An ensemble of 48 physically perturbed model estimates of the 1/8° terrestrial water budget over the conterminous United States, 1980–2015

Hui Zheng[1], Wenli Fei[1,2], Zong-Liang Yang[3], Jiangfeng Wei[3,4], Long Zhao[3,5], Lingcheng Li[3,6], and Shu Wang[7]

[1]Key Laboratory of Regional Climate-Environment Research for Temperate East Asia, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, 100029, China
[2]College of Earth and Planetary Sciences, University of Chinese Academy of Sciences, Beijing, 100049, China
[3]Department of Geological Sciences, John A. and Katherine G. Jackson School of Geosciences, the University of Texas at Austin, Austin, Texas, 78705, USA
[4]Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters/Key Laboratory of Meteorological Disaster, Ministry of Education/International Joint Research Laboratory on Climate and Environment Change, Nanjing University of Information Science and Technology, Nanjing, 210044, China
[5]School of Geographical Sciences, Southwest University, Chongqing, 400715, China
[6]Pacific Northwest National Laboratory, Richland, Washington, 99354, USA
[7]State Key Laboratory of Operation and Control of Renewable Energy & Storage Systems, China Electric Power Research Institute, Beijing, 100192, China

Correspondence: Zong-Liang Yang (liang@jsg.utexas.edu)

Abstract.

Terrestrial water budget (TWB) data over large domains are of high interest for various hydrological applications. Spatiotemporally continuous and physically consistent estimations of TWB rely on land surface models (LSMs). As an augmentation of the operational North American Land Data Assimilation System Phase 2 (NLDAS-2) four-LSM ensemble, this paper describes a dataset simulated from an ensemble of 48 physics configurations of the Noah LSM with multi-physics options (Noah-MP). The 48 Noah-MP physics configurations are selected to give a representative cross-section of commonly used LSMs for parameterizing runoff, atmospheric surface layer turbulence, soil moisture limitation on photosynthesis, and stomatal conductance.

The dataset spans from 1980 to 2015 over the conterminous United States (CONUS) at a monthly temporal resolution and a 1/8° spatial resolution. The dataset variables include total evapotranspiration and its constituents (canopy evaporation, soil evaporation, and transpiration), runoff (the surface and subsurface components), as well as terrestrial water storage (snow water equivalent, four-layer soil water content from the surface down to 2 m, and the groundwater storage anomaly). The dataset is available at https://doi.org/10.5281/zenodo.7109816 (Zheng et al., 2022). Evaluations carried out in this study and previous investigations show that the ensemble performs well in reproducing the observed terrestrial water storage, snow water equivalent, soil moisture, and runoff. Noah-MP complements the NLDAS models well, and adding Noah-MP consistently improves the NLDAS estimations of the above variables in most areas of CONUS. Besides, the perturbed-physics ensemble facilitates the identification of model deficiencies. The parameterizations of shallow snow, spatially varying groundwater dynamics, and near-surface atmospheric turbulence should be improved in future model versions.

1 Introduction

Estimates of the terrestrial water budgets (TWBs)—evapotranspiration, runoff, terrestrial water storage, and their constituents—over continental domains are of high interest for a broad range of hydrological applications. Publicly available data have been applied to investigate the state of the terrestrial water cycle (Trenberth and Fasullo, 2013a; Rodell et al., 2015; Scanlon et al., 2018; Yin and Roderick, 2020); to understand the interactions among hydrological processes, vegetation, climate, and human activities (Trenberth and Fasullo, 2013b; LaFontaine et al., 2015; Ward et al., 2014; Levia et al., 2020); to examine the availability and variability of water resources and use (Wu et al., 2021; Hejazi et al., 2014; Scanlon et al., 2012; Voss et al., 2013; Lv et al., 2019; Le et al., 2011; Rodell et al., 2009); and to assess the risk of extreme events such as droughts (Peters-Lidard et al., 2021; Prudhomme et al., 2014; Dai, 2013; Su et al., 2021) and floods (Emerton et al., 2017; Lin et al., 2018).

As the applications have expanded, the availability of TWB estimates has increased rapidly (Peters-Lidard et al., 2018; Saxe et al., 2021; Zhang et al., 2018). Commonly used estimation methods include remote sensing, in situ observations, and model simulations (Saxe et al., 2021; McCabe et al., 2017; Pan et al., 2012; Gao et al., 2010; Trenberth et al., 2007). Among these methods, land surface models (LSMs) are apt for continuously producing physically consistent TWBs over a large domain and long period, and their characteristics are particularly favorable for certain circumstances. For instance, LSMs can estimate various TWB components simultaneously; whereas, for some components, such as runoff (Lin et al., 2019; Beck et al., 2017), root-zone soil moisture (Xia et al., 2015b, a), and transpiration (Lian et al., 2018), direct remote sensing is either unavailable or highly uncertain. Additionally, LSMs are valuable in remote or topographically complex regions because of the sparseness of in situ observations (Kim et al., 2021). Estimations based on remote sensing and in situ observations are often impeded by scale mismatches and observation gaps, whereas these issues are rarely an impairment for LSM simulations. Besides, LSM simulations can complement remote sensing and in situ observations well. Combinations of estimates from different techniques can improve the estimation accuracy (Zhang et al., 2018; Pan et al., 2012; Zhao and Yang, 2018), while comparisons between model-simulated estimates and observations can reveal the impacts of human activities (Zaussinger et al., 2019) and underground processes (Zheng et al., 2020).

Several operational LSM simulation systems have been set up over different regions of the globe (Xia et al., 2019; Shi et al., 2011; Carrera et al., 2015; Rodell et al., 2004). The systems combine an ensemble of LSMs to utilize the competitive strengths of different LSMs and eliminate the weakness associated with individual ones. Among them, the North American Land Data Assimilation System (NLDAS) (Xia et al., 2012a, b; Mitchell et al., 2004) stands as a pioneering and successful one. The NLDAS Phase 2 (NLDAS-2) operates over the conterminous United States (CONUS) from 1979 to near real time at a spatial resolution of 1/8°. The system generates a set of multi-source synthesized data of surface meteorology, vegetation, and soils, and uses them to drive an ensemble of four different LSMs. The four LSMs—namely Noah version 2.8 (Ek et al., 2003; Chen and Dudhia, 2001a, b; Chen et al., 1997), Variable Infiltration Capacity (VIC) version 4.0.3 (Liang et al., 1994), Mosaic (Koster and Suarez, 1992), and Sacramento Soil Moisture Accounting (SAC) model (Burnash et al., 1973), were selected to give a good cross-section of the diverse range of LSMs (Mitchell et al., 2004). The models have varying strengths and weaknesses in process parameterizations and modeling skills (Kumar et al., 2017). A combination of multiple models

can produce an aggregated estimate that outperforms most of the individual constituents (Fei et al., 2021; Beck et al., 2017; Guo et al., 2007; Ajami et al., 2007). An ensemble can also quantify the estimation uncertainty resulting from different model formulations (Troin et al., 2021; Cloke and Pappenberger, 2009). Evaluations of the NLDAS-2 four-LSM ensemble estimates have shown satisfactory performance in matching the observed evapotranspiration (ET) (Zhang et al., 2020; Xia et al., 2012b; Kumar et al., 2018), runoff (Xia et al., 2012a), and soil moisture (Xia et al., 2015b, a).

We have enriched the NLDAS-2 four-model ensemble with 48 perturbed-physics configurations of the Noah LSM with multi-physics options (Noah-MP) (Fei et al., 2021; Zheng et al., 2020, 2019). Noah-MP has more physically realistic representations of vertical stratification than the NLDAS-2 models have. A column of land in Noah-MP consists of a vegetation canopy layer, three snowpack layers, four soil layers, and a groundwater component (Niu et al., 2011; Yang et al., 2011). Conceptual (e.g., the five water tanks of SAC) and lumped (e.g., the combined vegetation–soil surface layer of Noah) representations of the stratification of vegetation and soil, as used in the NLDAS-2 models, are minimized. Moreover, Noah-MP has a more comprehensive representation of various land surface processes that are evident at different depths. The modeled processes include snow accumulation and ablation, infiltration, percolation, retention, freeze–thaw of snow or soil water, groundwater recharge/discharge, and energy constraints (Niu et al., 2011). These improvements in vertical stratification and process parameterizations are expected to better estimate TWBs. Indeed, previous comparisons between Noah-MP and the four NLDAS-2 LSMs have shown that Noah-MP is comparable or better when it comes to estimating soil moisture (Cai et al., 2014b), runoff (Cai et al., 2014a; Fei et al., 2021), and ET (Zhang et al., 2020).

Our enrichment also features a single-model perturbed-physics ensemble, which is different from the widely used multi-model ensemble approach. The Noah-MP ensemble is constructed by shuffling the available parameterization options of several selected processes. The ensemble size grows exponentially as a multiplication of the available parameterization options of different processes (Yang et al., 2011; Zhang et al., 2016; Gan et al., 2019). A large ensemble should give a broad cross-section of feasible model formulations to account for the model uncertainty in TWB estimation (Telteu et al., 2021; Mitchell et al., 2004) and is critical for a statistically reliable estimation of the probability of hydrological events such as floods and droughts (Troin et al., 2021). The single-model perturbed-physics ensemble also facilitates uncertainty attribution and reduction. The ensemble consists of pairs that are different in the parameterization of one process and the same for another. Variance analysis of the ensemble can quantify the contribution of the parameterization of a process and compare the relative importance of two processes (Zheng et al., 2019; Clark et al., 2011). Such quantification could inform further model development to reduce the model uncertainty. However, there are also pitfalls unique to the single-model perturbed-physics ensemble. Fei et al. (2021) found that the ensemble members generated by naive perturbation of the Noah-MP physics are not independent enough from each other. The low independence hinders the skill gained from the ensembling method. The finding suggests that advanced techniques of physics perturbations should be developed to maximize the ensemble skill and minimize the ensemble size. An open-accessible dataset should facilitate the research that leverages the advantages and addresses the issues of the perturbed-physics ensemble.

We have previously evaluated the runoff and compared it with NLDAS-2 (Fei et al., 2021). This paper describes the estimation of all the TWB variables, along with the description of the spread among the ensemble members, the difference with the

Table 1. The dataset variables.

Symbol	Units	Description
	surface water budget	
E	$\mathrm{kg\,m^{-2}\,s^{-1}}$	total evapotranspiration
E_{can}	$\mathrm{kg\,m^{-2}\,s^{-1}}$	evaporation of canopy interception
E_{gnd}	$\mathrm{kg\,m^{-2}\,s^{-1}}$	direct evaporation from the ground
E_{tran}	$\mathrm{kg\,m^{-2}\,s^{-1}}$	transpiration
R	$\mathrm{kg\,m^{-2}\,s^{-1}}$	total runoff
R_{srf}	$\mathrm{kg\,m^{-2}\,s^{-1}}$	surface runoff
R_{und}	$\mathrm{kg\,m^{-2}\,s^{-1}}$	subsurface runoff
W	$\mathrm{kg\,m^{-2}}$	terrestrial water storage
W_{snow}	$\mathrm{kg\,m^{-2}}$	snow water equivalent
W_{gw}	$\mathrm{kg\,m^{-2}}$	groundwater storage
$w_{soil,i}$	$\mathrm{m^3\,m^{-3}}$	volumetric soil water content
z_{snow}	m	snow depth
	auxiliary variables	
X	-	land–water mask

NLDAS models, and the performance with reference to various observations are presented. The paper is organized as follows. Section 2 presents the information necessary for using the dataset, including the dataset variables, file organization, and the source data and models used for data generation. Section 3 describes the intercomparison methods, evaluation metrics, and reference datasets. Section 4 presents the results and discussion. Finally, after stating the data availability in Section 5, Section 6 draws conclusions.

2 Data description

The dataset contains gridded water budget variables over CONUS. Section 2.1 describes the dataset variables and their physical relationships. The 48 Noah-MP physics configurations used to create the dataset are detailed in Section 2.2. Section 2.3 briefly covers the atmospheric forcing, the static parameters of vegetation and soil, and the simulation settings.

2.1 Dataset variables

Table 1 lists the dataset variables. The variables are available at each 1/8° grid point in NLDAS-2, indicated by a land–water mask (X). The surface water budgets of each grid cell are represented as follows.

Neglecting horizontal water exchange between adjacent grids, the water budget closure can be obtained among the precipitation (P; $\mathrm{kg\,m^{-2}\,s^{-1}}$), ET ($E$), runoff ($R$), and terrestrial water storage change ($W' = \frac{dW}{dt}$; $\mathrm{kg\,m^{-2}\,s^{-1}}$) (Zheng et al., 2020):

$$P = E + R + W', \tag{1}$$

where precipitation (P) is from NLDAS-2 (described in Section 1) and used as the model input.

Noah-MP resolves the components of the water budget closure equation (1). ET (E) consists of canopy evaporation (E_{can}), ground evaporation (E_{gnd}), and transpiration (E_{tran}):

$$E = E_{can} + E_{gnd} + E_{tran}. \tag{2}$$

Runoff (R) has a surface (R_{srf}) and subsurface (R_{sub}) component:

$$R = R_{srf} + R_{sub}. \tag{3}$$

Terrestrial water storage (TWS; W) is the sum of snow water equivalent (SWE; W_{snow}), groundwater storage in unconfined aquifers (W_{gw}), and soil water content in the four model layers ($W_{soil,i}$):

$$W = W_{snow} + W_{gw} + \sum_{i=1}^{N_{soil}} W_{soil,i}, \tag{4}$$

where $N_{soil} = 4$ is the number of soil layers. Soil water storage ($W_{soil,i}$) is not included in the dataset but can be calculated from the volumetric water content ($w_{soil,i}$) as follows:

$$W_{soil,i} = \rho_{wat} w_{soil,i} \Delta z_{soil,i} \quad \text{for } i = 1, \cdots, N_{soil}, \tag{5}$$

where $\rho_{wat} = 1000\,\mathrm{kg\,m^{-3}}$ is the water density; and $\Delta z_{soil,1} = 0.1\,\mathrm{m}$, $\Delta z_{soil,2} = 0.3\,\mathrm{m}$, $\Delta z_{soil,3} = 0.6\,\mathrm{m}$, and $\Delta z_{soil,4} = 1\,\mathrm{m}$ are the thicknesses of the four soil layers.

2.2 The 48 Noah-MP physics configurations

The Noah-MP LSM version 3.6 is used. We perturbed the parameterization of runoff, stomatal conductance, soil moisture stress factor, and near-surface atmospheric turbulence. The processes are selected as they directly control runoff generation and evapotranspiration. Their importance has been shown in global simulations (Yang et al., 2011). The perturbation creates an ensemble of 48 members (48 = 4 runoff $\times$ 2 stomatal conductance $\times$ 3 soil moisture stress $\times$ 2 turbulence). Limited by computational resources, we did not perturb the parameterization of the cryosphere hydrological processes such as snow albedo (Chen et al., 2014; He et al., 2019) and rain–snow partitioning (Wang et al., 2019), which may limit the usage of the dataset in cryosphere hydrology. Appendix A details the formulation of and parameters in the selected parameterization. The parameters use the Noah-MP default values.

2.3 Domain, temporal span, atmospheric forcings, and static parameters

The simulation domain covers the CONUS (25°–53°N, 125°–67°W) at a spatial resolution of 0.125°.

The hourly NLDAS-2 atmospheric forcings were used to drive the 48 Noah-MP configurations. This study used seven forcing variables: downward solar radiation, downward longwave radiation, air temperature, surface pressure, specific humidity, wind speed, and precipitation rate. The static datasets, including topography (https://ldas.gsfc.nasa.gov/nldas/elevation), predominant vegetation class (https://ldas.gsfc.nasa.gov/nldas/vegetation-class), and soil texture type (https://ldas.gsfc.nasa.gov/nldas/soils), are also the same as those in NLDAS-2. We used the default Noah-MP lookup tables to convert the input vegetation and soil types to parameter values.

The simulation spans 36 years from January 1980 to December 2015 at a time step of 15 minutes. The initial states on 1 January 1980 were obtained by cycling the year 1979 one hundred times.

3 Intercomparison and evaluation methods

The evaluations and intercomparisons in this paper are performed for 12 River Forecast Centers (RFCs): Northeast (NE), Mid-Atlantic (MA), Ohio (OH), Lower Mississippi (LM), Southeast (SE), North Central (NC), Northwest (NW), Arkansas (AB), Missouri (MB), West Gulf (WG), California–Nevada (CN), and Colorado (CB). Figure S1 displays the geographical delineation of the RFCs. More details on the RFCs, such as their multi-year average precipitation, potential evaporation, and topography, can be found in Fei et al. (2021, Figure 1).

The intercomparison and evaluations were conducted at different timescales—the long-term climatological mean, annual cycle, and interannual anomaly. Section 3.1 details how the temporal variations and ensemble spread are derived for the timescales. We utilized the Taylor diagram and Taylor skill score (TSS) to measure the performance of Noah-MP against various reference datasets. The evaluation methods are shown in Section 3.2, and the reference datasets are described in Section 3.4. In addition to the intercomparisons and evaluations, Section 3.3 introduces the Sobol' sensitivity index for the process sensitivity analysis.

3.1 Temporal variability and ensemble spread

We calculated the total temporal variability (σ_{total}), the variability of the annual cycle (σ_{ancy}), and the interannual variability (σ_{anom}) for each ensemble member and the ensemble arithmetic mean following Dirmeyer et al. (2006):

$$\sigma_{total} = \sqrt{\frac{1}{12Y}\sum_{y,m}(x_{y,m} - x_{clim})^2}, \tag{6}$$

$$\sigma_{ancy} = \sqrt{\frac{1}{12}\sum_{m}(x_{ancy,m} - x_{clim})^2}, \tag{7}$$

$$\sigma_{anom} = \sqrt{\frac{1}{12Y}\sum_{y,m}(x_{y,m} - x_{ancy,m})^2}, \tag{8}$$

where x_{clim}, x_{ancy}, and x_{anom} are the climatology, annual cycle, and interannual anomaly of the monthly time series $x_{y,m}$ (month m of the year y; $m = 1, \cdots, 12$; $y = 1, \cdots, Y$).

The ensemble spread for the total time series ($\breve{\sigma}_{total}$), annual cycle ($\breve{\sigma}_{ancy}$), and interannual anomaly ($\breve{\sigma}_{anom}$) are derived also following Dirmeyer et al. (2006):

$$\breve{\sigma}_{total} = \frac{1}{T}\sum_{t=1}^{T}\sigma(x_{y,m}), \tag{9}$$

$$\breve{\sigma}_{ancy} = \frac{1}{12}\sum_{m=1}^{12}\sigma(x_{ancy}), \tag{10}$$

$$\breve{\sigma}_{anom} = \frac{1}{12Y}\sum_{y=1}^{Y}\sum_{m=1}^{12}\sigma(x_{anom}), \tag{11}$$

where $\sigma(x) = \sqrt{[\sum_{n=1}^{N}(x-\breve{x})^2]/N}$ denotes the standard deviation across the ensemble members at each time step. $\breve{x}$ denotes the arithmetic ensemble mean.

We calculated the ratio ($r = \breve{\sigma}/\sigma$) of ensemble spread $\breve{\sigma}$ to temporal variability σ for the total monthly time series (with a subscript of "total"), annual cycle ("ancy"), and interannual anomaly ("anom"). The ratio enables the intercomparison of the ensemble spread among the selected timescales. A grade is assigned according to the ratio following Dirmeyer et al. (2006): a grade "A" for $r < 0.316$; "B" for $0.316 \leq r < 1$; "C" for $1 \leq r < 3.16$; "D" for $3.16 \leq r < 10$; and "E" for $r > 10$. A lower (higher) r or a higher (lower) grade denotes a lower (higher) ensemble spread.

