Quantifying Contributions of Uncertainties in Physical Parameterization Schemes and Model Parameters to Overall Errors in Noah-MP Dynamic Vegetation Modeling

Jianduo Li1, Fei Chen2, Xingjie Lu3, Wei Gong4,5, Guo Zhang1, and Yanjun Gan6

1State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing, China, 2National Center for Atmospheric Research, Boulder, CO, USA, 3School of Atmospheric Sciences, Sun Yat-sen University, Guangzhou, China, 4State Key Laboratory of Earth Surface Processes and Resource Ecology, Faculty of Geographical Science, Beijing Normal University, Beijing, China, 5Institute of Land Surface System and Sustainable Development, Faculty of Geographic Science, Beijing Normal University, Beijing, China, 6Department of Civil Engineering, University of Texas at Arlington, Arlington, TX, USA

Abstract Quantifying contributions of errors in model structure and parameters to biases in a land surface model (LSM) is critical for model improvement. This paper investigated the uncertainties in parameterizations and parameters in the Noah with multiparameterization (Noah-MP) LSM with dynamic vegetation using eddy flux data. First, we conducted full factorial experiments of eight land subprocesses, followed by sensitivity analysis (SA) to identify the subprocesses for which possible parameterizations made significant difference. Then, based on the full factorial experiments and SA results, we selected the statistically optimal parameterizations combination and the most biased parameterizations combination. Lastly, we calibrated the parameters in two selected parameterizations combinations. The results showed that five subprocesses—surface exchange coefficient, soil moisture β threshold, radiation transfer, runoff and groundwater, and surface resistance to evaporation—had significant influence on the model performances, and the interactions were generally low but contributed up to 80% of the variation in the performance at some sites. In the optimization period, following the criterion by Moriasi et al. (2007, https://doi.org/10.13031/2013.23153), parameter optimization improved the performance of both parameterizations combinations at most sites to be satisfactory, and the superiority between two parameterizations combinations was preserved; in the validation period, adjusting the parameterizations was more robust than parameter optimization in improving LSMS. Finally, we found that uncertainty in soil parameters was much higher than that in vegetation parameters because the optimal soil parameters were significantly different among sites with the same soil types and recommended that spatially calibrating the soil parameters was a major issue for Noah-MP dynamic vegetation modeling.

Plain Language Summary Imperfect model structure and improper parameter values are two primary causes of errors in land surface models, but the uncertainties in these two sources always interacting have rarely been analyzed together. This study identified the key land subprocesses based on sensitivity analysis and found out the “best” and “worst” parameterizations combinations for these key subprocesses. Then, the key parameters, selected based on previous work, were calibrated in two selected parameterizations combinations, respectively. The effectiveness and robustness of tuning the parameterizations and calibrating the parameters were compared. Many interesting findings were obtained to guide further land model development, and generally, combining the adjustments of both parameterizations and parameters takes full advantages of each other and provides the most effective way to reduce modeling errors. This study is also helpful for quantifying and reducing the errors in weather and climate models.

1. Introduction

Reliable simulation of the biological, physical, and hydrological processes in a land surface model (LSM) is critical for improving the accuracy of both numerical weather forecasts and climate predictions (Friedlingstein et al., 2013; Gao et al., 2016). Most published studies suggest that transpiration accounts for the majority of water vapor flux from terrestrial land to the atmosphere despite significant
uncertainties (Good et al., 2015; Jasechko et al., 2013; Miralles et al., 2011; Wei et al., 2017). Therefore, accurately representing the seasonality of vegetation is crucial for modeling the seasonal and annual exchanges of energy, water, and carbon between the land surface and the atmosphere (Arora, 2002; Prentice et al., 2007, 2015).

Implementing a dynamic vegetation model can potentially be beneficial to a more physically realistic and bioclimatically consistent model framework (Alo & Wang, 2010) and predicting the responses of vegetation to climate change at various temporal and spatial scales (Tang & Bartlein, 2008; Woodward & Lomas, 2004). In the 1990s, some LSMs predicted canopy amount and leaf area by adopting a submodel of vegetation dynamics, which consisted of a semiprocess-based photosynthesis model and an empirical respiration model (e.g., Dickinson et al., 1998; Foley et al., 1996; Thornton et al., 2002). Jiang et al. (2009) suggested that incorporating vegetation dynamics into the Noah LSM via a regional climate model is especially beneficial for seasonal precipitation forecasting in transition zones. Van den Hoof et al. (2011) developed a version of the joint UK land environment simulator (JULES) with a dynamic crop growth submodel, which was found to significantly improve the simulated land fluxes over cropland and well capture the spatial and temporal variabilities in growth conditions in Europe. However, because the dynamic vegetation models were developed with limited information, their predictions now are highly uncertain. In fact, vegetation dynamics has already been one of the largest sources of uncertainty in Earth system models (Purves & Pacala, 2008; Sitch et al., 2008). Several studies on LSMs reported that vegetation dynamics still need to be improved, especially regarding vegetation seasonality and long-term changes at regional and local scales (e.g., Anav et al., 2013; Bonan & Levis, 2006; Cadule et al., 2010; Murray-Tortarolo et al., 2013; Randerson et al., 2009; Richardson et al., 2012; Sitch et al., 2015).

