JGR Atmospheres

RESEARCH ARTICLE
10.1029/2020JD033539

Special Section:
Community Earth System Model version 2 (CESM2) Special Collection

# Validation of the Community Land Model Version 5 Over the Contiguous United States (CONUS) Using In Situ and Remote Sensing Data Sets

**Yanyan Cheng[1], Maoyi Huang[1,2], Bowen Zhu[1,3], Gautam Bisht[1], Tian Zhou[1], Ying Liu[1], Fengfei Song[1], and Xiaogang He[4,5]**

[1]Atmospheric Sciences and Global Change Division, Pacific Northwest National Laboratory, Richland, WA, USA, [2]Now at Office of Science and Technology Integration, National Weather Service, National Oceanic and Atmospheric Administration, Silver Spring, MD, USA, [3]State Key Laboratory of Remote Sensing Science, Jointly Sponsored by Beijing Normal University and Institute of Remote Sensing and Digital Earth of Chinese Academy of Sciences, Beijing, China, [4]Water in the West, Woods Institute for the Environment, Stanford University, Stanford, CA, USA, [5]Now at Department of Civil and Environmental Engineering, National University of Singapore, Singapore

**Abstract** The Community Land Model (CLM) is an effective tool to simulate the biophysical and biogeochemical processes and their interactions with the atmosphere. Although CLM version 5 (CLM5) constitutes various updates in these processes, its performance in simulating energy, water and carbon cycles over the contiguous United States (CONUS) at scales which land surface changes and hydrometeorological and hydroclimatological applications are more locally relevant is yet to be assessed. In this study, we conducted three simulations at 0.125° during 1979–2018 over the CONUS using different configurations of CLM, namely CLM5-biogeochemistry (CLM5BGC), CLM4.5BGC, and CLM5-satellite phenology (CLM5SP). We validated and compared their simulations against multiple remote-sensed and in situ data sets. Overall, the parametric and structural updates (e.g., carbon cost for nitrogen uptake, variable soil thickness, dry surface layer) in CLM5 improve its ability in capturing terrestrial biogeochemical dynamics. The low evapotranspiration in CLM5BGC is associated with biases in simulating vegetation phenological characteristics rather than soil water limitations. The mismatch between CLM5BGC-simulated peak leaf area index and reference data can be attributed to CLM5BGC's inability in simulating phenology of trees and grasses. The differences between CLM-simulated irrigation and reference estimates can be attributed to differences between processes represented in models and in reality, and uncertainties in input and validation data sets. Evaluation against observations at small catchments suggest that hydrologic parameters needed to be calibrated to improve simulations of runoff, especially subsurface runoff. Additional efforts are needed to incorporate spatially distributed plant phenology and physiology parameters and regional-specific agricultural management practices (e.g., planting and harvest).

**Key Points:**

- Parametric and structural updates in Community Land Model version 5 (CLM5) improve its ability in capturing terrestrial biogeochemical dynamics
- Low evapotranspiration in CLM5-biogeochemistry can be attributed to biases in simulating plant phenology rather than soil water limitations
- Efforts are needed to improve runoff and incorporate spatially distributed vegetation parameters and agricultural management practices

**Supporting Information:**

- Supporting Information S1

**Correspondence to:**
Y. Cheng,
yanyan.cheng@nus.edu.sg

**Citation:**
Cheng, Y., Huang, M., Zhu, B., Bisht, G., Zhou, T., Liu, Y., et al. (2021). Validation of the Community Land Model version 5 over the contiguous United States (CONUS) using in situ and remote sensing data sets. *Journal of Geophysical Research: Atmospheres*, *126*, e2020JD033539. https://doi.org/10.1029/2020JD033539

Received 16 JUL 2020
Accepted 12 JAN 2021

**Author Contributions:**
**Conceptualization:** Yanyan Cheng, Maoyi Huang
**Data curation:** Yanyan Cheng, Bowen Zhu, Ying Liu
**Formal analysis:** Yanyan Cheng, Xiaogang He
**Funding acquisition:** Maoyi Huang
**Investigation:** Yanyan Cheng, Gautam Bisht, Xiaogang He
**Methodology:** Yanyan Cheng, Maoyi Huang, Gautam Bisht



## 1. Introduction

Land surface processes can alter regional and global climates (Cheruy et al., 2014; Lin et al., 2017; Van Weverberg et al., 2018) through physical exchanges (e.g., surface albedo and radiative forcing, boundary layer profiles of temperature and water vapor, near-surface momentum) (Cox et al., 1999; Crossley et al., 2000; Dickinson, 1983; Mahmood et al., 2014; Pielke et al., 2011; Pitman et al., 2009) and biogeochemical feedbacks (e.g., carbon emissions) (Bonan, 1995; Cox et al., 2000; P. Lawrence et al., 2012; Unger, 2014). For physical exchanges, biases in simulating land surface fluxes and states (e.g., sensible/latent heat partitioning, irrigation) and their interactions with the atmosphere (e.g., soil moisture-temperature feedback, evapotranspiration-temperature feedback) are reported to play an important role in the persistent warm temperature and dry precipitation biases in Earth system model (ESM) simulations over the Central United States (Cheruy et al., 2014; Klein et al., 2006; Morcrette et al., 2018; Qian et al., 2013; Van Weverberg et al., 2018). For biogeochemical feedbacks, land surfaces contribute a large portion of greenhouse gases (e.g., $CO_2$ and $CH_4$), and it is still debatable whether the terrestrial biosphere is a net carbon sink or source (Berenguer et al., 2014; Crossley et al., 2000; Schaphoff et al., 2006). Modeling the physical, biogeochemical,

**Resources:** Yanyan Cheng, Maoyi Huang, Bowen Zhu, Tian Zhou, Ying Liu
**Software:** Yanyan Cheng, Maoyi Huang, Gautam Bisht
**Supervision:** Maoyi Huang
**Validation:** Yanyan Cheng, Maoyi Huang, Bowen Zhu
**Visualization:** Yanyan Cheng, Fengfei Song, Xiaogang He
**Writing – original draft:** Yanyan Cheng
**Writing – review & editing:** Yanyan Cheng, Maoyi Huang, Fengfei Song, Xiaogang He

and ecosystem dynamics of land surface processes is thus crucial for a comprehensive understanding of land-atmosphere interactions (Bonan, 1995; Bonan & Doney, 2018).

The Community Land Model (CLM) is the land component of the Community Earth System Model (CESM) and has been adopted as the land component in the Norwegian Earth System Model (NorESM) and the Euro-Mediterranean Center on Climate Change coupled Earth System model (CMCC-ESM2). These models contribute to the Coupled Model Intercomparison Project Phase 5 (CMIP5) and CMIP6. CLM can simultaneously simulate biogeophysical, biogeochemical, and ecological processes in the terrestrial environment and is an effective tool to quantify and predict the Earth's carbon, water, and energy budgets over a wide range of spatial (e.g., watershed, regional, continental, and global) and temporal (e.g., from half-hourly to decade) scales (Bonan & Doney 2018; G. B. Bonan et al. 2002; Getirana et al., 2014; Haddeland et al., 2011; Rodell et al., 2004). These capabilities are essential for quantifying the role of terrestrial systems in modulating land surface fluxes and their interactions with boundary layer dynamics, convection, cloud formation, and mesoto large-scale circulations in the climate system (Devanand et al., 2020; Koster et al., 2014; Ma et al., 2018; Qian et al., 2013; Yang et al., 2019).

Over the past decades, CLM has been widely used to improve our understanding of terrestrial energy, water, and carbon cycle dynamics and their interactions (Green et al., 2019; Koven et al., 2017; Lei et al., 2014; Li et al., 2015; Mcguire et al., 2018), the impact of land use and land cover change on climate, carbon, water, and extremes (P. Lawrence & Chase, 2010; P. Lawrence et al., 2018b; Mahowald et al., 2016) and many others. CLM has evolved from version 2 to version 5 with enhancement in various model capabilities. Compared to CLM3, CLM3.5 exhibited substantial improvements in simulating energy and water cycle dynamics, such as partitioning of the ET components (i.e., transpiration, canopy evaporation, and soil evaporation), total water storage and vegetation biogeography (Oleson et al., 2008). An additional biogeochemical model which couples carbon and nitrogen cycles with biophysical, urban, and watershed processes leads to improved performance in simulating snow, soil temperature, river discharge, and surface albedo in CLM4 and CLM4.5 (D. Lawrence et al., 2011). The latest version of CLM (CLM5) has been augmented continuously in several ways driven by various scientific topics, such as the need for an improved understanding of terrestrial energy, carbon, and nitrogen cycle dynamics and a better assessment for the response of terrestrial ecosystems to land use/land cover change and climate change (D. Lawrence et al., 2019). In addition, CLM5 is the first version of CLM that includes transient representation of managed agriculture (e.g., time-varying irrigation) (D. Lawrence et al., 2019). Due to the profound impacts of agricultural management practices (e.g., irrigation, cover crop) on climate (e.g., summer heat extremes and winter warming) (Alter et al., 2018; Bagley et al., 2015; Davin et al., 2014; Lombardozzi et al., 2018; Mueller et al., 2017; Thiery et al., 2017), the new representation of agriculture management in CLM5 could affect energy, water, and carbon fluxes from the land surface and feedback to regional and global climates (D. Lawrence et al., 2019).

With increased complexity of parameterizations in land surface models (LSMs), comprehensive validation of LSMs in various aspects become more and more important to guide future directions of model development. For example, a groundwater-focused multimodel comparison study highlighted the importance of better representations of subsurface hydrological processes in LSMs, including CLM4 (Rashid et al., 2019). Swenson and Lawrence (2015) found that water storage dynamics were sensitive to soil layer depth in CLM4.5. Findings from this and other validation studies (e.g., Decker & Zeng, 2009; Gochis et al., 2010; Gulden et al., 2007) inspire a series of follow-up model development. For instance, a spatially explicit soil thickness product (Pelletier et al., 2016) has been implemented in CLM5 to replace uniform soil layers, which significantly improved water and energy simulations.

Although CLM5-simulated global energy, water, and carbon budgets have been benchmarked against metrics included in the International Land Model Benchmarking (ILAMB) system (Collier et al., 2018; D. Lawrence et al., 2019), CLM5 has not yet been systematically evaluated over the conterminous United States (CONUS), especially at a high spatial resolution (e.g., 0.125°), at which scales land surface heterogeneity and certain physical processes can be appropriately represented. For example, the central US has been identified as a hotspot for warm-season land-atmosphere coupling (Devanand et al., 2020; Koster et al., 2014; Mei & Wang, 2012), especially over human-modified landscapes (e.g., irrigated and rainfed croplands). To alleviate the warm-dry biases in current weather and climate model simulations over this region, fine resolution simulations (e.g., 0.125°) are required to adequately capture changes in land surface fluxes and properties (e.g.,

spatial variabilities of irrigation operations and agricultural system expansions) as well as impacts of land surface processes on convective-scale dynamics (e.g., development of summertime mesoscale convective systems) (Cheruy et al., 2014; Devanand et al., 2020; Klein et al., 2006; Van Weverberg et al., 2018).

Additionally, available data sets for model validation exist over the CONUS. Over the last 2 decades, the blooming of satellite remote sensing, which monitors various water and carbon fluxes and states globally, greatly aids the evaluation and improvement of CLM. P. Lawrence and Chase (2007) develops new land surface parameters in CLM3 that consistent with retrievals from the Moderate Resolution Imaging Spectroradiometer (MODIS) to enable a consistent historical vegetation mapping, leading to improvement in simulating precipitation and near surface air temperature globally. The Gravity Recovery and Climate Experiment (GRACE) enables the tracking of dynamics in terrestrial water storage (TWS) changes (Bonsor et al., 2018; Rodell & Famiglietti, 2002; Rodell et al., 2007; Syed et al., 2008). Through the comparison with TWS from GRACE, Swenson and Lawrence (2014) reports that the improvement of soil evaporation parameterization greatly reduced biases in simulated seasonal cycle of TWS over semiarid regions. Furthermore, multiple ground-based, site- and local-level measurements over the CONUS complement the large-scale, space-based remote-sensing data, particularly for the subsurface hydrological processes and local agriculture management practices (Dirmeyer et al., 2018; Turner et al., 2005).

The objective of this study is to perform a comprehensive validation of a few key variables that are crucial for understanding skills and biases in land surface water and energy budgets in a few widely used CLM versions at a resolution that matters for high-resolution applications of Earth system models (i.e., 0.125°), using the best available in situ measurements and satellite-based products. It should be noted that previous studies by Ma et al. (2017) and Zheng et al. (2019) have conducted a systematic evaluation study over the CONUS at a 0.125° resolution based on Noah-MP (Noah LSM with multiparameterization options). Although Noah-MP and CLM share similar development pathways (Niu et al., 2011), CLM5 has many additional features (D. Lawrence et al., 2019) and must be comprehensively revalidated for regional-scale applications. We emphasize our validation in the following aspects: (1) evaluate various simulated fluxes and states related to energy, water, and carbon cycles (e.g., latent heat [LE], sensible heat [SH], evapotranspiration [ET], terrestrial water storage anomaly [TWSA], soil moisture, runoff, gross primary production [GPP], leaf area index [LAI], and irrigation) across various spatial scales (e.g., continental, catchment, county, and point) and (2) identify model enhancement and shortcomings related to model parameterization (e.g., soil, crop, and hydrology), model structures (e.g., carbon and nitrogen cycling), and agricultural management practices (e.g., planting, harvest, and irrigation) for potential future model development.

## 2. Model Description

In this section, we will describe the CLM model used in this study including its key updates in the latest version, as well as the designed model experiments.

### 2.1. The CLM

The CLM is the land component of the CESM. It represents several aspects of the land surface including land surface heterogeneity and ecosystem structure, and consists of components or submodules related to biogeophysics, biogeochemistry, hydrology, human dimensions, and ecosystem dynamics (D. Lawrence et al., 2019; Oleson et al., 2013). To represent land surface heterogeneity, CLM classifies each grid cell into multiple land units. Each of these units consists of multiple snow/soil columns, which are occupied with different plant functional types (PFTs).

CLM is capable of simulating energy, water, and carbon cycle dynamics. In the hydrology module, CLM parameterizes water state variables including canopy water (i.e., interception, throughfall, and canopy drip), snow water, evaporation, soil ice, soil water, surface and subsurface runoff, and water table depth. Processes like snow accumulation and melt, water transfer between snow layers, infiltration and redistribution within the soil column, groundwater discharge and recharge are simulated in CLM to update the hydrological variables. In the biogeochemistry module, all the state variables in natural vegetation, crops, litter, and soil organic matter within the terrestrial carbon and nitrogen cycles are fully prognostic. In the crop module, the phenology for each vegetation type, such as plant growth and senescence, is also prognostic,

**Table 1**
*Summary of the Three Simulations Conducted in This Study*

| Experiment name | Domain | Spatial resolution | Simulation period | LAI calculation | BGC module | Crop module | Soil depth | Soil layers | Irrigation |
|---|---|---|---|---|---|---|---|---|---|
| **CLM4.5BGC** (CLM4.5 in the prognostic vegetation and biogeochemistry mode) | CONUS: −125°W to −67°W, 25°N to 53°N | 0.125° | 1979–2018 | Prognostic simulated using the BGC module in CLM4.5 | • Vertically resolved soil C and N | 4 crop types (temperate corn, temperature soybean, spring wheat, and cotton) | Spatially uniform, 3.4 m | 10 layers | On |
| **CLM5BGC** (CLM5 in the prognostic vegetation and biogeochemistry (BGC) mode) | | | | Prognostic simulated using the BGC module in CLM5 | • Carbon cost for nitrogen update<br>• Flexible leaf C:N ratios | 8 crop types (temperate corn, temperature soybean, spring wheat, cotton, rice, sugarcane, tropical corn, and tropical soybean) | Spatially variable, 0.4–8.5 m | 20 layers | On |
| **CLM5SP** (CLM5 in the prescribed satellite phenology mode) | | | | Prescribed using satellite data | – | 2 crop types (c3 unmanaged rainfed and irrigated crop) | Spatially variable, 0.4–8.5 m | 20 layers | On |

and is dynamically related to soil and air temperature, soil water availability, daylength, and agricultural management practices. The planting date of crops is determined based on heat accumulation using thresholds measured by growing degree-days. Harvest is assumed to occur as soon as the maximum growing degree-day required for crop maturity is reached or the number of days past the planting date reaches a crop-specific maximum. Irrigation is activated for the irrigated areas in each grid cell (see Text S1 and S2 in Supplementary Materials for more details).

### 2.2. Key Updates of CLM5 Compared to CLM4.5

There are many new and updated processes and parameterizations in CLM5 relative to CLM4.5. Here we only summarize a few key updates relevant to this study. For a full overview of CLM4.5/CLM5 and their differences, interested readers are referred to Oleson et al. (2013) and D. Lawrence et al. (2019).

#### 2.2.1. Hydrology

In CLM4.5, the soil depth is a spatially uniform value (i.e., 3.4 m) and the soil profile is divided into 10 layers (Table 1). To capture the potential variability of soil and snow related variables, the thickness of each soil column can vary in space in CLM5 (Pelletier et al., 2016), ranging from 0.4 to 8.5 m and is explicitly discretized into 20 hydrologically and biogeochemically active layers and five bedrock layers (Table 1). The unconfined aquifer in CLM4.5 is replaced with a zero-flux boundary condition and an explicit simulation of both saturated and unsaturated zones in CLM5. A revised soil evaporation parameterization that accounts for the rate of water vapor diffusion through a dry surface layer is implemented in CLM5 (Swenson & Lawrence, 2014). Prior study reported that this dry surface layer-based soil resistance scheme restricts soil evaporation and improves ET seasonality (Swenson & Lawrence, 2014).

### 2.3. Biogeochemistry

The biogeochemistry (BGC) module in CLM5 builds upon the implementation of carbon-nitrogen cycle coupling in both CLM4 and CLM4.5 (D. Lawrence et al., 2011; Thornton et al., 2007). More recently, the Fixation and Update of Nitrogen (FUN) model has been incorporated into CLM5 to account for the carbon cost for plant nitrogen uptake. The implementation of the FUN model in CLM5 adds the capability describing the costs of nitrogen acquisition from the environment and control on the flexibility of the plant C:N

ratios. Flexible plant carbon: nitrogen (C:N) ratios are also introduced in CLM5 to replace the static plant C:N ratios in CLM4.5 to allow plants to adjust their C:N ratios (D. Lawrence et al., 2019).