3.2 Taylor diagram and skill score

The Taylor diagram (Taylor, 2001) is a graphical representation of how closely a model simulation matches observations in terms of correlation coefficient (R), normalized unbiased root mean square error (nuRMSE), and normalized standard deviation ($\hat{\sigma}_f$). The TSS is an index that measures the distance between a model simulation and the observations in the Taylor diagram. The TSS is defined as follows:

$$TSS = \frac{4(1+R)}{(\hat{\sigma}_f + \frac{1}{\hat{\sigma}_f})^2(1+R_0)} \tag{12}$$

$$\hat{\sigma}_f = \frac{\sigma_f}{\sigma_o} \tag{13}$$

where σ_f and σ_o are the standard deviations of the model simulation and the observation, and R_0 is the maximum correlation coefficient attainable (in this study, $R_0 = 1$). The value range of TSS is from 0 to 1. A higher TSS indicates a higher overall performance of model prediction with reference to the observations.

3.3 Sobol's total sensitivity index

The sensitivity of the Noah-MP ensemble to a physical process can be quantified by the Sobol' total sensitivity index (Sobol′, 1993; Zheng et al., 2019). The Sobol' total sensitivity index measures the proportion of the variance of different processes to the total variance, which is defined as follows:

$$S_\Lambda = \frac{\mathrm{E}_{\sim\Lambda}(\mathrm{Var}_\Lambda(Y|\sim\Lambda))}{\mathrm{Var}(Y)}, \tag{14}$$

where S_Λ is the Sobol' total sensitivity index for one process Λ; $\sim\Lambda$ represents the other processes except for Λ; Y represents the 48 Noah-MP ensemble members; $\mathrm{Var}(Y)$ is the variance of Y; $\mathrm{Var}_\Lambda(Y|\sim\Lambda)$ denotes the variance among different parameterization schemes of the process Λ, and $\mathrm{E}(\sim\Lambda)$ denotes the arithmetic average across all combinations of the other processes except for Λ. Detailed calculations and examples can be found in Zheng et al. (2019, Appendix A).

3.4 Reference data

3.4.1 Terrestrial water storage

We used the $1^\circ\times1^\circ$ monthly Gravity Recovery and Climate Experiment (GRACE) land water-equivalent-thickness surface-mass anomaly, level-3, Release 6.0, version 04, as the reference of TWS (W in Table 1). The GRACE products from different processing centers are slightly different. To reduce the noise of the estimates (Sakumura et al., 2014), we used the arithmetic average of the products from three centers: (1) GeoForschungsZentrum Potsdam (or the German Research Center for Geosciences) (https://podaac.jpl.nasa.gov/dataset/TELLUS_GRAC_L3_GFZ_RL06_LND_v04), (2) the Center for Space Research at the University of Texas, Austin (https://podaac.jpl.nasa.gov/dataset/TELLUS_GRAC_L3_CSR_RL06_LND_v04), and (3) NASA's Jet Propulsion Laboratory (https://podaac.jpl.nasa.gov/dataset/TELLUS_GRAC_L3_JPL_RL06_LND_v04). The GRACE satellites began orbiting Earth on 17 March 2002. We selected the period from 2003 to 2015 for the evaluation. There are 14 missing values during the period, which were filled with a simple linear interpolation.

The GRACE products experience signal leakages between land and water (Save et al., 2016). Such signal leakage could impact the estimation in the RFCs that are adjacent to the Great Lakes (Ma et al., 2017) and oceans. To alleviate the impacts of the Great Lakes, we corrected the GRACE TWS estimates in the NC, OH, and NE RFCs using the water level variations observed by the National Oceanic and Atmospheric Administration (NOAA). Appendix B details the algorithm.

3.4.2 Soil moisture

We used the daily North American Soil Moisture Database (NASMD, http://soilmoisture.tamu.edu) (Quiring et al., 2016) as the reference for the simulated soil moisture ($W_{soil,i}$ in Table 1), similar to previous NLDAS evaluations (Xia et al., 2015b, a). NASMD assembles the soil moisture time series at multiple depths of more than 2200 stations of 24 networks with quality control. The observation depth varies with the network. We interpolated the observations to the centers of the Noah-MP soil layers, which are $0.05\,\mathrm{m}$, $0.25\,\mathrm{m}$, $0.7\,\mathrm{m}$, and $1.5\,\mathrm{m}$, respectively. The interpolation is performed only when a valid observation is exactly at, or two valid observations exist above and below, the given depth; otherwise, a missing value is given. We excluded the soil layers with more than 50% missing values to minimize the impacts of missing values on the evaluation, after which 264 $w_{soil,1}$, 214 $w_{soil,2}$, 95 $w_{soil,3}$, and 23 $w_{soil,4}$ valid time series remained. Daily data from 1996 to 2013 were then aggregated into monthly values. For any month, if less than 10 days of valid data are available, a missing value is assigned.

3.4.3 Snow water equivalent

We used the daily Snow Data Assimilation System (SNODAS; https://nsidc.org/data/G02158/versions/1) as the reference of SWE (W_{snow} in Table 1). SNODAS is a data assimilation system developed by the NOAA National Weather Service's National Operational Hydrologic Remote Sensing Center. This system aims to provide a physically consistent framework to combine snow modeling and observations from satellites, airborne platforms, and ground stations (National Operational Hydrologic Remote Sensing Center, 2004). The original spatial resolution is 1 km$\times$1 km, and we aggregated the data to the 0.125° NLDAS grids. SNODAS began on 2003-09-28, and we selected the period from September 2004 to August 2015 that spans 11 whole snow seasons. Clow et al. (2012) showed that the SNODAS SWE performs well in the forest areas of the Colorado Rocky Mountains but performs poorly in the alpine areas.

3.4.4 Evapotranspiration

We used plot-scale AmeriFlux observations and two gridded products as the reference ET (E in Table1). The two gridded products are derived from different methods, and a common evaluation period from 1982 to 2015 is selected for this study. The gridded products have different spatial resolutions. We downscaled the data to the NLDAS grids and then aggregated them to the 12 RFCs.

We selected 25 AmeriFlux sites (https://ameriflux.lbl.gov). The 25 selected sites were selected because they have the longest observation periods for seven major land cover types (i.e., evergreen forest, deciduous forest, mixed forest, shrubland, savanna, grassland, and cropland). Figure S1 and Table S1 detail the selected sites. The data have been widely used in LSM evaluations (Cai et al., 2014a; Zhang et al., 2020). AmeriFlux provides hourly or half-hourly latent heat measurements. We converted the measurement to ET by dividing the latent heat of water vaporization ($2.5104 \times 10^6\,\mathrm{J\,kg^{-1}}$). The hourly or half-hourly ET is then aggregated into monthly values. In the process of aggregation, if there were less than 8 valid hours in a day, a missing day was marked; if there were fewer than 10 valid days in a month, a missing month was assigned; if there was less than 50% valid months, the time series was dropped. In this study, the data serve as the ground truth for the gridded ET products and model estimation.

The first gridded ET product is FLUXNET multi-tree ensemble (MTE) (Jung et al., 2009) (https://www.bgc-jena.mpg.de/geodb/projects/Home.php). FLUXNET MTE ET is a monthly data produced from the FLUXNET eddy covariance measurements, remote sensing, and meteorological data using the multi-tree ensemble statistical method (Jung et al., 2009). The product is widely used in LSM evaluations (Cai et al., 2014a; Ma et al., 2017; Xia et al., 2016; Jung et al., 2019; Fang et al., 2020; Zhang et al., 2020; Pan et al., 2020). FLUXNET MTE ET has a spatial resolution of 0.5°$\times$0.5°. We remap the data to the 0.125°$\times$0.125° NLDAS grids with a first-order conservative method. We use the FLUXNET ET as the reference for the annual cycle, since it replicates the AmeriFlux observations best among the two gridded products (Table S2–S4).

The second gridded ET product is the Global Land Evaporation Amsterdam Model (GLEAM), version 3.3a (https://www.gleam.eu), which is another widely used ET product (Xu et al., 2019). GLEAM estimates transpiration, canopy evaporation, soil evaporation, open-water evaporation, and sublimation separately and then sums them as ET. The method aims to maximize

the utilization of satellite information. The product estimates monthly ET at a spatial resolution of $0.25° \times 0.25°$. We bilinearly interpolated the data to the NLDAS grids. The GLEAM ET is used as the reference for the interannual anomaly, as it better replicates the AmeriFlux-observed anomaly than FLUXNET (Table S2–S4).

3.5 NLDAS ensemble

We used three NLDAS-2 models—namely Noah-2.8, VIC-4.0.3, and Mosaic—as the benchmark of the Noah-MP ensemble. Their outputs can be publicly downloaded from the NASA Goddard Earth Sciences Data and Information Services Center (https://disc.gsfc.nasa.gov/datasets?keywords=NLDAS). More information on the NLDAS-2 models, the forcing and static datasets, and simulation settings can be found in Xia et al. (2012b, a). The NLDAS-2 datasets have been proven to perform soundly for regional hydrological simulations (Xia et al., 2012b, a, 2016, 2015b, a) and are widely used as a benchmark for LSM evaluations (Cai et al., 2014b; Fei et al., 2021).

4 Results and discussions

In this section, we begin by intercomparing the ensemble spread of the dataset variables (Section 4.1). Then, Sections 4.2–4.5 examines the performance and ensemble spread of the TWS anomaly (TWSA), SWE, soil moisture, and ET, in comparison with the NLDAS ensemble. The performance of runoff can be found in our previous paper Fei et al. (2021) and Zheng et al. (2020).

4.1 Comparison of the ensemble spread

We aggregated the dataset variables across CONUS. Table 2 summarizes the ensemble spread and temporal variability of the ensemble mean for all the dataset variables. Among the variables, runoff (including surface and subsurface components) shows the largest spread, accounting for one-fifth to one-third of the climatological mean. The ensemble spread is comparable to or larger than the temporal variability. The magnitude is similar to that estimated in GSWP-2. The spread of ET and TWS is significantly smaller than those observed in GSWP-2. The high consistency among the Noah-MP configurations could be a sign of the limited sampling of available process parameterizations but could also be a result of continuous model improvements. ET might be the former case, since the ensemble does not perturb several important processes such as roughness sublayer (Abolafia-Rosenzweig et al., 2021) and plant hydraulics (Li et al., 2021). TWS is likely the latter case, since the parameters of different groundwater schemes are calibrated using GRACE (Niu et al., 2007). The spread of snow water equivalent depicts the smallest spread, which is 2.5% of the climatological mean. The small spread reflects a limited sampling of the cryosphere processes as discussed in Section 2.2. For soil moisture, the ensemble spread is marginal at the surface and increases significantly in the deep layers. The difference hints that the controlling processes vary with depth. The surface soil moisture is tightly controlled by the atmospheric forcings, whereas the spread of the subsurface soil moisture hints at the complex interplay among various land surface processes (e.g., root water uptake and subsurface runoff) (Koster, 2015).

Table 2. The climatological mean, ensemble spread, and temporal variability of different dataset variables. $\bar{\bar{x}}$ denotes the climatological average of the ensemble mean. $\breve{\sigma}_{ancy}$, $\breve{\sigma}_{anom}$, and $\breve{\sigma}_{total}$ denote the spread of the 48 Noah-MP configurations in simulating the multi-year averaged annual cycle, interannual anomaly, and total 36-year monthly time series, respectively. σ_{ancy}, σ_{anom}, and σ_{total} denote the temporal variability of the annual cycle, interannual anomaly, and the total time series, respectively. The rating of the ratio $\breve{\sigma}/\sigma$ is defined in Section 3.1. The units of the variables are presented in Table 1. $\tilde{W}$ ($\tilde{W}_{gw}$) denotes the terrestrial water storage (groundwater) anomaly ($\mathrm{kg\,m^{-2}}$), whereas W' (W'_{gw}) denotes the monthly terrestrial water storage (groundwater) change ($\mathrm{kg\,m^{-2}\,s^{-1}}$).

Variables	$\breve{\sigma}_{ancy}$	$\breve{\sigma}_{anom}$	$\breve{\sigma}_{total}$	$\bar{\bar{x}}$	$\frac{\breve{\sigma}_{total}}{\bar{\bar{x}}}$ (%)	σ_{ancy}	σ_{anom}	σ_{total}	$r \sim \frac{\breve{\sigma}_{ancy}}{\sigma_{ancy}}$	$r \sim \frac{\breve{\sigma}_{anom}}{\sigma_{anom}}$	$r \sim \frac{\breve{\sigma}_{total}}{\sigma_{total}}$
E ($\times 10^{-6}$)	1.19	0.266	1.22	17.5	7.0	11.8	1.18	11.8	A~0.102	A~0.225	A~0.103
E_{can} ($\times 10^{-6}$)	0.207	0.0417	0.207	1.52	14	1.11	0.308	1.15	A~0.187	A~0.136	A~0.180
E_{gnd} ($\times 10^{-6}$)	0.733	0.143	0.743	7.96	9.3	3.64	0.778	3.72	A~0.201	A~0.185	A~0.200
E_{tran} ($\times 10^{-6}$)	1.33	0.206	1.35	7.99	17	7.87	0.666	7.90	A~0.170	A~0.310	A~0.171
R ($\times 10^{-6}$)	1.22	0.369	1.26	6.73	19	3.13	0.369	1.26	B~0.390	A~0.206	B~0.349
R_{srf} ($\times 10^{-6}$)	0.682	0.225	0.701	2.09	34	0.690	0.578	0.900	B~0.988	B~0.389	B~0.778
R_{sub} ($\times 10^{-6}$)	1.00	0.343	1.04	4.64	22	2.47	1.31	2.80	B~0.406	A~0.261	B~0.371
W' ($\times 10^{-6}$)	1.10	0.347	1.14	-0.0331	/	8.71	2.29	9.00	A~0.126	A~0.152	A~0.127
W'_{gw} ($\times 10^{-6}$)	0.144	0.116	0.186	$-0.005\,71$	/	1.59	0.630	1.71	A~0.091	A~0.184	A~0.109
$\tilde{W}$	5.57	3.43	6.48	0	/	45.0	18.1	48.4	A~0.124	A~0.190	A~0.134
$\tilde{W}_{gw}$	0.671	0.855	1.09	0	/	8.18	6.11	10.2	A~0.082	A~0.140	A~0.106
W_{snow}	0.125	0.127	0.173	6.84	2.5	7.56	3.80	8.47	A~0.017	A~0.033	A~0.020
$w_{soil,1}$ ($\times 10^{-3}$)	6.66	0.947	6.65	227	2.9	24.2	10.2	26.2	A~0.273	A~0.093	A~0.254
$w_{soil,2}$ ($\times 10^{-3}$)	8.40	1.24	8.48	239	3.5	19.4	8.14	21.0	B~0.434	A~0.152	B~0.404
$w_{soil,3}$ ($\times 10^{-3}$)	11.9	1.83	12.1	233	5.2	22.7	8.78	24.4	B~0.525	A~0.208	B~0.496
$w_{soil,4}$ ($\times 10^{-3}$)	14.6	1.84	14.7	250	5.9	16.0	7.09	17.5	B~0.911	A~0.259	B~0.839

Table 2 also shows the comparison between the annual cycle and interannual anomaly for different variables. For runoff and soil moisture, the ensemble spread in the annual cycle is larger (i.e., $\breve{\sigma}_{ancy} \geq \sigma_{anom}$). Whereas for transpiration (E_{tran}), groundwater storage ($\hat{W}_{gw}$ and W'_{gw}), and snow water equivalent (W_{snow}), the spread is larger for the interannual anomaly (i.e., $\breve{\sigma}_{ancy} \leq \sigma_{anom}$). The dynamics at different times are modulated by different processes (Dickinson et al., 2003). Since the timescale of the interannual anomaly is smaller than that of the annual cycle, the land surface memory and the associated land model parameterizations may contribute a larger part to the annual cycle, whereas the interannual anomaly is more tightly controlled by atmospheric forcing. The smaller ensemble spread in the annual cycle than in the interannual anomaly is a sign of insufficient representation of the physics uncertainty.

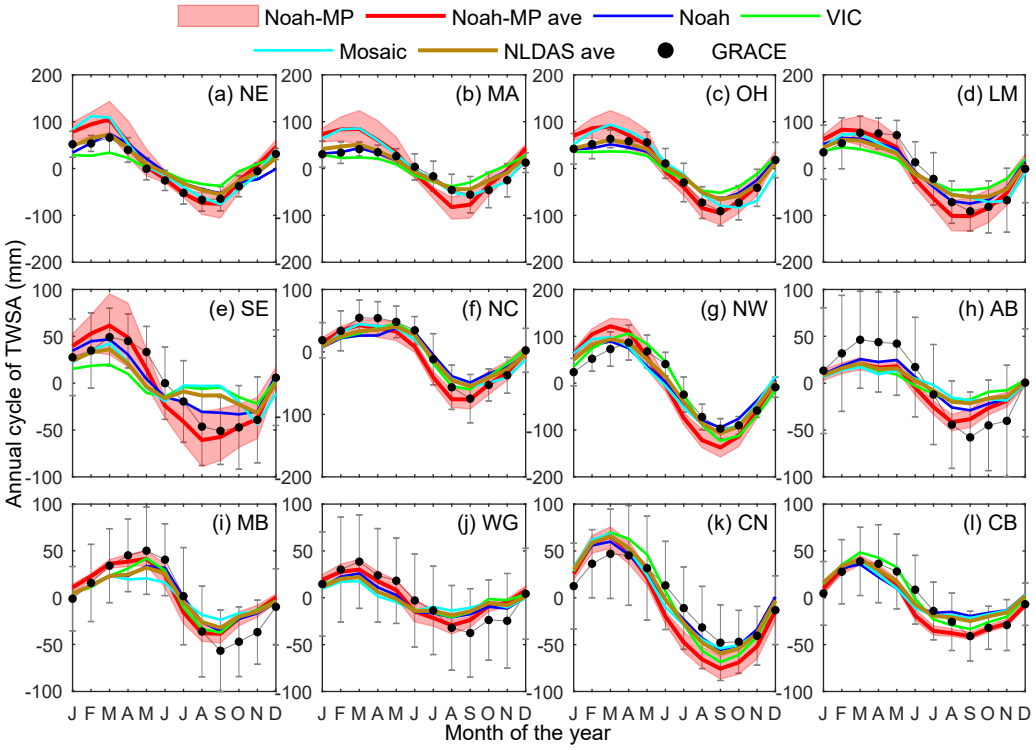

Figure 1. Annual cycle of TWSA from the model estimates and GRACE in the 12 RFCs for the period 2003–2015. The TWSA is calculated from the monthly TWS by subtracting the 13-year average (2003 to 2015). Black dots denote the GRACE observations. Error bars show the standard deviation of the year-to-year differences. The shaded areas denote the range between the maxima and minima of the 48 Noah-MP estimates. The solid red line denotes the Noah-MP multi-physics ensemble mean. The three NLDAS models (Noah, Mosaic, VIC) and their ensemble means are denoted by the blue, green, cyan, and dark golden lines, respectively. The 12 RFCs are sorted based on climatic aridity, i.e., the most humid RFC in the top left and the driest RFC in the bottom right.