The Noah with multiparameterization (Noah-MP) LSM has a significant advantage of taking into account model uncertainties because it incorporates different parameterization schemes (PSs) for various physical subprocesses in the same model dynamic structure (Niu et al., 2011; Yang et al., 2011), and therefore, with a specific combination of PSs, it can be regarded as a surrogate of another LSM. Noah-MP has been widely used as a stand-alone LSM, and it was also employed as one of the available LSM options in the Weather Research and Forecasting (WRF) model (Skamarock et al., 2008), which has a large worldwide community of registered users. A number of researches have showed the advantage of the Noah-MP LSM in reproducing the water and energy fluxes between the land surface and atmosphere (Barlage et al., 2015; Cai et al., 2014; Chen et al., 2014; Gao et al., 2015; Salamanca et al., 2018; Zhuo et al., 2019). However, among these studies, the Noah-MP LSM used default parameterizations in which the dynamic vegetation was turned off due to its significant error. For example, Ma et al. (2017) found that the Noah-MP LSM with dynamic vegetation (Dickinson et al., 1998; Yang & Niu, 2003) overestimated gross primary productivity (GPP) by 40% and evapotranspiration by 22% over the whole CONUS domain. Therefore, the main goal of this study is to quantify and attribute the errors in the Noah-MP LSM with dynamic vegetation for its future improvement and development.

There are three primary causes of errors in LSMs (Clark et al., 2011; Duan et al., 2006; Wang et al., 2009): (1) an imperfect model structure; (2) inappropriate model parameter values (PVs); and (3) incorrect model inputs, including meteorological forcing data and initial surface conditions. Obviously, parts of the errors in the Noah-MP LSM with dynamic vegetation are model structural errors caused by missing physical features. For example, Wang et al. (2018) recently implemented a vegetation optimality model of root dynamics in the Noah-MP LSM to model the extraction of groundwater through phreatophytic roots, which substantially improved the simulation of surface energy and water fluxes at a hyperarid site. Gim et al. (2017) developed a new PS to characterize the allocation of assimilated carbon to plant parts, including leaves and fine roots. In particular, interactive nutrient cycling has not been included in most dynamic vegetation models, although it was found that the long-term plant response to elevated CO₂ is likely affected by nutrients and its impact on plant carbon allocation (Zaehle et al., 2014).

In addition, the land model errors caused by the first two sources can also be reduced by tuning the PSs (Gan et al., 2019) and PVs (Wang et al., 2001, 2007), respectively. For example, Hong et al. (2014) and Zhang et al. (2016) used a natural selection approach to obtain the optimal combination of PSs in the Noah-MP LSM for the Han River basin in South Korea and a cropland site on the Tibetan Plateau in China, respectively. Li et al. (2018) found that up to 65% of model errors can be reduced by parameter
optimization for most of the plant functional types (PFTs) in two LSMs. Before optimizing the PSs or PVs, sensitivity analysis (SA), that is, the study of how a given model depends on the factors fed into it (Saltelli et al., 2008), is typically used to screen out the most sensitive subprocesses (Li et al., 2019; Zheng et al., 2019) or parameters (Gan et al., 2015; Li et al., 2013, 2016; Lu et al., 2013) in LSMs.

However, the uncertainties in the PSs and PVs always interacting has largely been ignored, which causes the optimal values of parameters, especially key parameters, to vary with the model structures. Although the Noah-MP LSM was frequently used to study the sensitivity of the simulated surface fluxes to particular subprocesses (Gan et al., 2019; Li et al., 2019; Zhang et al., 2016; Zheng et al., 2019), few studies focused on parameter optimization for the Noah-MP LSM in which all the PSs for each subprocess share the same PVs.

Therefore, the main goal of this study is to assess and reduce the errors in the Noah-MP LSM with dynamic vegetation relative to the background structural and parametric uncertainties. The objectives are to (1) quantify the impacts of the PS uncertainties on model performance; (2) investigate the optimal combination of PSs for the Noah-MP LSM with dynamic vegetation; and (3) calibrate the parameters in different combinations of PSs.