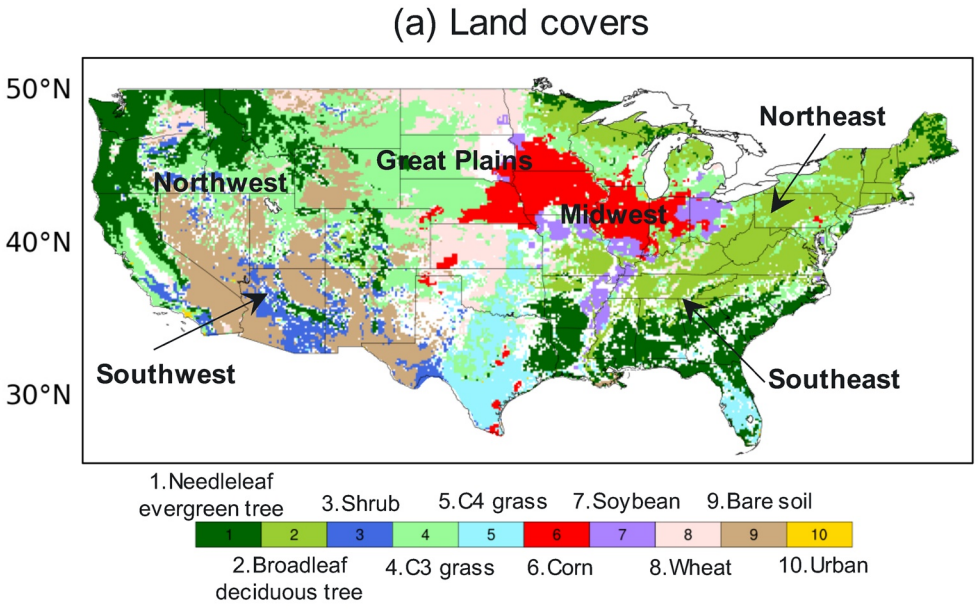
### 2.4. Agricultural Management Practice

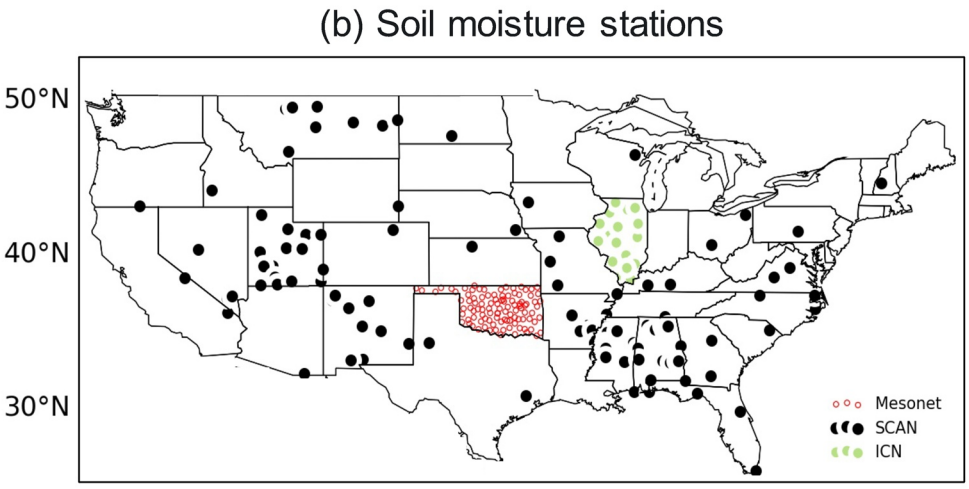
The number of crop types represented in the crop model increases from four in CLM4.5 to eight in CLM5 (Badger & Dirmeyer, 2015; Levis et al., 2018) (Table 1). For irrigated areas, CLM4.5 divides the total cropland area of each grid cell into irrigation and nonirrigation areas based on the Global Map of Irrigated Area (GMIA) data set. GMIA characterizes the fractions of areas equipped for irrigation around year 2000 at a 5 arcmin spatial resolution by combining agricultural censuses and geographical information on irrigation croplands based on surveys and remote sensing (Siebert et al., 2005). In CLM5, irrigated and rainfed fractions in each grid cell are obtained from MIRCA2000 (Portmann et al., 2010). MIRCA2000 provides monthly irrigated and rainfed crop areas of 26 crop types in year 2000 at a 5 arcmin spatial resolution by combining agricultural censuses of harvest area of crops, crop calendars, remotely sensed cropland extent, and GMIA. Compared to GMIA, MIRCA2000 maximizes the consistency between different subnational statistics and considers more factors that could affect the actual irrigation area, such as crop rotation and water shortage. In CLM4.5, irrigation is needed when water is limited for photosynthesis based on the soil water stress function (Equation S1 in supporting information). In CLM5, irrigation is needed when the available soil water is below a specified threshold (Equation S4 in supporting information). For completeness, we also provide a summary of key differences between CLM4.5 and CLM5 irrigation schemes in the supporting information (Text S1 and S2).

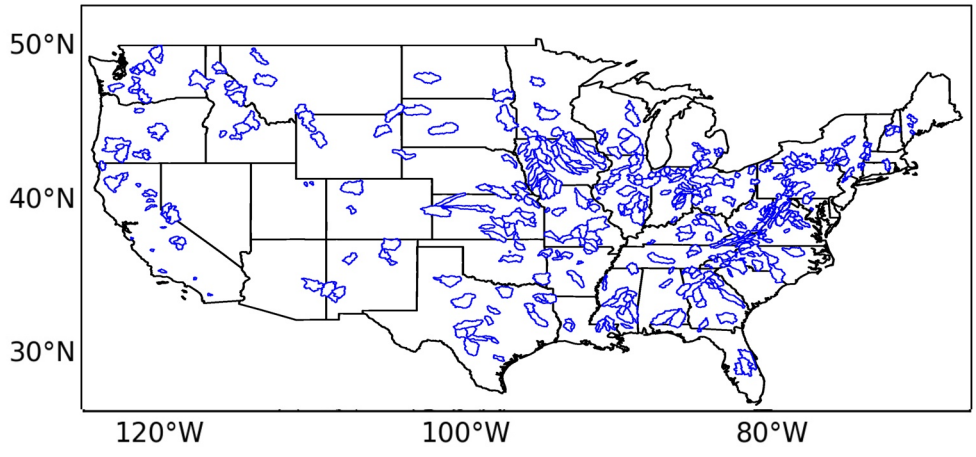
### 2.5. CLM Configurations

CLM can be used with either prognostic biogeochemistry (BGC) mode or prescribed satellite phenology (SP) mode. In the BGC mode, CLM uses a fully prognostic treatment for terrestrial carbon and nitrogen simulations to predict all state variables in vegetation, litter, and soil organic matter within each soil column. Plants compete for nutrients in the soil. These prognostic carbon and nitrogen variables are utilized by the biophysical module to simulate hydrological and energy budget terms. In the SP mode, LAI is prescribed using present day satellite-based data. There is no carbon-nitrogen cycling (e.g., no leaf nitrogen and soil carbon) in the SP mode and photosynthesis is not limited by leaf nutrients. In this study, we conduct three CLM simulations over the CONUS: (1) CLM4.5 in the prognostic vegetation and biogeochemistry mode (CLM4.5BGC), (2) CLM5 in the biogeochemistry mode (CLM5BGC), and (3) CLM5 in the prescribed satellite vegetation phenology mode (CLM5SP) (Table 1), to explore the impacts of different model configurations (CLM5BGC vs. CLM5SP) and model structural evolutions (CLM4.5BGC vs. CLM5BGC) on water, carbon and energy cycle dynamics.

We use hourly meteorological forcing obtained from the National Land Data Assimilation System phase 2 (NLDAS-2) at 0.125° to drive CLM5 simulations from 1979 to 2018. The NLDAS-2 forcing is derived from the 32-km and 3-h North American Regional Reanalysis (NARR) and bias-corrected by additional observed data (e.g., the monthly Parameter elevation Regression on Independent Slopes Model [PRISM]) (Cosgrove et al., 2003; Daly et al., 1994; Xia et al., 2012). It consists of air pressure, air temperature, wind speed, specific humidity, solar radiation, longwave radiation, and precipitation. CLM5 is configured to run over the CONUS domain (−125°W to −67°W, 25°N to 53°N, 464 × 224 grid cells) at a 0.125° spatial resolution and a 30-min time step. Land surface parameters, such as the fractions of each land unit type (lake, glacier, urban, natural vegetation, and crop) of a grid cell, soil properties (e.g., soil color, soil texture, and soil organic matter density), and PFT characteristics (e.g., canopy top and bottom heights), are aggregated from high-resolution input data sets that are derived from various sources (e.g., the International Geosphere-Biosphere Program, the Global Land One-km Base Elevation Project) (D. Lawrence et al., 2018a). More specifically, the land cover information (i.e., the percentage of PFTs) is derived from a combination of the 2001 MODIS Vegetation Continuous Field (VCF), MODIS land cover product, and 1992–1993 AVHRR Continuous Field Tree Cover Project data using the method proposed in P. Lawrence and Chase (2007) (Figure 1a). Irrigation is turned on in all three simulations (Table 1). The 40 years NLDAS-2 forcing data were recycled for 800 years for the carbon and nitrogen pools to reach equilibrium.


(a) Land covers
50°N
40°N
30°N
Northeast
Great Plains
Northwest
Midwest
Southwest
Southeast
1.Needleleaf evergreen tree
2.Broadleaf deciduous tree
3.Shrub
4.C3 grass
5.C4 grass
6.Corn
7.Soybean
8.Wheat
9.Bare soil
10.Urban


(b) Soil moisture stations
50°N
40°N
30°N
Mesonet
SCAN
ICN


(c) MOPEX catchments
50°N
40°N
30°N
120°W
100°W
80°W

## 3. Validation Data Sets

In this study, a suite of remote-sensing and upscaled data sets and in situ observations are utilized to validate CLM5 simulations (Table 2). Specifically, we evaluate the model skills in capturing latent heat (LE)/evapotranspiration (ET), sensible heat (SH), total water storage anomaly (TWSA), runoff, soil moisture, irrigation, gross primary production (GPP), net ecosystem exchange (NEE), and leaf area index (LAI). In this section we describe these reference data sets as well as the evaluation strategies. These selected reference data sets have been widely employed to evaluate LSM simulations (e.g., Collier et al., 2018; Ma et al., 2017; Xia et al., 2018; Zheng & Yang, 2016; Zheng et al., 2019).

### 3.1. Remote Sensing and Data-Driven Upscaled Products

Monthly TWSA from 2002 to 2014 at a 1° spatial resolution is obtained from the Level-3 GRACE (Landerer & Swenson, 2012; Swenson et al., 2006; Tapley et al., 2004). TWSA, calculated as the difference between transient monthly TWS and the 2002–2014 time-mean baseline, is used to evaluate CLM5-simulated changes in total water storage. MODIS LAI (MOD15A2) at a spatial resolution of 500 m and an 8-day time scale from 2001 to 2018 is used to evaluate model simulated LAI. Monthly MODIS GPP (MOD17A2, 0.05°, 2000–2015) (Zhao & Running, 2006, 2010; Zhao et al., 2005), monthly FLUXNET multitree-ensemble (MTE) GPP (0.5°, 1982–2011) (Jung et al., 2010, 2011), and monthly solar-induced chlorophyll fluorescence (SIF) GPP (0.5°, 2001–2018) (Li & Xiao, 2019) are used to evaluated model simulated GPP. The MTE product integrates FLUXNET measurements with surface meteorological observations and geospatial information from remote sensing using a machine learning based approach (i.e., model tree ensemble) (Jung et al., 2009).

The monthly MODIS ET from 2000 to 2014 at 0.05° (Mu et al, 2007, 2009, 2011; Zhao & Running, 2006; Zhao et al., 2005), daily Global Land Evaporation Amsterdam Model (GLEAM) land-surface evaporation data from 1980 to 2017 over the CONUS at 0.25° (Martens et al., 2017; Miralles et al., 2011), and monthly FLUXNET MTE ET from 1982 to 2011 at 0.5° (Jung et al., 2010, 2011) are used to evaluate CLM-simulated ET. The GLEAM ET and MTE ET are selected as they have a longer record and recent evaluations showed GLEAM ET is superior to other remote-sensed ET products (Michel et al., 2016; Miralles et al., 2016).

The diurnal cycles of land surface fluxes provide useful information for model reliability (He et al., 2015; Robock et al., 2003) and it is quite important for the development of summertime mesoscale convective systems (MCSs), when the large-scale atmospheric forcing is weak (Song et al., 2019). Upscaled half-hourly LE, SH, GPP, and NEE from 2001 to 2014 at a 0.5° spatial resolution from Bodesheim et al. (2018) (B2018, Table 2) are used to evaluate diurnal cycles of CLM simulated land-atmosphere fluxes. This data set is upscaled from site-level flux tower measurements to global-scale gridded estimates by integrating half-hourly in situ observations from FLUXNET flux tower sites and gridded remote-sensing data and meteorological forcing at global scales (Bodesheim et al., 2018). It is recommended to primarily use monthly averaged diurnal cycles because it is more robust (Bodesheim et al., 2018).

### 3.2. In Situ Observations

#### 3.2.1. Soil Moisture and Runoff

In situ point observations of 123 Soil Climate Analysis Network (SCAN) sites over 2000–2012 (Schaefer et al., 2007), 103 Oklahoma Mesonet sites over 2000–2012 (Scott et al., 2013), and 18 Illinois Climate Network (ICN) sites over 2003–2012 (Robock et al., 2000) are used for soil moisture comparison (Figure 1b, Table 2). Each soil moisture data set contains daily observations at various soil depths (Table 2). Detailed information for these soil moisture observations including quality control strategies can be found in Schaefer et al. (2007), Scott et al. (2013), and Robock et al. (2000).

**Figure 1.** Maps showing (a) spatial distribution of dominant land cover types over the contiguous United States (CONUS), (b) locations of 103 Soil Climate Analysis Network (SCAN), 123 Oklahoma Mesonet, and 18 Illinois Climate Network (ICN) soil moisture measurement stations, and (c) 336 Model Parameter Estimation Experiment (MOPEX) catchments.

**Table 2**
*Summary of Data Sources Used for CLM Validation*

| Variable | Data source | Data period | Spatial resolution | Temporal resolution | Data source and website | Reference |
|---|---|---|---|---|---|---|
| LE | Upscaled diurnal cycles of energy fluxes | 2001–2014 | 0.5° | Half-hourly | https://www.bgc-jena.mpg.de/geodb/projects/FileDetails.php | Bodesheim et al. (2018) |
| SH | Upscaled diurnal cycles of energy fluxes | 2001–2014 | 0.5° | Half-hourly | https://www.bgc-jena.mpg.de/geodb/projects/FileDetails.php | Bodesheim et al. (2018) |
| GPP | Upscaled diurnal cycles of carbon fluxes | 2001–2014 | 0.5° | Half-hourly | https://www.bgc-jena.mpg.de/geodb/projects/FileDetails.php | Bodesheim et al. (2018) |
| NEE | Upscaled diurnal cycles of carbon fluxes | 2001–2014 | 0.5° | Half-hourly | https://www.bgc-jena.mpg.de/geodb/projects/FileDetails.php | Bodesheim et al. (2018) |
| ET | GLEAM | 1980–2017 | 0.25° | Daily | www.GLEAM.eu | Martens et al. (2017) and Miralles et al. (2011) |
| ET | FLUXNET MTE | 1982–2011 | 0.5° | Monthly | Max-Planck Institute for Biogeochemistry: https://www.bgc-jena.mpg.de/geodb/projects/FileDetails.php | Jung et al. (2010, 2011) |
| ET | MODIS (MOD16A2) | 2000–2014 | 0.05° | Monthly | http://files.ntsg.umt.edu | Mu et al. (2007, 2009, 2011), Zhao et al. (2005), and Zhao and Running, 2006) |
| GPP | MODIS (MOD17A2) | 2000–2015 | 0.05° | Monthly | http://files.ntsg.umt.edu | Zhao and Running (2010) and Zhao et al. (2005, 2006) |
| LAI | MODIS (MOD15A2) | 2001–2018 | 500 m | 8 days | https://search.earthdata.nasa.gov/search | https://search.earthdata.nasa.gov/search |
| GPP | SIF | 2001–2018 | 0.05° | Monthly | http://data.globalecology.unh.edu | Li and Xiao (2019) |
| GPP | FLUXNET MTE | 1982–2011 | 0.5° | Monthly | Max-Planck Institute for Biogeochemistry: https://www.bgc-jena.mpg.de/geodb/projects/FileDetails.php | Jung et al. (2010, 2011) |
| TWSA | GRACE (Spherical harmonics-based) | 2004–2010 | 1° | Monthly | https://grace.jpl.nasa.gov | Landerer and Swenson (2012), Swenson et al. (2006), and Tapley et al. (2004) |
| Irrigation | USGS | 2005 | US County | Yearly | https://water.usgs.gov/watuse/data | Kenny et al. (2009) |
| Runoff | USGS | 1979–2008 | 336 basins of MOPEX | Daily | https://www.nws.noaa.gov/ohd/mopex/mo_datasets.htm | Duan et al. (2006) |
| SM | Oklahoma Mesonet | 2000–2012 | 103 sites, at 5, 25, 60, and 75 cm soil depths | Daily | http://nationalsoilmoisture.com/About.html | Scott et al. (2013) |
| SM | SCAN | 2000–2012 | 123 sites, at 5, 10, 20, 51, and 102 cm soil depths | Daily | https://www.wcc.nrcs.usda.gov/scan/ | Schaefer et al. (2007) |
| SM | ICN | 2003–2012 | 18 sites, at 5, 20, 50, 100, and 150 cm soil depths | Daily | http://nationalsoilmoisture.com/About.html | Robock et al. (2000) |

Abbreviation: LE, latent heat; SH, sensible heat; NEE, net ecosystem exchange; GPP, gross primary production; LAI, leaf area index; ET, evapotranspiration; TWSA, total water storage anomaly; SM, soil moisture; MOPEX, Model Parameter Estimation Experiment.

