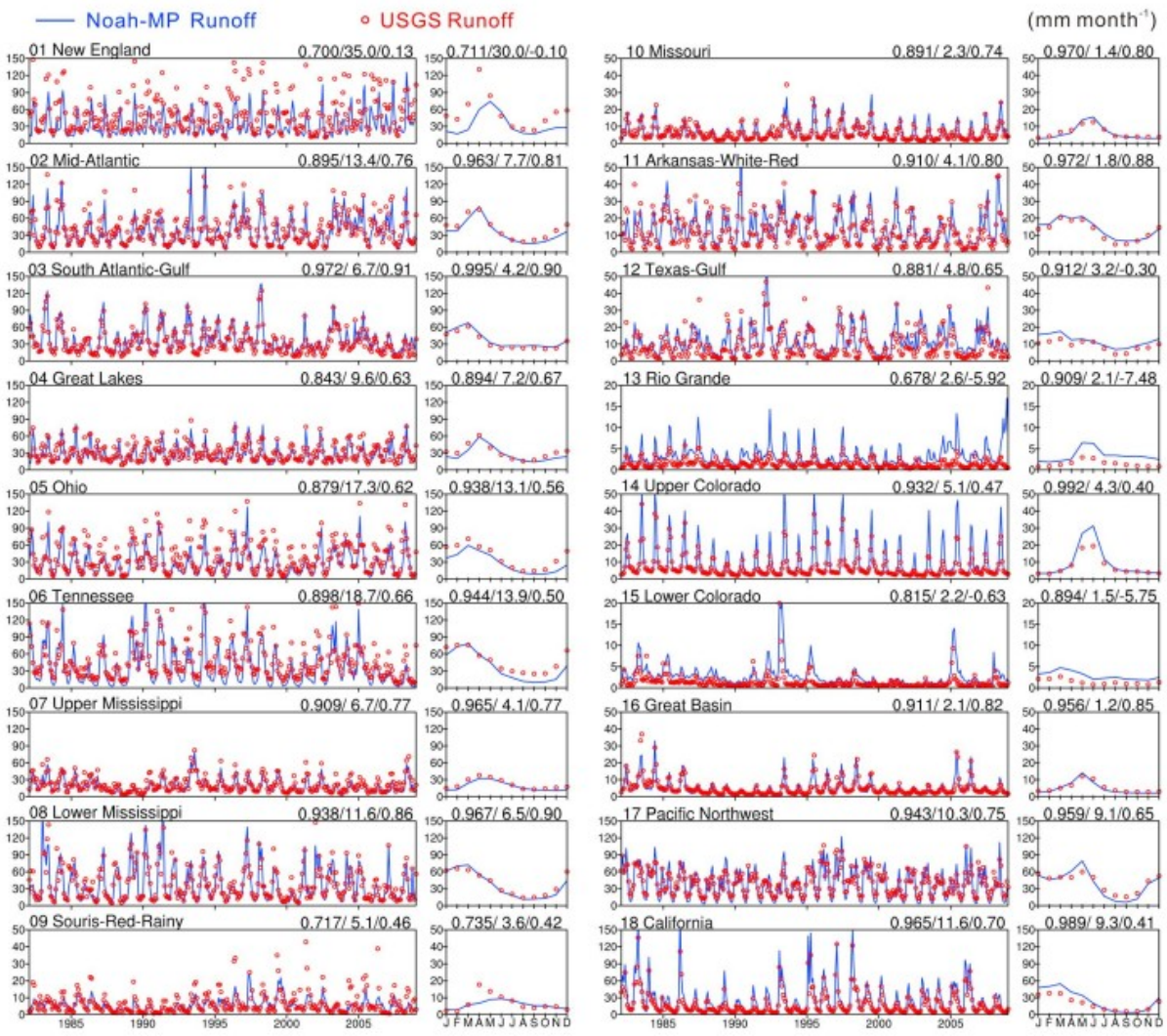

Figure 15

The Noah-MP simulated (blue line) and the USGS measured (red circle) monthly runoff and their multiyear mean annual cycle during 1982–2008 over the 18 HUC2 regions. Also shown on the top of each panel are modeling metrics in the order of *R*, RMSE, and NSE.

3.8 Comparisons Between Noah-MP and the NLDAS-2 LSMs

We also compared Noah-MP to the NLDAS-2 LSMs, that is, Mosaic, VIC, SAC, and Noah and their MME (Xia, Mitchell, Ek, Sheffield, et al., 2012). Using the same reference data products, we computed the NSE values of the monthly outputs of R_n during 1984–2007, ET and runoff during 1982–2008 for four NLDAS-2 LSMs, MME, and Noah-MP over the 18 HUC2 regions (Figure 16). Note that the Noah-MP simulation results of ET in the prescribed vegetation experiment (EXP_LAI) are also included.