4.2 Terrestrial water storage anomaly

Figure 1 shows the annual cycle of the TWSA estimated from GRACE, Noah-MP, and NLDAS. Figure 2 presents the TSS. In the 12 RFCs over CONUS, the TWSA peaks in spring, declines rapidly in summer, reaches a minimum in autumn, and recovers in winter. In terms of the timing of the peak and trough, Noah-MP and the NLDAS models perform similarly. In terms of the amplitude of variation, Noah-MP generally produces higher values in all RFCs. Previous studies have attributed this difference to the inclusion of a bucket groundwater component in Noah-MP (Cai et al., 2014b; Ma et al., 2017). However, we found the Noah-MP configurations without a groundwater component can still produce a higher amplitude, especially considering the structural similarity between Noah-MP and Noah. Further investigation of the model difference is necessary.

Figure 2 shows the Taylor diagram for the annual cycle of TWSA. The Noah-MP configurations generally outperform the NLDAS models in most of the RFCs, which results in the superior performance of the ensemble mean (shown in Table S5).

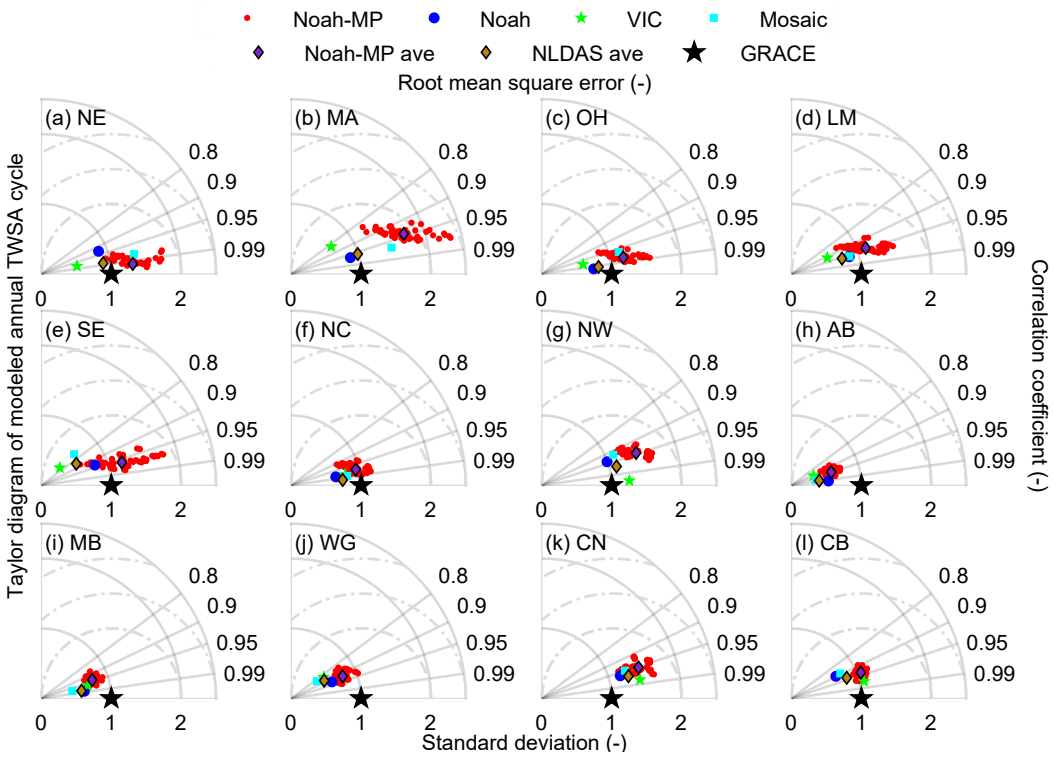

Figure 2. Taylor diagram of TWSA's annual cycle for the 48 Noah-MP configurations (red dots) and three NLDAS models (blue dots for Noah, green stars for VIC, and cyan square for Mosaic) at the 12 RFCs. Black stars denote the observations. The ensemble mean of Noah-MP and NLDAS is presented by purple and dark golden diamonds, respectively. The distance between the model and observation presents the nuRMSE. The radial lines show the correlation coefficient, while the distance to the origin along the line denotes the normalized variability.

Detailed examination of the TSS reveals that Noah-MP and NLDAS have similar correlation coefficients. Their difference is manifested in the modeled standard deviation (i.e., the amplitude of variation). In NE, MA, NW, and CN, Noah-MP underperforms compared with NLDAS, mainly due to overestimating the standard deviation. However, the interpretation of the overestimation is multifaceted. First, Noah-MP could overestimate the variability due to unsuitable parameters. For instance, specific yield is an important parameter for the simulation of groundwater storage and water level (Lv et al., 2021). The parameter is calibrated as 0.2 by global simulations (Niu et al., 2007). The globally calibrated value may be overestimated in the RFCs, leading to an overestimation in TWSA. Second, the GRACE data could underestimate the temporal variability at these coastal RFCs due to signal leakage from the ocean (Cai et al., 2014b). In AB and MB, Noah-MP performs slightly better than the NLDAS models, but both underestimate the standard deviation. The Ogallala Aquifer encompasses the two RFCs. Noah-MP can better present the groundwater changes than the NLDAS models due to an unconfined aquifer module. But Noah-MP does not include the confined aquifers, leading to an underestimation of the groundwater storage variability.

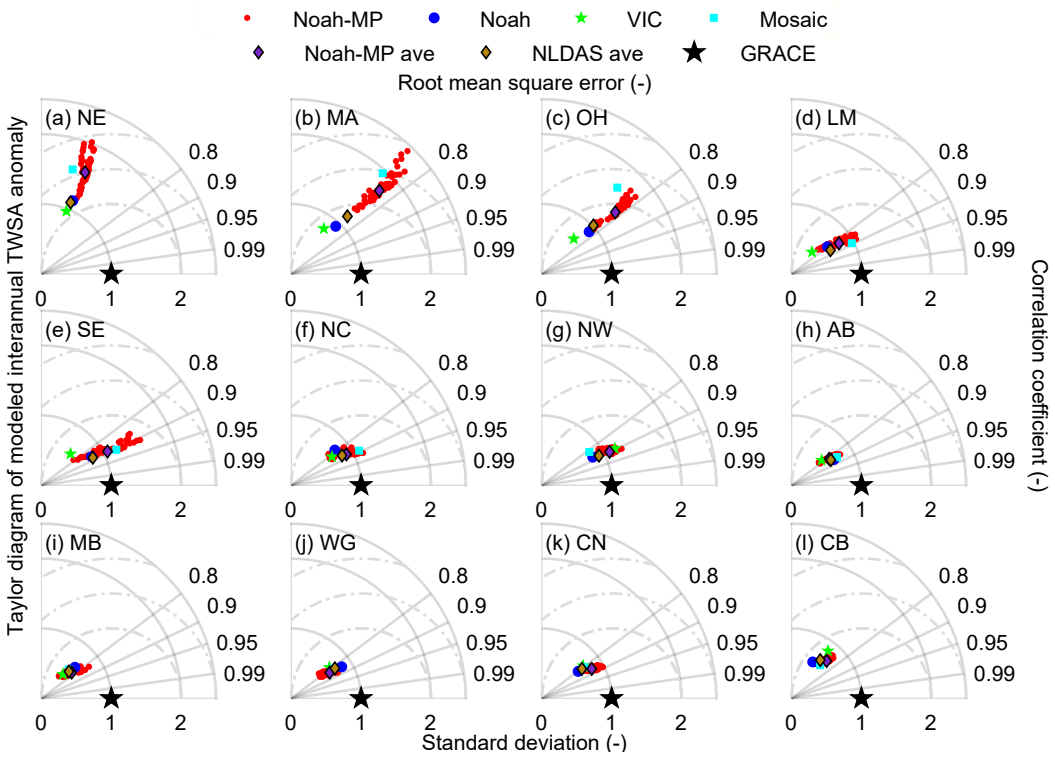

Figure 3. As in Figure 2, but for the interannual anomaly of TWSA.

Figure 3 shows the Taylor diagram for the interannual anomaly of TWSA. Compared with the annual cycle (Figure 2), both the Noah-MP configurations and the NLDAS models degrade significantly. Noah-MP still performs better than NLDAS in most RFCs. However, the superiority is marginal. In three RFCs—namely NE, MA, and OH—Noah-MP notably underperforms NLDAS. The underperformance is mainly due to higher variability than GRACE. Similar to the annual cycle, possible reasons include: (1) Noah-MP overestimated the variability due to unsuitable parameter values and (2) GRACE underestimated the variability due to the signal leaked from oceans.

4.3 Soil moisture

Figure 4 presents the time series of the surface (0–0.1 m) and root-zone (0–1.0 m) soil moisture in NC, NW, AB, WG, and CB. These RFCs were selected as they have more than 10 valid sites. Table S6 presents the corresponding TSSs. The Noah-MP configurations are consistent in estimating the surface soil moisture, having a spread remarkably smaller than that among the three NLDAS models. The NLDAS models have more diverse representations of soil moisture. Noah is the same as Noah-MP. Both solve Richards' equation to present the soil moisture dynamics in four layers (Niu et al., 2011). Mosaic also solves Richards' equation but at three soil layers. The top layer is further divided into tiles to better represent spatial heterogeneity(Koster and Suarez, 1992). VIC is different from them, utilizing a conceptual soil water tank (Liang et al., 1994).

It could be that Noah-MP underestimated the ensemble spread, especially when considering its inability to represent the subgrid heterogeneity. The NLDAS models could also overestimate the spread when considering the conceptual representation of VIC. The spread among the Noah-MP configurations increases significantly from the surface (Figure 4e) to the root zone (Figure 4k). The ensemble spread in the root-zone soil moisture reflects the difference in modeling root-water uptake for plant transpiration and soil-bottom drainage as described in Appendix A. Further investigation (Figures S2–S5) shows that, in the deep layers (the third and fourth layers), Noah-MP has a comparable or greater spread than NLDAS.

Comparison between Noah-MP and Noah shows that they perform similarly in AB (Figures 4e and 4f) and WG (Figures 4g and 4h) but are different in NC (Figures 4a and 4b), NW (Figures 4c and 4d), and CB (Figures 4i and 4j). The similarity in AB and WG is reasonable, since the two models have similar soil layer structures and parameterizations. The dissimilarity in NC, NW, and CB is most pronounced in winter. It could result from the different snow parameterizations in Noah-MP and Noah, which is investigated in Section 4.4.

In the RFCs and soil layers examined in Figure 4, the Noah-MP ensemble mean performs similarly or better than the NLDAS ensemble mean except in AB and CB. The superiority of the Noah-MP in simulating soil moisture is also reported in previous evaluations (Cai et al., 2014b). In AB and CB, individual NLDAS models do not show a consistent superiority over Noah-MP. In AB, the best NLDAS model is VIC in winter and Noah in summer. The performance of VIC in winter corresponds to the best-performing snow estimation (Figure 8h). In CB, the best NLDAS model is Noah in winter and Mosaic in summer. Mosaic carefully considers the subgrid variability of soil moisture, which could lead to better skill in RFCs with complex topography such as CB and NW. The NLDAS ensemble mean takes the advantage of wintertime soil moisture from VIC in AB and summertime soil moisture from Mosaic in CB. Both the advantage of VIC and Mosaic come from the representation of subgrid heterogeneity.

Figures 5 and 6 compare the TSS between the NLDAS and Noah-MP ensemble mean at each NAMSD site for the annual cycle and interannual anomaly, respectively. The comparison varies significantly with site and soil layer depth, revealing two major patterns. First, similar to Figure 4, NLDAS tends to outperform Noah-MP in AB and CB. The high skill of the NLDAS ensemble mean is likely a result of a high ensemble skill gain (Fei et al., 2021) related to the diversity among the NLDAS models. Noah-MP has both a low ensemble spread and an inadequate representation of subgrid heterogeneity in the two RFCs. Second, Noah-MP outperforms NLDAS in other RFCs. In NC, OH, and MA, the low performance of NLDAS is related to the anomaly in wintertime soil moisture (Figure S2). The anomaly suggests that the NLDAS models generally have difficulty in modeling snow and snow–soil moisture interactions (refer to Section 4.4 for more information). On the other hand, Noah-MP has a better snow module, leading to a higher soil moisture estimation skill.

To maximize the utilization of the NLDAS model diversity and Noah-MP physics improvements, we combine the Noah-MP ensemble mean and the three NLDAS models. The right-hand columns of Figures 5 and 6 show that the four estimates' arithmetic average outperforms the three-model NLDAS ensemble mean at most NASMD sites. The outperformance suggests an added value of the Noah-MP data. If the Noah-MP ensemble mean already outperforms the NLDAS ensemble mean, the added value appears in almost every site. If the Noah-MP ensemble mean underperforms the NLDAS ensemble mean, the added value can still show up at approximately one-third (one-fourth) of the sites for the annual cycle (interannual anomaly).

4.4 Snow water equivalent

Figure 7a presents the spatial patterns of the multi-year averaged SWE (W_{snow}) from SNODAS. Snow is mainly distributed in the northeast (NE, NC, OH, and MA) and in the mountains of the west (the Cascade Mountains, Rocky Mountains, and Sierra Nevada in NW, AB, MB, WG, CN, and CB). Figure 7b (7c) shows the geographical difference between the Noah-MP (NLDAS) ensemble mean and SNODAS. Both Noah-MP and NLDAS exhibit a considerable underestimation in most areas of CONUS. However, the underestimation of Noah-MP is generally smaller. Figure 7d confirms that both Noah-MP and the NLDAS models tend to underestimate SWE in most areas but exhibit an overestimation when snow is extremely thick (SWE is greater than $400\,\mathrm{mm}$). Noah-MP performs better than the NLDAS models in most cases. The superiority of Noah-MP is likely attributable to the three-layer snowpack module, which can represent snow dynamics better in a wide range of snow depth than the single-layer Noah and Mosaic snow module and the quasi-two-layer VIC snow module. A careful consideration of the surface energy balance can also contribute to Noah-MP's superiority over VIC and Mosaic. Consequently, Noah-MP captures the spatial patterns better than NLDAS, with a spatial correlation of 0.87 versus 0.43. Further examination reveals that the superiority of Noah-MP appears in all elevation bands and is the most significant between 1000 to 2000 m with a spatial correlation of 0.85 versus 0.38 (0.77 versus 0.76 below 1000 m, and 0.89 versus 0.75 above 2000 m).

Figure 8 compares the annual cycle estimated from SNODAS, NLDAS, and Noah-MP. The annual cycle in the 12 RFCs exhibits a similar pattern: it accumulates in winter, peaks in spring, and melts from late spring to summer. The snow season in the northeastern RFCs (i.e., NE, MA, OH, and NC) spans from October to May, whereas the snow season is longer in the mountainous RFCs of the west (i.e., NW, AB, MB, WG, CN, and CB), lasting to June. From the comparison between Noah-MP and NLDAS, we make three observations. First, consistent with Figure 7, the NLDAS models underestimate the SWE in all RFCs. Among the three NLDAS models, Noah performs the best in the northeast (e.g., NE, MA, and OH), whereas VIC shows some advantages in the western mountainous RFCs (e.g., NW, CN, and CB). Possible reasons for the Noah superiority in the northeast include the careful consideration of the surface energy balance, whereas, in the western mountains, the elevation bands of VIC can better capture the spatial heterogeneity. Second, Noah-MP outperforms the NLDAS models in almost all RFCs, except for showing a marginal degradation in AB and MB. In these two RFCs, VIC outperforms not only the NLDAS models but also Noah-MP. The terrain is hilly, and the elevation-banded parameterization enables a better representation of subgrid snow variability. Another reason is related to the shallow snow in these two RFCs. Noah-MP has known been experiencing negative biases with shallow snow (i.e., AB, MB, and WG). The bias is attributable to the shallow snow albedo as discussed in Dang et al. (2019) and Wang et al. (2020). Third, the ensemble spread among the Noah-MP configurations is small. The estimates of SWE and its uncertainty should be improved in the future by considering processes such as rain–snow partitioning Wang et al. (2019), snow albedo (Wang et al., 2020; Dang et al., 2019), and roughness length (He et al., 2019).

Figure 9 shows the TSS of the NLDAS and Noah-MP ensembles in estimating the annual cycle and interannual anomaly of SWE. Table S7 summarizes the skill scores for the 12 RFCs. The annual cycle and interannual anomaly exhibit similar spatial patterns. NLDAS performs well in most parts of the CONUS, with TSSs higher than 0.75. In MB, TSS reaches a minimum of 0.60. In comparison with NLDAS, Noah-MP performs notably better in the east and west but marginally worse in parts of the

central CONUS. The location of Noah-MP's underperformance coincides well with shallow snow and relatively flat terrain. The coincidence hints at the weakness of Noah-MP in simulating shallow snow and the advantage of VIC in representing subgrid snow variability as discussed previously.

We further averaged the 48 Noah-MP configurations and added their average to the three-model NLDAS ensemble. Figures 9e and 9f show that the four-estimate ensemble mean outperforms the three-model NLDAS ensemble mean in nearly all areas of CONUS, again proving the added value of the data provided in this paper.

4.5 Evapotranspiration

Figure 10 compares the annual cycle estimated from FLUXNET MTE, NLDAS, and Noah-MP in the 12 RFCs. We choose FLUXNET MTE as the reference here since its performance is superior when compared to AmeriFlux (Tables S2–S4). In all the 12 RFCs, ET peaks during summer and is lowest during winter. Noah-MP successfully captures the timing of the peak in humid RFCs (i.e., NE, MA, OH, LM, SE, NC, and NW) but shows a 1-month lead in a few semi-arid and arid RFCs (i.e., MB, WG, and CN). The average of the three NLDAS models better reproduces the timing of the peak, but the models differ from each other significantly. Among the Noah-MP and NLDAS ensembles, VIC and Mosaic are notably different. VIC exhibits a systematic underestimation, while Mosaic shows an overall overestimation. The 48 Noah-MP configurations and Noah perform closely during autumn and winter, whereas their differences are pronounced during spring and summer. During spring and summer, Noah is the closest to FLUXNET MTE in most RFCs except NE and MA, whereas Noah-MP constantly overestimates the ET in all RFCs.

The overestimation of Noah-MP was investigated by separately comparing the three components of ET (i.e., transpiration, canopy evaporation, and ground evaporation) with GLEAM (Figure S6). Figure 11 shows the overestimation of total ET is closely linked to the overestimation of ground evaporation, which could be partially attributable to the overly high roughness length for heat and water, as described in Appendix A8 and A9. Besides, the lack of a litter layer (Decker et al., 2017) in Noah-MP could also play a part.

Figure 12 evaluates Noah-MP and NLDAS using the 25 AmeriFlux sites. The NLDAS ensemble mean outperforms the Noah-MP ensemble mean for the annual cycle, and this outperformance results from two causes. First, an NLDAS member, Noah, performs the closest to the observations, as shown in Figures 12b and 10. Second, the three NLDAS models are remarkably different from each other. The diversity of the ensemble gives a higher skill gain by combining them, as shown by the difference between the ensemble mean skill (Figures 12a) and the median TSS (Figures 12b). On the other hand, the Noah-MP configurations are too similar to each other, and all have a positive bias (Figure 10). However, for the interannual anomaly, the Noah-MP ensemble mean slightly outperforms the NLDAS ensemble mean (Figure 12c). Figure 12d shows that the Noah-MP configurations marginally outperform the NLDAS models. Among the NLDAS models, VIC performs the best, and Noah does not exhibit the same superiority shown in the annual cycle. The difference between the NLDAS ensemble mean performance (Figure 12c) and median performance (Figure 12d) is marginal, suggesting that the NLDAS ensemble skill gains are not notable for the interannual anomaly.