Here, as a follow-up study by Li et al. (2019), we conducted physics-ensemble simulations of eight subprocesses, including soil moisture β threshold (BTR), runoff and groundwater (RUN), surface exchange coefficient (SFC), supercooled liquid water in frozen soil (SUP), frozen soil permeability (FRO), radiation transfer (RAD), snow albedo (SNO), and surface resistance to evaporation (SRE) at multiple FLUXNET sites. We computed and evaluated the variance-based sensitivity indices (Homma & Saltelli, 1996; Sobol’, 2001), including the total effects and interactions, of these eight subprocesses for simulated sensible heat fluxes (SH), latent heat fluxes (LH), net absorbed radiation (Rnet), and GPP. Furthermore, we intercompared the PSs to determine not only the optimal combination of model parameterizations but also the combination of PSs that has a significant difference from the optimal combination. The parameters in these two combinations of PSs were then calibrated by a popular automatic optimization method, the shuffled complex evolution (SCE-UA) algorithm (Duan et al., 1993; Naeini et al., 2019). The model errors caused by different sources and the robustness of model improvement were examined and compared.

The remainder of this paper is organized as follows. Section 2 briefly summarizes the Noah-MP LSM with dynamic vegetation, selected PSs, and eddy covariance flux data; then, section 3 describes the various methods used in this study. Section 4 presents the results, and section 5 shows some associated discussions and conclusions.

2. Model and Data

As a follow-up on our previous work, this study used almost identical LSM and data sets as Li et al. (2019). Brief descriptions were given as follows.

2.1. Eddy Covariance Flux Data

This study selected the Tier-1 sites from the FLUXNET2015 data release (Baldocchi et al., 2001). The eddy covariance flux data at these sites contained hourly meteorological variables, including precipitation, surface atmospheric pressure, air temperature, air specific humidity, wind speed, downward shortwave radiation, and downward longwave radiation, which were used to drive the Noah-MP LSM, along with hourly SH, LH, Rnet, and GPP, which were used to evaluate model performance.

Two criteria were used to select the flux sites: (1) The measurements should cover at least two consecutive years. For each site, we chose the first year as the optimization period in which the PSs or PVs were tuned to reduce the model errors, and the remaining years were defined as the validation period in which the Noah-MP LSM was run with the optimal PSs or PVs to examine the robustness of the improvement. (2) In any given year, the proportion of missing data should be less than 15%. The meteorological forcing variables were gap filled using ERA-Interim (Dee et al., 2011; Vuichard & Papale, 2015) estimates provided as part of the FLUXNET2015 data set.

Consequently, we selected 92 flux sites across 11 PFTs and 10 soil texture types (STTs) (see Table A1 in the appendix for details of these sites). The PFTs, provided by FLUXNET2015, included evergreen needleleaf forest (ENF), evergreen broadleaf forest (EBF), deciduous broadleaf forest (DBF), mixed forest (MF),

LI ET AL. 3 of 23
closed shrubland (CSH), open shrubland (OSH), woody savanna (WS), savanna (SAV), grassland (GL), permanent wetland (PWL), and cropland (CRL). The STTs, classified using the soil particle size distribution data set developed by Shangguan et al. (2014), included sand (SD), loamy sand (LSD), sandy loam (SDL), silty loam (STL), loam (L), sandy clay loam (SDCL), silty clay loam (STCL), clay loam (CL), silty clay (STC), and clay (C).

2.2. The Noah-MP LSM With Dynamic Vegetation

The Noah-MP LSM (version 4.0.1) was developed using multiple parameterization options for key land-atmosphere interaction processes, such as vegetation phenology, canopy stomatal resistance, runoff, and groundwater (Chen et al., 1996, 1997; Chen & Dudhia, 2001; Ek et al., 2003; Koren et al., 1999; Schaake et al., 1996). Details concerning the PSs for each subprocess can be found in Niu et al. (2011). The PSs selected to conduct physical ensemble simulations in this study are shown in Table 1.

The dynamic vegetation parameterization in the Noah-MP LSM required the use of the Ball-Berry scheme (Ball et al., 1987). It incorporated the leaf area index (LAI) growth model of Dickinson et al. (1998). This LAI growth model has two major parts: one is a stomatal conductance-photosynthesis part, which is based on Collatz et al. (1991), with photosynthesis being calculated for both sunlit and shaded leaves, and a dynamic leaf part. Another is concerned with the changes in carbon allocation, respiration, and vegetation phenology. The Noah-MP LSM added a stem carbon-balance equation for simulating stem-rich plants (Yang & Niu, 2003). Thus, the Noah-MP LSM accounts for biochemical processes, including carbon assimilation through photosynthesis, allocation of assimilated carbon to various carbon pools (i.e., leaf, stem, wood, root, and soil), and respiration from each of the carbon pools. The leaf carbon mass, $C_{\text{leaf}}$ (g m$^{-2}$), is computed from the following:

$$\frac{\partial C_{\text{leaf}}}{\partial t} = F_{\text{leaf}} A - (S_{\text{cd}} + T_{\text{leaf}} + R_{\text{leaf}}) C_{\text{leaf}}$$