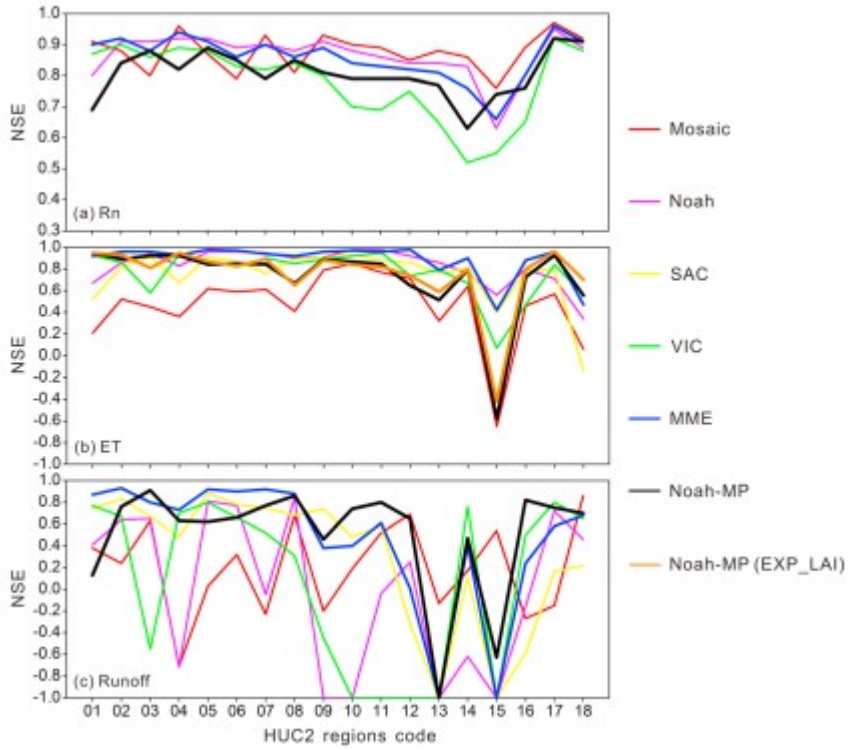

Figure 16

The Nash-Sutcliffe efficiency (NSE) of the simulated (a) R_n, (b) ET, and (c) runoff for the NLDAS-2 LSMs (Mosaic, Noah, SAC, and VIC), their multimodel ensembles mean (MME) and Noah-MP in the 18 HUC2 regions. The NSE values for ET from the prescribed vegetation experiment (EXP_LAI) of Noah-MP are also shown. Note that NSE values are calculated using the 288 (324) monthly model outputs and corresponding reference products during 1984–2007 (1982–2008) for R_n (ET and runoff).

The NSE values of the modeled R_n resulting from the various LSMs show a more considerable scatter over the western arid regions, suggesting a greater challenge for modeling surface albedo that is more affected by the coupled dynamics of vegetation, snow cover, and soil moisture over these regions. Noah-MP performs worse than Noah and MME but better than VIC more obviously in the western arid regions (Figure 16a). However, it should be noted that the low-resolution NASA/GEWEX SRB radiation product tends to bear large uncertainties since the albedo values in the cold season are unreasonably low in the high-elevation and high-latitude regions (see Figure S2). Hence, we argue that further assessment of the models needs more reliable, high-resolution reference data products before further developments of model representations of the coupled dynamics of vegetation and snow cover.

In terms of ET, Noah-MP shows a relatively worse performance than Noah, VIC, SAC, and MME but better than Mosaic over the majority of the HUC2 regions (Figure 16b). All the models fail to simulate ET over the Lower Colorado (region “15”). As mentioned previously, such an insufficient skill of Noah-MP in modeling ET is likely attributed to the deficiency in its dynamic vegetation model. Compared to the experiment with dynamic vegetation

model, EXP_LAI (with prescribed LAI) improved the simulation with higher NSE values over 14 HUC2 regions. However, it is uncertain how much the NSE values of these models would vary after taking into consideration of the unrealistically low interannual variability in the FLUXNET MTE products.

The NSE values of the modeled runoff by all the models show a much greater scatter than $R_n$and ET, reflecting a greater uncertainty in representing runoff processes in LSMs. Noah-MP performs better than the NLDAS-2 LSMs in simulating runoff over the majority of the HUC2 regions except the New England, Mid-Atlantic, Ohio, Tennessee, and Upper Mississippi regions. This is consistent with Xia, Cosgrove, Mitchell, Peters-Lidard, Ek, Brewer, et al. (2016), which demonstrates that Noah-MP improved the runoff modeling in terms of timing and magnitude over the majority of area within CONUS due to the advanced schemes of snow, runoff, and surface energy exchange processes when compared to the research version of Noah, Noah-I. Because the USGS runoff data, which are derived from the river discharge measured with streamflow gauges, may be the most reliable with the lowest uncertainty among all the reference data products, the evaluation of the modeled runoff is regarded as the most stringent in this study.