Figure 13 examines the ensemble spread of Noah-MP and NLDAS. The ensemble spread is normalized by the temporal variability calculated using the FLUXNET MTE ET. NLDAS has a significant spread in the southeast and west in all seasons, while spring shows the largest value. As seen in Figure 10, the NLDAS ensemble spread mainly reflects the differences between VIC and Mosaic. The Noah-MP ensemble has a notably smaller spread than NLDAS. The Noah-MP ensemble spread is manifested in spring and summer in the southeastern (SE, LM, and WG) and western (CN, CB, and NW) RFCs.

We can decompose the Noah-MP ensemble spread and pinpoint the dominant process using Sobol' sensitivity analysis (Zheng et al., 2019). Figure 14 delineates the Sobol' total sensitivity index of total ET to the four processes described in Appendix A. In spring and summer, for the regions where the Noah-MP configurations show significant spread (SE, LM, WG, CB, CN, and NW) (Figure 13), ET is most sensitive to the parameterization of stomatal conductance (Figures 14e and 14i) and then to the β factor (Figures 14f and 14j). However, for regions with positive biases (NC, OH, and LM, as shown in Figures 11b and 11c), the Noah-MP estimation is more sensitive to the turbulence parameterizations (Figures 14c and 14g). During autumn and winter, the parameterizations of stomatal conductance (Figures 14m and 14q) and β-factor (Figures 14n and 14r) still have significant impacts on the estimation of ET, and these impacts could be a result of the "memory" of TWS (Zheng et al., 2019). Besides these two processes, the runoff parameterization is dominant during autumn in the east (Figure 14p), and the turbulence parameterization is dominant during winter (Figure 14s).

5 Code and data availability

The dataset is freely available for download from the Zenodo online repository at https://doi.org/10.5281/zenodo.7109816 (Zheng et al., 2022). The dataset (along with datasets on which it is based) is subject to a Creative Commons BY (attribution) license agreement (https://creativecommons.org/licenses, last access: 2021-08-16).

6 Conclusions

This paper describes a 1/8° dataset of the TWB over the CONUS from 1980 to 2015 simulated from an ensemble of 48 perturbed-physics configurations of Noah-MP. This Noah-MP multi-physics ensemble features an enrichment of the NLDAS-2 four-model ensemble and brings convenience for multi-model comparison. The dataset has already been used in the monitoring of groundwater storage change (Rateb et al., 2020), the analysis of LSM parameterization sensitivity (Zheng et al., 2019), the development of model evaluation method (Zheng et al., 2020), and hydrological ensemble simulations (Fei et al., 2021). This paper details the Noah-MP parameterizations employed and evaluates the estimated TWSA, soil moisture, SWE, and ET in comparison with the NLDAS ensemble.

The spread of the ensemble estimation is the largest for the runoff. The spread in surface runoff accounts for 34% of its climatological mean. The spread is comparable to the previous estimates for multi-model ensembles (Dirmeyer et al., 2006). The ensemble spread in snow water equivalent is the smallest, 2.5% of its climatological mean. The ensemble has not included different parameterizations of several snow processes such as rain–snow partitioning, snow albedo, and roughness length,

which could lead to an underestimation of the ensemble spread. The underestimation of the ensemble spread becomes more apparent when the Noah-MP ensemble mean is biased (Figure 8). The bias is more pronounced relative to the NLDAS models in parts of AB and MB where snow is shallow and the terrain is hilly (Figure 9). VIC performs better there, suggesting the importance of considering subgrid variability.

Evaluation against various reference data shows that Noah-MP generally performs better than the NLDAS models. The augmented three-layer snow module of Noah-MP significantly improves the estimation of snow and also wintertime soil moisture. On the other hand, the NLDAS models outperform Noah-MP in AB and CB for surface soil moisture (Figures 4 to 6), in AB and MB for snow (Figures 8 and 9), and in OH, LM, and NC for the annual cycle of ET (Figure 10). The outperformance of NLDAS is likely attributable to the consideration of subgrid variability in soil moisture with Mosaic and in snow with VIC. The Noah-MP ensemble could be improved by increasing the spatial resolution or developing parameterizations of subgrid heterogeneity. Noah-MP also underestimates the temporal variability of TWS in coastal RFCs (Figures 2 and 3). Correction of the ocean signal leakage in the GRACE data and representation of the spatial variability of unconfined aquifers' parameters should be beneficial. For the annual cycle of ET, there is a systematic overestimation in spring and summer. Sobol' sensitivity analysis of the Noah-MP ensemble reveals that the bias is mainly related to the parameterization of turbulence. We have examined the code and found that the implementation is inconsistent with the literature. The parameterization of the roughness length of heat and water vapor likely contributes to the ET overestimation.

The Noah-MP ensemble shares the same atmospheric forcing and static parameters with the NLDAS models. The similarity enables the comparison between the multi-model and perturbed-physics ensembling methods as shown in Fei et al. (2021). Besides, Noah-MP complements the NLDAS models well. Adding Noah-MP to the NLDAS ensemble can consistently improve the TWB variables in most areas of CONUS.

The Noah-MP model has been undergoing rapid development. New components such as plant hydraulics (Li et al., 2021), roughness sublayer (Abolafia-Rosenzweig et al., 2021), crops (Liu et al., 2016), and dynamic rooting depth (Liu et al., 2020) have been added. New schemes for the processes such as rain–snow partitioning (Wang et al., 2019) have been included. Parameterizations of surface roughness length (He et al., 2019; Zhang et al., 2021), snow albedo (Wang et al., 2020), and vertical soil layers (Zhao et al., 2022; Shellito et al., 2020) have been refined. The dataset can be improved by using the updated model and including more perturbations.

Appendix A: Formulation of the used Noah-MP parameterization schemes

A1 SIMGM runoff parameterization scheme

SIMGM is a TOPMODEL-based runoff model (Niu et al., 2007). The scheme parameterizes runoff (R_{srf} and R_{sub}) as an exponential function of groundwater table depth (z_{wt}, m, positive down) as follows.

$$R_{srf} = Q_{soil,srf}[(1 - f_{frz,1})f_{sat} + f_{frz,1}], \tag{A1}$$

$$f_{sat} = f_{sat,max} \exp[-0.5 f(z_{wt} - z_{bot})], \tag{A2}$$

$$R_{sub} = [1 - \max_{i=1,\ldots,N_{soil}} (f_{frz,i})] R_{sub,max} \exp[-\Lambda - f(z_{wt} - z_{bot})], \tag{A3}$$

where $Q_{soil,srf}$ is the water incident on the soil surface (the sum of precipitation throughfall, snowmelt, and dewfall; $\mathrm{kg\,m^{-2}\,s^{-1}}$); $f_{frz,i}$ is the fractional frozen area of the i-th soil layer ($\mathrm{m^2\,m^{-2}}$), which is parameterized using the frozen water content of the soil layer following Niu and Yang (2006); f_{sat} is the saturation fraction of the grid cell ($\mathrm{m^2\,m^{-2}}$); and z_{bot} is the depth of the soil column bottom (2 m in this study), and z_{wt} is the groundwater table depth (m), which is converted from the groundwater storage by a specific-yield parameter. The groundwater storage is predicted using a dynamic groundwater model interacting with the soil column bottom (Niu et al., 2007).

The scheme has four calibratable parameters: (1) $f_{sat,max}$, the maximum saturation area fraction ($\mathrm{m^2\,m^{-2}}$), which is defined as the cumulative distribution function of the topographic index when the grid-cell-mean water table depth is zero; (2) f, a runoff decay factor (unitless); (3) $R_{sub,max}$, the maximum subsurface runoff when the grid-cell-mean water table depth is zero ($\mathrm{kg\,m^{-2}\,s^{-1}}$); and (4) Λ, the grid-cell-mean topographic index (unitless). In this study, the parameters have the following values: $f_{sat,max} = 0.38\,\mathrm{m^2\,m^{-2}}$, $f = 6$, $R_{sub,max} = 5\,\mathrm{kg\,m^{-2}\,s^{-1}}$, and $\Lambda = 10.5$.

A2 SIMTOP runoff parameterization scheme

SIMTOP is also a TOPMODEL-based runoff parameterization scheme, the same as SIMGM (equations (A1)–(A3)). The major difference between SIMTOP and SIMGM is that SIMTOP parameterizes the groundwater table depth (z_{wt}) using the soil liquid water content by assuming the water head is at equilibrium throughout the soil column down to the water table (Niu et al., 2005). Although SIMTOP and SIMGM share the same conceptual model of runoff generation, implementation differences exist. First, in contrast to equations (A2) and (A3), SIMTOP does not use the soil column bottom depth (z_{bot}) in calculating the saturation area fraction (f_{sat}) and subsurface runoff:

$$f_{sat} = f_{sat,max} \exp(-0.5 f z_{wt}), \tag{A4}$$

$$R_{sub} = [1 - \max_{i=1,\cdots,N_{soil}} (f_{frz,i})] R_{sub,max} \exp(-\Lambda - f z_{wt}). \tag{A5}$$

Second, parameter values are slightly different for the runoff decay factor and maximum subsurface runoff: $f = 2$, and $R_{sub,max} = 4\,\mathrm{kg\,m^{-2}\,s^{-1}}$.

A3 NOAHR runoff parameterization scheme

NOAHR parameterizes surface runoff (R_{srf}) as infiltration excess:

$$R_{srf} = Q_{soil,srf} - Q_{soil,in}, \tag{A6}$$

where $Q_{soil,i}$ is the infiltration into the soil ($\mathrm{kg\,m^{-2}\,s^{-1}}$). The infiltration is derived from the approximate solution to the Richards equation following Philip (1969) with additional considerations of the spatial variability of precipitation and infiltration capacity. By assuming exponential and independent distributions of precipitation and infiltration capacity within a model grid cell, NOAHR formulates the soil infiltration as follows:

$$Q_{soil,in} = Q_{soil,srf} \frac{I_c}{Q_{soil,srf}\Delta t + I_c}, \tag{A7}$$

$$I_c = w_d[1 - \exp(-K_{\Delta t}\Delta t)], \tag{A8}$$

$$w_d = \sum_{i=1}^{N_{soil}} \rho_{wat}(w_{sat,i} - w_{soil,i})\Delta z_{soil,i}, \tag{A9}$$

where I_c is the soil infiltration capacity of the model grid cell ($\mathrm{kg\,m^{-2}}$), w_d is the water deficit of the soil column ($\mathrm{kg\,m^{-2}}$), and Δt is the model time step (s). Following Chen and Dudhia (2001a), the parameter is assumed as propositional to the saturated hydraulic conductivity of the first soil layer ($K_{sat,1}$; $\mathrm{kg\,m^{-2}\,s^{-1}}$):

$$K_{\Delta t} = \frac{K_{\Delta t,ref}}{k_{ref}} K_{sat,1}, \tag{A10}$$

where $K_{\Delta t,ref}$ and k_{ref} are two parameters. In Noah-MP (and Noah), $K_{\Delta t,ref} = \frac{3}{86400}\,\mathrm{s^{-1}}$, and $k_{ref} = 2 \times 10^{-3}\,\mathrm{kg\,m^{-2}\,s^{-1}}$. $K_{sat,1}$ is assigned using a soil parameter lookup table according to the soil texture type.

NOAHR assumes free drainage at the soil column bottom. The subsurface runoff is calculated as

$$R_{sub} = \alpha_{slope} K_{soil,N_{soil}}, \tag{A11}$$

where α_{slope} is the terrain slope index, which is arbitrarily given as 0.1 in the adopted version of Noah-MP. $K_{soil,N_{soil}}$ is the hydraulic conductivity of the bottom soil layer, which is parameterized following Clapp and Hornberger (1978).

A4 BATS runoff parameterization scheme

The BATS scheme parameterizes surface runoff (R_{srf}) as a function of soil wetness (Yang and Dickinson, 1996):

$$R_{srf} = Q_{soil,srf}\left[(1 - f_{frz,1})f_{sat} + f_{frz,1}\right], \tag{A12}$$

$$f_{sat} = \theta^4, \tag{A13}$$

$$\theta = \frac{\sum_{i=1}^{N_{soil}} \frac{w_{soil,i}}{w_{sat,i}} \Delta z_{soil,i}}{\sum_{i=1}^{N_{soil}} \Delta z_{soil,i}}, \tag{A14}$$

where θ is the averaged wetness throughout the soil column ($\mathrm{m^3\,m^{-3}}$).

Similar to NOAHR, the BATS scheme also assumes a free drainage boundary condition at the soil column bottom. Subsurface runoff (R_{sub}) is parameterized as follows:

$$R_{sub} = \left(1 - \max_{i=1,\cdots,N_{soil}}(f_{frz,i})\right) K_{soil,N_{soil}}. \tag{A15}$$

A5 Ball–Berry scheme of stomatal resistance

Leaf stomata are the small pores typically found on the underside of leaves. They control the gas exchange of CO_2, H_2O, and O_2 between the internal leaf structure and the external atmosphere. In LSMs, the opening and closing of the stomata are characterized by stomatal conductance.

The Ball–Berry scheme for parameterizing the stomatal conductance (g_s) for H_2O is as follows:

$$g_s = m\frac{A}{c_s}\frac{e_s}{e_i}P_{atm} + b, \tag{A16}$$

where g_s is the leaf stomatal conductance ($\mu\mathrm{mol\,m^{-2}\,s^{-1}}$), m is a vegetation-type dependent parameter (unitless), A is the leaf photosynthesis rate, c_s is the CO_2 partial pressure at the leaf surface (Pa), e_s is the water vapor pressure at the leaf surface (Pa), e_i is the saturated water vapor at the stomata (Pa), P_{atm} is the ambient air pressure (Pa), and b is the stomatal conductance at zero photosynthesis ($\mu\mathrm{mol\,m^{-2}\,s^{-1}}$). The parameters m and b are assigned from a lookup table using the vegetation type.

A6 Jarvis scheme of stomatal resistance

The Jarvis scheme for parameterizing the canopy resistance (R_c) based on the product of four stress factors ($\mathrm{s\,m^{-1}}$) is calculated as follows (Chen et al., 1996; Sellers et al., 1996; Jacquemin and Noilhan, 1990; Jarvis, 1976):

$$R_c = R_{c,min}\frac{1}{f_1 f_2 f_3 \beta}, \tag{A17}$$

$$f_1 = \frac{\frac{R_{c,min}}{R_{c,max}} + f}{1+f}, \tag{A18}$$

$$f = 0.55\frac{2R_g}{R_{gl}}, \tag{A19}$$

$$f_2 = \frac{1}{1 + h_s[q_{sat}(T_l) - q_a]}, \tag{A20}$$

$$f_3 = 1 - 0.0016(T_{ref} - T_l)^2, \tag{A21}$$

where f_1, f_2 and f_3 are the stress factors of solar radiation, vapor pressure deficit, and air temperature, respectively (unitless), which are unitless and range from 0 to 1; β is the soil moisture stress factor, which is detailed in Section A7; R_g is the incoming solar radiation ($\mathrm{W\,m^{-2}}$) for unit leaf area index; T_l is leaf temperature (K). $q_{sat}(T_l)$ ($\mathrm{kg\,kg^{-1}}$) and q_a are the saturated humidity at the temperature of T_l ($\mathrm{kg\,kg^{-1}}$) and ambient humidity, respectively. In literature (Chen et al., 1996; Jacquemin and Noilhan, 1990), $q_{sat}T_l$ and q_a are specific humidity, whereas in Noah-MP v3.6, they are implemented as mixing ratio.

The scheme has five parameters: $R_{c,min}$, the minimum stomatal resistance ($\mathrm{s\,m^{-1}}$) per unit leaf area index; $R_{c,max}$, the maximum resistance; R_{gl}, a radiation scaling factor (unitless); h_s, a humidity scaling factor (unitless); T_{ref}, the optimum temperature (K). Among these parameters, $R_{c,min}$, R_{gl}, and h_s are assigned using a vegetation-parameter lookup table, while $R_{c,max}$ and T_{ref} are assigned assumedly to $5000\,\mathrm{s\,m^{-1}}$ and $298\,\mathrm{K}$, respectively.

A7 Three soil moisture stress factor schemes

The NOAHB scheme parameterizes the soil moisture stress factor controlling transpiration (β factor) as a function of soil moisture, which is calculated as follows:

$$\beta = \sum_{i=1}^{N_{root}} \frac{\Delta z_{soil,i}}{z_{root}} \min(1, \frac{\theta_i - \theta_{wilt}}{\theta_{ref} - \theta_{wilt}}), \tag{A22}$$

where N_{root} is the total number of soil layers that contain roots, z_{root} is the total depth of the root-zone layer (m), and θ_i is the volumetric soil moisture of the i-th soil layer ($\mathrm{m^3\,m^{-3}}$). NOAHB has two parameters: θ_{ref}, the field capacity ($\mathrm{m^3\,m^{-3}}$); and θ_{wilt}, the wilting volumetric soil moisture ($\mathrm{m^3\,m^{-3}}$).

The CLM scheme (Oleson et al., 2004) parameterizes β as a function of soil matric potential, which is calculated as follows:

$$\beta = \sum_{i=1}^{N_{root}} \frac{\Delta z_{soil,i}}{z_{root}} \min(1, \frac{\psi_{wilt} - \psi_i}{\psi_{wilt} - \psi_{sat}}), \tag{A23}$$

where ψ_i is the water pressure head of the i-th soil layer (m), and ψ_i is converted from θ_i using the formula of Clapp and Hornberger (1978). CLM has two parameters: ψ_{sat}, the saturated water pressure head (m); and ψ_{wilt}, the wilting pressure head (m).

The SSiB scheme (Xue et al., 1991) also parameterizes the β factor as a function of the soil pressure head, similar to CLM. However, the formula is different, as follows:

$$\beta = \sum_{i=1}^{N_{root}} \frac{\Delta z_{soil,i}}{z_{root}} \min\left[1, 1 - \exp\left(-c_2 \ln(\frac{\psi_{wilt}}{\psi_i})\right)\right]. \tag{A24}$$

SSiB has two parameters: ψ_{wilt}, the wilting pressure head (m); and c_2, a unitless coefficient.

In Noah-MP version 3.6, the parameters θ_{sat}, θ_{wilt}, and ψ_{sat} are assigned using a soil parameter lookup table (Chen and Dudhia, 2001a, Table 2); ψ_{wilt} is $-10\,\mathrm{m}$, independent of vegetation and soil types (Niu et al., 2011); c_2 is assumed constant at 5.8, whereas in SSiB, this parameter varies with vegetation type (Xue et al., 1991, Table 2).

A8 Chen97 near-surface turbulence scheme

The Chen97 scheme (Chen et al., 1997) parameterizes the surface exchange coefficient for heat (C_h) as follows:

$$C_h = \kappa^2 \left[\ln(\frac{z}{z_{0m}}) - \Psi_m(\frac{z}{L}) + \Psi_m(\frac{z_{0m}}{L})\right]^{-1} \left[\ln(\frac{z}{z_{0h}}) - \Psi_h(\frac{z}{L}) + \Psi_h(\frac{z_{0h}}{L})\right]^{-1}, \tag{A25}$$

where $\kappa = 0.4$ is the von Kármán constant; L is the Monin–Obukhov (M–O) length (m); z is the reference height (m); Ψ_m and Ψ_h are the similarity theory-based stability functions for momentum and heat, respectively; z_{0m} is the roughness length for momentum (m) and depends on the land cover/land-use type; and z_{0h} is the roughness length for heat (m). Niu et al. (2011) parameterized $z_{0h} = z_{0m} \exp(-\kappa C \sqrt{Re^*})$, where $C = 0.1$ and Re^* is the roughness Reynolds number. However, in the code of Noah-MP version 3.6, $z_{0h} = z_{0m}$.