For streamflow evaluation, an examination of model performance over 336 Model Parameter Estimation Experiment (MOPEX) catchments (Figure 1c) (Duan et al., 2006; Ren et al., 2016) is performed. Comparisons are made against 30-year (1979–2008) USGS gauge observed surface and subsurface streamflow. The baseflow was separated using the one-parameter recursive filter developed by Lyne and Hollick (1979). The MOPEX data set (Duan et al., 2006) and the corresponding flow separation method have been widely used/validated by previous studies (Arnold & Allen, 1999; Brooks et al., 2011; Nathan & McMahon, 1990; Voepel et al., 2011).

#### 3.2.2. Irrigation

Many studies have found that irrigation could substantially alter land surface fluxes/states and interact with atmospheric processes (Leng et al., 2013; Qian et al., 2013; Thiery et al., 2017; Yang et al., 2019). County level irrigation water use from USGS (Kenny et al., 2009) is used to evaluate the model performance in simulating irrigation water use. Conducted every five years, the USGS irrigation estimates are one of the few comprehensive sources on regional and national irrigation water withdrawals. Data sources for the USGS irrigation withdrawals and irrigated acres include State and Federal crop reporting programs. Information on irrigated crop areas along with crop-specific water consumption coefficients or irrigation-system application rates were also used for estimating irrigation water use. Irrigated areas were reported by three types of irrigation methods: sprinkler, micro irrigation, and surface (flood) systems. Note that although climatic conditions (e.g., temperature and precipitation extremes) have a prominent effect on irrigation water withdrawals, their effects in any particular year cannot be associated readily with the aggregated irrigation data, and therefore are difficult to isolate from other factors that affect water use (Kenny et al., 2009).

### 3.3. Model Evaluation

We use root mean square error (RMSE), relative bias ($R_{\text{bias}}$), and anomaly correlation (AC) (detailed calculations can be found in Text S3 in the supporting information) to evaluate the performance of the three CLM simulations (i.e., CLM5BGC, CLM5SP, and CLM4.5BGC, Table 1). The AC is calculated as the correlation of the anomalies after subtracting the respective mean annual climatology from observations and simulations to remove the influence of seasonality (Equation S11). These statistics are calculated using the domain averaged values. The RMSE is used to evaluate the overall model error, $R_{\text{bias}}$ is used to evaluate model biases (low or high) relative to the reference data sets, and AC is used to evaluate model skill in temporal variability.

Since some remote-sensing and upscaled data have coarser spatial resolutions (e.g., 1° for GRACE TWSA, 0.25° for GLEAM ET, 0.5° for MTE ET) than CLM simulations (0.125°), we aggregated CLM outputs to the same spatial resolution of the corresponding reference data using grid box average values. As GPP, ET, and LAI from MODIS have a finer (e.g., 500 m, 0.05°) resolution than CLM simulation results, we transformed MODIS products to the same spatial resolution of CLM results (i.e., 0.125°). The temporal resolutions at which the comparisons are conducted are based on the timescales of the remote-sensing or in situ data (e.g., daily for soil moisture, half-hourly for diurnal cycles of energy and carbon fluxes, monthly for TWSA, Table 2).

When compared with in situ observations, for soil moisture, we compare the model results from the grids that corresponding to the locations of the soil moisture measurement sites (Figure 1b). After identifying the corresponding grids, simulated soil moistures at different soil depth are evaluated at a daily timescale. For irrigation, since the land cover map used in the model simulation is derived from MODIS products corresponding to year 2000, the irrigation water use data for year 2005 is selected for model evaluation as it is the census data closest to the year 2000 from USGS. Model outputs at a 0.125° resolution are aggregated to the county level to compare with the observed irrigation water use data from USGS for 2005.

As there is no common period for all the variables (Table 2), evaluation is conducted over different periods. We use the two tailed t-test to assess the statistical significance (at the 5% significance level) of the bias between simulation and reference data sets as well as bias between different CLM experiments.

**Table 3**
*Evaluation Metrics (Root Mean Square Error [RMSE], Relative Bias [$R_{bias}$], and Anomaly Correction [AC]) for Daily Soil Moisture (SM) Averaged Over Stations of SCAN, Mesonet, and ICN (Station Locations Are Plotted in Figure 1a), Runoff Averaged Over the MOPEX Catchments (Locations Are Plotted in Figure 1b), ET of GLEAM, FLUXNET MTE, and MODIS, GPP of MODIS, SIF, and MTE, and LAI of MODIS*

| Evaluation matrix | | RMSE | | | $R_{bias}$ | | | AC | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Model | | CLM4.5BGC | CLM5BGC | CLM5SP | CLM4.5BGC | CLM5BGC | CLM5SP | CLM4.5BGC | CLM5BGC | CLM5SP |
| Upscaled LE ($W/m^2$) | | 15.424 | 18.610 | 17.098 | −14.205 | −22.902 | −20.013 | 0.971 | 0.968 | 0.968 |
| Upscaled SH ($W/m^2$) | | 18.181 | 15.952 | 16.694 | −5.872 | 5.688 | 1.430 | **0.975** | **0.976** | **0.973** |
| Upscaled GPP ($g\ C/m^2/d$) | | 2.068 | 1.438 | – | −99.677 | −80.314 | – | 0.899 | 0.928 | – |
| Upscaled NEE ($g\ C/m^2/d$) | | 2.451 | 1.443 | – | −48.840 | −25.183 | – | 0.940 | 0.950 | – |
| SCAN SM ($m^3/m^3$) | 0–5 cm | 0.054 | 0.090 | 0.077 | 16.901 | 38.183 | 30.876 | **0.705** | 0.720 | 0.718 |
| | 5–10 cm | 0.038 | 0.068 | 0.057 | 6.952 | 25.823 | 19.022 | **0.694** | 0.736 | **0.732** |
| | 10–20 cm | 0.035 | 0.057 | 0.045 | 5.009 | 19.256 | 12.782 | **0.701** | **0.775** | **0.761** |
| | 20–51 cm | 0.042 | 0.055 | 0.045 | 7.644 | 16.277 | 10.450 | **0.610** | 0.688 | **0.679** |
| | 51–102 cm | 0.052 | 0.044 | 0.050 | −8.132 | −1.575 | −6.284 | **0.417** | **0.480** | **0.462** |
| Mesonet SM ($m^3/m^3$) | 0–5 cm | 0.033 | 0.045 | 0.039 | −7.050 | 14.034 | 5.208 | **0.933** | 0.901 | **0.906** |
| | 5–25 cm | 0.028 | 0.031 | 0.029 | −7.546 | 8.320 | −0.077 | **0.919** | **0.904** | **0.932** |
| | 25–60 cm | 0.019 | 0.026 | 0.024 | −4.209 | 5.932 | −1.230 | **0.866** | **0.871** | **0.927** |
| | 60–75 cm | 0.014 | 0.036 | 0.027 | −0.484 | 10.612 | 3.849 | **0.765** | **0.790** | **0.837** |
| ICN SM ($m^3/m^3$) | 0–5 cm | 0.073 | 0.100 | 0.099 | 6.660 | 21.817 | 19.862 | **0.536** | **0.464** | **0.468** |
| | 5–10 cm | 0.061 | 0.068 | 0.068 | −3.258 | 8.936 | 7.057 | **0.599** | **0.566** | **0.577** |
| | 10–20 cm | 0.043 | 0.043 | 0.042 | −4.279 | 5.681 | 4.023 | **0.746** | **0.753** | **0.757** |
| | 20–50 cm | 0.055 | 0.035 | 0.038 | −13.141 | −6.615 | −7.361 | 0.764 | **0.796** | **0.792** |
| | 50–100 cm | 0.047 | 0.043 | 0.045 | −2.891 | 3.523 | 3.215 | **0.710** | **0.719** | **0.671** |
| | 100–150 cm | 0.079 | 0.110 | 0.110 | −17.511 | −24.484 | −24.647 | 0.673 | 0.531 | 0.603 |
| MOPEX runoff (mm/d) | Total | 0.693 | 0.563 | 0.663 | −49.7 | −32.7 | −45.1 | **0.893** | **0.868** | **0.857** |
| | Surface | 0.250 | 0.194 | 0.172 | −38.8 | −7.7 | −1.0 | **0.554** | 0.621 | 0.690 |
| | Subsurface | 0.523 | 0.476 | 0.559 | −58.0 | −46.4 | −60.0 | **0.878** | **0.848** | **0.865** |
| ET (mm/d) | GLEAM | 0.23 | 0.37 | 0.28 | −4.8 | −20.5 | −11.1 | **0.965** | **0.969** | **0.972** |
| | MTE | 3.862 | 5.031 | 4.099 | 5.8 | −10.8 | −0.05 | **0.991** | **0.996** | **0.996** |
| | MODIS | 0.2135 | 0.2246 | 0.272 | 4.6 | −13.2 | −2.7 | 0.983 | **0.984** | **0.988** |
| GPP ($g\ C/m^2/d$) | MODIS | 1.4254 | 1.2538 | – | −44.0658 | −37.9265 | – | **0.9382** | **0.9669** | – |
| | SIF | 1.8739 | 1.7025 | – | −51.574 | −45.9632 | – | **0.9185** | **0.9462** | – |
| | MTE | 1.4812 | 1.2955 | – | −44.1876 | −37.7079 | – | **0.9291** | **0.9619** | – |
| MODIS LAI ($m^2/m^2$) | | 0.223 | 0.266 | 0.202 | 3.453 | −4.791 | 13.835 | **0.947** | **0.810** | **0.994** |

*Note*. Bold AC numbers indicate significance to 95%.

## 4. Results

In the following sections, a quantitative evaluation of CLM performance in terms of energy partitioning, water budget, and phenological characteristics is presented through comparisons against various reference data sets shown in Table 2. The RMSE, $R_{bias}$, and AC results used to evaluate the model performance are given in Table 3. The reasons for biases in simulating energy, water, and carbon fluxes are discussed. The performance improvement or degradation of CLM5BGC relative to CLM5SP and CLM4.5BGC is presented.

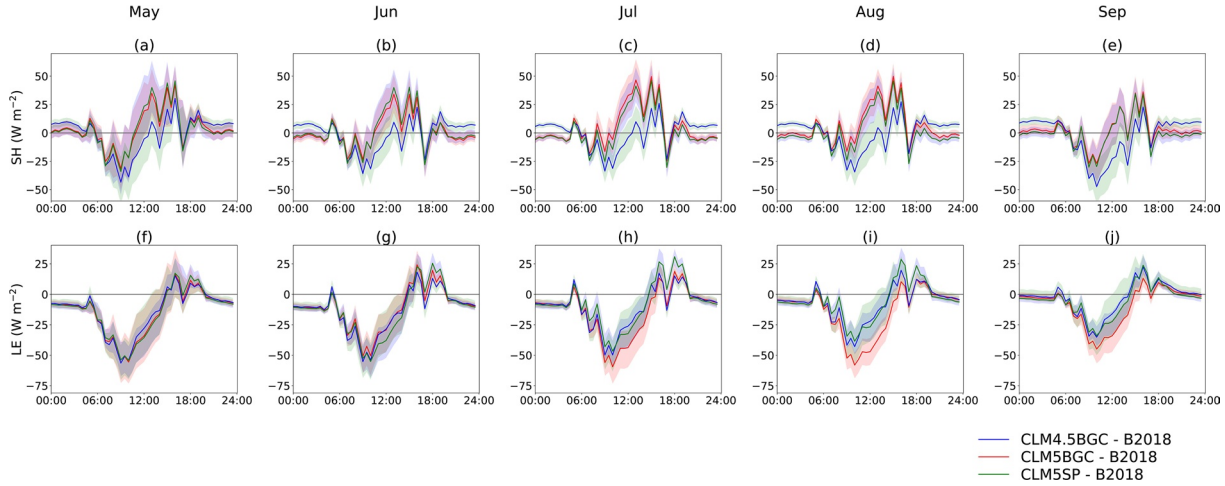



**Figure 2.** Differences in the diurnal cycle of energy fluxes between Bodesheim et al., 2018 (B2018, upscaled diurnal data) and three Community Land Model (CLM) simulations (CLM4.5BGC, CLM5BGC, and CLM5SP) for (a–e) sensible heat (SH, first row) and (f–j) latent heat (LE, second row along with their 5% and 95% percentiles (shaded areas) from May to September (first to fifth column), averaged over the CONUS during the study period 2001–2014.

### 4.1. Sensible and Latent Heat Fluxes

Figure 2 shows the evaluation of diurnal patterns of domain-averaged simulated SH and LE fluxes against the upscaled diurnal cycles of land-atmosphere fluxes from Bodesheim et al. (2018). In general, compared to the mean diurnal cycle of upscaled SH fluxes, CLM4.5BGC simulates a low SH and CLM5BGC and CLM5SP simulate a high SH during mid of the day. The $R_{bias}$/AC values with respect to upscaled SH are −5.9%/0.98%, 5.7%/0.98%, and 1.4%/0.97 for CLM4.5BGC, CLM5BGC, and CLM5SP, respectively. All three CLM simulations have lower LE relative to the upscaled LE during the growing season (Figures 2f–2j), especially for July and August (Figures 2h–2i). Such low LE is more pronounced in CLM5BGC compared to CLM4.5BGC and CLM5SP ($R_{bias}$ values are −22.9%, −14.2%, and −20.0% for CLM5BGC, CLM4.5BGC, and CLM5SP, respectively), especially over the northwest, southwest, and northeast US (Figure 3k). In summary, compared to CLM4.5, CLM5 partitions more total surface heat flux into the sensible than the latent heat flux, especially during mid of the day. Such biases are consistent the biases in ET that will be discussed in Section 4.2.1.

### 4.2. Water Budget Components

For the overall water budget, compared to the reference data sets, all three CLM simulations have lower ET, overestimated soil moisture and irrigation, and largely underestimated total runoff at the MOPEX catchments, especially for subsurface runoff (Table 4). CLM5SP outperforms CLM4.5BGC and CLM5BGC in ET simulations, while CLM5BGC outperforms CLM4.5BGC and CLM5SP in simulating runoff. All simulations exhibit similar performance in soil moisture simulations.

#### 4.2.1. Evapotranspiration

While the three CLM simulations reasonably capture the spatial patterns and gradients of ET (e.g., higher ET in eastern US than western US, Figures 3e, 3j, and 3o) compared to those from MODIS, GLEAM, and MTE (Figures 3b–3d), almost all simulations have lower ET than the three reference data sets over the northwest, northeast, and southeast US (Figures 3f–3s, Table S1) where are mainly covered by forest, shrub, and grasses (Figure 1a), consistent with the diurnal cycle comparison results (Figures 3f, 3k, and 3p).

The area-weighted mean seasonal values from the CLM simulations are compared with those from the reference data sets to calculate the mean bias averaged over the entire CONUS (Figure 4a). For the mean seasonal cycles, CLM5BGC-simulated ET is consistently lower than those from MODIS, GLEAM, MTE, upscaled data, CLM4.5BGC, and CLM5SP, particularly during the growing season from June to August (Figure 4a), consistent with the diurnal scale results (Figures 2f–2j). When compared with MODIS/GLEAM/MTE ET, the $R_{bias}$ values are 4.6%/−4.8%/5.8%, −13.2%/−20.5%/−10.8%, and −2.7%/−11.1%/−0.05% for

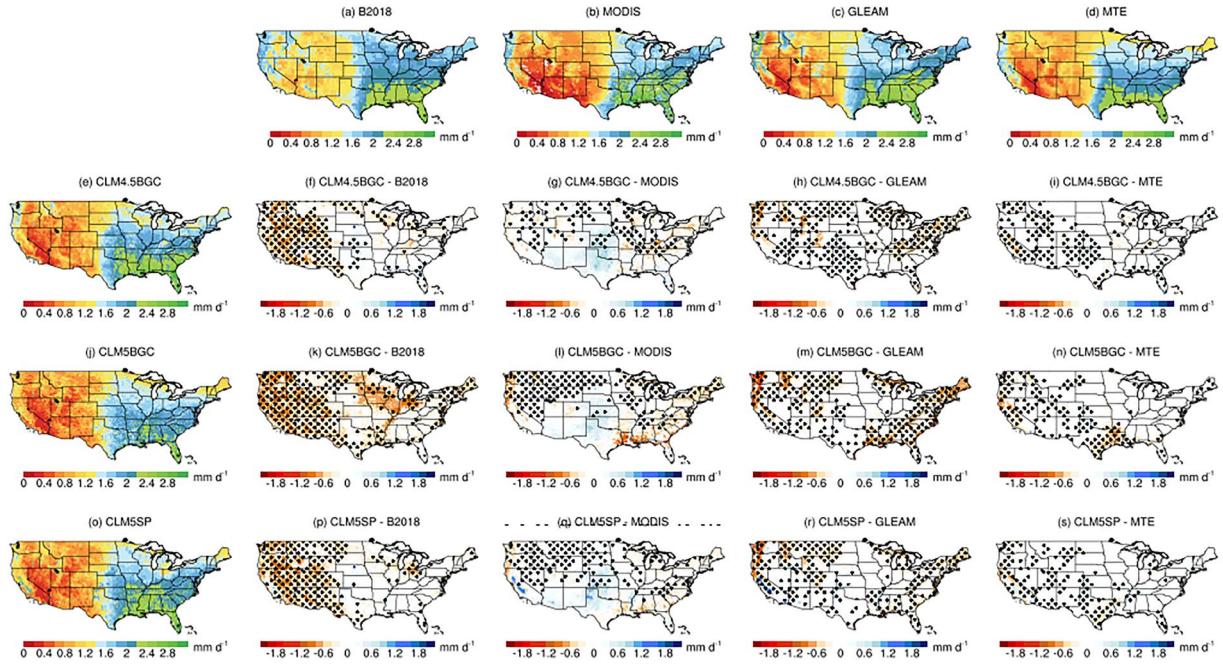



**Figure 3.** Spatial distributions of evapotranspiration (ET) for (a) B2018 (upscaled diurnal flux), (b) Moderate Resolution Imaging Spectroradiometer (MODIS), (c) Global Land Evaporation Amsterdam Model (GLEAM), (d) FLUXNET MTE, (e) CLM4.5BGC, (j) CLM5BGC, (o) CLM5SP; (f, k, and p) biases between the three CLM simulations and B2018 (second column), (g, l, and q) biases between the three CLM simulations and MODIS (third column), (h, m, and r) biases between the three CLM simulations and GLEAM (fourth column), and (i, n, and s) biases between the three CLM simulations and multitree-ensemble (MTE) (fifth column). CLM results were aggregated from 0.125° to 0.25° to compare with GLEAM ET and aggregated to 0.5° to compare with MET ET and upscaled diurnal ET. MODIS ET was aggregated from 0.05° to 0.125° to compare with CLM results. Black dots denote statistically significant changes at the 5% significance level.