4 Conclusions

In this study, we evaluated the Noah-MP's performance in simulating the major land-atmosphere energy, water, and carbon exchanging fluxes and runoff, TWSA, and SCF against a wide variety of ground- and satellite-based data sets over the 18 HUC2 regions of CONUS across multiple hydroclimatic regimes. The results are summarized as below:

1. Noah-MP performs very well in simulating R_n over the southeastern CONUS with a smaller relative bias within ±10%, while it yields much smaller R_n than does the NASA/GEWEX SRB with a large relative bias of approximately −40% in the Rocky Mountains and northeastern CONUS with substantial snow cover in cold seasons. However, the comparison of cold season albedo between the model and GEWEX/NASA SRB indicates that Noah-MP's R_n may be more plausible over such regions. Comparison of the modeled SCF to the MODIS SCF implies that the modeled surface albedo controlled by the coupled dynamics of vegetation and snow cover is favorable, although further elaborate assessments of Noah-MP against more reliable, high-resolution data, for example, MODIS surface albedo and LAI products are still needed.
2. Compared to the FLUXNET MTE GPP, ET, and H products, Noah-MP simulates a much larger GPP over the central and eastern CONUS with a relative bias being about 50%, resulting in excessive ET and thus less H, especially in spring and early summer over the same regions. Without the dynamic vegetation model, Noah-MP yields more accurate ET with a relative bias dropping from 22% to 4% over the whole CONUS domain. This suggests that the dynamic vegetation model in Noah-MP should be improved to rigorously represent the carbon partitioning into shoot and

root, root dynamics, and the feedbacks to photosynthesis. Progresses in refining carbon allocation schemes (Gim et al., 2017), applying the temperature limitation on photosynthesis (Schaefer et al., 2012), and introducing the nitrogen limitation (Cai et al., 2016; Stöckli et al., 2008) will facilitate simulations of both carbon and water fluxes across multiple scales.

3. Noah-MP is capable of reproducing the monthly TWSA over most of 18 HUC2 regions (10 regions have NSE values >0.6 at a monthly scale) except those severely affected by either anthropogenic activities including irrigation and impoundments or significant water storage changes over water bodies. Adding the lake water storage changes of the Michigan Lake with an area of 57,756 km^2 to the simulated TWSA reduces the modeling biases in the Great Lakes region to a large extent. Such an improvement is also true over the Great Basin region with consideration of the effect of the Great Salt Lake, though at a lesser extent because of its smaller water surface area. Therefore, incorporating modules describing human-induced disturbances and lake/reservoir water dynamics into Noah-MP would definitely improve its ability in hydrological applications over heavily managed river basins.

4. The modeled SCF agrees well with that of the MODIS derived with relative biases varying from 4% in the accumulation phase to 14% in the melting phase. The present study is consistent with previously reported good performances of Noah-MP in snow depth and SWE evaluations (Chen et al., 2014; Niu et al., 2011; Yang et al., 2011). This is likely attributed to the more sophisticated schemes of SCF (Niu & Yang, 2007), snow interception by the vegetation canopy (Niu & Yang, 2004), and the multilayer structure for the snowpack on the ground (Niu et al., 2011).

5. Despite the spatially constant runoff parameters, the Noah-MP simulated runoff generally agrees with the USGS-measured results in both timing and magnitude with NSE values of 13 HUC2 regions being >0.6 during 1982–2008 at a monthly scale. Though beyond the scope of this study, a thorough calibration of model parameters (distributed or lumped) including not only the runoff parameters but also snow, soil, and vegetation properties for a specific river basin will further improve the runoff simulations.

6. Noah-MP shows a better performance in runoff simulations when compared to the NLDAS-2 LSMs over most of the HUC2 regions. Meanwhile, its ability to model R_n is comparable to the NLDAS-2 LSMs regardless of the uncertainties in the referencing data. However, the NSE values of the modeled ET by Noah-MP are relatively lower than most NLDAS-2 LSMs. Such a deficiency highlights an urgent need of refinements of the dynamic vegetation model in Noah-MP despite the large uncertainties in the FLUXNET MTE products with respect to its interannual variability. Because the USGS runoff data may be the most

reliable with the lowest uncertainty among all the reference data products, the evaluation of the modeled runoff is regarded as the most stringent in this study.

One of the most prominent merits of Noah-MP, compared to other NLDAS-2 LSMs, is that the former provides multiple options of parameterization schemes for key physical processes (Niu et al., 2011). However, the main purpose of this study is not to seek the best combinations of schemes or to improve the hydroclimatological applications over CONUS through physically based ensemble simulations; rather, we intended to investigate the model's strengths and weaknesses, thereby providing guidance for future model developments in representing key terrestrial ecohydrological processes that may be missing in the current LSMs. Though the combination of the schemes used in this study may not be the best, for the aim of identifying the model weaknesses, we included the vegetation dynamic model, which is not well developed in not only Noah-MP but also other LSMs. The present study suggests that future model developments should focus more on improving the representations of vegetation dynamics, lake water storage dynamics, and human activities including irrigation and impoundments.

Acknowledgments

This research was jointly funded by the National Key Research and Development Program of China (2017YFA0603101), China Postdoctoral Science Foundation (2017LH032 and 2017M620069), National Natural Science Foundation of China (41661144025 and 41430748), NASA MAP Program (80NSSC17K0352), and the University of Arizona Germinating Research Program Success: Faculty Seed Grants. We are grateful to the Editor and anonymous reviewers for their comments that greatly improved the earlier manuscript. All data used in this study can be freely accessed from the cited websites described in the text. We appreciate all organizations and persons for making such data available to the public. Noah-MP simulation outputs are available from the corresponding authors.