where $A$ is the total carbon assimilation rate of the sunlit and shaded leaves (g m$^{-2}$ s$^{-1}$); $F_{\text{leaf}}$ is the fraction of the assimilated carbon allocated to the leaf; $S_{\text{cd}}$ is the death rate due to cold and drought stresses; $T_{\text{leaf}}$ is the rate of leaf turnover due to senescence, herbivory, or mechanical loss; and $R_{\text{leaf}}$ is the leaf respiration rate, including maintenance and growth respiration. $F_{\text{leaf}}$ is parameterized as a function of the LAI, and the LAI is converted from $C_{\text{leaf}}$ using a specific leaf area (m$^2$ g$^{-1}$), which is a vegetation-type-dependent parameter. The vegetation greenness fraction is then simply converted from the LAI:

| Subprocess | The studied options |
|------------|---------------------|
| BTR | 1: The Noah type (Chen & Dudhia, 2001)  
2: The Community Land Model (CLM) type (Oleson et al., 2004)  
3: The simplified biosphere model (SSiB) type (Xue et al., 1991) |
| RUN | 1: TOPMODEL-based runoff scheme with simple groundwater (Niu et al., 2007)  
3: The Noah type (Schaake et al., 1996)  
4: The biosphere atmosphere transfer scheme (BATS) type (Dickinson & Kennedy, 1993; Yang & Dickinson, 1996) |
| SFC | 1: The Monin-Obukhov scheme (Brutsaert, 1982)  
2: The Noah type (Chen et al., 1997) |
| SUP | 1: General form of the freezing-point depression equation (Niu & Yang, 2006)  
2: Koren’s iteration (Koren et al., 1999) |
| FRO | 1: Hydraulic properties from total soil water and ice (Niu & Yang, 2006)  
2: Nonlinear effects, less permeable (Koren et al., 1999) |
| RAD | 1: Canopy gaps from 3-D structure and solar zenith angle (Niu & Yang, 2004)  
2: No canopy gap (Niu & Yang, 2004)  
3: Gaps from vegetated fraction (Niu & Yang, 2004) |
| SNO | 1: The BATS type (Dickinson & Kennedy, 1993)  
2: The Canadian land surface scheme (CLASS) type (Verseghy, 1991) |
| SRE | 1: Sakaguchi and Zeng’s scheme (Sakaguchi & Zeng, 2009)  
2: Seller’s scheme (Sellers et al., 1992)  
4: Different surface resistance for non-snow and snow |
Table 2
Selected Parameters, Physical Meanings and Ranges for Parameter Optimization

| Parameter | Physical meaning | Range | Reference |
|-----------|------------------|-------|-----------|
| BB        | Clapp-Hornberger b parameter | SD: [1.41, 4.17]; LSD: [2.31, 6.21]; SDL: [3.34, 6.14]; STL: [3.61, 7.05]; L: [3.59, 6.91]; SDCL: [4.39, 10.16]; STCL: [4.39, 13.09]; CL: [4.43, 11.91]; STC: [6.12, 14.66]; C: [7.62, 15.48] | Cosby et al. (1984) |
| MAXSMC    | Porosity, saturated value of soil moisture (volumetric) | SD: [0.266, 0.412]; LSD: [0.349, 0.493]; SDL: [0.346, 0.522]; STL: [0.422, 0.53]; L: [0.365, 0.513]; SDCL: [0.356, 0.452]; STCL: [0.418, 0.51]; CL: [0.411, 0.519]; STC: [0.406, 0.53]; C: [0.433, 0.503] | Cosby et al. (1984) |
| VCMX25    | Maximum rate of carboxylation at 25 °C (μmol CO2/m2/s) | [10, 100] | Lu et al. (2013) |
| LTOVRC    | Leaf turnover (s⁻¹) | [0, 1] | Arsenault et al. (2018) |
| RMF25     | Leaf maintenance respiration at 25 °C (μmol CO2/m2/s) | [0, 4] | Arsenault et al. (2018) |
| SLA       | Single-side leaf area per kg (m²/kg) | [10, 80] | Arsenault et al. (2018) |
| QE25      | Quantum efficiency at 25 °C (μmol CO₂/μmol photon) | ENF, EBF, DBF, MF, CSH, OSH, WS: [0.04, 0.12]; SAV, GL, PWL, CRL: [0.06, 0.18] | Li et al. (2016) |

\[
F_{\text{veg}} = 1 - e^{-0.52\text{LAI}}
\] (2)

This study designed a full factorial experiment taking on all possible combinations of these schemes to conduct a physical ensemble simulation, so that the total ensemble size for each site was 1,296. The ensembles allowed for the calculations of the effect of each subprocess and their interactions.

The parameters to be optimized in this study are shown in Table 2. It has been reported that for LSMs, the key parameters of simulated carbon and water fluxes are consistent (Li et al., 2016). Therefore, this study did not carry out the SA for the model parameters but selected the key parameters and defined their ranges based on previous studies. For example, Arsenault et al. (2018) found that the SH, LH, and net ecosystem exchange were sensitive to the parameters BB (Clapp-Hornberger b parameter), MAXSMC (saturated value of soil moisture), VCMX25 (maximum rate of carboxylation at 25°C), LTOVRC (leaf turnover), RMF25 (leaf maintenance respiration at 25°C), SLA (single-side leaf area per kg) and QE25 (quantum efficiency at 25°C) for the Noah-MP LSM with dynamic vegetation, which were generally consistent with Lu et al. (2013), Li et al. (2016), and Cuntz et al. (2016).