CLM4.5BGC, CLM5BGC, and CLM5SP, respectively. The lowest CLM5BGC-simulated ET among the three simulations compared to ET from MODIS, GLEAM, and MTE is consistent with the lowest CLM5BGC-simulated diurnal LE as discovered earlier (Section 4.1, Figures 2f–2j). More discussions on the causes for ET underestimation in CLM5BGC are provided in Section 5.2.

#### 4.2.2. Irrigation

Compared to USGS estimates, all three CLM simulations overestimate the magnitude of irrigation (domain-averaged irrigation amount is 38.4, 67.2, 69.3, and 130.7 mm/yr for USGS, CLM4.5BGC, CLM5BGC, and CLM5SP, respectively). The simulated irrigation amounts differ between CLM5SP and CLM5BGC as their specific crop types are different (Table 1). We will further discuss the potential causes for the mismatch in Section 5.3.1.1. Nevertheless, CLM5 can better capture the spatial patterns of irrigation water use at the county level in 2005 compared to CLM4.5, such as in the major irrigated regions located in the western US (e.g., California, Idaho) (Kenny et al., 2009) (Figure 5), either due to better calibration of CLM5 or updates of irrigation trigger to be soil water deficit in CLM5 (Text S1 and S2). Therefore, the low ET is not caused by irrigation.

#### 4.2.3. Total Water Storage Anomaly

Figure 4b shows the evaluation of monthly TWSA when GRACE-based data are used as the reference. Although CLM5BGC tends to simulate a higher annual peak of TWSA and a steeper decline in TWSA following the peak compared to that of GRACE, the seasonal variability and amplitude of TWSA are well simulated by CLM5BGC (Figure 4b), consistent with previous studies that track the seasonal water storage fluctuations globally (Scanlon et al., 2019). CLM5BGC outperforms both CLM4.5BGC and CLM5SP in capturing variations of TWSA (Figure 4b; RMSE values for the entire time series are 1.634, 1.975, and 1.882 for CLM5BGC, CLM4.5BGC, and CLM5SP, respectively). These results are consistent with the better performance of global-scale TWS that associated with the implementation of a dry surface layer-based soil

**Table 4**
*Water Cycle Variables for Soil Moisture (SM) Averaged Over Stations of SCAN, Mesonet, and ICN, Runoff Averaged Over MOPEX catchments, and ET of GLEAM, FLUXNET MTE, and MODIS*

| | | Reference | CLM4.5BGC | CLM5BGC | CLM5SP |
|---|---|---|---|---|---|
| GLEAM ET (mm/yr) | | 525 | 497 | 416 | 463 |
| FLUXNET MTE ET (mm/yr) | | 463 | 497 | 416 | 463 |
| MODIS ET (mm/yr) | | 476 | 497 | 416 | 463 |
| SCAN SM ($m^3/m^3$) | 0–5 cm | 0.206 | 0.246 | 0.291 | 0.275 |
| | 5–10 cm | 0.231 | 0.247 | 0.291 | 0.275 |
| | 10–20 cm | 0.250 | 0.264 | 0.300 | 0.284 |
| | 20–51 cm | 0.265 | 0.287 | 0.310 | 0.295 |
| | 51–102 cm | 0.317 | 0.292 | 0.313 | 0.298 |
| Mesonet SM ($m^3/m^3$) | 0–5 cm | 0.258 | 0.238 | 0.292 | 0.269 |
| | 5–25 cm | 0.277 | 0.255 | 0.299 | 0.276 |
| | 25–60 cm | 0.284 | 0.270 | 0.299 | 0.279 |
| | 60–75 cm | 0.279 | 0.277 | 0.308 | 0.289 |
| ICN SM ($m^3/m^3$) | 0–5 cm | 0.293 | 0.312 | 0.356 | 0.351 |
| | 5–10 cm | 0.325 | 0.315 | 0.354 | 0.348 |
| | 10–20 cm | 0.332 | 0.318 | 0.351 | 0.345 |
| | 20–50 cm | 0.362 | 0.313 | 0.337 | 0.334 |
| | 50–100 cm | 0.447 | 0.317 | 0.338 | 0.337 |
| | 100–150 cm | 0.441 | 0.365 | 0.334 | 0.333 |
| MOPEX runoff (mm/yr) | Total | 414 | 208 | 279 | 228 |
| | Surface | 135 | 83 | 125 | 134 |
| | Subsurface | 275 | 116 | 148 | 110 |

resistance scheme in CLM (Swenson & Lawrence, 2014), demonstrating that the newly incorporated dry surface layer expression in CLM5 is effective in improving TWS simulations.

#### 4.2.4. Runoff

CLM4.5 and CLM5 largely underestimate total runoff at MOPEX catchments (Figure 1c), especially for subsurface runoff ($R_{bias}$ values for subsurface runoff are −32.7%, −49.7%, and −45.1% for CLM5BGC, CLM4.5BGC, and CLM5SP, respectively, Table 3). This is within our expectation, as large-scale LSMs are

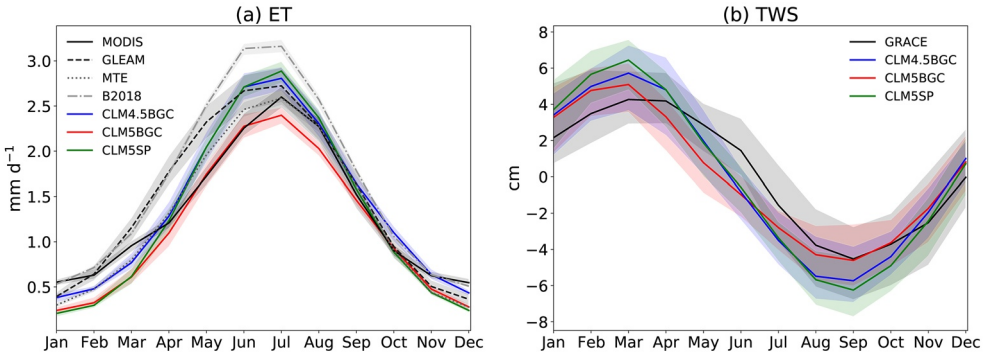



**Figure 4.** Mean monthly remote-sensed and simulated variables that are averaged over the entire CONUS for (a) ET between MODIS, GLEAM, FLUXNET MTE, B2018 (upscaled diurnal data), and CLM and (b) total water storage anomaly (TWSA) between GRACE and CLM. The shaded areas represent the 5% and 95% percentiles of the reference data and model predictions, to indicate interannual variability for each month.

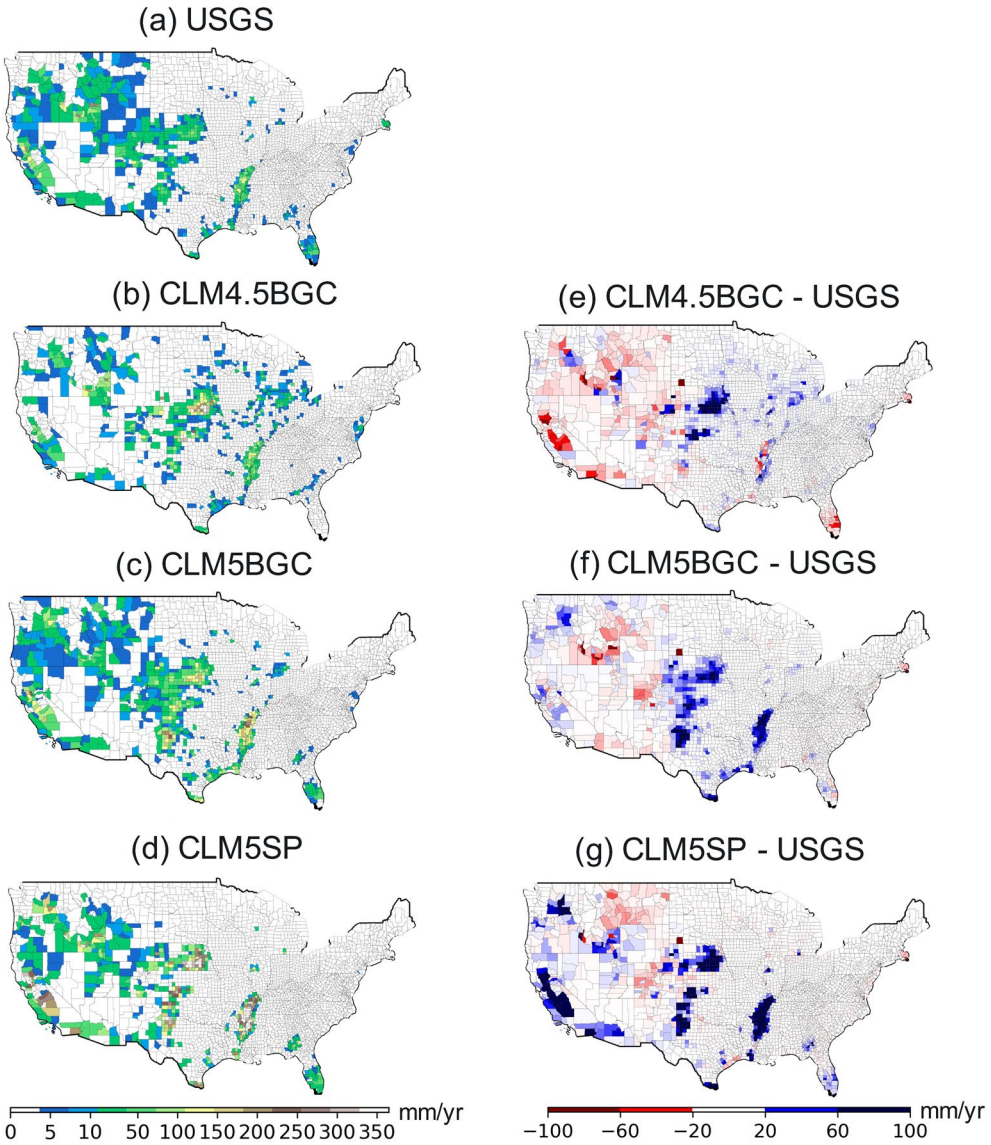



**Figure 5.** (a–d) Spatial distributions of irrigation water use (mm/yr) in 2005 at county scale from USGS, CLM4.5BGC, CLM5BGC, and CLM5SP (left column) and (e–g) difference between the three CLM simulations and USGS estimates (right column). CLM results were aggregated from 0.125° to U.S. county scale to compare with the USGS data.

usually tuned at coarse spatial resolutions, and therefore may not perform well at smaller scales. More discussions are provided in Sections 5.4.3 and 5.4.4.

#### 4.2.5. Soil Moisture

Statistical metrics for observed and simulated daily soil moistures averaged over 123 SCAN stations, 103 Oklahoma Mesonet stations, and 18 ICN stations for multiple soil layers are summarized in Table 3. We find that soil moisture is overestimated in the top 50 cm soil layers but underestimated in the deeper soil layers (50–150 cm) for all three CLM simulations when compared to the reference data sets (Tables 3 and 4, Figures S3–S5). In particular, CLM5BGC is wettest in the top 50 cm soil layers among the three CLM simulations. It is important to note that the observed soil moisture is measured at a point, while the model simulated soil moisture is averaged over a 0.125° × 0.125° pixel, in which subgrid variability in vegetation and soil texture could be very different from the site condition. In addition, the soil hydraulic information used in LSMs usually only consider soil textures but overlook the effect of soil structures such as biopores

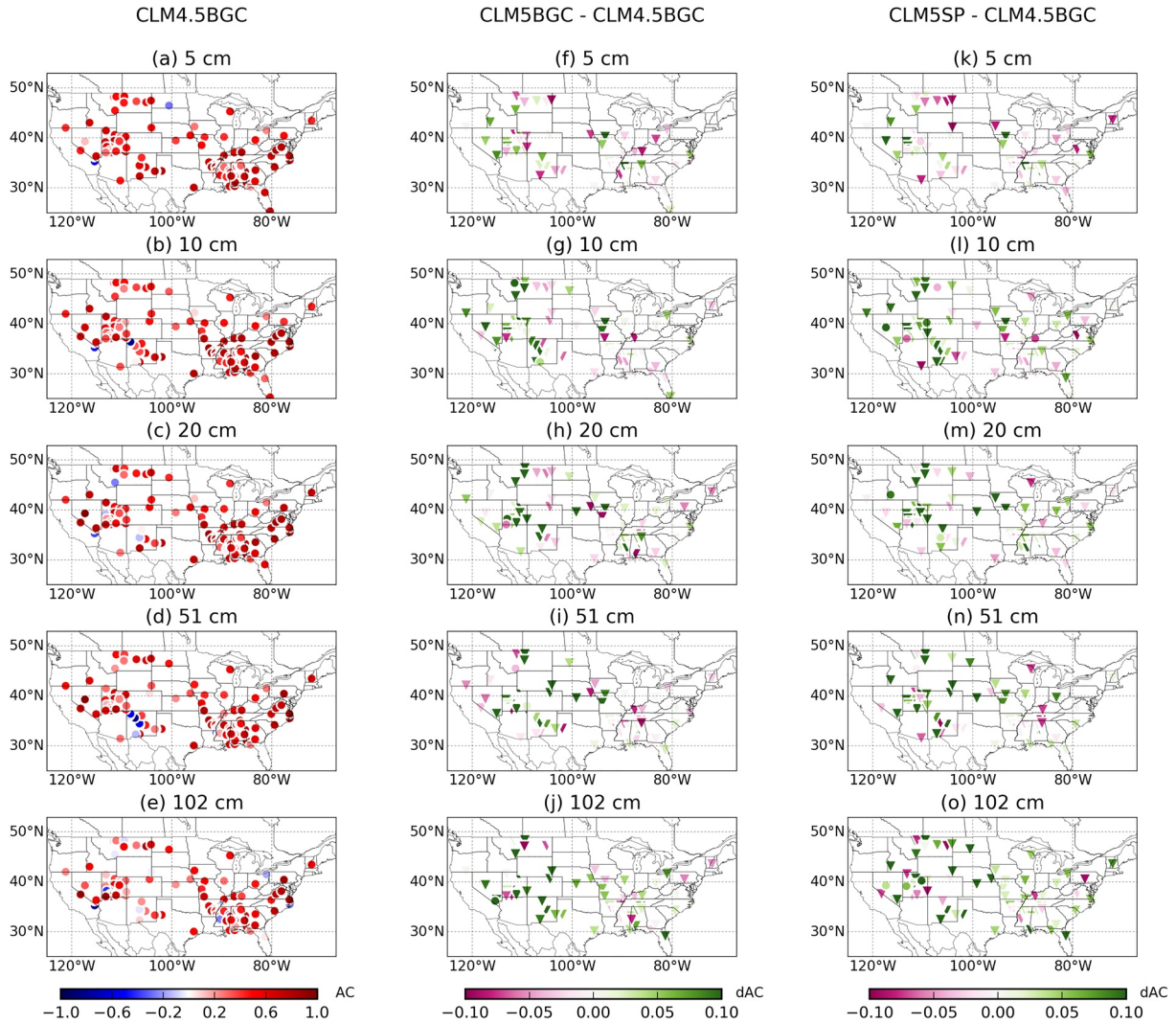



**Figure 6.** (a–e) Spatial distributions of anomaly correlation (AC) between daily observations (123 SCAN soil moisture stations) and CLM4.5BGC (left column), (f–j) differences in AC between CLM5BGC and CLM4.5BGC (middle column), and (k–o) differences in AC between CLM5SP and CLM4.5BGC (right column) for 5 cm (first row), 10 cm (second row), 20 cm (third row), 51 cm (fourth row), and 102 cm (fifth row) soil depths. Triangle shapes denote statistically significant changes at the 5% significance level.

and soil aggregates that created by biological activities (Cheng et al, 2017, 2018, 2019; Fatichi et al., 2020). These differences could partly explain the mismatch between observed and simulated soil moisture.

There could be two major possible pathways that may lead to the low CLM5BGC-simulated ET that identified in Sections 4.1 and 4.2.1: one is limitation in soil water supply, and the other is low plant phenology characteristics such as low photosynthesis rate and LAI. However, the overestimation of soil moisture and underestimation of runoff in CLM5BGC do not lead to increase in ET, suggesting that the low CLM5BGC-simulated ET may not be caused by soil water availability but rather due to biases in simulating plant phenology. This is consistent with our findings in Sections 4.1 and 4.2.1 that CLM5SP which uses prescribed satellite vegetation phenology can better capture LE/ET than CLM5BGC (Figures 2–4). Therefore, we will discuss the model performance in simulating carbon fluxes (GPP and NEE) and plant phenology characteristics (LAI) in the following sections to pinpoint other potential causes.

We use the difference of AC (dAC) to evaluate the relative improvement or deterioration of CLM5 in capturing the spatial soil moisture dynamics compared to CLM4.5. CLM4.5BGC reasonably simulates soil moisture compared with the daily observation data from SCAN (Figures 6a–6e). Spatially, CLM5BGC and

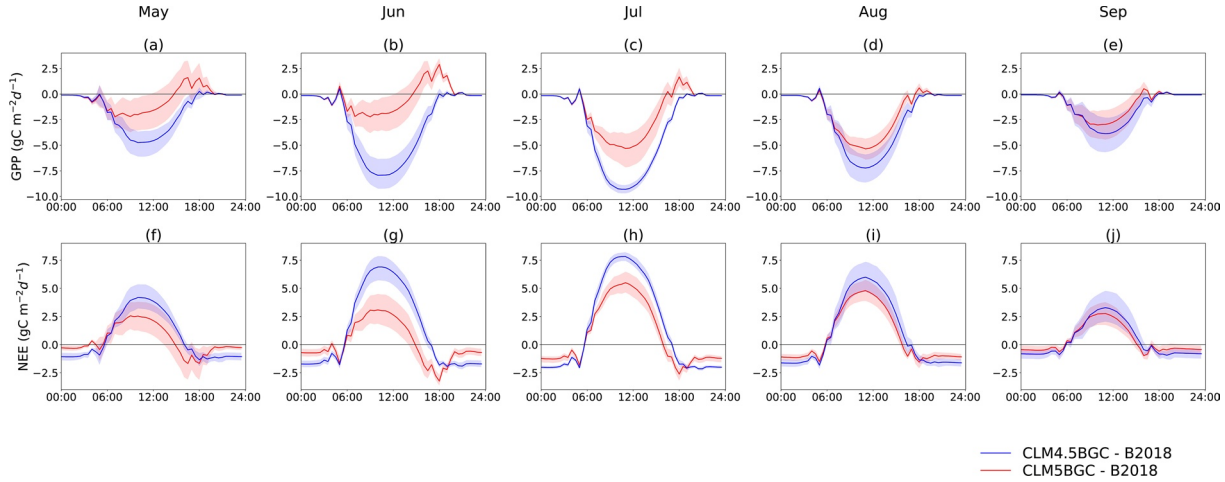



**Figure 7.** Differences in the diurnal cycle of carbon fluxes between B2018 (upscaled diurnal data) and three CLM simulations (CLM4.5BGC, CLM5BGC, and CLM5SP) for (a–e) gross primary production (GPP, first row) and (f–j) net ecosystem exchange (NEE, second row) along with their 5% and 95% percentiles (shaded areas) from May to September (first to fifth column), averaged over the CONUS during the study period 2001–2014.