References

Anav, A., Friedlingstein, P., Beer, C., Ciais, P., Harper, A., Jones, C., … Papale, D. (2015). Spatiotemporal patterns of terrestrial gross primary production: A review. *Reviews of Geophysics*, 53, 785– 818. https://doi.org/10.1002/2015RG000483

Arsenault, K. R., Houser, P. R., & De Lannoy, G. J. M. (2014). Evaluation of the MODIS snow cover fraction product. *Hydrological Processes*, 28(3), 980– 998. https://doi.org/10.1002/hyp.9636

Ashfaq, M., Ghosh, S., Kao, S.-C., Bowling, L. C., Mote, P., Touma, D., … Diffenbaugh, N. S. (2013). Near-term acceleration of hydroclimatic change in the western U.S. *Journal of Geophysical Research: Atmospheres*, 118, 10,676– 10,693. https://doi.org/10.1002/jgrd.50816

Barlage, M., Tewari, M., Chen, F., Miguez-Macho, G., Yang, Z.-L., & Niu, G.-Y. (2015). The effect of groundwater interaction in North American regional climate simulations with WRF/Noah-MP. *Climatic Change*, 129(3–4), 485– 498. https://doi.org/10.1007/s10584-014-1308-8

Best, M. J., Abramowitz, G., Johnson, H. R., Pitman, A. J., Balsamo, G., Boone, A., … Decharme, B. (2015). The plumbing of land surface models: Benchmarking model performance. *Journal of Hydrometeorology*, 16(3), 1425– 1442. https://doi.org/10.1175/JHM-D-14-0158.1

Bonan, G. B., Lawrence, P. J., Oleson, K. W., Levis, S., Jung, M., Reichstein, M., … Swenson, S. C. (2011). Improving canopy processes in the Community Land Model version 4 (CLM4) using global flux fields empirically inferred from FLUXNET data. *Journal of Geophysical Research*, 116, G02014. https://doi.org/10.1029/2010JG001593

Brakebill, J., Wolock, D., & Terziotti, S. (2011). Digital hydrologic networks supporting applications related to spatially referenced regression modeling. *Journal of the American Water Resources Association*, 47(5), 916– 932. https://doi.org/10.1111/j.1752-1688.2011.00578.x

Cai, X., Yang, Z.-L., David, C. H., Niu, G.-Y., & Rodell, M. (2014). Hydrological evaluation of the Noah-MP land surface model for the Mississippi River Basin. *Journal of Geophysical Research: Atmospheres*, 119, 23– 38. https://doi.org/10.1002/2013JD020792

Cai, X., Yang, Z. L., Fisher, J. B., Zhang, X., Barlage, M., & Chen, F. (2016). Integration of nitrogen dynamics into the Noah-MP land surface model v1.1 for climate and environmental predictions. *Geoscientific Model Development*, 9(1), 1– 15. https://doi.org/10.5194/gmd-9-1-2016

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

Castle, S. L., Reager, J. T., Thomas, B. F., Purdy, A. J., Lo, M.-H., Famiglietti, J. S., & Tang, Q. (2016). Remote detection of water management impacts on evapotranspiration in the Colorado River Basin. *Geophysical Research Letters*, 43, 5089– 5097. https://doi.org/10.1002/2016GL068675

Chen, F., Barlage, M., Tewari, M., Rasmussen, R., Jin, J., Lettenmaier, D., … Lin, C. (2014). Modeling seasonal snowpack evolution in the complex terrain and forested Colorado Headwaters region: A model intercomparison study. *Journal of Geophysical Research: Atmospheres*, 119, 13,795– 13,819. https://doi.org/10.1002/2014JD022167

Chen, F., & Dudhia, J. (2001). Coupling an advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Monthly Weather Review*, 129(4), 569– 585. https://doi.org/10.1175/1520-0493(2001)129%3C0569:CAALSH%3E2.0.CO;2

Chen, F., Mitchell, K., Schaake, J., Xue, Y., Pan, H., Koren, V., … Betts, A. (1996). Modeling of land-surface evaporation by four schemes and comparison with FIFE observations. *Journal of Geophysical Research*, 101(D3), 7251– 7268. https://doi.org/10.1029/95JD02165

Clapp, R. B., & Hornberger, G. M. (1978). Empirical equations for some soil hydraulic properties. *Water Resources Research*, 14(4), 601– 604. https://doi.org/10.1029/WR014i004p00601

Clark, M. P., Bierkens, M. F. P., Samaniego, L., Woods, R. A., Uijlenhoet, R., Bennett, K. E., … Peters-Lidard, C. D. (2017). The evolution of process-based hydrologic models: Historical challenges and the collective quest for physical realism. *Hydrology and Earth System Sciences*, 21(7), 3427– 3440. https://doi.org/10.5194/hess-21-3427-2017

Clark, M. P., Fan, Y., Lawrence, D. M., Adam, J. C., Bolster, D., Gochis, D. J., … Kumar, M. (2015). Improving the representation of hydrologic processes in Earth system models. *Water Resources Research*, 51, 5929– 5956. https://doi.org/10.1002/2015WR017096