A9 M–O near-surface turbulence scheme

The M–O scheme is based on the M–O similarity theory (Brutsaert, 1982), which parameterizes C_h as follows:

$$C_h = \kappa^2 \left[\ln(\frac{z-d_0}{z_{0m}}) - \Psi_m(\frac{z-d_0}{L})\right]^{-1} \left[\ln(\frac{z-d_0}{z_{0h}}) - \Psi_h(\frac{z-d_0}{L})\right]^{-1} \tag{A26}$$

where $z_{0h} = z_{0m}$. d_0 is the zero-displacement height (m),

$$d_0 = 0.64 z_{ct}, \tag{A27}$$

where z_{ct} is the canopy top height (m).

Appendix B: Estimation of the terrestrial water storage for the RFCs neighboring the Great Lakes

The TWS estimation for the NC, OH, and NE RFCs is performed in two steps: (1) aggregate the GRACE TWS over both the RFC land area and neighboring lakes (lakes Superior, Michigan, and Huron for NC; Erie for OH; and Ontario for NE); and (2) subtract the lake water storage anomaly from the aggregated TWS. The lake water storage in the second step is calculated as the product of the observed water level and the lake area.

The lake water level is an arithmetic average of selected NOAA in situ observations (https://tidesandcurrents.noaa.gov/stations.html?type=Water+Levels). For Lake Superior, five observation stations were selected: Point Iroquois, Marquette C.G., Ontonagon, Duluth, and Grand Marais. For Lake Michigan, seven stations were selected: Ludington, Holland, Calumet Harbor, Milwaukee, Kewaunee, Sturgeon Bay Canal, and Port Inland. For Lake Huron, five stations were selected: Lakeport, Harbor Beach, Essexville, Mackinaw City, and De Tour Village. For Lake Erie, eight stations were selected: Buffalo, Sturgeon Point, Erie, Fairport, Cleveland, Marblehead, Toledo, and Fermi Power Plant. And for Lake Ontario, four stations were selected: Cape Vincent, Oswego, Rochester, and Olcott.

The lake area is estimated from the lake boundary data provided by the United States Geological Survey (https://www.sciencebase.gov/catalog/item/530f8a0ee4b0e7e46bd300dd). Only the area within the United States is considered, which is within a 150 km radius from the studied RFCs. The lake areas are calculated as follows: 52441 km^2 for Lake Superior within the USA, 57509 km^2 for Lake Michigan, 23185 km^2 for Lake Huron within the USA, 25494 km^2 for Lake Erie, and 18871 km^2 for Lake Ontario. Month-to-month variations in lake area are neglected in this study for simplicity.

Author contributions. ZLY initiated and funded the study. HZ conducted the simulation and generated the data. WF analyzed the data and created the figures. JW, LZ, LL, and SW contributed to the validation of the data. All authors contributed to creating the dataset and drafting the paper.

Competing interests. The authors declare that they have no conflict of interest

Disclaimer. The data are provided as is with no warranties.

Financial support. This research has been supported by the National Natural Science Foundation of China (grants 42075165, 41375088, and 41605062) and the Beijing Natural Science Foundation (8204072).

References

Abolafia-Rosenzweig, R., He, C., Burns, S. P., and Chen, F.: Implementation and Evaluation of a Unified Turbulence Parameterization throughout the Canopy and Roughness Sublayer in Noah-MP Snow Simulations, Journal of Advances in Modeling Earth Systems, 13, e2021MS002 665, https://doi.org/10.1029/2021MS002665, 2021.

Ajami, N. K., Duan, Q., and Sorooshian, S.: An Integrated Hydrologic Bayesian Multimodel Combination Framework: Confronting Input, Parameter, and Model Structural Uncertainty in Hydrologic Prediction, Water Resources Research, 43, W01 403, https://doi.org/10.1029/2005WR004745, 2007.

Beck, H. E., van Dijk, A. I. J. M., de Roo, A., Dutra, E., Fink, G., Orth, R., and Schellekens, J.: Global Evaluation of Runoff from 10 State-of-the-Art Hydrological Models, Hydrology and Earth System Sciences, 21, 2881–2903, https://doi.org/10.5194/hess-21-2881-2017, 2017.

Brutsaert, W.: Evaporation into the Atmosphere: Theory, History, and Applications, Springer, Dordrecht, https://doi.org/10.1007/978-94-017-1497-6, 1982.

Burnash, R. J. C., Ferral, R. L., and McGuire, R. A.: A Generalized Streamflow Simulation System: Conceptual Modeling for Digital Computers, Technical Report, Joint Federal-State River Forecast Center, U.S. National Weather Service and California Department of Water Resources, Sacramento, California, USA, 1973.

Cai, X., Yang, Z.-L., David, C. H., Niu, G.-Y., and Rodell, M.: Hydrological Evaluation of the Noah-MP Land Surface Model for the Mississippi River Basin, Journal of Geophysical Research: Atmospheres, 119, 23–38, https://doi.org/10.1002/2013JD020792, 2014a.

Cai, X., Yang, Z.-L., Xia, Y., Huang, M., Wei, H., Leung, L. R., and Ek, M. B.: Assessment of Simulated Water Balance from Noah, Noah-MP, CLM, and VIC over CONUS Using the NLDAS Test Bed, Journal of Geophysical Research: Atmospheres, 119, 13 751–13 770, https://doi.org/10.1002/2014JD022113, 2014b.

Carrera, M. L., Bélair, S., and Bilodeau, B.: The Canadian Land Data Assimilation System (CaLDAS): Description and Synthetic Evaluation Study, Journal of Hydrometeorology, 16, 1293–1314, https://doi.org/10.1175/JHM-D-14-0089.1, 2015.

Chen, F. and Dudhia, J.: Coupling an Advanced Land Surface–Hydrology Model with the Penn State–NCAR MM5 Modeling System. Part I: Model Implementation and Sensitivity, Monthly Weather Review, 129, 569–585, https://doi.org/10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2, 2001a.

Chen, F. and Dudhia, J.: Coupling an Advanced Land Surface–Hydrology Model with the Penn State–NCAR MM5 Modeling System. Part II: Preliminary Model Validation, Monthly Weather Review, 129, 587–604, https://doi.org/10.1175/1520-0493(2001)129<0587:CAALSH>2.0.CO;2, 2001b.

Chen, F., Mitchell, K. E., Schaake, J., Xue, Y., Pan, H. L., Koren, V., Duan, Q., Ek, M., and Betts, A. K.: Modeling of Land Surface Evaporation by Four Schemes and Comparison with FIFE Observations, Journal of Geophysical Research: Atmospheres, 101, 7251–7268, https://doi.org/10.1029/95JD02165, 1996.

Chen, F., Janjić, Z., and Mitchell, K.: Impact of Atmospheric Surface-Layer Parameterizations in the New Land-Surface Scheme of the NCEP Mesoscale Eta Model, Boundary-Layer Meteorology, 85, 391–421, https://doi.org/10.1023/A:1000531001463, 1997.

Chen, F., Barlage, M., Tewari, M., Rasmussen, R., Jin, J., Lettenmaier, D. P., Livneh, B., Lin, C., Miguez-Macho, G., Niu, G.-Y., Wen, L., and Yang, Z.-L.: Modeling Seasonal Snowpack Evolution in the Complex Terrain and Forested Colorado Headwaters Region: A Model Intercomparison Study, Journal of Geophysical Research: Atmospheres, 119, 2014JD022 167, https://doi.org/10.1002/2014JD022167, 2014.

Clapp, R. B. and Hornberger, G. M.: Empirical Equations for Some Soil Hydraulic Properties, Water Resources Research, 14, 601–604, https://doi.org/10.1029/WR014i004p00601, 1978.

Clark, M. P., Kavetski, D., and Fenicia, F.: Pursuing the Method of Multiple Working Hypotheses for Hydrological Modeling, Water Resources Research, 47, 1–16, https://doi.org/10.1029/2010WR009827, 2011.

Cloke, H. and Pappenberger, F.: Ensemble Flood Forecasting: A Review, Journal of Hydrology, 375, 613–626, https://doi.org/10.1016/j.jhydrol.2009.06.005, 2009.

Clow, D. W., Nanus, L., Verdin, K. L., and Schmidt, J.: Evaluation of SNODAS Snow Depth and Snow Water Equivalent Estimates for the Colorado Rocky Mountains, USA, Hydrological Processes, 26, 2583–2591, https://doi.org/10.1002/hyp.9385, 2012.

Dai, A.: Increasing Drought under Global Warming in Observations and Models, Nature Climate Change, 3, 52–58, https://doi.org/10.1038/nclimate1633, 2013.

Dang, C., Zender, C. S., and Flanner, M. G.: Intercomparison and Improvement of Two-Stream Shortwave Radiative Transfer Schemes in Earth System Models for a Unified Treatment of Cryospheric Surfaces, The Cryosphere, 13, 2325–2343, https://doi.org/10.5194/tc-13-2325-2019, 2019.

Decker, M., Or, D., Pitman, A. J., and Ukkola, A.: New Turbulent Resistance Parameterization for Soil Evaporation Based on a Pore-Scale Model: Impact on Surface Fluxes in CABLE, Journal of Advances in Modeling Earth Systems, 9, 220–238, https://doi.org/10.1002/2016MS000832, 2017.

Dickinson, R. E., Wang, G., Zeng, X., and Zeng, Q.: How Does the Partitioning of Evapotranspiration and Runoff between Different Processes Affect the Variability and Predictability of Soil Moisture and Precipitation?, Advances in Atmospheric Sciences, 20, 475–478, https://doi.org/10.1007/BF02690805, 2003.

Dirmeyer, P. A., Gao, X., Zhao, M., Guo, Z., Oki, T., and Hanasaki, N.: GSWP-2: Multimodel Analysis and Implications for Our Perception of the Land Surface, Bulletin of the American Meteorological Society, 87, 1381–1398, https://doi.org/10.1175/BAMS-87-10-1381, 2006.

Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G., and Tarpley, J. D.: Implementation of Noah Land Surface Model Advances in the National Centers for Environmental Prediction Operational Mesoscale Eta Model, Journal of Geophysical Research: Atmospheres, 108, 8851, https://doi.org/10.1029/2002JD003296, 2003.

Emerton, R. E., Cloke, H. L., Stephens, E. M., Zsoter, E., Woolnough, S. J., and Pappenberger, F.: Complex Picture for Likelihood of ENSO-driven Flood Hazard, Nature Communications, 8, 14 796, https://doi.org/10.1038/ncomms14796, 2017.

Fang, B., Lei, H., Zhang, Y., Quan, Q., and Yang, D.: Spatio-Temporal Patterns of Evapotranspiration Based on Upscaling Eddy Covariance Measurements in the Dryland of the North China Plain, Agricultural and Forest Meteorology, 281, 107 844, https://doi.org/10.1016/j.agrformet.2019.107844, 2020.

Fei, W., Zheng, H., Xu, Z., Wu, W.-Y., Lin, P., Tian, Y., Guo, M., She, D., Li, L., Li, K., and Yang, Z.-L.: Ensemble Skill Gains Obtained from the Multi-Physics versus Multi-Model Approaches for Continental-Scale Hydrological Simulations, Water Resources Research, 57, e2020WR028 846, https://doi.org/10.1029/2020wr028846, 2021.

Gan, Y., Liang, X.-Z., Duan, Q., Chen, F., Li, J., and Zhang, Y.: Assessment and Reduction of the Physical Parameterization Uncertainty for Noah-MP Land Surface Model, Water Resources Research, 55, 5518–5538, https://doi.org/10.1029/2019WR024814, 2019.

Gao, H., Tang, Q., Ferguson, C. R., Wood, E. F., and Lettenmaier, D. P.: Estimating the Water Budget of Major US River Basins via Remote Sensing, International Journal of Remote Sensing, 31, 3955–3978, https://doi.org/10.1080/01431161.2010.483488, 2010.

Guo, Z., Dirmeyer, P. A., Gao, X., and Zhao, M.: Improving the Quality of Simulated Soil Moisture with a Multi-Model Ensemble Approach, Quarterly Journal of the Royal Meteorological Society, 133, 731–747, https://doi.org/10.1002/qj.48, 2007.

He, C., Chen, F., Barlage, M., Liu, C., Newman, A., Tang, W., Ikeda, K., and Rasmussen, R.: Can Convection-Permitting Modeling Provide Decent Precipitation for Offline High-Resolution Snowpack Simulations over Mountains?, Journal of Geophysical Research: Atmospheres, 124, 12 631–12 654, https://doi.org/10.1029/2019JD030823, 2019.

Hejazi, M. I., Edmonds, J., Clarke, L., Kyle, P., Davies, E., Chaturvedi, V., Wise, M., Patel, P., Eom, J., and Calvin, K.: Integrated Assessment of Global Water Scarcity over the 21st Century under Multiple Climate Change Mitigation Policies, Hydrology and Earth System Sciences, 18, 2859–2883, https://doi.org/10.5194/hess-18-2859-2014, 2014.

Jacquemin, B. and Noilhan, J.: Sensitivity Study and Validation of a Land Surface Parameterization Using the HAPEX-MOBILHY Data Set, Boundary-Layer Meteorology, 52, 93–134, https://doi.org/10.1007/BF00123180, 1990.

Jarvis, P. G.: The Interpretation of the Variations in Leaf Water Potential and Stomatal Conductance Found in Canopies in the Field, Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 273, 593–610, https://doi.org/10.1098/rstb.1976.0035, 1976.

Jung, M., Reichstein, M., and Bondeau, A.: Towards Global Empirical Upscaling of FLUXNET Eddy Covariance Observations: Validation of a Model Tree Ensemble Approach Using a Biosphere Model, Biogeosciences, 6, 2001–2013, https://doi.org/10.5194/bg-6-2001-2009, 2009.

Jung, M., Koirala, S., Weber, U., Ichii, K., Gans, F., Camps-Valls, G., Papale, D., Schwalm, C., Tramontana, G., and Reichstein, M.: The FLUXCOM Ensemble of Global Land-Atmosphere Energy Fluxes, Scientific Data, 6, 74, https://doi.org/10.1038/s41597-019-0076-8, 2019.

Kim, R. S., Kumar, S., Vuyovich, C., Houser, P., Lundquist, J., Mudryk, L., Durand, M., Barros, A., Kim, E. J., Forman, B. A., Gutmann, E. D., Wrzesien, M. L., Garnaud, C., Sandells, M., Marshall, H.-P., Cristea, N., Pflug, J. M., Johnston, J., Cao, Y., Mocko, D., and Wang, S.: Snow Ensemble Uncertainty Project (SEUP): Quantification of Snow Water Equivalent Uncertainty across North America via Ensemble Land Surface Modeling, The Cryosphere, 15, 771–791, https://doi.org/10.5194/tc-15-771-2021, 2021.

Koster, R. D.: "Efficiency Space": A Framework for Evaluating Joint Evaporation and Runoff Behavior, Bulletin of the American Meteorological Society, 96, 393–396, https://doi.org/10.1175/BAMS-D-14-00056.1, 2015.

Koster, R. D. and Suarez, M. J.: Modeling the Land Surface Boundary in Climate Models as a Composite of Independent Vegetation Stands, Journal of Geophysical Research: Atmospheres, 97, 2697–2715, https://doi.org/10.1029/91JD01696, 1992.

Kumar, S., Holmes, T., Mocko, M. D., Wang, S., and Peters-Lidard, C.: Attribution of Flux Partitioning Variations between Land Surface Models over the Continental U.S., Remote Sensing, 10, 751, https://doi.org/10.3390/rs10050751, 2018.

Kumar, S. V., Wang, S., Mocko, D. M., Peters-Lidard, C. D., and Xia, Y.: Similarity Assessment of Land Surface Model Outputs in the North American Land Data Assimilation System, Water Resources Research, 53, 8941–8965, https://doi.org/10.1002/2017WR020635, 2017.

LaFontaine, J. H., Hay, L. E., Viger, R. J., Regan, R. S., and Markstrom, S. L.: Effects of Climate and Land Cover on Hydrology in the Southeastern U.S.: Potential Impacts on Watershed Planning, Journal of the American Water Resources Association, 51, 1235–1261, https://doi.org/10.1111/1752-1688.12304, 2015.

Le, P. V. V., Kumar, P., and Drewry, D. T.: Implications for the Hydrologic Cycle under Climate Change Due to the Expansion of Bioenergy Crops in the Midwestern United States, Proceedings of the National Academy of Sciences, 108, 15 085–15 090, https://doi.org/10.1073/pnas.1107177108, 2011.

Levia, D. F., Creed, I. F., Hannah, D. M., Nanko, K., Boyer, E. W., Carlyle-Moses, D. E., van de Giesen, N., Grasso, D., Guswa, A. J., Hudson, J. E., Hudson, S. A., Iida, S., Jackson, R. B., Katul, G. G., Kumagai, T., Llorens, P., Ribeiro, F. L., Pataki, D. E., Peters, C. A., Carretero,

D. S., Selker, J. S., Tetzlaff, D., Zalewski, M., and Bruen, M.: Homogenization of the Terrestrial Water Cycle, Nature Geoscience, 13, 656–658, https://doi.org/10.1038/s41561-020-0641-y, 2020.

Li, L., Yang, Z.-L., Matheny, A. M., Zheng, H., Swenson, S. C., Lawrence, D. M., Barlage, M., Yan, B., McDowell, N. G., and Leung, L. R.: Representation of Plant Hydraulics in the Noah-MP Land Surface Model: Model Development and Multiscale Evaluation, Journal of Advances in Modeling Earth Systems, 13, e2020MS002 214, https://doi.org/10.1029/2020ms002214, 2021.

Lian, X., Piao, S., Huntingford, C., Li, Y., Zeng, Z., Wang, X., Ciais, P., McVicar, T. R., Peng, S., Ottlé, C., Yang, H., Yang, Y., Zhang, Y., and Wang, T.: Partitioning Global Land Evapotranspiration Using CMIP5 Models Constrained by Observations, Nature Climate Change, 8, 640–646, https://doi.org/10.1038/s41558-018-0207-9, 2018.

Liang, X., Lettenmaier, D. P., Wood, E. F., and Burges, S. J.: A Simple Hydrologically Based Model of Land Surface Water and Energy Fluxes for General Circulation Models, Journal of Geophysical Research: Atmospheres, 99, 14 415–14 428, https://doi.org/10.1029/94JD00483, 1994.

Lin, P., Hopper, L. J., Yang, Z.-L., Lenz, M., and Zeitler, J. W.: Insights into Hydrometeorological Factors Constraining Flood Prediction Skill during the May and October 2015 Texas Hill Country Flood Events, Journal of Hydrometeorology, 19, 1339–1361, https://doi.org/10.1175/JHM-D-18-0038.1, 2018.

Lin, P., Pan, M., Beck, H. E., Yang, Y., Yamazaki, D., Frasson, R., David, C. H., Durand, M., Pavelsky, T. M., Allen, G. H., Gleason, C. J., and Wood, E. F.: Global Reconstruction of Naturalized River Flows at 2.94 Million Reaches, Water Resources Research, 55, 6499–6516, https://doi.org/10.1029/2019WR025287, 2019.