3. Methods

3.1. RMSE—Observations Standard Deviation Ratio

The root-mean-square error (RMSE)-observations standard deviation ratio (RSR) has been widely used to evaluate the performance of hydrological models (Ang & Oeurng, 2018; Golmohammadi et al., 2014) and LSMs (Li et al., 2018, 2019). It was developed by Moriasi et al. (2007) to combine both the RMSE error index and the additional observational information recommended by Legates and McCabe (1999).

For a land flux \( Y \) at each site, the RSR is defined as the ratio of the RMSE to standard deviation of observations:

\[
\text{RSR}_Y = \frac{\sqrt{\sum_{i=1}^{n} (Y_{i,\text{sim}} - Y_{i,\text{obs}})^2}}{\sqrt{\sum_{i=1}^{n} (Y_{i,\text{obs}} - \overline{Y_{\text{obs}}})^2}}
\] (3)

where \( Y_{i,\text{sim}} \) and \( Y_{i,\text{obs}} \) are the simulated and observed \( Y \) at the \( i \)th time step, respectively; \( \overline{Y_{\text{obs}}} \) is the mean of the observed \( Y \); and \( n \) is the number of time steps.

To evaluate multiple variables simultaneously, the weighted mean values of the RSRs were calculated, which were RSR\(_K\) for \( K \) model outputs \( Y_1, \ldots, Y_K \), and RSR\(_M\) for \( K \) model outputs at \( N \) sites:
\[ RSR_K = \sum_{j=1}^{K} \omega_j RSR_{Kj} \]  
(4)

and

\[ RSR_M = \sum_{i=1}^{N} \omega_i RSR_{Mi} \]  
(5)

where the weight \( \omega_j \) was set to be 1/4 for \( RSR_K \) and \( \omega_i \) was set to be 1/92 for \( RSR_M \), as there were four variables (i.e., SH, LH, Rnet, and GPP) and 92 flux sites.

The RSR index is very similar to the well-known Nash–Sutcliffe efficiency coefficient (Nash & Sutcliffe, 1970), and therefore, many literatures can be referred to reflect model performance (e.g., Gupta & Kling, 2011). This study followed the criteria given by Moriasi et al. (2007) to evaluate the model performance, which were as follows: The model prediction was very good if \( 0 \leq RSR \leq 0.5 \), good if \( 0.5 < RSR \leq 0.6 \), satisfactory if \( 0.6 < RSR \leq 0.7 \), and unsatisfactory if \( RSR > 0.7 \).

### 3.2. Variance-Based SA and Subprocess Screening

The field of SA contains a broad family of methods. Overall, they can be categorized into local SA that explores the changes of model response by varying one factor while keeping other factors constant and global SA that seeks to characterize the properties of a response surface across the entire factor space (Gan et al., 2014; Razavi & Gupta, 2015). Obviously, the global SA methods are more appropriate for this study than the local ones, because for a land subprocess, each of the PSs and all its possible combinations with PSs for other subprocesses should be comprehensively examined to confirm whether it is significantly different from the other PSs. This is also why we designed a full factorial experiment taking on all possible combinations of these PSs to conduct a physical ensemble simulation.

Razavi and Gupta (2015) reviewed that the global SA approaches were evolved from model-based methods that assumed particular model forms (typically linear or polynomial) for the underlying response surfaces, to model-free methods that make minimal assumptions regarding the functional form of the response surface. Due to the characteristics of high dimension and nonlinearity for land surface modeling, this study adopted a model-free SA method to screen out the key land surface subprocesses from the insensitive ones.

The derivative-based Morris (Campolongo et al., 2007; Morris, 1991), the variance-based Sobol’ (Sobol’, 1990, 2001), and the variogram-based analysis of response surfaces (VARS) (Razavi & Gupta, 2016a, 2016b) are model-free global SA methods and have been widely used for complex dynamical system models. Among them, the Sobol’ method quantifies the contributions of each factor and each possible two or higher-order interaction to the total variance as the sensitivity indices. The only shortcoming of the Sobol’ method is possibly that it suffers from poor computational efficiency, particularly as the problem dimension grows. For example, Rosolem et al. (2012) used 45,000 model runs to assess the Sobol’ sensitivities of 42 parameters in the Simple Biosphere 3 (SiB3) model.

As each subprocess in the Noah-MP LSM contained three PS options at most, it was affordable for us to carry out the Sobol’ SA based on a full factorial experiment to identify the key subprocesses and investigate the interactions of these subprocesses on LSM behavior.