Collatz, G. J., Ball, J. T., Grivet, C., & Berry, J. A. (1991). Physiological and environmental regulation of stomatal conductance, photosynthesis and transpiration: A model that includes a laminar boundary layer. *Agricultural and Forest Meteorology*, 54(2-4), 107– 136. https://doi.org/10.1016/0168-1923(91)90002-8

Collatz, G. J., Ribas-Carbo, M., & Berry, J. A. (1992). Coupled photosynthesis-stomata1 conductance model for leaves of C4 plants. *Australian Journal of Plant Physiology*, 19(5), 519– 538. https://doi.org/10.1071/PP9920519

Cosgrove, B., Gochis, D. J., Clark, E., Cui, Z., Dugger, A., Feng, X., … Khan, S. (2017). Continental-scale operational hydrologic modeling: Version 1.0 of the National Water Model. Paper presented at 97th American Meteorological Society Annual Meeting, American Meteorological Society, Seattle, WA. Retrieved from https://ams.confex.com/ams/97Annual/webprogram/Paper314045.html

Cuntz, M., Mai, J., Samaniego, L., Clark, M., Wulfmeyer, V., Branch, O., … Thober, S. (2016). The impact of standard and hard-coded parameters on the hydrologic fluxes in the Noah-MP land surface model. *Journal of Geophysical Research: Atmospheres*, 121, 10,676– 10,700. https://doi.org/10.1002/2016JD025097

Deardorff, J. W. (1978). Efficient prediction of ground surface temperature and moisture, with inclusion of a layer of vegetation. *Journal of Geophysical Research*, 83(C4), 1889– 1903. https://doi.org/10.1029/JC083iC04p01889

Demaria, E. M. C., Roundy, J. K., Wi, S., & Palmer, R. N. (2016). The effects of climate change on seasonal snowpack and the hydrology of the northeastern and upper Midwest United States. *Journal of Climate*, 29(18), 6527– 6541. https://doi.org/10.1175/JCLI-D-15-0632.1

Dickinson, R. E., Shaikh, M., Bryant, R., & Graumlich, L. (1998). Interactive canopies for a climate model. *Journal of Climate*, 11(11), 2823– 2836. https://doi.org/10.1175/1520-0442(1998)011%3C2823:ICFACM%3E2.0.CO;2

Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., … Tarpley, J. D. (2003). Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *Journal of Geophysical Research*, 108(D22), 8851. https://doi.org/10.1029/2002JD003296

Farquhar, G. D., von Caemmerer, S., & Berry, J. A. (1980). A biochemical model of photosynthetic CO2 assimilation in leaves of C3 species. *Planta*, 149(1), 78– 90. https://doi.org/10.1007/BF00386231

C. C. Faunt (Ed.) (2009). *Groundwater availability of the Central Valley aquifer, California*. *US Geological Survey Professional Paper, 1766*, 225.

Frankenberg, C., Fisher, J. B., Worden, J., Badgley, G., Saatchi, S. S., Lee, J.-E., … But, A. (2011). New global observations of the terrestrial carbon cycle from GOSAT: Patterns of plant fluorescence with gross primary productivity. *Geophysical Research Letters*, 38, L17706. https://doi.org/10.1029/2011GL048738

Fu, Q., & Feng, S. (2014). Responses of terrestrial aridity to global warming. *Journal of Geophysical Research: Atmospheres*, 119, 7863– 7875. https://doi.org/10.1002/2014JD021608

Gao, Y., Li, K., Chen, F., Jiang, Y., & Lu, C. (2015). Assessing and improving Noah-MP land model simulations for the central Tibetan Plateau. *Journal of Geophysical Research: Atmospheres*, 120, 9258– 9278. https://doi.org/10.1002/2015JD023404

Gayler, S., Wohling, T., Grzeschik, M., Ingwersen, J., Wizemann, H.-D., Warrach-Sagi, K., … Wulfmeyer, V. (2014). Incorporating dynamic root growth enhances the performance of Noah-MP at two contrasting winter wheat field sites. *Water Resources Research*, 50, 1337– 1356. https://doi.org/10.1002/2013WR014634

Gim, H.-J., Park, S. K., Kang, M., Thakuri, B. M., Kim, J., & Ho, C.-H. (2017). An improved parameterization of the allocation of assimilated carbon to plant parts in vegetation dynamics for Noah-MP. *Journal of Advances in Modeling Earth Systems*, 9(4), 1776– 1794. https://doi.org/10.1002/2016MS000890

Griffin, D., & Anchukaitis, K. J. (2014). How unusual is the 2012-2014 California drought? *Geophysical Research Letters*, 41, 9017– 9023. https://doi.org/10.1002/2014GL062433

Gupta, H. V., Perrin, C., Blöschl, G., Montanari, A., Kumar, R., Clark, M., & Andréassian, V. (2014). Large-sample hydrology: A need to balance depth with breadth. *Hydrology and Earth System Sciences*, 18(2), 463– 477. https://doi.org/10.5194/hess-18-463-2014