Liu, X., Chen, F., Barlage, M., Zhou, G., and Niyogi, D. S.: Noah-MP-Crop: Introducing Dynamic Crop Growth in the Noah-MP Land Surface Model, Journal of Geophysical Research: Atmospheres, 121, 13 953–13 972, https://doi.org/10.1002/2016JD025597, 2016.

Liu, X., Chen, F., Barlage, M., and Niyogi, D.: Implementing Dynamic Rooting Depth for Improved Simulation of Soil Moisture and Land Surface Feedbacks in Noah-MP-Crop, Journal of Advances in Modeling Earth Systems, 12, e2019MS001 786, https://doi.org/10.1029/2019MS001786, 2020.

Lv, M., Ma, Z., Li, M., and Zheng, Z.: Quantitative Analysis of Terrestrial Water Storage Changes under the Grain for Green Program in the Yellow River Basin, Journal of Geophysical Research: Atmospheres, 124, 1336–1351, https://doi.org/10.1029/2018JD029113, 2019.

Lv, M., Xu, Z., Yang, Z.-L., Lu, H., and Lv, M.: A Comprehensive Review of Specific Yield in Land Surface and Groundwater Studies, Journal of Advances in Modeling Earth Systems, 13, e2020MS002 270, https://doi.org/10.1029/2020MS002270, 2021.

Ma, N., Niu, G.-Y., Xia, Y., Cai, X., Zhang, Y., Ma, Y., and Fang, Y.: 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, 12 245–12 268, https://doi.org/10.1002/2017JD027597, 2017.

McCabe, M. F., Rodell, M., Alsdorf, D. E., Miralles, D. G., Uijlenhoet, R., Wagner, W., Lucieer, A., Houborg, R., Verhoest, N. E. C., Franz, T. E., Shi, J., Gao, H., and Wood, E. F.: The Future of Earth Observation in Hydrology, Hydrology and Earth System Sciences, 21, 3879–3914, https://doi.org/10.5194/hess-21-3879-2017, 2017.

Mitchell, K. E., Lohmann, D., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., Cosgrove, B. A., Sheffield, J., Duan, Q., Luo, L., Higgins, R. W., Pinker, R. T., Tarpley, J. D., Lettenmaier, D. P., Marshall, C. H., Entin, J. K., Pan, M., Shi, W., Koren, V., Meng, J., Ramsay, B. H., and Bailey, A. A.: The Multi-Institution North American Land Data Assimilation System (NLDAS): Utilizing Multiple GCIP Products and Partners in a Continental Distributed Hydrological Modeling System, Journal of Geophysical Research: Atmospheres, 109, D07S90, https://doi.org/10.1029/2003JD003823, 2004.

Niu, G.-Y. and Yang, Z.-L.: Effects of Frozen Soil on Snowmelt Runoff and Soil Water Storage at a Continental Scale, Journal of Hydrometeorology, 7, 937–952, https://doi.org/10.1175/JHM538.1, 2006.

Niu, G.-Y., Yang, Z.-L., Dickinson, R. E., and Gulden, L. E.: A Simple TOPMODEL-based Runoff Parameterization (SIMTOP) for Use in Global Climate Models, Journal of Geophysical Research: Atmospheres, 110, D21 106, https://doi.org/10.1029/2005JD006111, 2005.

Niu, G.-Y., Yang, Z.-L., Dickinson, R. E., Gulden, L. E., and Su, H.: Development of a Simple Groundwater Model for Use in Climate Models and Evaluation with Gravity Recovery and Climate Experiment Data, Journal of Geophysical Research, 112, D07 103, https://doi.org/10.1029/2006JD007522, 2007.

Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M., and Xia, Y.: The Community Noah Land Surface Model with Multiparameterization Options (Noah-MP): 1. Model Description and Evaluation with Local-Scale Measurements, Journal of Geophysical Research: Atmospheres, 116, D12 109, https://doi.org/10.1029/2010JD015139, 2011.

Oleson, K. W., Dai, Y., Bonan, G. B., Bosilovich, M., Dirmeyer, P. A., Hoffman, F. M., Houser, P. R., Levis, S., Niu, G.-Y., Thornton, P. E., Vertenstein, M., Yang, Z.-L., and Zeng, X.: Technical Description of the Community Land Model (CLM), Tech. rep., National Center for Atmospheric Research, Boulder, Colorado, https://doi.org/10.5065/D6N877R0, 2004.

Pan, M., Sahoo, A. K., Troy, T. J., Vinukollu, R. K., Sheffield, J., and Wood, E. F.: Multisource Estimation of Long-Term Terrestrial Water Budget for Major Global River Basins, Journal of Climate, 25, 3191–3206, https://doi.org/10.1175/JCLI-D-11-00300.1, 2012.

Pan, S., Pan, N., Tian, H., Friedlingstein, P., Sitch, S., Shi, H., Arora, V. K., Haverd, V., Jain, A. K., Kato, E., Lienert, S., Lombardozzi, D., Nabel, J. E. M. S., Ottlé, C., Poulter, B., Zaehle, S., and Running, S. W.: Evaluation of Global Terrestrial Evapotranspiration Using State-of-the-Art Approaches in Remote Sensing, Machine Learning and Land Surface Modeling, Hydrology and Earth System Sciences, 24, 1485–1509, https://doi.org/10.5194/hess-24-1485-2020, 2020.

Peters-Lidard, C. D., Hossain, F., Leung, L. R., McDowell, N., Rodell, M., Tapiador, F. J., Turk, F. J., and Wood, A.: 100 Years of Progress in Hydrology, Meteorological Monographs, 59, 25.1–25.51, https://doi.org/10.1175/AMSMONOGRAPHS-D-18-0019.1, 2018.

Peters-Lidard, C. D., Mocko, D. M., Su, L., Lettenmaier, D. P., Gentine, P., and Barlage, M.: Advances in Land Surface Models and Indicators for Drought Monitoring and Prediction, Bulletin of the American Meteorological Society, 102, E1099–E1122, https://doi.org/10.1175/BAMS-D-20-0087.1, 2021.

Philip, J. R.: Theory of Infiltration, in: Advances in Hydroscience, vol. 5, pp. 215–296, Elsevier, https://doi.org/10.1016/B978-1-4831-9936-8.50010-6, 1969.

Prudhomme, C., Giuntoli, I., Robinson, E. L., Clark, D. B., Arnell, N. W., Dankers, R., Fekete, B. M., Franssen, W., Gerten, D., Gosling, S. N., Hagemann, S., Hannah, D. M., Kim, H., Masaki, Y., Satoh, Y., Stacke, T., Wada, Y., and Wisser, D.: Hydrological Droughts in the 21st Century, Hotspots and Uncertainties from a Global Multimodel Ensemble Experiment, Proceedings of the National Academy of Sciences, 111, 3262–3267, https://doi.org/10.1073/pnas.1222473110, 2014.

Quiring, S. M., Ford, T. W., Wang, J. K., Khong, A., Harris, E., Lindgren, T., Goldberg, D. W., and Li, Z.: The North American Soil Moisture Database: Development and Applications, Bulletin of the American Meteorological Society, 97, 1441–1459, https://doi.org/10.1175/BAMS-D-13-00263.1, 2016.

Rateb, A., Scanlon, B. R., Pool, D. R., Sun, A., Zhang, Z., Chen, J., Clark, B., Faunt, C. C., Haugh, C. J., Hill, M., Hobza, C., McGuire, V. L., Reitz, M., Schmied, H. M., Sutanudjaja, E. H., Swenson, S., Wiese, D., Xia, Y., and Zell, W.: Comparison of Groundwater Storage Changes from GRACE Satellites with Monitoring and Modeling of Major U.S. Aquifers, Water Resources Research, 56, e2020WR027 556, https://doi.org/10.1029/2020wr027556, 2020.

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J. K., Walker, J. P., Lohmann, D., and Toll, D.: The Global Land Data Assimilation System, Bulletin of the American Meteorological Society, 85, 381–394, https://doi.org/10.1175/BAMS-85-3-381, 2004.

Rodell, M., Velicogna, I., and Famiglietti, J. S.: Satellite-Based Estimates of Groundwater Depletion in India, Nature, 460, 999–1002, https://doi.org/10.1038/nature08238, 2009.

Rodell, M., Beaudoing, H. K., L'Ecuyer, T. S., Olson, W. S., Famiglietti, J. S., Houser, P. R., Adler, R., Bosilovich, M. G., Clayson, C. A., Chambers, D., Clark, E., Fetzer, E. J., Gao, X., Gu, G., Hilburn, K., Huffman, G. J., Lettenmaier, D. P., Liu, W. T., Robertson, F. R., Schlosser, C. A., Sheffield, J., and Wood, E. F.: The Observed State of the Water Cycle in the Early Twenty-First Century, Journal of Climate, 28, 8289–8318, https://doi.org/10.1175/JCLI-D-14-00555.1, 2015.

Sakumura, C., Bettadpur, S., and Bruinsma, S.: Ensemble Prediction and Intercomparison Analysis of GRACE Time-Variable Gravity Field Models, Geophysical Research Letters, 41, 1389–1397, https://doi.org/10.1002/2013GL058632, 2014.

Save, H., Bettadpur, S., and Tapley, B. D.: High-Resolution CSR GRACE RL05 Mascons, Journal of Geophysical Research: Solid Earth, 121, 7547–7569, https://doi.org/10.1002/2016JB013007, 2016.

Saxe, S., Farmer, W., Driscoll, J., and Hogue, T. S.: Implications of Model Selection: A Comparison of Publicly Available, Conterminous US-extent Hydrologic Component Estimates, Hydrology and Earth System Sciences, 25, 1529–1568, https://doi.org/10.5194/hess-25-1529-2021, 2021.

Scanlon, B. R., Faunt, C. C., Longuevergne, L., Reedy, R. C., Alley, W. M., McGuire, V. L., and McMahon, P. B.: Groundwater Depletion and Sustainability of Irrigation in the US High Plains and Central Valley, Proceedings of the National Academy of Sciences, 109, 9320–9325, https://doi.org/10.1073/pnas.1200311109, 2012.

Scanlon, B. R., Zhang, Z., Save, H., Sun, A. Y., Schmied, H. M., van Beek, L. P. H., Wiese, D. N., Wada, Y., Long, D., Reedy, R. C., Longuevergne, L., Döll, P., and Bierkens, M. F. P.: Global Models Underestimate Large Decadal Declining and Rising Water Storage Trends Relative to GRACE Satellite Data, Proceedings of the National Academy of Sciences, 115, E1080–E1089, https://doi.org/10.1073/pnas.1704665115, 2018.

Sellers, P. J., Randall, D. A., Collatz, G. J., Berry, J. A., Field, C. B., Dazlich, D. A., Zhang, C., Collelo, G. D., and Bounoua, L.: A Revised Land Surface Parameterization (SiB2) for Atmospheric GCMs. Part I: Model Formulation, Journal of Climate, 9, 676–705, https://doi.org/10.1175/1520-0442(1996)009<0676:ARLSPF>2.0.CO;2, 1996.

Shellito, P. J., Kumar, S. V., Santanello, J. A., Lawston-Parker, P., Bolten, J. D., Cosh, M. H., Bosch, D. D., Holifield Collins, C. D., Livingston, S., Prueger, J., Seyfried, M., and Starks, P. J.: Assessing the Impact of Soil Layer Depth Specification on the Observability of Modeled Soil Moisture and Brightness Temperature, Journal of Hydrometeorology, 21, 2041–2060, https://doi.org/10.1175/JHM-D-19-0280.1, 2020.

Shi, C., Xie, Z., Qian, H., Liang, M., and Yang, X.: China Land Soil Moisture EnKF Data Assimilation Based on Satellite Remote Sensing Data, Science China Earth Sciences, 54, 1430–1440, https://doi.org/10.1007/s11430-010-4160-3, 2011.

Sobol′, l. M.: Sensitivity Estimates for Nonlinear Mathematical Models, Mathematical Modelling and Computational Experiment, 1, 407–414, 1993.

Su, L., Cao, Q., Xiao, M., Mocko, D. M., Barlage, M., Li, D., Peters-Lidard, C. D., and Lettenmaier, D. P.: Drought Variability over the Conterminous United States for the Past Century, Journal of Hydrometeorology, 22, 1153–1168, https://doi.org/10.1175/JHM-D-20-0158.1, 2021.

Taylor, K. E.: Summarizing Multiple Aspects of Model Performance in a Single Diagram, Journal of Geophysical Research: Atmospheres, 106, 7183–7192, https://doi.org/10.1029/2000JD900719, 2001.

Telteu, C.-E., Müller Schmied, H., Thiery, W., Leng, G., Burek, P., Liu, X., Boulange, J. E. S., Andersen, L. S., Grillakis, M., Gosling, S. N., Satoh, Y., Rakovec, O., Stacke, T., Chang, J., Wanders, N., Shah, H. L., Trautmann, T., Mao, G., Hanasaki, N., Koutroulis, A., Pokhrel, Y., Samaniego, L., Wada, Y., Mishra, V., Liu, J., Döll, P., Zhao, F., Gädeke, A., Rabin, S. S., and Herz, F.: Understanding Each Other's Models: An Introduction and a Standard Representation of 16 Global Water Models to Support Intercomparison, Improvement, and Communication, Geoscientific Model Development, 14, 3843–3878, https://doi.org/10.5194/gmd-14-3843-2021, 2021.

Trenberth, K. E. and Fasullo, J. T.: North American Water and Energy Cycles, Geophysical Research Letters, 40, 365–369, https://doi.org/10.1002/grl.50107, 2013a.

Trenberth, K. E. and Fasullo, J. T.: Regional Energy and Water Cycles: Transports from Ocean to Land, Journal of Climate, 26, 7837–7851, https://doi.org/10.1175/JCLI-D-13-00008.1, 2013b.

Trenberth, K. E., Smith, L., Qian, T., Dai, A., and Fasullo, J.: Estimates of the Global Water Budget and Its Annual Cycle Using Observational and Model Data, Journal of Hydrometeorology, 8, 758–769, https://doi.org/10.1175/JHM600.1, 2007.

Troin, M., Arsenault, R., Wood, A. W., Brissette, F., and Martel, J.-L.: Generating Ensemble Streamflow Forecasts: A Review of Methods and Approaches over the Past 40 Years, Water Resources Research, 57, e2020WR028 392, https://doi.org/10.1029/2020WR028392, 2021.

Voss, K. A., Famiglietti, J. S., Lo, M., de Linage, C., Rodell, M., and Swenson, S. C.: Groundwater Depletion in the Middle East from GRACE with Implications for Transboundary Water Management in the Tigris-Euphrates-Western Iran Region, Water Resources Research, 49, 904–914, https://doi.org/10.1002/wrcr.20078, 2013.

Wang, W., Yang, K., Zhao, L., Zheng, Z., Lu, H., Mamtimin, A., Ding, B., Li, X., Zhao, L., Li, H., Che, T., and Moore, J. C.: Characterizing Surface Albedo of Shallow Fresh Snow and Its Importance for Snow Ablation on the Interior of the Tibetan Plateau, Journal of Hydrometeorology, 21, 815–827, https://doi.org/10.1175/JHM-D-19-0193.1, 2020.

Wang, Y.-H., Broxton, P., Fang, Y., Behrangi, A., Barlage, M., Zeng, X., and Niu, G.-Y.: A Wet-Bulb Temperature-Based Rain-Snow Partitioning Scheme Improves Snowpack Prediction over the Drier Western United States, Geophysical Research Letters, 46, 13 825–13 835, https://doi.org/10.1029/2019gl085722, 2019.

Ward, P. J., Jongman, B., Kummu, M., Dettinger, M. D., Sperna Weiland, F. C., and Winsemius, H. C.: Strong Influence of El Niño Southern Oscillation on Flood Risk around the World, Proceedings of the National Academy of Sciences, 111, 15 659–15 664, https://doi.org/10.1073/pnas.1409822111, 2014.

Wu, W.-Y., Yang, Z.-L., and Barlage, M.: The Impact of Noah-MP Physical Parameterizations on Modeling Water Availability during Droughts in the Texas–Gulf Region, Journal of Hydrometeorology, 22, 1221–1233, https://doi.org/10.1175/JHM-D-20-0189.1, 2021.

Xia, Y., Mitchell, K., Ek, M., Cosgrove, B. A., Sheffield, J., Luo, L., Alonge, C., Wei, H., Meng, J., Livneh, B., Duan, Q., and Lohmann, D.: Continental-Scale Water and Energy Flux Analysis and Validation for North American Land Data Assimilation System Project Phase 2 (NLDAS-2): 2. Validation of Model-Simulated Streamflow, Journal of Geophysical Research: Atmospheres, 117, D03 110, https://doi.org/10.1029/2011JD016051, 2012a.

Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B. A., Wood, E. F., Luo, L., Alonge, C., Wei, H., Meng, J., Livneh, B., Lettenmaier, D. P., Koren, V., Duan, Q., Mo, K. C., Fan, Y., and Mocko, D.: Continental-Scale Water and Energy Flux Analysis and Validation for the North American Land Data Assimilation System Project Phase 2 (NLDAS-2): 1. Intercomparison and Application of Model Products, Journal of Geophysical Research: Atmospheres, 117, D03 109, https://doi.org/10.1029/2011JD016048, 2012b.

Xia, Y., Ek, M. B., Wu, Y., Ford, T., and Quiring, S. M.: Comparison of NLDAS-2 Simulated and NASMD Observed Daily Soil Moisture. Part II: Impact of Soil Texture Classification and Vegetation Type Mismatches, Journal of Hydrometeorology, 16, 1981–2000, https://doi.org/10.1175/JHM-D-14-0097.1, 2015a.

Xia, Y., Ek, M. B., Wu, Y., Ford, T., and Quiring, S. M.: Comparison of NLDAS-2 Simulated and NASMD Observed Daily Soil Moisture. Part I: Comparison and Analysis, Journal of Hydrometeorology, 16, 1962–1980, https://doi.org/10.1175/JHM-D-14-0096.1, 2015b.

Xia, Y., Cosgrove, B. A., Mitchell, K. E., Peters-Lidard, C. D., Ek, M. B., Brewer, M., Mocko, D., Kumar, S. V., Wei, H., Meng, J., and Luo, L.: Basin-Scale Assessment of the Land Surface Water Budget in the National Centers for Environmental Prediction Operational and Research NLDAS-2 Systems, Journal of Geophysical Research: Atmospheres, 121, 2750–2779, https://doi.org/10.1002/2015JD023733, 2016.

Xia, Y., Hao, Z., Shi, C., Li, Y., Meng, J., Xu, T., Wu, X., and Zhang, B.: Regional and Global Land Data Assimilation Systems: Innovations, Challenges, and Prospects, Journal of Meteorological Research, 33, 159–189, https://doi.org/10.1007/s13351-019-8172-4, 2019.

Xu, T., Guo, Z., Xia, Y., Ferreira, V. G., Liu, S., Wang, K., Yao, Y., Zhang, X., and Zhao, C.: Evaluation of Twelve Evapotranspiration Products from Machine Learning, Remote Sensing and Land Surface Models over Conterminous United States, Journal of Hydrology, 578, 124 105, https://doi.org/10.1016/j.jhydrol.2019.124105, 2019.