Here, the use of the SA aims to analyze the contributions of the uncertainties in the PSs of eight physical subprocesses, that is, \( X_i \) (\( i = 1, \cdots, 8 \)), to the RSRs of model output \( Y \). For any set of subprocesses \( X \), the variance in \( Y \), \( V_Y \), can be written as

\[ V_Y = V_X[E(Y|X)] + E_X[V(Y|X)], \]  
(6)

where \( V_X[E(Y|X)] \) is the variance in the conditional expectation and \( E_X[V(Y|X)] \) is the residual part (Saltelli et al., 2008).

First, we calculated the sensitivity indices, including (1) the first-order sensitivity,

\[ S_i = \frac{V_X[E(Y|X_i)]}{V_Y}, \]  
(7)

which is the main effect of \( X_i \) on the output (the fractional contribution of \( X_i \) to the variance in \( Y \)); (2) the second-order sensitivity,
Figure 1. The total effects of PS uncertainties on the performances of SH, LH, Rnet, and GPP at each site.
which is the interactive effect (the proportion of the variation in $Y$ due to $X_i$ and $X_j$ but cannot be explained by the sum of the first-order effects of $X_i$ and $X_j$); and (3) the total effect,

$$S_{T_i} = \frac{EX_i[V(Y|X_{-i})]}{V_Y},$$

which is the sum of all sensitivity indices (first and higher orders) involving $X_i$, where $X_{-i}$ is all subprocesses other than $X_i$ (Homma & Saltelli, 1996).

Having obtained the total effects, we then used the Pareto rank-based approach introduced by Rosolem et al. (2012) to differentiate the most sensitive subprocesses from the insensitive ones for the RSR_K defined by Equation 4. This approach assigns a Pareto rank $r$ (Gupta et al., 1998) to each subprocess on the basis of a simultaneous maximization of the total effect indices for all of the objective functions. Further details can be found in Rosolem et al. (2012) and Li et al. (2019).

### 3.3. The SCE-UA Method

The SCE-UA method is a globally based automatic search algorithm proposed by Duan et al. (1993) that was designed to combine the strengths of global and local search methods, such as genetic algorithms and the simplex method, with the concepts of complex partitioning and complex shuffling.

Duan et al. (1994) conducted a detailed investigation on the selection of the SCE-UA algorithmic parameters and found that the most important parameter in SCE-UA was the number of complexes, which was dependent on the complexity of the problem. As suggested by Duan et al. (1994), this study set a value of two for the complex number to calibrate the seven parameters in the Noah-MP LSM. The SCE-UA algorithm has three independent convergence criteria. The first one is the maximum number of model evaluations during optimization (set to 3,000). The second one is the percentage of objective function improvement during the last 10 evolution loops (set to be 0.1). The third is the percentage of the normalized geometric range of the population (set to be 0.001). The searching process will stop if the population distribution range is within a small percentage such that all of the population members do not have significant differences.

Figure 2. Percentages of the subprocesses in each rank group for different seasons: (a) spring, (b) summer, (c) autumn, and (d) winter.
Figure 3. Impact of subprocess interactions on the RSRs of simulated SH (a), LH (b), Rnet (c), and GFP (d). Each pair of subprocesses is connected by an underscore (_) to represent the interaction between them.

Figure 4. Box plots of the mean RSRs for all sites considering the scheme combinations of BTR_RUN (a); BTR_SFC (b); BTR_RAD (c); BTR_SRE (d); SFC_RAD (e); SFC_SRE (f); and the individual options for SUP, FRO, and SNO (g).
Numerous researchers have investigated the use of the SCE-UA for watersheds model calibration purposes (Hogue et al., 2000; Luce & Cundy, 1994). The SCE-UA has been shown to be consistently more efficient and robust when it was compared against a variety of search methods (Franchini et al., 1998; Gan & Biftu, 1996; Kuczera, 1997; Thyer et al., 1999). Naeini et al. (2019) reviewed the history of the development and application of the SCE-UA method and showed that it had great potential to become more prominent in the field of optimization.

4. Results

4.1. Impacts of PS Uncertainty

Figure 1 shows the total effects of the PS uncertainty in each subprocess on the RSRs of SH, LH, Rnet, and GPP at the sites, as sorted by PFTs. It can be found that, for each site, the performance of each simulated flux was dominated by one or two subprocesses, with values of total effects higher than 0.5. However, the total effects and the dominant subprocesses may be different across the sites, even for those within the same PFTs. Generally, the results showed that (1) the RSRs of SH and Rnet were the most sensitive to the SFC and RAD subprocesses; (2) the RSRs of LH and GPP were the most sensitive to the BTR, RUN, SFC, and RAD subprocesses; (3) the SNO and FRO subprocesses appeared to be sensitive only at those sites with high elevations (e.g., US-GLE, CN-Dan, IT-MBo, and CN-Ha2), while the SUP subprocess was generally insensitive at all sites; and (4) for LH and GPP, the SFC subprocess was more sensitive at the forest-type sites, and for SH and Rnet, the RAD subprocess was more sensitive at the forest-type sites, especially the ENF sites.