Guttman, N. B., & Quayle, R. G. (1996). A historical perspective of U.S. climate divisions. *Bulletin of the American Meteorological Society*, 77(2), 293– 303. https://doi.org/10.1175/1520-0477(1996)077%3C0293:AHPOUC%3E2.0.CO;2

Hall, D. K., & Riggs, G. A. (2007). Accuracy assessment of the MODIS snow products. *Hydrological Processes*, 21(12), 1534– 1547. https://doi.org/10.1002/hyp.6715

Jung, M., Reichstein, M., Ciais, P., Seneviratne, S. I., Sheffield, J., Goulden, M. L., … Cescatti, A. (2010). Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Nature*, 467(7318), 951– 954. https://doi.org/10.1038/nature09396

Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Altaf Arain, M., … Bernhofer, C. (2011). Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. *Journal of Geophysical Research*, 116, G00J07. https://doi.org/10.1029/2010JG001566

Koster, R. D., Salvucci, G. D., Rigden, A. J., Jung, M., Collatz, G. J., & Schubert, S. D. (2015). The pattern across the continental United States of evapotranspiration variability associated with water availability. *Frontiers in Earth Science*, 3. https://doi.org/10.3389/feart.2015.00035

Kumar, S. V., Zaitchik, B. F., Peters-Lidard, C. D., Rodell, M., Reichle, R., Li, B., … Mocko, D. (2016). Assimilation of gridded GRACE terrestrial water storage estimates in the North American Land Data Assimilation System. *Journal of Hydrometeorology*, 17(7), 1951– 1972. https://doi.org/10.1175/JHM-D-15-0157.1

Landerer, F. W., & Swenson, S. C. (2012). Accuracy of scaled GRACE terrestrial water storage estimates. *Water Resources Research*, 48, W04531. https://doi.org/10.1029/2011WR011453

Luo, Y. Q., Randerson, J. T., Abramowitz, G., Bacour, C., Blyth, E., Carvalhais, N., … Dalmonech, D. (2012). A framework for benchmarking land models. *Biogeosciences*, 9(10), 3857– 3874. https://doi.org/10.5194/bg-9-3857-2012

Luo, S., Sun, Z., Zheng, X., Rikus, L., & Franklin, C. (2016). Evaluation of ACCESS model cloud properties over the Southern Ocean area using

multiple-satellite products. *Quarterly Journal of the Royal Meteorological Society*, 142(694), 160– 171. https://doi.org/10.1002/qj.2641

Ma, N., Zhang, Y., Xu, C.-Y., & Szilagyi, J. (2015). Modeling actual evapotranspiration with routine meteorological variables in the data-scarce region of the Tibetan Plateau: Comparisons and implications. *Journal of Geophysical Research: Biogeosciences*, 120, 1638– 1657. https://doi.org/10.1002/2015JG003006

Manabe, S. (1969). Climate and ocean circulation. Part I: The atmospheric circulation and the hydrology of the Earth's surface. *Monthly Weather Review*, 97(11), 739– 774. https://doi.org/10.1175/1520-0493(1969)097%3C0739:CATOC%3E2.3.CO;2

Mao, J., Thornton, P. E., Shi, X., Zhao, M., & Post, W. M. (2012). Remote sensing evaluation of CLM4 GPP for the period 2000 to 2009. *Journal of Climate*, 25(15), 5327– 5342. https://doi.org/10.1175/JCLI-D-11-00401.1

Marston, L., Konar, M., Cai, X., & Troy, T. J. (2015). Virtual groundwater transfers from overexploited aquifers in the United States. *Proceedings of the National Academy of Sciences of the United States of America*, 112(28), 8561– 8566. https://doi.org/10.1073/pnas.1500457112

Mathis, M., Elizalde, A., Mikolajewicz, U., & Pohlmann, T. (2015). Variability patterns of the general circulation and sea water temperature in the North Sea. *Progress in Oceanography*, 135, 91– 112. https://doi.org/10.1016/j.pocean.2015.04.009

Mitchell, K. E., Lohmann, D., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., … Sheffield, J. (2004). The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. *Journal of Geophysical Research*, 109, D07S90. https://doi.org/10.1029/2003JD003823

Niu, G.-Y., & Yang, Z.-L. (2004). Effects of vegetation canopy processes on snow surface energy and mass balances. *Journal of Geophysical Research*, 109(D23), D23111. https://doi.org/10.1029/2004JD004884

Niu, G.-Y., & Yang, Z.-L. (2007). An observation-based formulation of snow cover fraction and its evaluation over large North American river basins. *Journal of Geophysical Research*, 112, D21101. https://doi.org/10.1029/2007JD008674

Niu, G.-Y., Yang, Z.-L., Dickinson, R. E., & Gulden, L. E. (2005). A simple TOPMODEL-based runoff parameterization (SIMTOP) for use in global climate models. *Journal of Geophysical Research*, 110, D21106. https://doi.org/10.1029/2005JD006111

Niu, G.-Y., Yang, Z.-L., Dickinson, R. E., Gulden, L. E., & Su, H. (2007). Development of a simple groundwater model for use in climate models and evaluation with Gravity Recovery and Climate Experiment

data. *Journal of Geophysical Research*, 112, D07103. https://doi.org/10.1029/2006JD007522

Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., … Manning, K. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research*, 116, D12109. https://doi.org/10.1029/2010JD015139

Niu, G.-Y., & Zeng, X. (2012). Earth system model, modeling the land component of. In P. J. Rasch (Ed.), *Climate change modeling methodology: Selected entries from the encyclopedia of sustainability science and technology* (pp. 139– 168). New York: Springer. https://doi.org/10.1007/978-1-4614-5767-1_7

Oubeidillah, A. A., Kao, S. C., Ashfaq, M., Naz, B. S., & Tootle, G. (2014). A large-scale, high-resolution hydrological model parameter data set for climate change impact assessment for the conterminous US. *Hydrology and Earth System Sciences*, 18(1), 67– 84. https://doi.org/10.5194/hess-18-67-2014

Ozdogan, M., & Gutman, G. (2008). A new methodology to map irrigated areas using multi-temporal MODIS and ancillary data: An application example in the continental U.S. *Remote Sensing of Environment*, 112(9), 3520– 3537. https://doi.org/10.1016/j.rse.2008.04.010

Penman, H. L. (1948). Natural evaporation from open water, bare soil and grass. *Proceedings of the Royal Society of London, Series A*, 193(1032), 120– 145. https://doi.org/10.1098/rspa.1948.0037

Piao, S., Yin, G., Tan, J., Cheng, L., Huang, M., Li, Y., … Mao, J. (2015). Detection and attribution of vegetation greening trend in China over the last 30 years. *Global Change Biology*, 21(4), 1601– 1609. https://doi.org/10.1111/gcb.12795

Pitman, A. J. (2003). The evolution of, and revolution in, land surface schemes designed for climate models. *International Journal of Climatology*, 23(5), 479– 510. https://doi.org/10.1002/joc.893

Rice, J. S., Emanuel, R. E., & Vose, J. M. (2016). The influence of watershed characteristics on spatial patterns of trends in annual scale streamflow variability in the continental U.S. *Journal of Hydrology*, 540, 850– 860. https://doi.org/10.1016/j.jhydrol.2016.07.006

Riggs, G. A., Hall, D. K., & Salomonson, V. V. (2006). MODIS snow products user guide for collection 5. Retrieved from http://modis-snow-ice.gsfc.nasa.gov/uploads/sug_c5.pdf

Roads, J., Lawford, R., Bainto, E., Berbery, E., Chen, S., Fekete, B., … Grundstein, A. (2003). GCIP water and energy budget synthesis (WEBS). *Journal of Geophysical Research*, 108(D16), 8609. https://doi.org/10.1029/2002JD002583

Sabo, J. L., Sinha, T., Bowling, L. C., Schoups, G. H. W., Wallender, W. W., Campana, M. E., … Fuller, P. L. (2010). Reclaiming freshwater sustainability in the Cadillac Desert. *Proceedings of the National Academy of Sciences of the United States of America*, 107(50), 21,263– 21,269. https://doi.org/10.1073/pnas.1009734108

Sakumura, C., Bettadpur, S., & Bruinsma, S. (2014). Ensemble prediction and intercomparison analysis of GRACE time-variable gravity field models. *Geophysical Research Letters*, 41, 1389– 1397. https://doi.org/10.1002/2013GL058632

Schaefer, K., Schwalm, C. R., Williams, C., Arain, M. A., Barr, A., Chen, J. M., … Dimitrov, D. (2012). A model-data comparison of gross primary productivity: Results from the North American carbon program site synthesis. *Journal of Geophysical Research*, 117, G03010. https://doi.org/10.1029/2012JG001960

Seaber, P. R., Kapinos, F. P., & Knapp, G. L. (1987). Hydrologic unit maps. *U.S. Geological Survey Water-Supply Paper, 2294*.

Sellers, P. J., Dickinson, R. E., Randall, D. A., Betts, A. K., Hall, F. G., Berry, J. A., … Denning, A. S. (1997). Modeling the exchanges of energy, water, and carbon between continents and the atmosphere. *Science*, 275(5299), 502– 509. https://doi.org/10.1126/science.275.5299.502

Seo, K.-W., Wilson, C. R., Famiglietti, J. S., Chen, J. L., & Rodell, M. (2006). Terrestrial water mass load changes from Gravity Recovery and Climate Experiment (GRACE). *Water Resources Research*, 42, W05417. https://doi.org/10.1029/2005WR004255

Smith, S. V., Renwick, W. H., Bartley, J. D., & Buddemeier, R. W. (2002). Distribution and significance of small, artificial water bodies across the United States landscape. *Science of the Total Environment*, 299(1-3), 21– 36. https://doi.org/10.1016/S0048-9697(02)00222-X

Stöckli, R., Lawrence, D. M., Niu, G. Y., Oleson, K. W., Thornton, P. E., Yang, Z. L., … Running, S. W. (2008). Use of FLUXNET in the Community Land Model development. *Journal of Geophysical Research*, 113, G01025. https://doi.org/10.1029/2007JG000562

Toure, A. M., Rodell, M., Yang, Z.-L., Beaudoing, H., Kim, E., Zhang, Y., & Kwon, Y. (2016). Evaluation of the snow simulations from the Community Land Model, version 4 (CLM4). *Journal of Hydrometeorology*, 17(1), 153– 170. https://doi.org/10.1175/JHM-D-14-0165.1

U.S. Department of Agriculture-National Agricultural Statistics Service (2013). Farm and ranch irrigation survey. 2012 Census of

Agriculture. National Agricultural Statistics Service. U.S. Department of Agriculture.