Xue, Y., Sellers, P. J., Kinter, J. L., and Shukla, J.: A Simplified Biosphere Model for Global Climate Studies, Journal of Climate, 4, 345–364, https://doi.org/10.1175/1520-0442(1991)004<0345:ASBMFG>2.0.CO;2, 1991.

Yang, Z.-L. and Dickinson, R. E.: Description of the Biosphere-Atmosphere Transfer Scheme (BATS) for the Soil Moisture Workshop and Evaluation of Its Performance, Global and Planetary Change, 13, 117–134, https://doi.org/10.1016/0921-8181(95)00041-0, 1996.

Yang, Z.-L., Niu, G.-Y., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Longuevergne, L., Manning, K., Niyogi, D., Tewari, M., and Xia, Y.: The Community Noah Land Surface Model with Multiparameterization Options (Noah-MP): 2. Evaluation over Global River Basins, Journal of Geophysical Research: Atmospheres, 116, D12 110, https://doi.org/10.1029/2010JD015140, 2011.

Yin, D. and Roderick, M. L.: Inter-Annual Variability of the Global Terrestrial Water Cycle, Hydrology and Earth System Sciences, 24, 381–396, https://doi.org/10.5194/hess-24-381-2020, 2020.

Zaussinger, F., Dorigo, W., Gruber, A., Tarpanelli, A., Filippucci, P., and Brocca, L.: Estimating Irrigation Water Use over the Contiguous United States by Combining Satellite and Reanalysis Soil Moisture Data, Hydrology and Earth System Sciences, 23, 897–923, https://doi.org/10.5194/hess-23-897-2019, 2019.

Zhang, B., Xia, Y., Long, B., Hobbins, M., Zhao, X., Hain, C., Li, Y., and Anderson, M. C.: Evaluation and Comparison of Multiple Evapotranspiration Data Models over the Contiguous United States: Implications for the next Phase of NLDAS (NLDAS-Testbed) Development, Agricultural and Forest Meteorology, 280, 107 810, https://doi.org/10.1016/j.agrformet.2019.107810, 2020.

Zhang, G., Chen, F., and Gan, Y.: Assessing Uncertainties in the Noah-MP Ensemble Simulations of a Cropland Site during the Tibet Joint International Cooperation Program Field Campaign, Journal of Geophysical Research: Atmospheres, 121, 9576–9596, https://doi.org/10.1002/2016JD024928, 2016.

Zhang, X., Chen, L., Ma, Z., and Gao, Y.: Assessment of Surface Exchange Coefficients in the Noah-MP Land Surface Model for Different Land-Cover Types in China, International Journal of Climatology, 41, 2638–2659, https://doi.org/10.1002/joc.6981, 2021.

Zhang, Y., Pan, M., Sheffield, J., Siemann, A. L., Fisher, C. K., Liang, M., Beck, H. E., Wanders, N., MacCracken, R. F., Houser, P. R., Zhou, T., Lettenmaier, D. P., Ma, Y., Pinker, R. T., Bytheway, J., Kummerow, C. D., and Wood, E. F.: A Climate Data Record (CDR) for the Global Terrestrial Water Budget: 1984–2010, Hydrology and Earth System Sciences, 22, 241–263, https://doi.org/10.5194/hess-22-241-2018, 2018.

Zhao, L. and Yang, Z.-L.: Multi-Sensor Land Data Assimilation: Toward a Robust Global Soil Moisture and Snow Estimation, Remote Sensing of Environment, 216, 13–27, https://doi.org/10.1016/j.rse.2018.06.033, 2018.

Zhao, L., Yang, K., He, J., Zheng, H., and Zheng, D.: Potential of Mapping Global Soil Texture Type from SMAP Soil Moisture Product: A Pilot Study, IEEE Transactions on Geoscience and Remote Sensing, 60, 1–10, https://doi.org/10.1109/TGRS.2021.3119667, 2022.

Zheng, H., Yang, Z.-L., Lin, P., Wei, J., Wu, W.-Y., Li, L., Zhao, L., and Wang, S.: On the Sensitivity of the Precipitation Partitioning into Evapotranspiration and Runoff in Land Surface Parameterizations, Water Resources Research, 55, 95–111, https://doi.org/10.1029/2017WR022236, 2019.

Zheng, H., Yang, Z.-L., Lin, P., Wu, W.-Y., Li, L., Xu, Z., Wei, J., Zhao, L., Bian, Q., and Wang, S.: Falsification-Oriented Signature-Based Evaluation for Guiding the Development of Land Surface Models and the Enhancement of Observations, Journal of Advances in Modeling Earth Systems, 12, e2020MS002 132, https://doi.org/10.1029/2020MS002132, 2020.

Zheng, H., Fei, W., Yang, Z.-L., Wei, J., Zhao, L., and Li, L.: An Ensemble of 48 Physically Perturbed Model Estimates of the 1/8° Terrestrial Water Budget over the Conterminous United States, 1980–2015, https://doi.org/10.5281/zenodo.7109816, 2022.

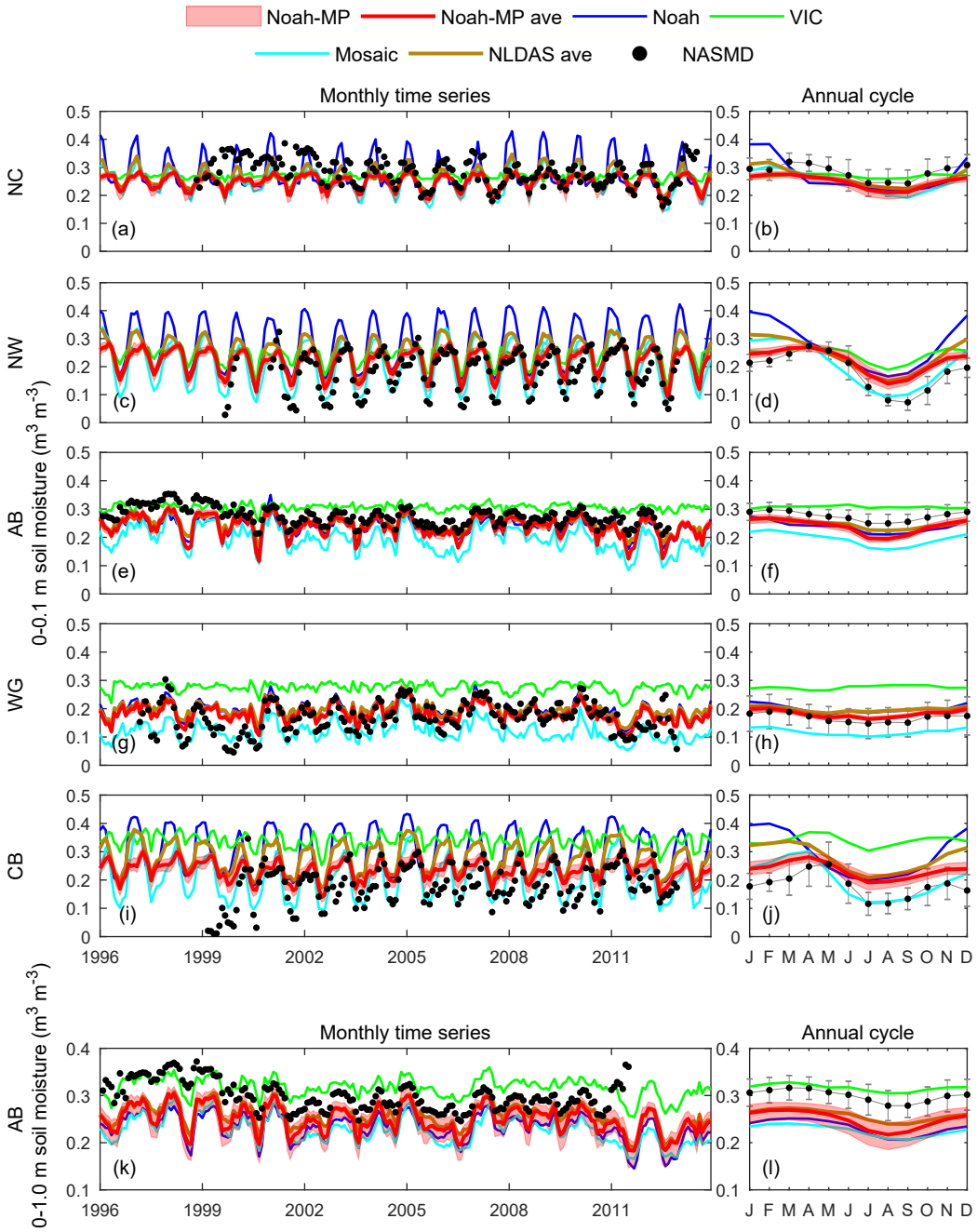

Figure 4. Monthly surface (0–0.1 m) and root-zone (0–1 m) soil moisture (left column) and the annual cycle (right column) from the Noah-MP ensemble, the NLDAS models, and the NASMD observations for the period 1996–2013. Only the RFCs with more than 10 observational sites are considered. Black dots denote the arithmetic average of the valid NASMD observations. Error bars in the right columns denote the standard deviation of the year-to-year differences. The shaded areas denote the range between the maxima and minima of the 48 Noah-MP estimates. The solid red line denotes the Noah-MP multi-physics ensemble mean. The three NLDAS models (Noah, Mosaic, VIC) and their ensemble mean are denoted by the blue, green, cyan, and dark golden lines, respectively. The five RFCs are sorted based on climatic aridity.

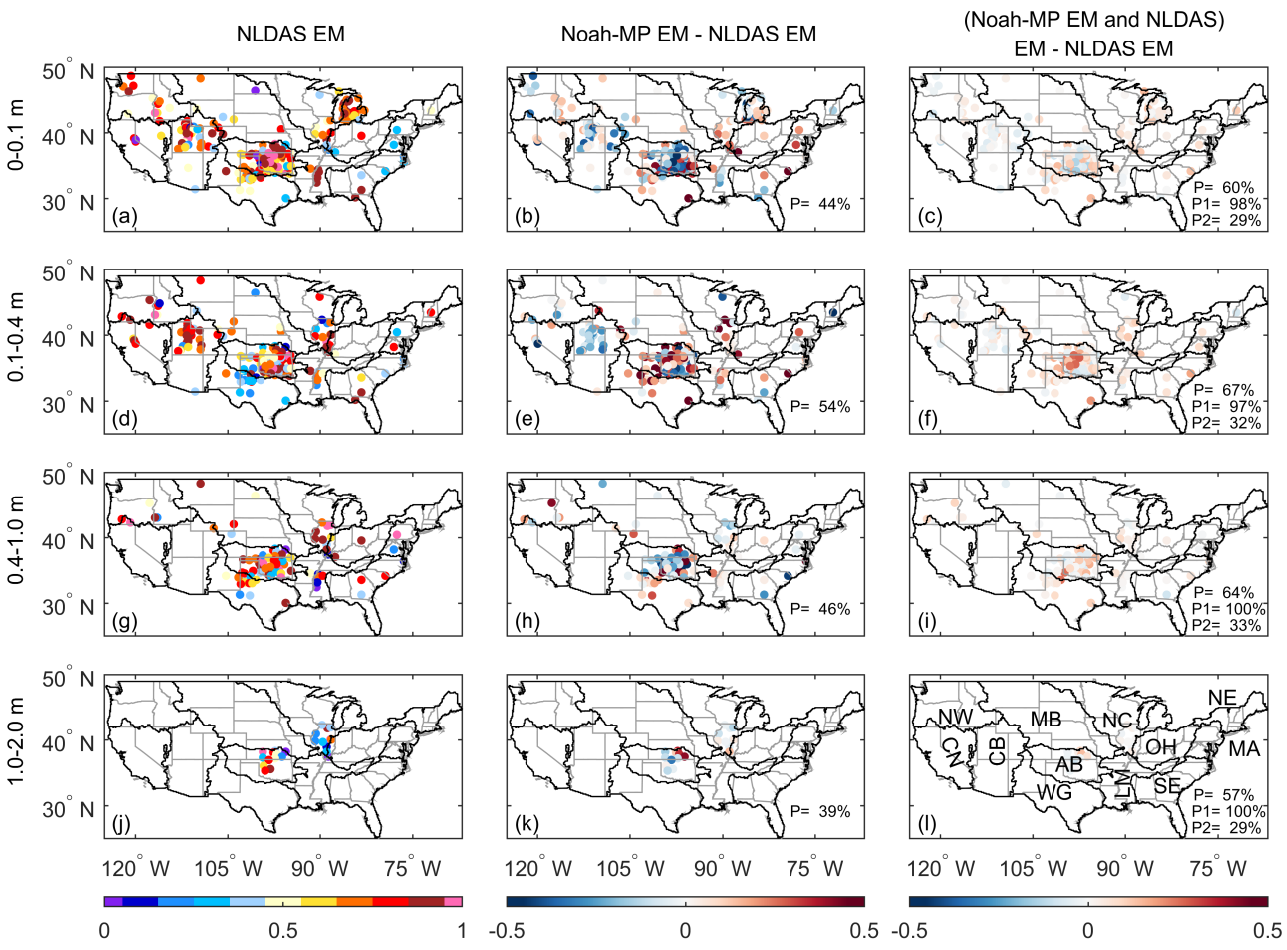

Figure 5. The first column shows the TSS of the NLDAS ensemble mean in simulating the annual cycle of soil moisture at different depths (0–0.1 m, 0.1–0.4 m, 0.4–1 m, 1–2 m). The second column presents the difference in TSS between the Noah-MP ensemble mean and the NLDAS ensemble mean. The third column depicts the TSS difference between the arithmetic average of the Noah-MP ensemble mean and the NLDAS models to the NLDAS three-model mean. P is the percentage of sites at which a higher TSS appears relative to the NLDAS ensemble mean. $P1$ ($P2$) is the percentage of sites at which a higher TSS appear given that the Noah-MP ensemble mean outperforms (underperforms) the NLDAS ensemble mean. The evaluation period is 1996–2013.

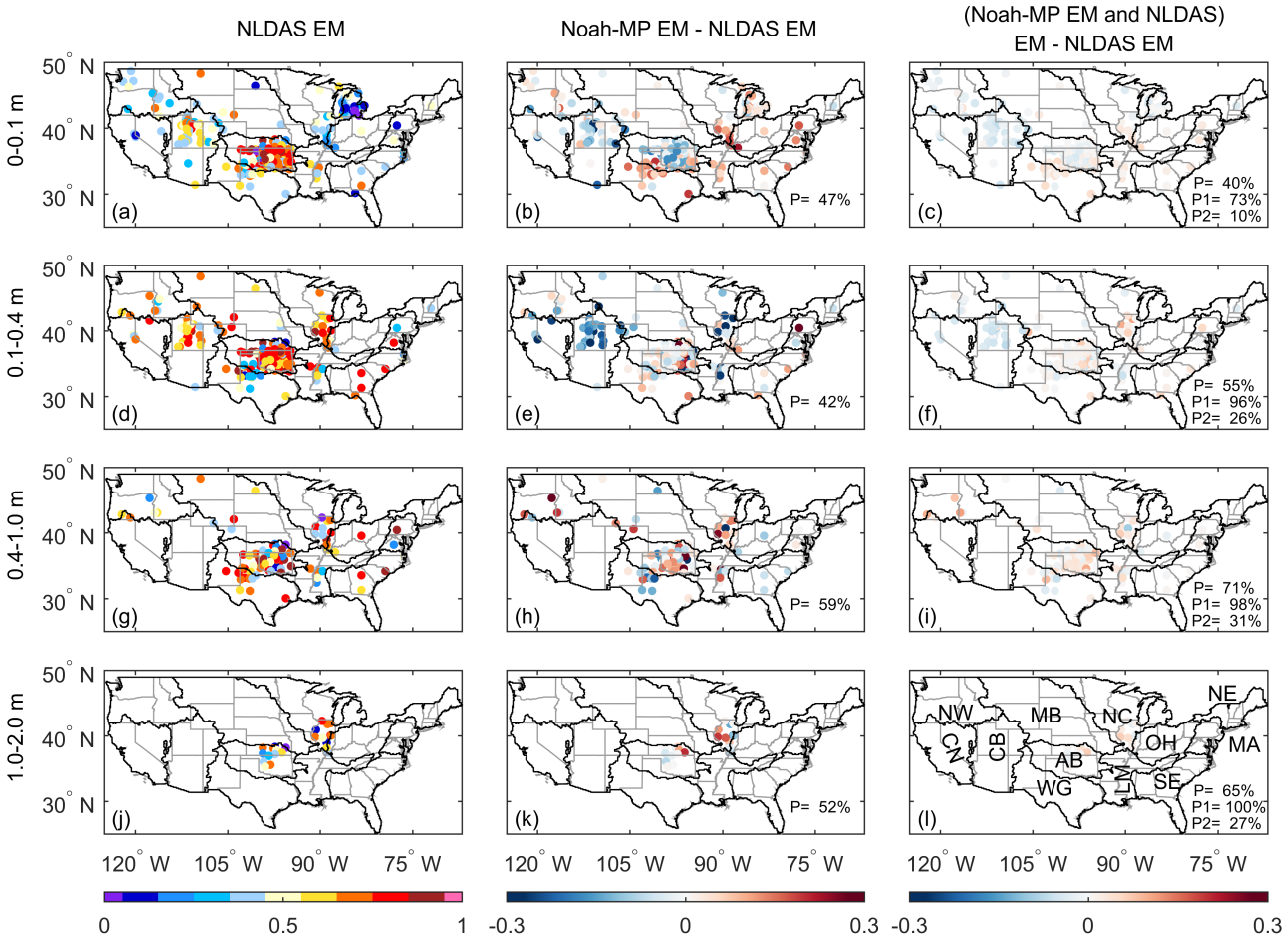

Figure 6. As in Figure 5, but for interannual anomaly.

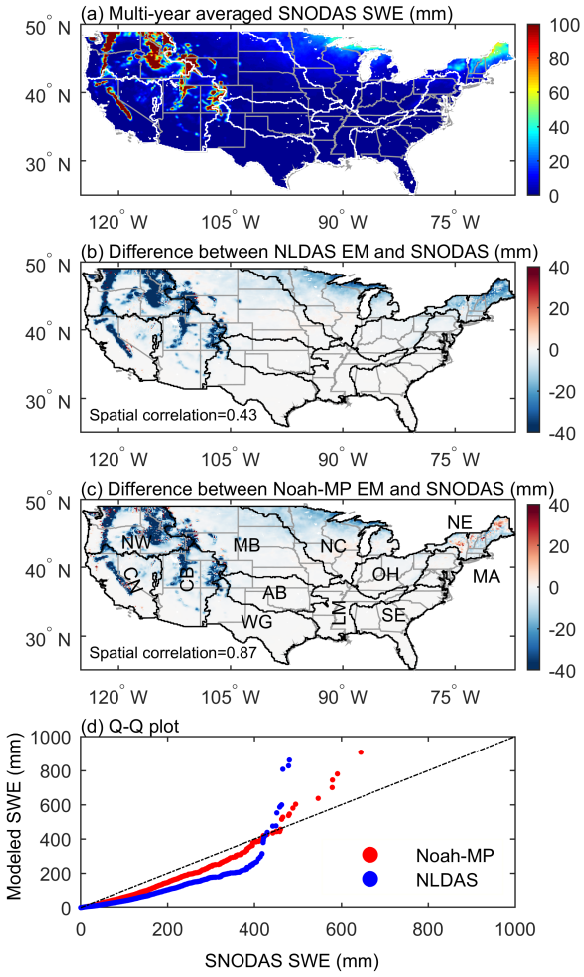

Figure 7. (a) Spatial distribution of the 11-year averaged (September 2004 to August 2015) SWE from SNODAS. (b) Difference between the NLDAS ensemble mean and SNODAS. (c) Difference between the Noah-MP ensemble mean and SNODAS. The spatial correlation coefficients between the NLDAS–Noah-MP ensemble mean and SNODAS are also presented.