As shown in Figure 2, for multiple surface fluxes at the flux sites, the percentages of the subprocesses in each Pareto rank group for different seasons were calculated. The results of the rank assignment were different among seasons, especially between the warm and cold seasons. For example, the SUP, FRO, and SNO subprocesses were assigned to rank 1 at much more sites in spring and winter than in summer and autumn. However, at most sites, SFC, BTR, and RAD were assigned to the rank 1 for all seasons, while SUP, FRO, and SNO were rarely in rank 1, indicating that the selection of these PSs has less impact on the performance of the Noah-MP LSM with respect to the global SH, LH, Rnet, and GPP.

Figure 3 presents the interactive effects among BTR, RUN, SFC, RAD, and SRE on the performances of the surface fluxes. Generally, the interactions of the key subprocesses were low for the flux sites, with values less than 10%. Relatively, the interactions between SFC and RAD (SFC_RAD) had higher impacts on the RSRs of SH, LH, and Rnet, and the interaction between BTR and RAD (BTR_RAD) had a higher impact on the RSR of the simulated GPP. However, for some sites, these interactions can be quite significant and much higher than the main effects of each subprocess. For example, the interactive effect of SFC_RAD at the IT-Lav ENF site on the RSR of SH was higher than 80%, and the interactive effect of BTR_SFC at the IT-SRo ENF site on the RSR of GPP was higher than 70%.

4.2. Selection of the PSs

Li et al. (2019) found that two interactive subprocesses should be jointly analyzed to identify the optimal PS. Therefore, we selected the optimal PS for the sensitive subprocesses by comparing the conditional box plots of the mean RSR at all sites with respect to the two given schemes for

Figure 5. Distribution of the RSRs calculated from physical ensemble simulations of each site, where combinations A and B use the default PVs and optimized PVs, respectively. The horizontal red dashed line corresponds to an RSR value of 0.7.
the subprocesses, with high interactions and each option for SUP, FRO, and SNO (see Figure 4).
BTR1 performed better than the other options of BTR (see Figures 4a–4d), RUN3 showed a slight superiority over RUN1 and RUN4 (see Figure 4a), and SFC1 performed better than SFC2 (see Figures 4b, 4e, and 4f). Figure 4g shows that SUP1, FRO2, and SNO2 were slightly better than SUP2, FRO1, and SNO1, respectively.

From the results, the performances of the schemes for each subprocess were closely interacted with the choices of the schemes for some other subprocesses. For example, SRE2 performed best if option 1 was chosen for SFC but performed worst if SFC2 was selected. Since BTR1 was shown to be the best choice and the BTR subprocess had higher sensitivity for the mean RSR at all sites, with the shortest lengths of the box plots, the optimal PSs for both the RAD and SRE subprocesses were selected as option 1 from Figures 4c and 4d.

Therefore, the recommended PSs—namely, global optimal schemes—for the Noah-MP LSM with dynamic vegetation were BTR1, RUN3, SFC1, SUP1, FRO2, RAD1, SNO2, and SRE1, which is known as “combination A.” To calibrate the parameters in different model structures, we also selected the PS combination with significant differences from the optimal combination, that is, BTR3, RUN1, SFC2, SUP2, FRO1, RAD2, SNO1, and SRE2, which is known as “combination B.”

4.3. Parameter Optimization for Different PS Combinations

For combinations A and B, the model parameters were calibrated by the SCE-UA method to minimize the mean RSRs of the four surface fluxes at the selected sites. Figure 5 presents a kernel density (estimation of the probability density function) plot of the RSRs for the physics ensemble simulations at each site, as associated with the RSRs calculated from combinations A and B with the default and optimized PVs.

Following the criterion by Moriasi et al. (2007) that model prediction is unsatisfactory if the RSR is greater than 0.7, it was found from the kernel density plots that for 66.3% of the selected sites, tuning the PSs can make the model performance satisfactory or better than satisfactory. For 85.9% of the sites, tuning the PVs for both combinations A and B can make the model performance satisfactory or better than satisfactory. For most sites, the superiority of combination A over combination B is preserved after parameter optimization. Even for combination B, parameter optimization is an effective way to improve the model performance from satisfactory or good to very good.
Then, we examined whether the optimal PVs significantly varied with different model structures. Figure 6 shows the differences between the optimal and default values of the vegetation (a) and soil (b) parameters for combinations A and B. In Figure 6, the difference was calculated as

\[ d = \frac{v_{\text{opt}} - v_{\text{def}}}{v_{\text{max}} - v_{\text{min}}} \]

where \( v_{\text{opt}}, v_{\text{def}}, v_{\text{max}}, \) and \( v_{\text{min}} \) were the optimized, default, maximum, and minimum values of one
parameter, respectively; the sites in Figure 6a was sorted by PFTs, and those in Figure 6b was sorted by STTs; Each subplot contains two bars: The top bar represents the difference between the optimal and default PVs for combination A, and the bottom bar represents that for combination B. It can be found that

(1) For the vegetation parameters, the optimal values for combination A were close to those for combination B; moreover, for each PFT, the optimal vegetation PVs were generally close (e.g., the parameter, QE25, was underestimated by the default values for the PFTs except for ENF), indicating that it was practical to use the same PFT-specific vegetation PVs for different model structures.