CLM5SP tend to have better performance over the southwest and northwest US and poorer skills in simulating soil moisture over Midwest and southeast United States than those of CLM4.5BGC (Figures 6f–6j). For different soil layers, generally, the simulation skills for CLM5BGC and CLM5SP have decreased for the shallow surface layers (i.e., 5 and 10-cm soil layers, Figures 6f–6g, Table 3) but increased for the deeper soil layers (i.e., 20, 51, and 102-cm soil layers, Figures 6h–6j, Table 3) compared to those of CLM4.5. Comparisons of daily soil moisture with measurements from ICN and the Oklahoma Mesonet networks exhibit similar trends as those at the SCAN sites (Table 3, Figures S1 and S2). The improvement of soil moisture simulation in deeper soil layers for CLM5BGC as compared to prior versions of CLM (Decker & Zeng, 2009; Oleson et al., 2008) suggests that the implementation of variable soil thickness (Brunke et al., 2016) and groundwater schemes (Swenson & Lawrence, 2015) are effective to improve soil moisture simulations.

### 4.3. Carbon Fluxes

CLM5 has lower GPP compared to the reference data in terms of diurnal cycles from May to September (Figures 7a–7e), especially during mid of the day. For spatial patterns, CLM5BGC-simulated GPP is lower than those of the reference data sets in northwest and southeast US (Figures 8k–8n and 9a, Table S1). These regions also have lower ET relative to the reference data sets (Figures 3j–3n, Table S1). These results indicate the biases in simulating GPP is responsible for the low ET for CLM5BGC. However, the simulated mean annual GPP from CLM5BGC is still improved over most regions of the CONUS compared to CLM4.5BGC-simulated GPP, especially over the Great Basin and Southwestern US (Figures 9a and 8). In addition, CLM5 has significantly improved skills than CLM4.5 for both GPP ($R_{\text{bias}}$ is −80.3% and −99.7% for CLM5BGC and CLM4.5BGC, respectively, Table 3) and NEE ($R_{\text{bias}}$ is −25.2% and −48.8% for CLM5BGC and CLM4.5BGC, respectively, Table 3), indicating updates in the biogeochemical modules (e.g., the FUN model and flexible plant C:N ratios as described in Section 2.2) in CLM5 have led to enhanced simulation skills for biogeochemistry fluxes.

### 4.4. Leaf Area Index

Figure 10 shows the comparison between the spatial mean annual climatology of three CLM-simulated LAIs against MODIS LAI over the CONUS. Since CLM5SP uses prescribed satellite LAI for simulation (Table 1), the CLM5SP-simulated LAI matches well with the MODIS data as expected (Figures 10p–10t, AC = 0.99, RMSE = 0.202), which leads to better simulated ET as discussed earlier (Figures 3o–3s and 4a). CLM5BGC and CLM4.5BGC capture the overall trends for spatial variability of LAI (AC/RMSE is 0.81/0.27 and 0.95/0.22 for CLM5BGC and CLM4.5BGC, respectively, Figures 10k and 10f). Specifically, LAIs in the eastern and northwestern US are greater than those in the southwest US and the Great Plains, mainly driven

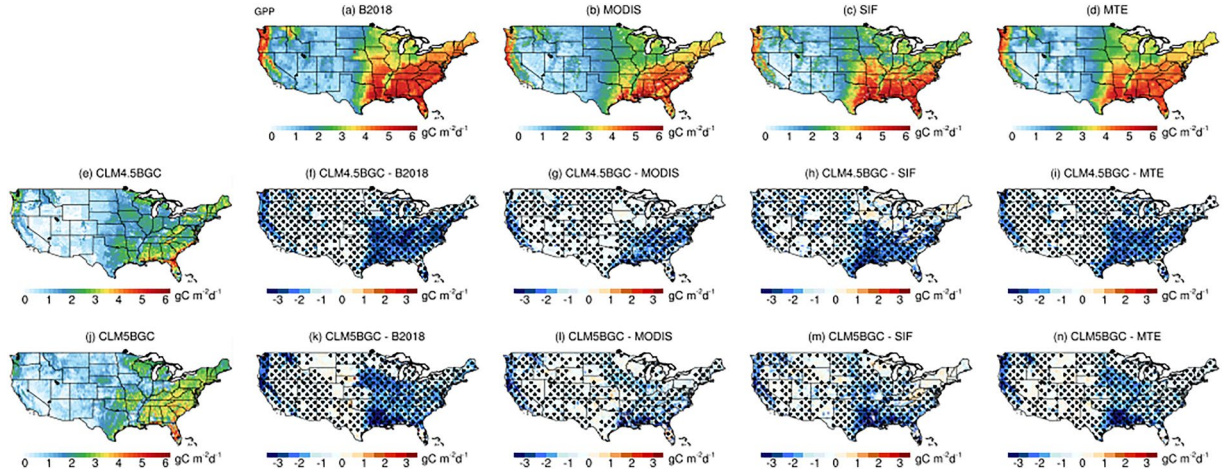



**Figure 8.** Spatial distributions of gross primary production (GPP) for (a) B2018 (upscaled diurnal data), (b) MODIS, (c) SIF, (d) MTE, (e) CLM4.5BGC, and (f) CLM5BGC, (f and k) biases between the two CLM simulations and B2018 (second column), (g and l) biases between the two CLM simulations and MODIS (third column), biases between the two CLM simulations and SIF (h and m, fourth column), and (i and n) biases between the two CLM simulations and MTE (fifth column). CLM results were aggregated from 0.125° to 0.5° to compare with SIF GPP, MET GPP, and B2018. MODIS GPP was aggregated from 0.05° to 0.125° to compare with CLM results. Black dots denote statistically significant changes at the 5% significance level.

by gradients in mean annual precipitation (Liu et al., 2018) and the forest land cover (Figure 1a). However, both CLM5BGC and CLM4.5BGC have difficulties in capturing the spatial mean annual magnitude of LAI. CLM5BGC-simulated LAI is smaller than that of MODIS in northwest, northeast, and southeast US (Figure 10k, Table S1), corresponding to where CLM5BGC-simulated ET and GPP are low (Figures 3j–3n and 8j–8n, Table S1), which further demonstrates that the low bias in simulating plant phenology is responsible for the low ET for CLM5BGC in these regions.

To examine the causes for the difference in spatial variability in mean annual LAI for CLM5BGC and CLM4.5BGC compared to that of MODIS, the spatial mean seasonal bias of CLM-simulated LAIs and the domain-averaged mean seasonal cycles of LAIs from MODIS and three CLM simulations are shown in Figures 10 and 9b, respectively. Note that the magnitudes of MODIS LAI remain high during the nongrowing season (Figure 9b). This is potentially associated with the MODIS reflectance calibration (Cohen et al., 2003) and the lack of seasonal variations of minimum canopy resistance and physiological temperatures in MODIS, especially during winter times (Sun et al., 2007; Zhu et al., 2020). CLM4.5BGC tends to have high LAI during the growing seasons. This is especially true for the Midwest (Figure 10i) which are mainly covered by crops (Figure 1a). CLM5BGC does not capture the strong seasonality (summer peak and winter trough) in observed LAI. First, it has a one-month earlier peak in LAI than the reference MODIS

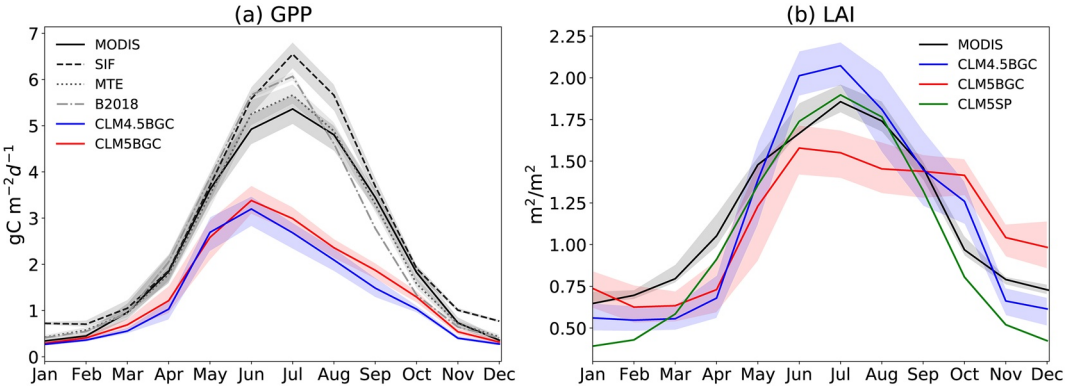



**Figure 9.** Mean monthly remote-sensed and simulated variables that are averaged over the entire CONUS for (a) GPP between MODIS, SIF, MTE, B2018, and CLM and (b) LAI between MODIS and CLM. The shaded areas represent the 5% and 95% percentiles of the reference data and model predictions, to indicate interannual variability for each month.

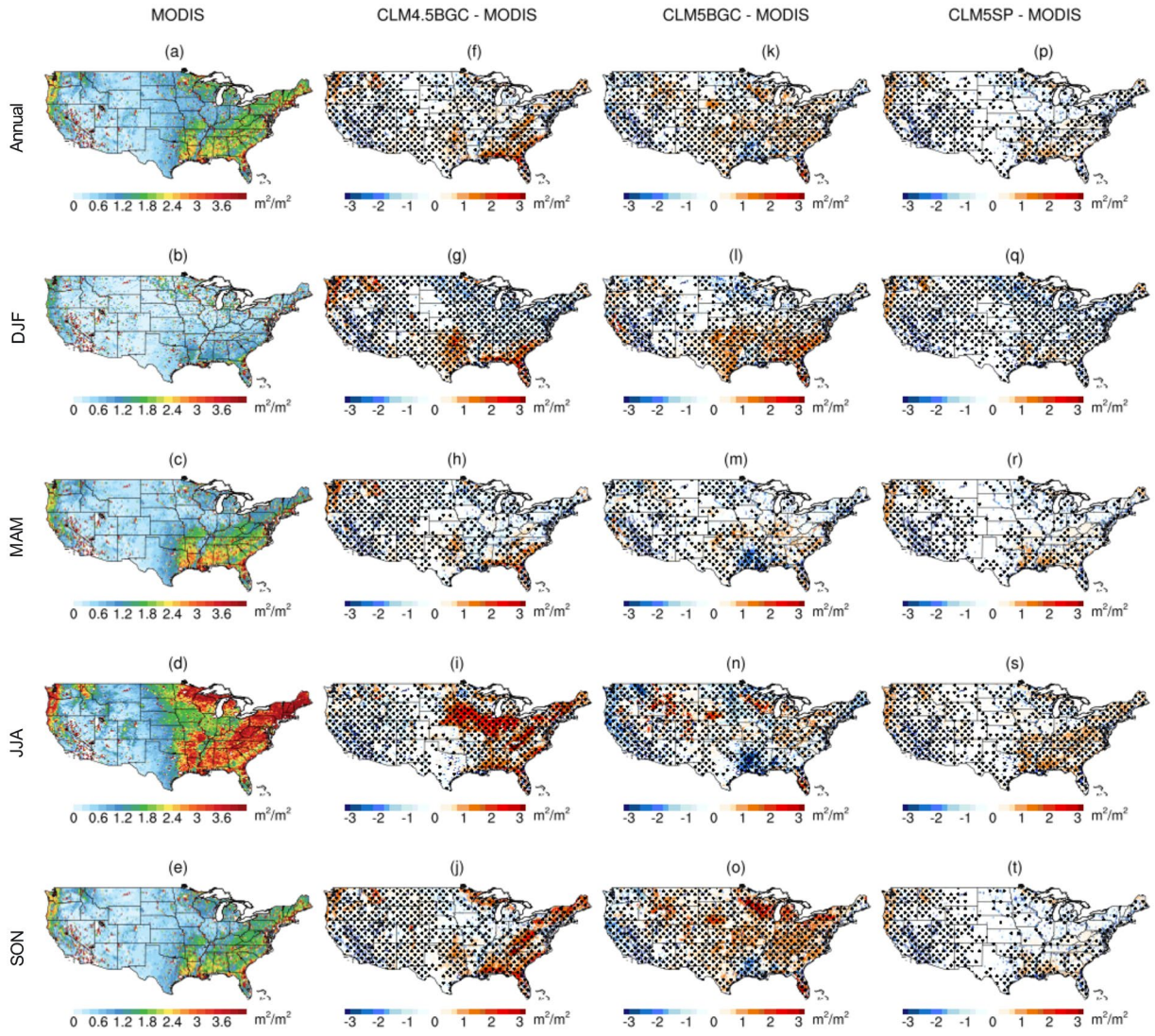



**Figure 10.** Spatial distributions of 18-year-average (2001–2018) leaf area index (LAI) for (a–e) MODIS (first column); (f–j) biases between CLM4.5BGC and MODIS (second column), (k–o) biases between CLM5BGC and MODIS (third column), (p–t) biases between CLM5SP and MODIS (fourth column), for mean annual (first row) and seasonal (second to fifth row) scales. MODIS products were aggregated from 500 m to 0.125° to compare with CLM results. Black dots denote statistically significant changes at the 5% significance level.

data. Specifically, the maximum LAI in MODIS and CLM5SP occurs in July, whereas CLM5BGC reaches its peak LAI around June. This is attributable to the agricultural management practices (e.g., planting and harvest) applied in CLM5 (see more discussions in Section 5.4.1.2). In addition, there is a positive bias for the mean seasonal LAI during November to December for CLM5BGC (Figure 9b), due to the higher simulated LAI over the Southeast US in winter compared to MODIS (Figure 10l). These regions are mainly covered by needleleaf evergreen trees, broadleaf deciduous trees, and C4 grasses (Figure 1a), indicating that the phenological simulations (e.g., carbon allocations) for forest and C4 grass needs to be improved in CLM5BGC, especially during cold seasons.

It should be noted that CLM4.5BGC-simulated GPP is lower than the reference data sets (Figure 9a), while CLM4.5BGC-simulated LAI is higher than the reference during the growing season (Figure 9b). This is mainly due to the high allocation of available carbon to leaf in CLM4.5BGC. In CLM, the carbon allocation routine determines the fate of newly assimilated carbon that comes from the calculation of photosynthesis (i.e., GPP), while CLM4.5BGC allocates more carbon to support the growth of leaf than other new tissues

(e.g., stem, root). A significant change of CLM5 relative to CLM4.5 is that allocation of carbon proceeds independently rather than in a sequential manner (D. Lawrence et al., 2019).

## 5. Discussion

### 5.1. Impact of Model Improvements in CLM5

Compared to earlier versions of CLM (e.g., CLM4.5BGC), CLM5BGC exhibits substantial improvements in simulating GPP and NEE (Figures 7–9), potentially due to its improved biogeochemical parameterizations, such as the FUN model and the flexible plant C:N ratios (Lawrence et al., 2018a, 2019) as described in Section 2.2. For hydrological variables, CLM5 well simulated the seasonal variability and amplitude of TWSA (Figure 4b), which can be attributed to several model improvements in CLM5. First, CLM5 can simulate the dynamics of different components of TWS (e.g., soil moisture, surface water, snow, ice, and groundwater) and can realistically represent key physical hydrological processes (e.g., surface water and groundwater) (Xia et al., 2017). Second, the soil-layer thickness extends from 3.4 m in CLM4.5–8.5 m in CLM5 and can vary in space. Therefore, both saturated and unsaturated zones are explicit modeled (Brunke et al., 2016). Last but not the least, there is an improved representation of the soil evaporation component of TWS in CLM5, which is based on a physically based formulation rather than a simplified empirical parameterization of soil resistance in CLM4.5 (Swenson & Lawrence, 2014). All of these improvements contribute to the enhanced capabilities of CLM5 in capturing interactions between surface and subsurface hydrological processes as compared to previous studies using earlier versions of CLM (Lei et al., 2014; Scanlon et al., 2019). In general, despite the mismatch of LAI during the cold season and the one-month advance for the peak time of LAI during the growing season as compared to the MODIS LAI, CLM5 adequately reproduced the spatial patterns and seasonal variabilities of the variables derived from in situ or satellite-based data, especially for SH, LE/ET, soil moisture, TWSA, and GPP, which showed great potential to inform water management and drought monitoring (Rodell et al., 2004).

### 5.2. Impact of Plant Phenology on Land Surface Fluxes

The magnitude of CONUS averaged ET is consistently low in CLM5BGC compared to the four selected reference data sets during the growing season, especially over the northwest, northeast, and southeast US (Sections 4.1 and 4.2.1, Figures 2, 3, and 4a, Table S1). There could be two major potential pathways that may lead to the low CLM5BGC-simulated ET: one is low soil evaporation due to limited soil water, and the other is low plant phenology characteristics. Instead of low soil moisture, we find that CLM5BGC produces high values of soil moisture, indicating that soil water supply is not the limiting factor for the low ET simulation. The robust low CLM5BGC-simulated GPP compared to the reference data sets supports the second pathway that the simulated phenology and physiology play a more important role for the low ET estimate. The low values of CLM5BGC-simulated GPP/LAI (Figures 8–10) modify the partitioning of energy at the land surface and result in an associated lower latent heat fluxes and higher sensible heat flux as we discovered in Sections 4.1 and 4.2.1, especially during mid of the day over the growing season (Figure 2). This is consistent with findings in Sections 4.1 and 4.2.1 that CLM5SP which uses prescribed satellite vegetation phenology can better capture LE/ET than CLM5BGC (Figures 2–4). These results demonstrate the impacts of vegetation phenology on altering surface energy and water fluxes. These impacts are profound on land surface processes (Foley et al., 2005) and can feedback to regional-to-global climate through land-atmosphere interactions (Searchinger et al., 2018).