Velpuri, N. M., Senay, G. B., Singh, R. K., Bohms, S., & Verdin, J. P. (2013). A comprehensive evaluation of two MODIS evapotranspiration products over the conterminous United States: Using point and gridded FLUXNET and water balance ET. *Remote Sensing of Environment*, 139, 35– 49. https://doi.org/10.1016/j.rse.2013.07.013

Wrzesien, M. L., Pavelsky, T. M., Kapnick, S. B., Durand, M. T., & Painter, T. H. (2015). Evaluation of snow cover fraction for regional climate simulations in the Sierra Nevada. *International Journal of Climatology*, 35(9), 2472– 2484. https://doi.org/10.1002/joc.4136

Xia, Y., Cosgrove, B. A., Mitchell, K. E., Peters-Lidard, C. D., Ek, M. B., Brewer, M., … Kumar, S. V. (2016). Basin-scale assessment of the land surface water budget in the National Centers for Environmental Prediction operational and research NLDAS-2 systems. *Journal of Geophysical Research: Atmospheres*, 121, 2750– 2779. https://doi.org/10.1002/2015JD023733

Xia, Y., Cosgrove, B. A., Mitchell, K. E., Peters-Lidard, C. D., Ek, M. B., Kumar, S., … Wei, H. (2016). Basin-scale assessment of the land surface energy budget in the National Centers for Environmental Prediction operational and research NLDAS-2 systems. *Journal of Geophysical Research: Atmospheres*, 121, 196– 220. https://doi.org/10.1002/2015JD023889

Xia, Y., Mitchell, K. E., Ek, M. B., Cosgrove, B. A., Sheffield, J., Luo, L., … Wei, H. (2012). Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 2. Validation of model-simulated streamflow. *Journal of Geophysical Research*, 117, D03110. https://doi.org/10.1029/2011JD016051

Xia, Y., Mitchell, K. E., Ek, M. B., Sheffield, J., Cosgrove, B. A., Wood, E., … Alonge, C. (2012). 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*, 117, D03109. https://doi.org/10.1029/2011JD016048

Xia, Y., Mocko, D., Huang, M., Li, B., Rodell, M., Mitchell, K. E., … Ek, M. B. (2017). Comparison and assessment of three advanced land surface models in simulating terrestrial water storage components over the United States. *Journal of Hydrometeorology*, 18(3), 625– 649. https://doi.org/10.1175/JHM-D-16-0112.1

Yan, B., & Dickinson, R. E. (2014). Modeling hydraulic redistribution and ecosystem response to droughts over the Amazon basin using Community Land Model 4.0 (CLM4). *Journal of Geophysical Research: Biogeosciences*, 119, 2130– 2143. https://doi.org/10.1002/2014JG002694

Yang, Z.-L., Niu, G.-Y., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., … Manning, K. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 2. Evaluation over global river basins. *Journal of Geophysical Research*, 116, D12110. https://doi.org/10.1029/2010JD015140

Zeng, Y., Xie, Z., Yu, Y., Liu, S., Wang, L., Zou, J., … Jia, B. (2016). Effects of anthropogenic water regulation and groundwater lateral flow on land processes. *Journal of Advances in Modeling Earth Systems*, 8(3), 1106– 1131. https://doi.org/10.1002/2016MS000646

Zhang, G., Chen, F., & Gan, Y. (2016). Assessing uncertainties in the Noah-MP ensemble simulations of a cropland site during the Tibet Joint International Cooperation program (JICA) field campaign. *Journal of Geophysical Research: Atmospheres*, 121, 9576– 9596. https://doi.org/10.1002/2016JD024928

Zhang, T., Stackhouse, P. W., Gupta, S. K., Cox, S. J., Colleen Mikovitz, J., & Hinkelman, L. M. (2013). The validation of the GEWEX SRB surface shortwave flux data products using BSRN measurements: A systematic quality control, production and application approach. *Journal of Quantitative Spectroscopy & Radiative Transfer*, 122, 127– 140. https://doi.org/10.1016/j.jqsrt.2012.10.004

Zhang, T., Stackhouse, P. W., Gupta, S. K., Cox, S. J., & Mikovitz, J. C. (2015). The validation of the GEWEX SRB surface longwave flux data products using BSRN measurements. *Journal of Quantitative Spectroscopy & Radiative Transfer*, 150, 134– 147. https://doi.org/10.1016/j.jqsrt.2014.07.013

Zheng, D., Van der Velde, R., Su, Z., Wen, J., Booij, M. J., Hoekstra, A. Y., & Wang, X. (2015). Under-canopy turbulence and root water uptake of a Tibetan meadow ecosystem modeled by Noah-MP. *Water Resources Research*, 51, 5735– 5755. https://doi.org/10.1002/2015WR017115