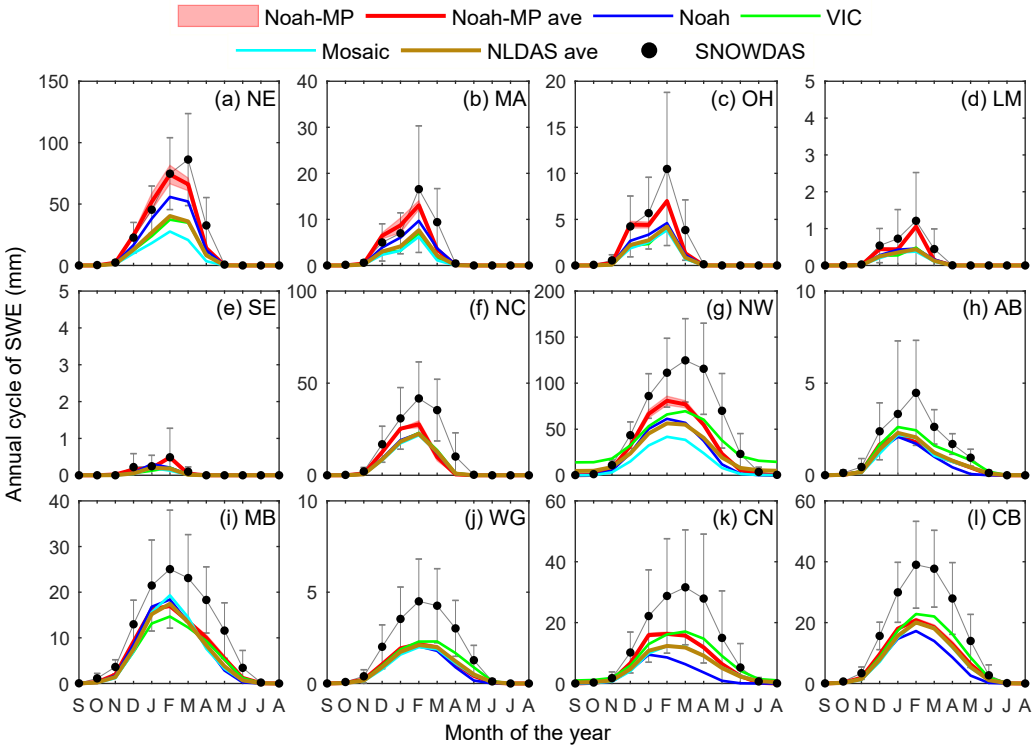

Figure 8. As in Figure 1, but for SWE between September 2004 and August 2015.

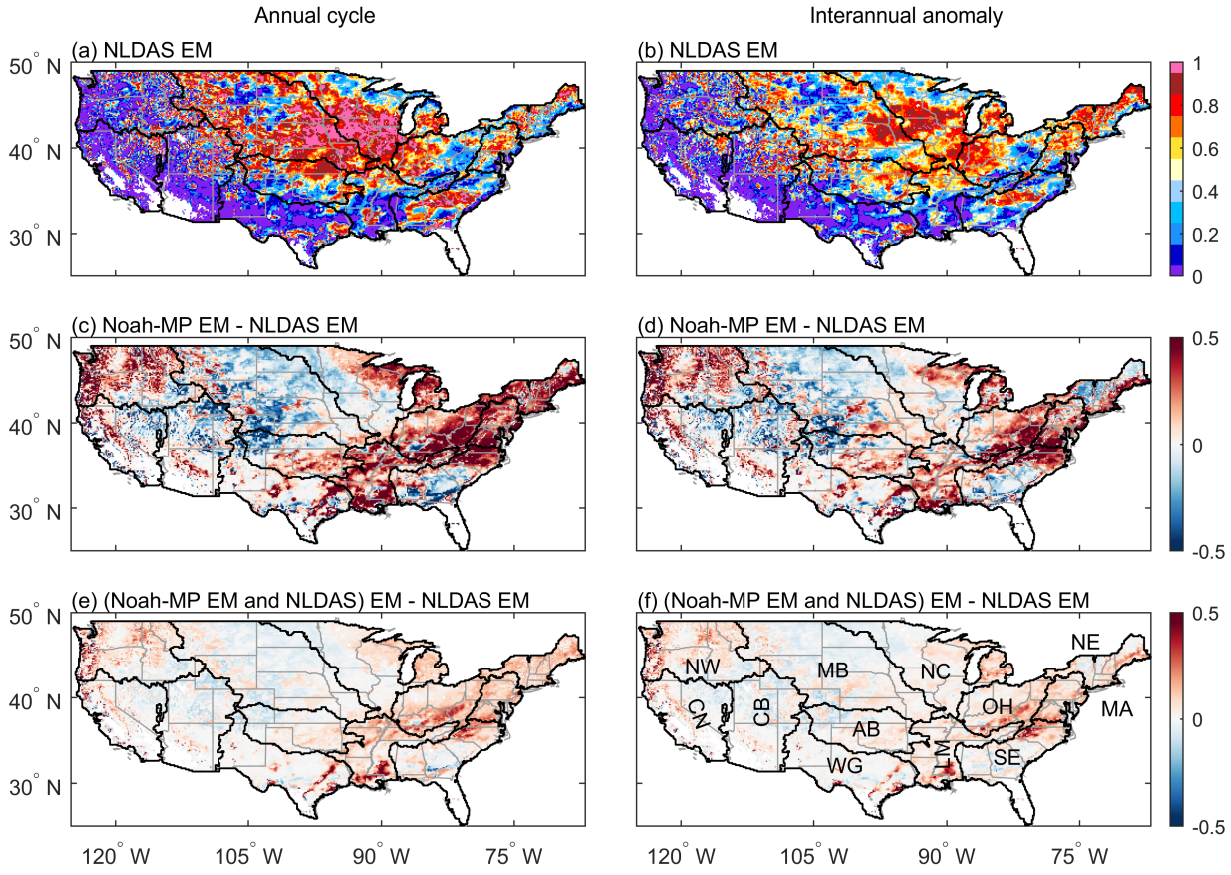

Figure 9. TSS of the ensemble means in simulating the annual cycle (left column) and interannual anomaly (right column) of SWE. (a) and (b) TSS of the NLDAS ensemble mean. (c) and (d) the difference in TSS between the Noah-MP ensemble mean and the NLDAS ensemble mean. (e) and (f) the difference between the arithmetic average of the Noah-MP ensemble mean and three NLDAS models to the NLDAS ensemble mean. The evaluation period is 2004–2015.

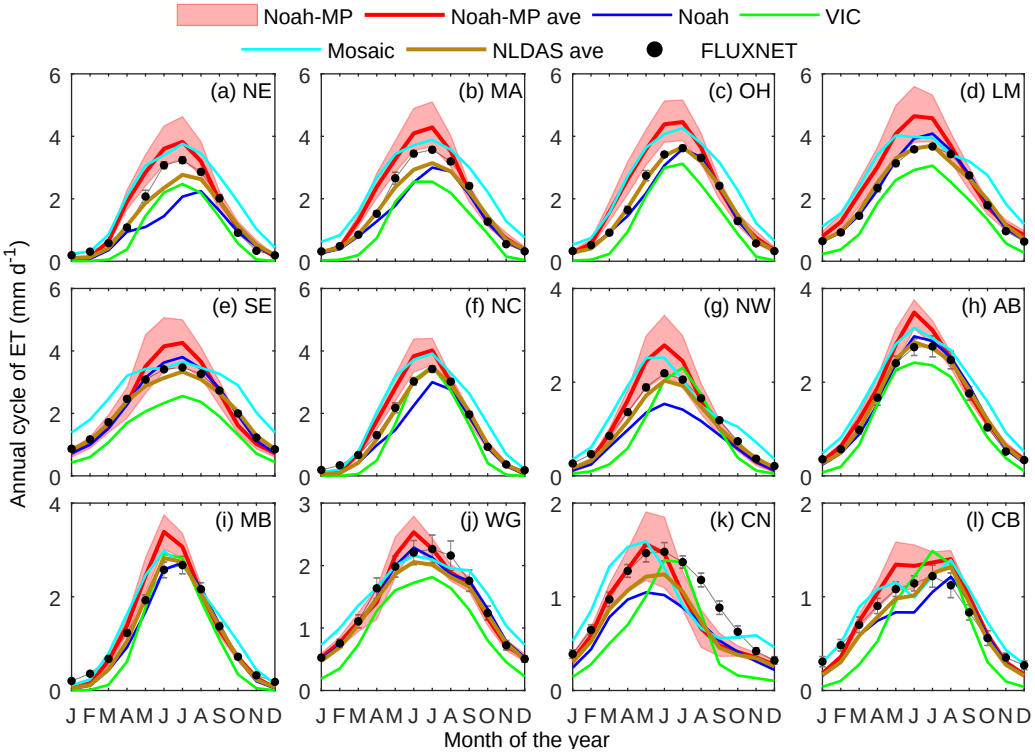

Figure 10. As in Figure 1, but for ET in the period of 1982–2011.

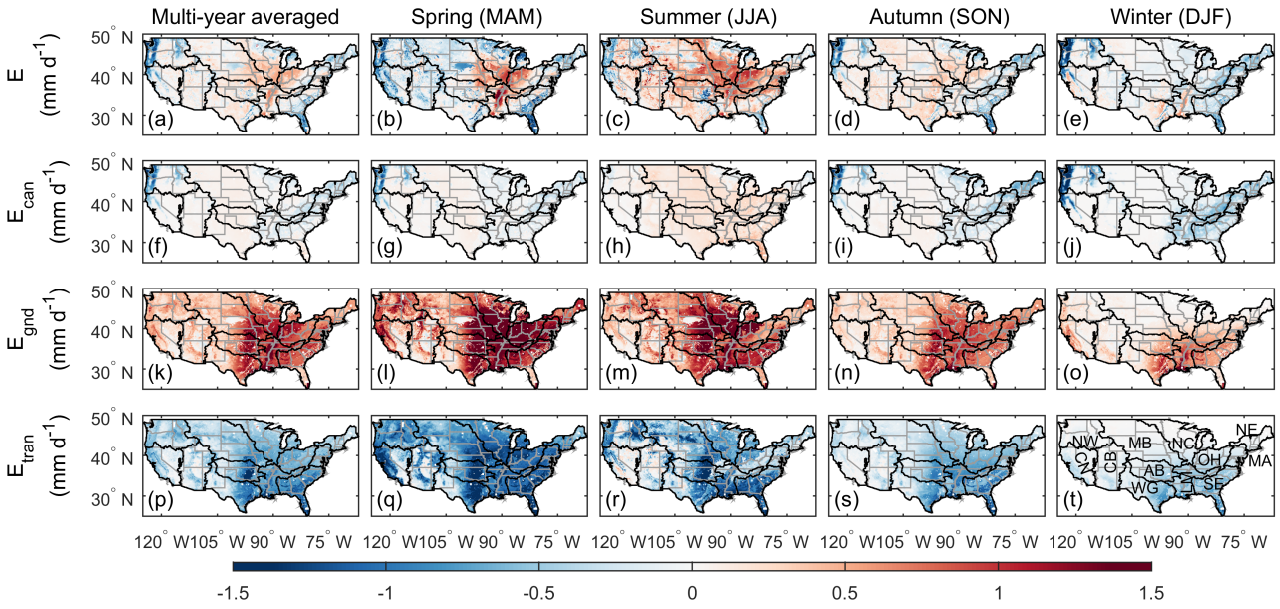

Figure 11. Differences between the Noah-MP ensemble mean and GLEAM in total ET (E), canopy evaporation (E_{can}), ground evaporation (E_{gnd}), and transpiration (E_{tran}). The units are millimeters per day ($\mathrm{mm\,day^{-1}}$).

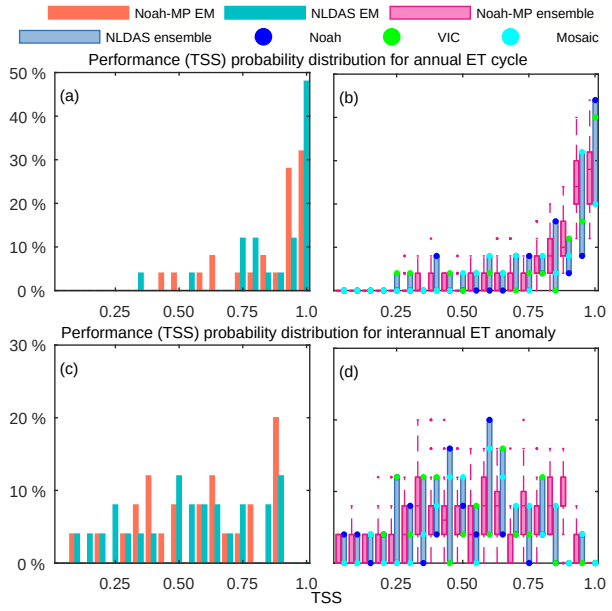

Figure 12. Probability distribution of ET's TSS for the annual cycle (a, b) and interannual anomaly (c, d). The left column compares the Noah-MP (orange) and NLDAS (cyan) ensemble means. The right column reveals the Noah-MP (magenta) and NLDAS (dark blue) ensembles. The upper, middle, and lower quantile lines of the magenta boxes show the 75th, 50th, and 25th percentile values of the Noah-MP ensemble. The upper, middle, and lower lines of the dark blue boxes show the three NLDAS models. The blue, green, and cyan dots denote Noah, VIC, and Mosaic, respectively. The evaluation period can be found in Table S1.

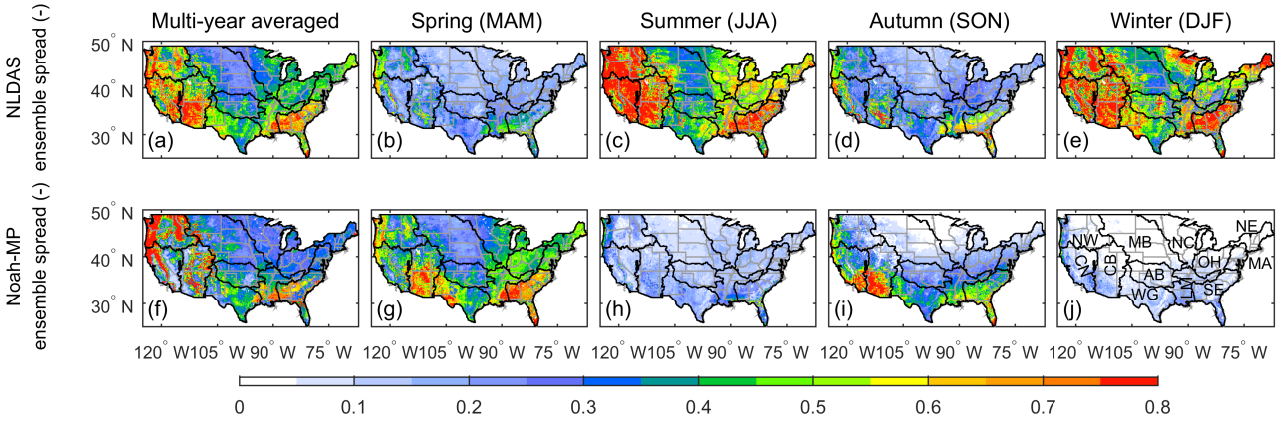

Figure 13. Normalized ensemble spread of the multi-year averaged annual (first row) and seasonal (second–fifth rows) ET from NLDAS (first column) and Noah-MP (second column). The ensemble spread is normalized by the temporal variability of the FLUXNET MTE ET calculated using equation (6).

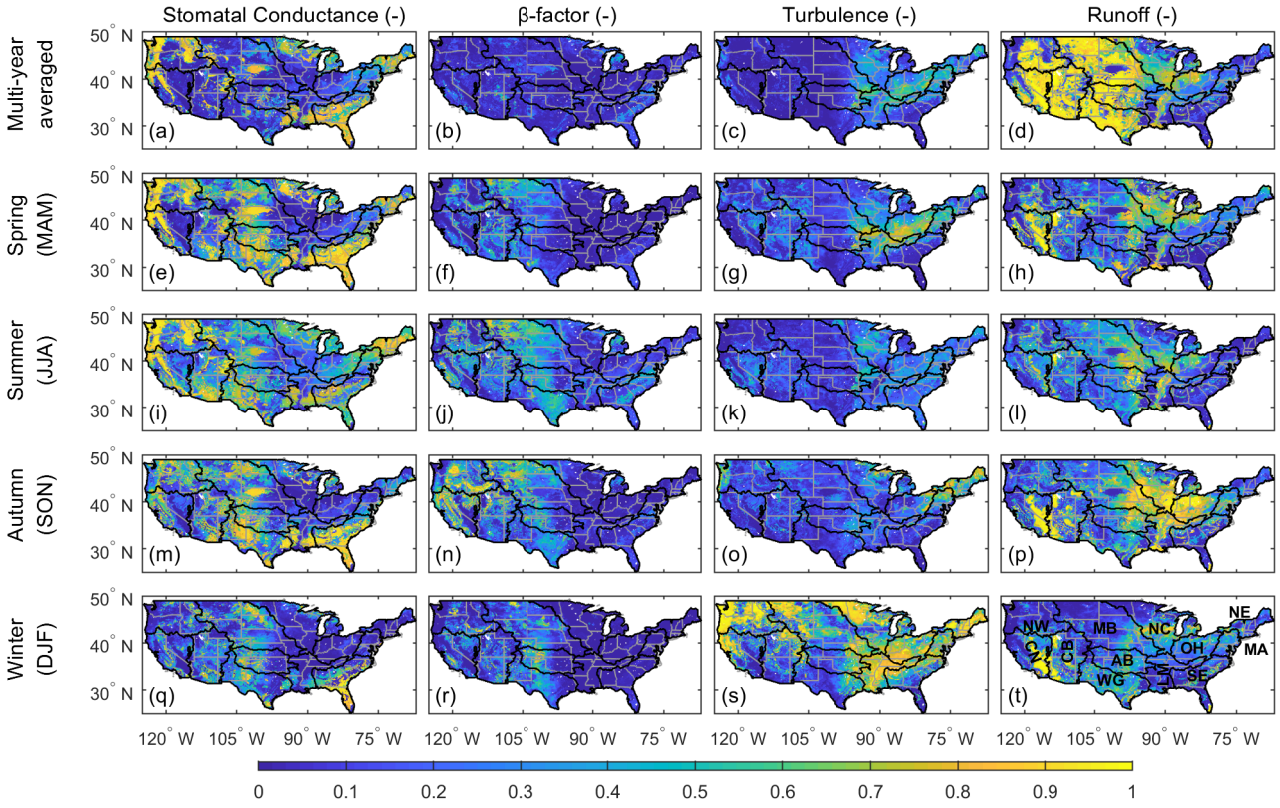

Figure 14. The Sobol' total sensitivity of the multi-year averaged and seasonal (spring—MAM, summer—JJA, autumn—SON, winter—DJF) ET to the four parameterizations: stomatal conductance, soil moisture limitation to transpiration (β factor), turbulence, and runoff. Higher values indicate higher sensitivities.