### 5.3. Uncertainties in the Reference Data Sets and Implications

It should be noted that there are some interproduct variabilities in the referenced ET and GPP estimates (Figures 3, 4, 8, and 9), since they are derived from different sources using different methods (Bodesheim et al., 2018; Jung et al., 2009, 2010, 2011; Li & Xiao, 2019; Miralles et al., 2011, 2016; Zhao & Running, 2006, 2010; Zhao et al., 2005). For example, MODIS used the Penman-Monteith algorithm to estimate ET (Mu et al., 2007, 2011). GLEAM estimated evaporation based on satellite forcing only (Miralles et al., 2011). Global ET estimates in FLUXNET MTE are derived by upscaling local eddy covariance estimates through machine learning techniques. Systematic and large uncertainties therefore exist, even across these reference

data sets. For instance, MODIS tends to have lower ET compared to other products due to higher evaporative stress in the Penman-Monteith model (Miralles et al., 2016). Uncertainties in FLUXNET MTE are mainly due to the uneven spatial distribution and limited number of flux towers that used for the MTE training. For GPP estimates, uncertainties in MODIS inputs (e.g., land cover, cloud contaminations, and GPP algorithm) and SIF characteristics (e.g., SIF-GPP relations) can further influence the estimated GPP (Zhao et al., 2005). All these uncertainties and interproduct variabilities highlight the importance of using multiple data sets to obtain more robust conclusions. In this study, despite these uncertainties in the referenced products, simulated ET (Figures 2–4) and GPP (Figures 7–9) are consistently lower in CLM5BGC. These robust results give confidence for the low CLM5BGC-simulated ET and GPP and provide additional evidence to attribute the low CLM5BGC-simulated ET to biases in simulating plant phenology (as discussed in Section 5.2).

## 5.4. Challenging Issues of Model Parameterization, Calibration, and Structures

### 5.4.1. Land Management Practices

#### *5.4.1.1. Irrigation*

There are several causes for the mismatch in irrigation amounts between the CLM simulations and the reference data set (Figure 5). First, we note that even though irrigation amount is expected to vary year to year, the area equipped with irrigation from GMIA or MIRCA2000 used as inputs for CLM remain constant in the simulation. Furthermore, large uncertainty remains in such data sets as they are estimated by combining data products from agricultural censuses at coarse spatial resolutions (e.g., counties) with remote-sensed land cover data sets (see detailed methodology in Portmann et al. [2010] and Siebert et al. [2005]). Therefore, the difference in simulated irrigation water use between CLM4.5 and CLM5 can be attributed to differences in physical processes and vegetation physiology and phenology represented in the models, while the differences between the CLM simulations and USGS estimates can be attributed to uncertainties in both difference between processes represented in the models and in reality, and uncertainties in input and validation data sets.

In addition, no calibration for irrigation is conducted in this study. The default parameters are determined based on climatology of water budget and benchmarked with the FAO data set at a global and annual scale (Shiklomanov, 2000), which may not be applicable for regional studies (Leng et al., 2013, 2017). Leng et al. (2013) reported that the irrigation amount simulated by CLM4 can be improved by tuning model parameters (e.g., weighted factor related to target soil moisture) and implementing a more accurate representation of the spatial distribution and intensity of irrigated areas. Additional water from irrigation can lead to increases in the soil water content and lower surface temperature and sensible/latent heat fluxes, with the potential to change boundary layer dynamics and regional scale precipitation patterns (Devanand et al., 2019; Qian et al., 2013; Thiery et al., 2017; Yang et al., 2019). It is therefore imperative to improve the simulation of irrigation which can significantly affect the simulated effects of irrigation on land-atmosphere exchange of water, carbon, and energy fluxes as well as regional/local climates.

#### *5.4.1.2. Phenology Stages*

In CLM5BGC, the planting date of crops is determined when the air temperature reaches a threshold. Harvest is simulated to occur when either the maximum growing degree days required for crop maturity are reached or the number of days past the planting date reaches a crop-specific maximum (D. Lawrence et al., 2018a). This approach fails to capture the local management practices as planting and harvest dates can vary from region to region and differ from crop to crop (Cheng et al., 2020; Sacks et al., 2010). As a result, the CLM-simulated peak LAI during the growing season is earlier than the peak LAI derived from MODIS by almost one month (Figure 9b). Future studies should draw attention to better represent these key phenological stages (e.g., spatially distributed planting and harvest dates) in CLM5 by taking full advantage of local observational data (e.g., US Department of Agriculture's Agricultural Statistics Service [NASS]) and global crop calendars (e.g., Portmann et al., 2010; Sack et al., 2010). The implementation of advanced crop models (e.g., the Agricultural Production Systems sIMulator [APSIM] model) which have more detailed crop growth processes and responses to environmental conditions warrants additional study (Peng et al., 2018).

#### 5.4.2. Phenology and Physiology Parameters

CLM5 has some deficiencies in simulating characteristics (e.g., LAI) of trees in northwest and southeast, C4 grass in southeast US (Figure 10l), and crops in the Midwest (Figure 10o), especially during the nongrowing seasons. There are two potential reasons that may explain the discrepancy for crop simulations. First, uniform parameters of physiology and phenology, such as photosynthesis capacity, crop phenology, and CN allocation, are applied to the same PFT even when they are growing under different climate conditions. For example, currently, the parameters for corn and soybean implemented in CLM5 are derived from studies conducted at a global scale (Levis et al., 2012), which may not be applicable at regional scales. More recently, Cheng et al. (2020) modified the default photosynthesis capacity parameter values for corn and soybean when applied CLM5 at an Illinois site, suggesting that these parameters need to be adjusted when applied to capture local observed phenology and physiology for crops. Moreover, a delayed end date of growing season for temperate grasses has been observed in Zhang et al. (2019). They adjusted the temperature threshold set for leaf offset and the carbon allocation strategy for grasses and found substantial improvement in the modeled phenology of temperate grassland. Adjusting key phenology and physiology parameters is expected to help resolve the deficiencies in simulating characteristics of crops, trees and grasses and improving the simulation of GPP/LAI/ET.

#### 5.4.3. Hydrological Parameters

Though multiple efforts have been made to improve simulations of terrestrial hydrological cycle (e.g., dry surface layer, groundwater dynamics) (Swenson & Lawrence, 2014, 2015; Swenson et al., 2012), there are still large discrepancies between model simulations and observations for hydrological variables (e.g., runoff) over the CONUS. As shown in previous studies, runoff partitioning and surface energy partitioning intrinsically closely interact with each other (Henderson-Sellers et al., 1995; Liang & Xie, 2003). For example, it has been documented that even though calibration of runoff parameters can partially improve runoff simulations, it may result in poorer simulation of other water budget fluxes (e.g., ET, soil moisture) (Hou et al., 2012). Thus, it is important to develop and apply suitable calibration schemes that may achieve satisfactory simulation results across various physical processes (e.g., surface water, groundwater) and hydrological variables (e.g., ET, runoff, soil moisture), especially when apply for small catchments. For example, Hou et al. (2012) designed an uncertainty quantification (UQ) framework for hydrologic parameter calibration in LSMs and reported possible ways for parameter inversion/calibration using available measurements of latent/sensible heat fluxes to obtain the optimal parameter set for CLM. This study provided guidance to reduce parameter set dimensionality and calibration which can be applied under different hydrologic and climatic regimes in LSMs. Huang et al. (2013) applied this UQ framework to investigate the sensitivity of runoff simulations to major hydrologic parameters in CLM across 20 MOPEX watersheds. They found the most significant parameters are those related to the subsurface runoff parameterizations and different hydrologic regimes have different types of parameter sensitivities. Ren et al. (2016) extended these two studies to classify basins based on hydrological parameter sensitivity, aiming at evaluating model parameter transferability across watersheds and reduce parameter calibration efforts. These approaches could provide insights for future research to calibrate relevant hydrological parameters (e.g., the maximum fractional saturated area, the decay factor representing the distribution of surface runoff, and the decay factor representing the distribution of subsurface runoff) and improve runoff simulations.

#### 5.4.4. Hydrological Processes

Moreover, CLM5 has not explicitly incorporated hillslope-scale terrain structures and processes, such as sunny and shady slopes and lateral ridge-to-valley flows (Fan et al., 2019). Chaney et al. (2018) proposed a statistical approach that grouped the hillslopes into natural clusters to parameterize the subgrid heterogeneity of LSM, which showed crucial implications for the evaluation and application of Earth system models. Swenson et al. (2019) implemented representative hillslopes into CLM5 to simulate hydrologically similar areas of a catchment and demonstrated its ability to reproduce the observed difference between ET in different portions of a catchment. Mizukami et al. (2016) developed a runoff routing tool named mizuRoute for continental domain applications and demonstrated its capability to capture spatially distributed streamflow. The introduction of hillslope hydrologic and runoff routing processes is expected to improve the streamflow simulation results at smaller catchments.

## 6. Conclusions

Understanding the role of multiscale land surface processes in modulating regional weather and climate is critical for weather forecast, hydrometeorological and hydroclimatological applications. This requires model simulations with fine resolutions (e.g., 0.125°) at which scales land surface changes and processes can be adequately captured. This study evaluates the performance of version 4.5 and version 5 of the CLM in simulating various land surface variables (e.g., energy, carbon, and water fluxes and state variables) over the CONUS at a 0.125° resolution and investigates the causes for simulation biases, benefited from high resolution (i.e., 0.125°) and long term (1979–2018) meteorological forcing and rich validation data sets over this region. Three configurations of CLM, namely CLM5-biogeochemistry (CLM5BGC), CLM4.5-biogeochemistry (CLM4.5BGC), and CLM5-satellite phenology (CLM5SP), were conducted and analyzed. Both remote-sensing, data-driven upscaled products (e.g., GLEAM ET, MTE ET, GRACE TWSA, MODIS LAI/GPP, upscaled diurnal cycles of LE, SH, GPP, and NEE), and in situ station data (e.g., USGS gauge streamflow, site-level soil moisture, county-scale irrigation) were used to perform the comprehensive validation and investigate the biases due to either water supply or plant phenology. The spatial distributions and seasonal variabilities of ET, TWSA, soil moisture, GPP, and irrigation at the CONUS-wide, county level, and point scales were reasonably captured by CLM5 and CLM4.5, indicating their abilities to capture energy, carbon, and water dynamics as well as the underlying physical processes. We found general improvements of CLM5 in simulating GPP, NEE, SH, TWSA, runoff, and irrigation than CLM4.5. These were achieved by CLM5's improvements in biogeochemical parameterizations (e.g., the FUN model that accounts for carbon cost for nitrogen uptake, flexible plant C:N ratios), hydrology (e.g., spatially variable soil thickness, explicit simulation for both saturated and unsaturated zones), agricultural management practices, and a consequent enhanced capability in capturing interactions between surface and subsurface land surface processes as well as energy and carbon cycles.

This study also revealed a number of biases of CLM5 and CLM4.5 in simulating magnitudes of energy, carbon, and water variables. These biases for carbon fluxes were mainly stemming from shortcomings in land management practices for crops, phenology and physiology parameters for trees and grasses, and hydrologic parameter values related to soil and hydrology. Specifically, due to a lack of spatial-explicit planting and harvest dates based on local data and advanced crop models with more detailed crop growth processes (e.g., the APSIM model), there was one-month advance for the peak LAI between CLM5BGC and MODIS. The low values of CLM5BGC-simulated LAI/GPP for trees and grasses during the growing season led to the low CLM5BGC-simulated ET over the northwest, northeast, and southeast US, especially during mid of the day over the growing season. The impacts of vegetation physiology and phenology on simulating energy fluxes could have significant impacts on land surface processes and feedback to local to global climates through land-atmosphere interactions. Furthermore, USGS gauge-based streamflows at 336 MOPEX catchments were underestimated as a result of inadequate calibration for runoff simulation and deficiency in integrating hillslope hydrologic processes. Caution should be drawn when applying CLM5 and CLM4.5-simulated runoff in small catchment, especially for subsurface runoff.

All these above-mentioned deficiencies of CLM5 call for future model development efforts. Regional-specific agricultural management practices, grid or climate-based plant parameters, and adjusted phenology and physiology parameters (e.g., carbon allocations) for trees and grasses should be implemented. Formal calibration along with sensitivity tests which can achieve satisfactory model performance in terms of multiple land surface processes and variables are needed.

## Data Availability Statement

The version 5 of CLM (CLM5) used in this study can be downloaded from https://github.com/ESCOMP/ctsm. The data sets for the model simulations in this study are available from the open-source repository https://doi.org/10.25584/im3clm/1673776. The NLDAS-2 forcing data are derived from NLDAS Primary Forcing Data L4 Hourly 0.125° × 0.125° V002 (https://disc.gsfc.nasa.gov/datasets/NLDAS_FORA0125_H_002/summary?keywords=NLDAS2). Remote sensing and in situ data sets used in this study can be obtained using the link listed in Table 1.

**Acknowledgments**
This research was supported by the US Department of Energy, Office of Science, as part of research in MultiSector Dynamics, Earth and Environmental System Modeling Program. PNNL is operated by Battelle Memorial Institute for the US DOE under contract DE-AC05-76RLO1830. The CLM simulations were performed using computing resources of the Pacific Northwest National Laboratory (PNNL) Institutional Computing (PIC) and the National Energy Research Supercomputing Center (NERSC), which is supported by the DOE Office of Science of the US Department of Energy under contract DE-AC0205CH11231. We would like to thank Stefan Kern (University of Hamburg) for providing the GLEAM data over the CONUS. The authors wish to express their great gratitude to the three anonymous reviewers for their constructive comments and suggestions, which significantly improve this paper.

## References

Alter, R. E., Douglas, H. C., Winter, J. M., & Eltahir, E. A. B. (2018). Twentieth century regional climate change during the summer in the Central United States attributed to agricultural intensification. *Geophysical Research Letters*, *45*(3), 1586–1594. https://doi.org/10.1002/2017GL075604

Arnold, J. G., & Allen, P. M. (1999). Automated methods for estimating baseflow and ground water recharge from stream flow records. *Journal of the American Water Resources Association*, *35*(2), 13353–13366. https://doi.org/10.1111/j.1752-1688.1999.tb03599.x

Badger, A. M., & Dirmeyer, P. A. (2015). Climate response to Amazon forest replacement by heterogeneous crop cover. *Hydrology and Earth System Sciences*, *19*(11), 4547–4557. https://doi.org/10.5194/hess-19-4547-2015

Bagley, J. E., Miller, J., & Bernacchi, C. J. (2015). Biophysical impacts of climate-smart agriculture in the Midwest United States. *Plant, Cell and Environment*, *38*(9), 1913–1930. https://doi.org/10.1111/pce.12485

Berenguer, E., Ferreira, J., Gardner, T. A., Aragão, L. E. O. C., De Camargo, P. B., Cerri, C. E., et al. (2014). A large-scale field assessment of carbon stocks in human-modified tropical forests. *Global Change Biology*, *20*(12), 3713–3726. https://doi.org/10.1111/gcb.12627

Bodesheim, P., Jung, M., Gans, F., Mahecha, M. D., & Reichstein, M. (2018). Upscaled diurnal cycles of land-atmosphere fluxes: A new global half-hourly data product. *Earth System Science Data*, *10*(3), 1327–1365. https://doi.org/10.5194/essd-10-1327-2018

Bonan, G. B. (1995). Land-atmosphere interactions for climate system models: Coupling biophysical, biogeochemical, and ecosystem dynamical processes. *Remote Sensing of Environment*, *51*(1), 57–73. https://doi.org/10.1016/0034-4257(94)00065-U

Bonan, G. B., & Doney, S. (2018). Climate, ecosystems, and planetary futures: The challenge to predict life in Earth system models. *Science*, *359*(6375), eaam8328. https://doi.org/10.1126/science.aam8328

Bonan, G. B., Oleson, K. W., Vertenstein, M., Levis, S., Zeng, X., Dai, Y., et al. (2002). The land surface climatology of the Community Land Model coupled to the NCAR Community Climate Model. *Journal of Climate*, *15*(22), 3123–3149. https://doi.org/10.1175/1520-0442(2002)015<3123:TLSCOT>2.0.CO;2

Bonsor, H. C., Shamsudduha, M., Marchant, B., MacDonald, A., & Taylor, R. (2018). Seasonal and decadal groundwater changes in African sedimentary aquifers estimated using GRACE products and LSMs. *Remote Sensing*, *10*(6), 904. https://doi.org/10.3390/rs10060904

Brooks, P. D., Troch, P. A., Durcik, M., Gallo, E., & Schlegel, M. (2011). Quantifying regional scale ecosystem response to changes in precipitation: Not all rain is created equal. *Water Resources Research*, *47*(7), 1–13. https://doi.org/10.1029/2010WR009762

Brunke, M., Broxton, P., Pelletier, J., Gochis, D., Hazenberg, P., Lawrence, D., et al. (2016). Implementing and evaluating variable soil thickness in the Community Land Model, version 4.5 (CLM4.5). *Journal of Climate*, *29*(9), 3441–3461. https://doi.org/10.1175/JCLI-D-15-0307.1

Chaney, N., Huijgevoort, M., Shevliakova, E., Malyshev, S., Milly, P., Gauthier, P., & Sulman, B. (2018). Harnessing big data to rethink land heterogeneity in Earth system models. *Hydrology and Earth System Sciences*, *22*(6), 3311–3330. https://doi.org/10.5194/hess-22-3311-2018

Cheng, Y., Huang, M., Chen, M., Guan, K., Bernacchi, C., Peng, B., & Tan, Z. (2020). Parameterizing perennial bioenergy crops in Version 5 of the Community Land Model based on site-level observations in the Central Midwestern United States. *Journal of Advances in Modeling Earth Systems*, *12*(1), 1–24. https://doi.org/10.1029/2019MS001719

Cheng, Y., Ogden, F., & Zhu, J. (2017). Earthworms and tree roots: A model study of the effect of preferential flow paths on runoff generation and groundwater recharge in steep, saprolitic, tropical lowland catchments. *Water Resources Research*, *53*(7), 5400–5419. https://doi.org/10.1002/2016WR020258

Cheng, Y., Ogden, F., & Zhu, J. (2019). Characterization of sudden and sustained base flow jump hydrologic behaviour in the humid seasonal tropics of the Panama Canal Watershed. *Hydrological Processes*, *34*(3), 569–582. https://doi.org/10.1002/hyp.13604

Cheng, Y., Ogden, F., Zhu, J., & Bretfeld, M. (2018). Land use dependent preferential flow paths affect hydrological response of steep tropical lowland catchments with saprolitic soils. *Water Resources Research*, *54*(8), 5551–5566. https://doi.org/10.1029/2017WR021875

Cheruy, F., Dufresne, J. L., Hourdin, F., & Ducharne, A. (2014). Role of clouds and land-atmosphere coupling in midlatitude continental summer warm biases and climate change amplification in CMIP5 simulations. *Geophysical Research Letters*, *41*(18), 6493–6500. https://doi.org/10.1002/2014GL061145

Cohen, W. B., Maiersperger, T. K., Yang, Z., Gower, S. T., Turner, D. P., Ritts, W. D., et al. (2003). Comparisons of land cover and LAI estimates derived from ETM + and MODIS for four sites in North America : A quality assessment of 2000/2001 provisional MODIS products. *Remote Sensing of Environment*, *88*(3), 233–255. https://doi.org/10.1016/j.rse.2003.06.006

Collier, N., Hoffman, F. M., Lawrence, D. M., Keppel-Aleks, G., Koven, C. D., Riley, W. J., et al. (2018). The International Land Model Benchmarking (ILAMB) system: Design, theory, and implementation. *Journal of Advances in Modeling Earth Systems*, *10*(11), 2731–2754. https://doi.org/10.1029/2018MS001354

Cosgrove, B. A., Lohmann, D., Mitchell, K. E., Houser, P. R., Wood, E. F., Schaake, J. C., et al. (2003). Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project. *Journal of Geophysical Research: Atmospheres*, *108*(D22), 8845. https://doi.org/10.1029/2002JD003118

Cox, P. M., Betts, R. A., Bunton, C. B., Essery, R. L. H., Rowntree, P. R., & Smith, J. (1999). The impact of new land surface physics on the GCM simulation of climate and climate sensitivity. *Climate Dynamics*, *15*(3), 183–203. https://doi.org/10.1007/s003820050276

Cox, P. M., Betts, R. A., Jones, C. D., & Spall, S. A. (2000). Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature*, *408*(6809), 184–187.

Crossley, J. F., Polcher, J., Cox, P. M., Gedney, N., & Planton, S. (2000). Uncertainties linked to land-surface processes in climate change simulations. *Climate Dynamics*, *16*(12), 949–961. https://doi.org/10.1007/s003820000092

Daly, C., Neilson, R. P., & Phillips, D. L. (1994). A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *Journal of Applied Meteorology*, *33*(2), 140–158. https://doi.org/10.1175/1520-0450(1994)033<0140:ASTMFM>2.0.CO;2

Davin, E. L., Seneviratne, S. I., Ciais, P., Olioso, A., & Wang, T. (2014). Preferential cooling of hot extremes from cropland albedo management. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(27), 9757–9761. https://doi.org/10.1073/pnas.1317323111

Decker, M., & Zeng, X. (2009). Impact of modified Richards equation on global soil moisture simulation in the Community Land Model (CLM3.5). *Journal of Advances in Modeling Earth Systems*, *1*(3). https://doi.org/10.3894/JAMES.2009.1.5

Devanand, A., Huang, M., Ashfaq, M., Barik, B., & Ghosh, S. (2019). Choice of irrigation water management practice affects Indian summer monsoon rainfall and its extremes. *Geophysical Research Letters*, *46*(15), 9126–9135. https://doi.org/10.1029/2019GL083875

Devanand, A., Huang, M., Lawrence, D. M., & Zarzycki, C. M. (2020). Land use and land cover change strongly modulates land-atmosphere coupling and warm-season precipitation over the Central United States in CESM2-VR. *Journal of Advances in Modeling Earth Systems*, *12*(9), e2019MS001925. https://doi.org/10.1029/2019MS001925

Dickinson, R. E. (1983). Land surface processes and climate—surface albedos and energy balance. *Advances in Geophysics*, *25*(C), 305–353. https://doi.org/10.1016/S0065-2687(08)60176-4

Dirmeyer, P. A., Chen, L., Wu, J., Shin, C. S., Huang, B., Cash, B. A., et al. (2018). Verification of land –atmosphere coupling in forecast models, reanalyses, and land surface models using flux site observations. *Journal of Hydrometeorology*, *19*(2), 375–392. https://doi.org/10.1175/JHM-D-17-0152.1

Duan, Q., Schaake, J., Andreassian, V., Franks, S., Goteti, G., Gupta, H. V., et al. (2006). Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops. *Journal of Hydrology*, *320*(1–2), 3–17. https://doi.org/10.1016/j.jhydrol.2005.07.031

Fan, Y., Clark, M., Lawrence, D. M., Swenson, S., Band, L. E., & Brantley, S. L. (2019). Hillslope hydrology in global change research and Earth system modeling. *Water Resources Research*, *55*(2), 1737–1772. https://doi.org/10.1029/2018WR023903

Fatichi, S., Or, D., Walko, R., Vereecken, H., Young, M. H., Ghezzehei, T. A., et al. (2020). Soil structure is an important omission in Earth System Models. *Nature Communications*, *11*(1), 1–11. https://doi.org/10.1038/s41467-020-14411-z

Foley, J. A., DeFries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., et al. (2005). Global consequences of land use. *Science*, *309*(5734), 570. https://doi.org/10.1126/science.1111772

Getirana, A. C., Dutra, E., Guimberteau, M., Kam, J., Li, H. Y., Decharme, B., et al. (2014). Water balance in the Amazon basin from a land surface model ensemble. *Journal of Hydrometeorology*, *15*(6), 2586–2614. https://doi.org/10.1175/JHM-D-14-0068.1

Gochis, D. J., Vivoni, E. R., & Watts, C. J. (2010). The impact of soil depth on land surface energy and water fluxes in the North American Monsoon region. *Journal of Arid Environments*, *74*(5), 564–571. https://doi.org/10.1016/j.jaridenv.2009.11.002

Green, J. K., Seneviratne, S. I., Berg, A. M., Findell, K. L., Hagemann, S., Lawrence, D. M., & Gentine, P. (2019). Large influence of soil moisture on long-term terrestrial carbon uptake. *Nature*, *565*(7740), 476–479. https://doi.org/10.1038/s41586-018-0848-x

Gulden, L. E., Rosero, E., Yang, Z. L., Rodell, M., Jackson, C. S., Niu, G. Y., et al. (2007). Improving land-surface model hydrology: Is an explicit aquifer model better than a deeper soil profile? *Geophysical Research Letters*, *34*(9), 1–5. https://doi.org/10.1029/2007GL029804

Haddeland, I., Clark, D. B., Franssen, W., Ludwig, F., Voß, F., Arnell, N. W., et al. (2011). Multimodel estimate of the global terrestrial water balance: Setup and first results. *Journal of Hydrometeorology*, *12*(5), 869–884. https://doi.org/10.1175/2011JHM1324.1

He, X., Kim, H., Kirstetter, P. E., Yoshimura, K., Chang, E. C., Ferguson, C. R., et al. (2015). The diurnal cycle of precipitation in regional spectral model simulations over West Africa: Sensitivities to resolution and cumulus schemes. *Weather and Forecasting*, *30*(2), 424–445. https://doi.org/10.1175/WAF-D-14-00013.1

Henderson-Sellers, A., Pitman, A. J., Love, P. K., Irannejad, P., & Chen, T. H. (1995). The project for intercomparison of land surface parameterization schemes (PILPS): Phases 2 and 3. *Bulletin of the American Meteorological Society*, *76*(4), 489–504. https://doi.org/10.1175/1520-0477(1995)076%3C0489:TPFIOL%3E2.0.CO;2

Hou, Z., Huang, M., Leung, L. R., Lin, G., & Ricciuto, D. M. (2012). Sensitivity of surface flux simulations to hydrologic parameters based on an uncertainty quantification framework applied to the Community Land Model. *Journal of Geophysical Research: Atmospheres*, *117*(D15), 1–18. https://doi.org/10.1029/2012JD017521

Huang, M., Hou, Z., Leung, L. R., Ke, Y., Liu, Y., Fang, Z., & Sun, Y. (2013). Uncertainty analysis of runoff simulations and parameter identifiability in the Community Land Model: Evidence from MOPEX basins. *Journal of Hydrometeorology*, *14*(6), 1754–1772. https://doi.org/10.1175/JHM-D-12-0138.1

Jung, M., Reichstein, M., & Bondeau, A. (2009). Towards global empirical upscaling of FLUXNET eddy covariance observations: Validation of a model tree ensemble approach using a biosphere model. *Biogeosciences*, *6*(10), 2001–2013. https://doi.org/10.5194/bg-6-2001-2009

Jung, M., Reichstein, M., Ciais, P., Seneviratne, S., Sheffield, J., Goulden, M., et al. (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., Cescatti, A., Richardson, A., Arain, A., et al. (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: Biogeosciences*, *116*(3), 1–16. https://doi.org/10.1029/2010JG001566

Kenny, J. F., Barber, N. L., Hutson, S. S., Linsey, K. S., Lovelace, J. K., & Maupin, M. A. (2009). Estimated use of water in the United States in 2005. *US Geological Survey Circular*, *No. 1344*, . http://water.usgs.gov/watuse

Klein, S. A., Jiang, X., Boyle, J., Malyshev, S., & Xie, S. (2006). Diagnosis of the summertime warm and dry bias over the U.S. Southern Great Plains in the GFDL climate model using a weather forecasting approach. *Geophysical Research Letters*, *33*(18), 1–6. https://doi.org/10.1029/2006GL027567

Koster, R. D., Guo, Z., Bonan, G., Chan, E., & Cox, P. (2014). Regions of strong coupling between soil moisture and precipitation. *Science*, *1138*(2004), 10–13. https://doi.org/10.1126/science.1100217

Koven, C. D., Hugelius, G., Lawrence, D. M., & Wieder, W. R. (2017). Higher climatological temperature sensitivity of soil carbon in cold than warm climates. *Nature Climate Change*, *7*(11), 817–822. https://doi.org/10.1038/nclimate3421

Landerer, F. W., & Swenson, S. C. (2012). Accuracy of scaled GRACE terrestrial water storage estimates. *Water Resources Research*, *48*(4), W04531. https://doi.org/10.1029/2011WR011453

Lawrence, D., Fisher, R., Koven, C., Oleson, K., Swenson, S., Bonan, G., et al. (2019). The Community Land Model version 5: Description of new features, benchmarking, and impact of forcing uncertainty. *Journal of Advances in Modeling Earth Systems*, *11*(12), 4245–4287. https://doi.org/10.1029/2018MS001583

Lawrence, D., Fisher, R., Koven, C., Oleson, K., Swenson, S., & Vertenstein, M. (2018a). *Technical description of version 5.0 of the Community Land Model (CLM)*. NCAR/TN-478+STR NCAR Technical Note (p. 350). Boulder, CO: National Center for Atmospheric Research (NCAR).Retrieved from https://doi.org/10.5065/D6RR1W7M

Lawrence, D., Oleson, K., Flanner, M., Thornton, P., Swenson, S., Peter, J., et al. (2011). Parameterization improvements and functional and structural advances in version 4 of the Community Land Model. *Journal of Advances in Modeling Earth Systems*, *3*(1), M03001. https://doi.org/10.1029/2011MS000045

Lawrence, P., & Chase, T. (2007). Representing a new MODIS consistent land surface in the Community Land Model (CLM 3.0). *Journal of Geophysical Research: Biogeosciences*, *112*(G1), G01023. https://doi.org/10.1029/2006JG000168

Lawrence, P., & Chase, T. (2010). Investigating the climate impacts of global land cover change in the community climate system model. *International Journal of Climatology*, *30*(13), 2066–2087. https://doi.org/10.1002/joc.2061

Lawrence, P., Feddema, J., Bonan, G., Meehl, G., O'Neill, B., Oleson, K., et al. (2012). Simulating the biogeochemical and biogeophysical impacts of transient land cover change and wood harvest in the Community Climate System Model (CCSM4) from 1850 to 2100. *Journal of Climate*, *25*(9), 3071–3095. https://doi.org/10.1175/JCLI-D-11-00256.1

Lawrence, P., Lawrence, D., & Hurtt, G. (2018b). Attributing the carbon cycle impacts of CMIP5 historical and future land use and land cover change in the Community Earth System Model (CESM1). *Journal of Geophysical Research: Biogeosciences*, *123*(5), 1732–1755. https://doi.org/10.1029/2017JG004348

Lei, H., Huang, M., Leung, L. R., Yang, D., Shi, X., Mao, J., et al. (2014). Sensitivity of global terrestrial gross primary production to hydrologic states simulated by the Community Land Model using two runoff parameterizations. *Journal of Advances in Modeling Earth Systems*, *6*(3), 658–679. https://doi.org/10.1002/2013MS000252

Leng, G., Huang, M., Tang, Q., Sacks, W. J., Lei, H., & Leung, L. R. (2013). Modeling the effects of irrigation on land surface fluxes and states over the conterminous United States: Sensitivity to input data and model parameters. *Journal of Geophysical Research: Atmospheres*, *118*(17), 9789–9803. https://doi.org/10.1002/jgrd.50792

Leng, G., Leung, L. R., & Huang, M. (2017). Significant impacts of irrigation water sources and methods on modeling irrigation effects in the ACME Land Model. *Journal of Advances in Modeling Earth Systems*, *9*(3), 1665–1683. https://doi.org/10.1002/2016MS000885

Levis, S., Badger, A., Drewniak, B., Nevison, C., & Ren, X. (2018). CLMcrop yields and water requirements: Avoided impacts by choosing RCP 4.5 over 8.5. *Climatic Change*, *146*(3–4), 501–515. https://doi.org/10.1007/s10584-016-1654-9

Levis, S., Gordon, B. B., Kluzek, E., Thornton, P. E., Jones, A., Sacks, W. J., & Kucharik, C. J. (2012). Interactive crop management in the Community Earth System Model (CESM1): Seasonal influences on land-atmosphere fluxes. *Journal of Climate*, *25*(14), 4839–4859. https://doi.org/10.1175/JCLI-D-11-00446.1

Li, H., Leung, R., Getirana, A., Huang, M., Wu, H., Xu, Y., et al. (2015). Evaluating global streamflow simulations by a physically based routing model coupled with the Community Land Model. *Journal of Hydrometeorology*, *16*(2), 948–971. https://doi.org/10.1175/JHM-D-14-0079.1

Li, X., & Xiao, J. (2019). Mapping photosynthesis solely from solar-induced chlorophyll fluorescence : A global, fine-resolution dataset of gross primary production derived from OCO-2. *Remote Sensing*, *11*(21), 2563. https://doi.org/10.3390/rs11212563

Liang, X., & Xie, Z. (2003). Important factors in land-atmosphere interactions: Surface runoff generations and interactions between surface and groundwater. *Global and Planetary Change*, *38*(1–2), 101–114. https://doi.org/10.1016/S0921-8181(03)00012-2

Lin, Y., Dong, W., Zhang, M., Xie, Y., Xue, W., Huang, J., & Luo, Y. (2017). Causes of model dry and warm bias over central U.S. and impact on climate projections. *Nature Communications*, *8*(1), 1–8. https://doi.org/10.1038/s41467-017-01040-2

Liu, M., Adam, J.C., Richey, A. S., Zhu, Z., & Myneni, R.B. (2018). Factors controlling changes in evapotranspiration, runoff, and soil moisture over the conterminous U.S. Accounting for vegetation dynamics. *Journal of Hydrology*, *565*(July), 123–137. https://doi.org/10.1016/j.jhydrol.2018.07.068

Lombardozzi, D. L., Bonan, G. B., Wieder, W., Grandy, A. S., Morris, C., & Lawrence, D. L. (2018). Cover crops may cause winter warming in snow-covered regions. *Geophysical Research Letters*, *45*(18), 9889–9897. https://doi.org/10.1029/2018GL079000

Lyne, V. D., & Hollick, M. (1979). Stochastic time-variable rainfall-runoff modelling. In *Institute of Engineers Australia National Conference* (pp. 89–93). Barton, Australia: Institute of Engineers Australia.

Ma, H., Klein, S., Xie, S., Zhang, C., Tang, S., Tang, Q., et al. (2018). CAUSES: On the role of surface energy budget errors to the warm surface air temperature error over the Central United States. *Journal of Geophysical Research: Atmospheres*, *123*(5), 2888–2909. https://doi.org/10.1002/2017JD027194

Ma, N., Niu, G. Y., Xia, Y., Cai, X., Zhang, Y., Ma, Y., & Fang, Y. (2017). A systematic evaluation of Noah-MP in simulating land-atmosphere energy, water, and carbon exchanges over the continental United States. *Journal of Geophysical Research: Atmospheres*, *122*(22), 12245–12268. https://doi.org/10.1002/2017JD027597

Mahmood, R., Pielke, R. A., Hubbard, K. G., Niyogi, D., Dirmeyer, P. A., Mcalpine, C., et al. (2014). Land cover changes and their biogeophysical effects on climate. *International Journal of Climatology*, *34*(4), 929–953. https://doi.org/10.1002/joc.3736

Mahowald, N. M., Randerson, J. T., Lindsay, K., Munoz, E., Doney, S. C., Lawrence, P., et al. (2016). Interactions between land use change and carbon cycle feedbacks. *Global Biogeochemical Cycles*, *31*(1), 96–113. https://doi.org/10.1002/2016GB005374

Martens, B., Miralles, D. G., Lievens, H., Van Der Schalie, R., De Jeu, R. A. M., Fernández-Prieto, D., et al. (2017). GLEAM v3: Satellite-based land evaporation and root-zone soil moisture. *Geoscientific Model Development*, *10*(5), 1903–1925. https://doi.org/10.5194/gmd-10-1903-2017

Mcguire, A. D., Lawrence, D. M., Koven, C., Clein, J. S., Burke, E., & Chen, G. (2018). Dependence of the evolution of carbon dynamics in the northern permafrost region on the trajectory of climate change. *Proceedings of the National Academy of Sciences*, *115*(15), 3882–3887. https://doi.org/10.1073/pnas.1719903115

Mei, R., & Wang, G. (2012). Summer land-atmosphere coupling strength in the United States: Comparison among observations, reanalysis data, and numerical models. *Journal of Hydrometeorology*, *13*(3), 1010–1022. https://doi.org/10.1175/JHM-D-11-075.1

Michel, D., Jiménez, C., Miralles, D. G., Jung, M., Hirschi, M., Ershadi, A., et al. (2016). The WACMOS-ET project – Part 1: Tower-scale evaluation of four remote-sensing-based evapotranspiration algorithms. *Hydrology and Earth System Sciences*, *20*(2), 803–822. https://doi.org/10.5194/hess-20-803-2016

Miralles, D. G., Holmes, T. R. H., De Jeu, R. A. M., Gash, J. H., Meesters, A. G. C. A., & Dolman, A. J. (2011). Global land-surface evaporation estimated from satellite-based observations. *Hydrology and Earth System Sciences*, *15*(2), 453–469. https://doi.org/10.5194/hess-15-453-2011

Miralles, D. G., Jiménez, C., Jung, M., Michel, D., Ershadi, A., Mccabe, M. F., et al. (2016). The WACMOS-ET project – Part 2: Evaluation of global terrestrial evaporation data sets. *Hydrology and Earth System Sciences*, *20*(2), 823–842. https://doi.org/10.5194/hess-20-823-2016

Mizukami, N., Clark, M. P., Sampson, K., Nijssen, B., Mao, Y., McMillan, H., et al. (2016). MizuRoute version 1: A river network routing tool for a continental domain water resources applications. *Geoscientific Model Development*, *9*(6), 2223–2228. https://doi.org/10.5194/gmd-9-2223-2016

Morcrette, C., Van Weverberg, K., Ma, H., Ahlgrimm, M., Bazile, E., Berg, L., et al. (2018). Introduction to CAUSES: Description of weather and climate models and their near-surface temperature errors in 5 day hindcasts near the Southern Great Plains. *Journal of Geophysical Research: Atmospheres*, *123*(5), 2655–2683. https://doi.org/10.1002/2017JD027199

Mu, Q., Jones, L. A., Kimball, J. S., McDonald, K. C., & Running, S. W. (2009). Satellite assessment of land surface evapotranspiration for the pan-Arctic domain. *Water Resources Research*, *45*(9), 1–20. https://doi.org/10.1029/2008WR007189

Mu, Q., Zhao, M., Heinsch, F. A., Liu, M., Tian, H., & Running, S. W. (2007). Evaluating water stress controls on primary production in biogeochemical and remote sensing based models. *Journal of Geophysical Research: Biogeosciences*, *112*(1), 1–13. https://doi.org/10.1029/2006JG000179

Mu, Q., Zhao, M., & Running, S. W. (2011). Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sensing of Environment*, *115*(8), 1781–1800. https://doi.org/10.1016/j.rse.2011.02.019

Mueller, N., Rhines, A., Butler, E., Ray, D., Siebert, S., Holbrook, M., & Huybers, P. (2017). Global relationships between cropland intensification and summer temperature extremes over the last 50 years. *Journal of Climate*, *30*(18), 7505–7528. https://doi.org/10.1175/JCLI-D-17-0096.1

Nathan, R. J., & McMahon, T. A. (1990). Evaluation of automated techniques for base flow and recession analyses. *Water Resources Research*, *26*(7), 1465–1473. https://doi.org/10.1029/WR026i007p01465

Niu, G. Y., Yang, Z. L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., et al. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research: Atmospheres*, *116*(12), 1–19. https://doi.org/10.1029/2010JD015139

Oleson, K., Lawrence, D. M., Bonan, G. B., Drewniak, B., Huang, M., Koven, C. D., et al. (2013). *Technical description of version 4.5 of the Community Land Model (CLM)* (No. NCAR/TN-503+STR). Boulder, CO: National Center for Atmospheric Research (NCAR). https://doi.org/10.5065/D6RR1W7M

Oleson, K., Niu, G., Yang, Z., Lawrence, D., Thornton, P., Lawrence, P., et al. (2008). Improvements to the community land model and their impact on the hydrological cycle. *Journal of Geophysical Research: Biogeosciences*, *113*(G1), G01021. https://doi.org/10.1029/2007JG000563

Pelletier, J. D., Broxton, P. D., Hazenberg, P., Zeng, X., Troch, P. A., Niu, G. Y., et al. (2016). A gridded global data set of soil, intact regolith, and sedimentary deposit thicknesses for regional and global land surface modeling. *Journal of Advances in Modeling Earth Systems*, *10*(11), 2731–2754. https://doi.org/10.1002/2013MS000282

Peng, B., Guan, K., Chen, M., Lawrence, D. M., Pokhrel, Y., Suyker, A., et al. (2018). Improving maize growth processes in the community land model: Implementation and evaluation. *Agricultural and Forest Meteorology*, *250–251*(May 2017), 64–89. https://doi.org/10.1016/j.agrformet.2017.11.012

Pielke, R. A., Pitman, A., Niyogi, D., Mahmood, R., McAlpine, C., Hossain, F., et al. (2011). Land use/land cover changes and climate: Modeling analysis and observational evidence. *Wiley Interdisciplinary Reviews: Climate Change*, *2*(6), 828–850. https://doi.org/10.1002/wcc.144

Pitman, A. J., De Noblet-Ducoudré, N., Cruz, F. T., Davin, E. L., Bonan, G. B., Brovkin, V., et al. (2009). Uncertainties in climate responses to past land cover change: First results from the LUCID intercomparison study. *Geophysical Research Letters*, *36*(14), 1–6. https://doi.org/10.1029/2009GL039076

Portmann, F. T., Siebert, S., & Döll, P. (2010). MIRCA2000—Global monthly irrigated and rainfed crop areas around the year 2000 : A new high-resolution data set for agricultural and hydrological modeling. *Global Biogeochemical Cycles*, *24*(1), GB1011. https://doi.org/10.1029/2008GB003435

Qian, Y., Huang, M., Yang, B., & Berg, L. K. (2013). A modeling study of irrigation effects on surface fluxes and land–air–cloud interactions in the Southern Great Plains. *Journal of Hydrometeorology*, *14*(3), 700–721. https://doi.org/10.1175/JHM-D-12-0134.1

Rashid, M., Chien, R. Y., Ducharne, A., Kim, H., Yeh, P. J. F., Peugeot, C., et al. (2019). Evaluation of Groundwater Simulations in Benin from the ALMIP2 Project. *Journal of Hydrometeorology*, *20*(2), 339–354. https://doi.org/10.1175/JHM-D-18-0025.1

Ren, H., Hou, Z., Huang, M., Bao, J., Sun, Y., Tesfa, T., & Leung, L. R. (2016). Classification of hydrological parameter sensitivity and evaluation of parameter transferability across 431 US MOPEX basins. *Journal of Hydrology*, *536*, 92–108. https://doi.org/10.1016/j.jhydrol.2016.02.042

Robock, A., Luo, L., Wood, E. F., Wen, F., Mitchell, K. E., Houser, P. R., et al. (2003). Evaluation of the North American Land Data Assimilation System over the southern Great Plains during the warm season. *Journal of Geophysical Research: Atmospheres*, *108*(22), 8846. https://doi.org/10.1029/2002jd003245

Robock, A., Vinnikov, K. Y., Srinivasan, G., Entin, J. K., Hollinger, S. E., Speranskaya, N. A., et al (2000). The global soil moisture data bank. *Bulletin of the American Meteorological Society*, *81*(6), 1281–1299. https://doi.org/10.1175/1520-0477(2000)081<1281:TGSMDB>2.3.CO;2

Rodell, M., Chen, J., Kato, H., Famiglietti, J., Nigro, J., & Wilson, C. (2007). Estimating groundwater storage changes in the Mississippi River basin (USA) using GRACE. *Hydrogeology Journal*, *15*(1), 159–166. https://doi.org/10.1007/s10040-006-0103-7

Rodell, M., & Famiglietti, J. (2002). The potential for satellite-based monitoring of groundwater storage changes using GRACE: The High Plains aquifer, Central US. *Journal of Hydrology*, *263*(1–4), 245–256. https://doi.org/10.1016/S0022-1694(02)00060-4

Rodell, M., Houser, P. R., Jambor, U. E. A., Gottschalck, J., Mitchell, K., Meng, C. J., et al. (2004). The global land data assimilation system. *Bulletin of the American Meteorological Society*, *85*(3), 381–394. https://doi.org/10.1175/BAMS-85-3-381

Sacks, W. J., Deryng, D., & Foley, J. A. (2010). Crop planting dates: An analysis of global patterns. *Global Ecology and Biogeography*, *19*(5), 607–620. https://doi.org/10.1111/j.1466-8238.2010.00551.x

Scanlon, B. R., Zhang, Z., Rateb, A., Sun, A., Wiese, D., Save, H., et al. (2019). Tracking seasonal fluctuations in land water storage using global models and GRACE satellites. *Geophysical Research Letters*, *46*(10), 5254–5264. https://doi.org/10.1029/2018GL081836

Schaefer, G. L., Cosh, M. H., & Jackson, T. J. (2007). The USDA natural resources conservation service soil climate analysis network (SCAN). *Journal of Atmospheric and Oceanic Technology*, *24*(12), 2073–2077. https://doi.org/10.1175/2007JTECHA930.1

Schaphoff, S., Lucht, W., Gerten, D., Sitch, S., Cramer, W., & Prentice, I. C. (2006). Terrestrial biosphere carbon storage under alternative climate projections. *Climatic Change*, *74*(1–3), 97–122. https://doi.org/10.1007/s10584-005-9002-5

Scott, B. L., Ochsner, T. E., Illston, B. G., Basara, J. B., & Sutherland, A. J. (2013). New soil property database improves Oklahoma Mesonet soil moisture estimates. *Journal of Atmospheric and Oceanic Technology*, *30*(11), 2585–2595. https://doi.org/10.1175/JTECH-D-13-00084.1

Searchinger, T. D., Wirsenius, S., Beringer, T., & Dumas, P. (2018). Assessing the efficiency of changes in land use for mitigating climate change. *Nature*, *564*(7735), 249–253. https://doi.org/10.1038/s41586-018-0757-z

Shiklomanov, I. A. (2000). Appraisal and assessment of world water resources. *Water International*, *25*(1), 11–32. https://doi.org/10.1080/02508060008686794

Siebert, S., Döll, P., Hoogeveen, J., Faures, J., Frenken, K., Feick, S., et al. (2005). Development and validation of the global map of irrigation areas. *Hydrology and Earth System Sciences Discussions, European Geosciences Union*, *2*(4), 1299–1327. https://hal.archives-ouvertes.fr/hal-00298682

Song, F., Feng, Z., Ruby Leung, L., Houze, R. A., Wang, J., Hardin, J., & Homeyer, C. R. (2019). Contrasting spring and summer large-scale environments associated with mesoscale convective systems over the U.S. Great Plains. *Journal of Climate*, *32*(20), 6749–6767. https://doi.org/10.1175/JCLI-D-18-0839.1

Sun, Z., Wang, Q., Ouyang, Z., Watanabe, M., & Matsushita, B. (2007). Evaluation of MOD16 algorithm using MODIS and ground observational data in winter wheat field in North China Plain. *Hydrological Processes: An International Journal*, *21*(9), 1196–1206. https://doi.org/10.1002/hyp

Swenson, S., Clark, M., Fan, Y., Lawrence, D., & Perket, J. (2019). Representing intrahillslope lateral subsurface flow in the Community Land Model. *Journal of Advances in Modeling Earth Systems*, *11*(12), 4044–4065. https://doi.org/10.1029/2019MS001833

Swenson, S., & Lawrence, D. (2014). Assessing a dry surface layer-based soil resistance parameterization for the Community Land Model using GRACE and FLUXNET-MTE data. *Journal of Geophysical Research: Atmospheres*, *119*(17), 10299–10312. https://doi.org/10.1002/2014JD022314

Swenson, S., & Lawrence, D. (2015). A GRACE-based assessment of interannual groundwater dynamics in the Community Land Model. *Water Resources Research*, *51*(11), 8817–8833. https://doi.org/10.1002/2015WR017582

Swenson, S., Lawrence, D., & Lee, H. (2012). Improved simulation of the terrestrial hydrological cycle in permafrost regions by the Community Land Model. *Journal of Advances in Modeling Earth Systems*, *4*(3), 1–15. https://doi.org/10.1029/2012MS000165

Swenson, S., Yeh, P., Wahr, J., & Famiglietti, J. (2006). A comparison of terrestrial water storage variations from GRACE with in situ measurements from Illinois. *Geophysical Research Letters*, *33*(16), L16401. https://doi.org/10.1029/2006GL026962

Syed, T. H., Famiglietti, J. S., Rodell, M., Chen, J., & Wilson, C. R. (2008). Analysis of terrestrial water storage changes from GRACE and GLDAS. *Water Resources Research*, *44*(2), W02433. https://doi.org/10.1029/2006WR005779

Tapley, B. D., Bettadpur, S., Watkins, M., & Reigber, C. (2004). The gravity recovery and climate experiment: Mission overview and early results. *Geophysical Research Letters*, *31*(9), 1–4. https://doi.org/10.1029/2004GL019920

Thiery, W., Davin, E. L., Lawrence, D. M., Hirsch, A. L., Hauser, M., & Seneviratne, S. I. (2017). Present-day irrigation mitigates heat extremes. *Journal of Geophysical Research: Atmospheres*, *122*(3), 1403–1422. https://doi.org/10.1002/2016JD025740

Thornton, P. E., Lamarque, J. F., Rosenbloom, N. A., & Mahowald, N. M. (2007). Influence of carbon-nitrogen cycle coupling on land model response to $CO_2$ fertilization and climate variability. *Global Biogeochemical Cycles*, *21*(4), 1–15. https://doi.org/10.1029/2006GB002868

Turner, D. P., Ritts, W. D., Cohen, W. B., Maeirsperger, T. K., Gower, S. T., Kirschbaum, A. A., et al. (2005). Site-level evaluation of satellite-based global terrestrial gross primary production and net primary production monitoring. *Global Change Biology*, *11*(4), 666–684. https://doi.org/10.1111/j.1365-2486.2005.00936.x

Unger, N. (2014). Human land-use-driven reduction of forest volatiles cools global climate. *Nature Climate Change*, *4*(10), 907–910. https://doi.org/10.1038/nclimate2347

Van Weverberg, K., Morcrette, C. J., Petch, J., Klein, S. A., Ma, H. Y., Zhang, C., et al. (2018). CAUSES: Attribution of surface radiation biases in NWP and climate models near the U.S. Southern Great Plains. *Journal of Geophysical Research: Atmospheres*, *123*(7), 3612–3644. https://doi.org/10.1002/2017JD027188

Voepel, H., Ruddell, B., Schumer, R., Troch, P. A., Brooks, P. D., Neal, A., et al. (2011). Quantifying the role of climate and landscape characteristics on hydrologic partitioning and vegetation response. *Water Resources Research*, *47*(8), 1–13. https://doi.org/10.1029/2010WR009944

Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., et al. (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: Atmospheres*, *117*(D3), D03109. https://doi.org/10.1029/2011JD016048

Xia, Y., Mocko, D., Huang, M., Li, B., Rodell, M., Mitchell, K. E., et al. (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

Xia, Y., Mocko, D., Wang, S., Pan, M., Kumar, S., Peters-Lidard, C., et al. (2018). Comprehensive evaluation of the variable infiltration capacity (VIC) model in the North American Land Data Assimilation System. *Journal of Hydrometeorology*, *19*(11), 1853–1879. https://doi.org/10.1175/JHM-D-18-0139.1

Yang, Z., Qian, Y., Liu, Y., Berg, L. K., Hu, H., Dominguez, F., et al. (2019). Irrigation impact on water and energy cycle during dry years over the United States using convection-permitting WRF and a dynamical recycling model. *Journal of Geophysical Research: Atmospheres*, *124*(21), 11220–11241. https://doi.org/10.1029/2019JD030524

Zhang, L., Lei, H., Shen, H., Cong, Z., Yang, D., & Liu, T. (2019). Evaluating the Representation of Vegetation Phenology in the Community Land Model 4.5 in a Temperate Grassland. *Journal of Geophysical Research: Biogeosciences*, *124*(2), 187–210. https://doi.org/10.1029/2018JG004866

Zhao, M., Heinsch, F. A., Nemani, R. R., & Running, S. W. (2005). Improvements of the MODIS terrestrial gross and net primary production global data set. *Remote Sensing of Environment*, *95*(2), 164–176. https://doi.org/10.1016/j.rse.2004.12.011

Zhao, M., & Running, S. W. (2006). Sensitivity of Moderate Resolution Imaging Spectroradiometer (MODIS) terrestrial primary production to the accuracy of meteorological reanalyses. *Journal of Geophysical Research*, *111*(G1), 1–13. https://doi.org/10.1029/2004JG000004

Zhao, M., & Running, S. W. (2010). Drought-induced reduction in global terrestrial net primary production from 2000 through 2009. *Science*, *329*(5994), 940–943. https://doi.org/10.1126/science.1192666

Zheng, H., & Yang, Z.-L. (2016). Effects of soil-type datasets on regional terrestrial water cycle simulations under different climatic regimes. *Journal of Geophysical Research: Atmospheres*, *121*(24), 14387–14402. https://doi.org/10.1002/2016JD025187

Zheng, H., Yang, Z.-L., Lin, P., Wei, J., Wu, W., Li, L., et al. (2019). On the sensitivity of the precipitation partitioning into evapotranspiration and runoff in land surface parameterizations. *Water Resources Research*, *55*(1), 95–111. https://doi.org/10.1029/2017WR022236

Zhu, B., Huang, M., Cheng, Y., Xie, X., Liu, Y., Zhang, X., et al. (2020). Effects of irrigation on water, carbon, and nitrogen budgets in a semi-arid watershed in the Pacific Northwest: A modelling study. *Journal of Advances in Modeling Earth Systems*, *12*(9), e2019MS001953. https://doi.org/10.1029/2019ms001953