(2) For the soil parameters, the difference between the optimal values for combinations A and B was significant, and for almost every STT, the obtained soil PVs from optimization were significantly different, indicating that the uncertainty in soil parameters caused a large portion of the parametric uncertainty and that the use of the same values of soil parameters among diverse model structures may increase the parameter errors.

### 4.4. Comparison of Improvements From Tuning PSs and PVs

In this section, we focus on comparing the improved model performance from tuning the PSs and PVs. For each site, the reference simulation is that using the combination of PSs with low performances (e.g., combination B for the ES-LgS site and combination A for the US-Myb site) without parameter optimization, which was then compared to the simulations with adjusted PSs and PVs at the seasonal scale to examine the detailed seasonal model improvement from these two simulations (see Figure 7 and Table 3) and during the optimization and validation periods to verify robustness (see Figure 8).

First, the sites for which tuning the PSs and PVs can make equivalent reductions in the annual RSR of multiple fluxes were selected. Then, for these sites, Figure 7 shows the reduced mean RSRs in different seasons by tuning the PSs, PVs, and both the PSs and PVs. It was clear that tuning both the PSs and PVs (i.e., calibrating the model parameters along with choosing better physics parameterizations) was the most effective in improving the seasonal land simulations. We also noticed that tuning only the PSs or PVs could hardly improve the seasonal simulations for some sites. For example, tuning the PSs cannot improve the winter simulations for the US-Me6 site, and tuning the PVs cannot improve the summer simulation for the AU-Das site.

Therefore, we counted the frequencies of the events where little improvement was made when only the PSs or PVs were tuned (Table 3). Here, we calculated the percentages of the reduced RSR, \( \frac{RSR_0 - RSR_1}{RSR_0} \times 100\%\), where \( RSR_0 \) is the RSR calculated from the reference experiment, and \( RSR_1 \) is the corresponding RSR after the PSs or PVs were tuned. It was assumed that the percentages of the reduced RSRs with a value of less than or equal to 10% meant that the model improvement made by tuning PSs or PVs was not significant.

Table 3 shows that (1) for most sites, either approach can improve the seasonal performance significantly; (2) generally, by tuning the PVs, the performance at more sites can be improved than by tuning the PSs between combinations A and B; (3) it can be inferred that a non-negligible part of the error in the winter simulations was attributed to missing physical features in the Noah-MP LSM, as both tuning the PSs and PVs lost efficacy at 34.78% of the sites.

Figure 8 can be divided into three areas: The first area is where the reduced RSRs during the validation period are greater than both the value of 0 and the corresponding values during the optimization period; the second area is where the reduced RSRs during the

| Percentage of the reduced RSR | Spring | Summer | Autumn | Winter |
|-------------------------------|--------|--------|--------|--------|
| \( RSR_{PS} \leq 10\%\), \( RSR_{PV} \leq 10\% \) | 6.52%  | 14.13% | 25.00% | 34.78% |
| \( RSR_{PS} \leq 10\%\), \( RSR_{PV} > 10\% \) | 40.22% | 25.00% | 20.65% | 16.30% |
| \( RSR_{PS} > 10\%\), \( RSR_{PV} \leq 10\% \) | 5.43%  | 7.61%  | 9.78%  | 8.70%  |
| \( RSR_{PS} > 10\%\), \( RSR_{PV} > 10\% \) | 47.83% | 53.26% | 44.57% | 40.22% |

Note: \( RSR_{PS} \) and \( RSR_{PV} \) are the percentages of the reduced RSRs by tuning PSs and PVs, respectively.

Figure 8. Plot of the reduced RSRs during the optimization period versus the reduced RSRs during the validation period by tuning the PSs and PVs, respectively.
validation period are greater than 0 but smaller than the corresponding values during the optimization period; and the third area is where the reduced RSRs during the validation period are smaller than 0. Thus, we found that (1) most of the sites were in the second area, indicating that by tuning both the PSs and PVs, the model improvement tended to degrade during the validation period; (2) generally, tuning the PSs made a more robust model improvement. By tuning the PVs, the degradations in the improvements were more significant than those by tuning the PSs, and there was one site at which tuning the parameter reduced the RSR by up to 1.1 in the optimization period, but use of the optimized PVs increased the RSR during the validation period.

5. Discussion and Conclusions

To understand and reduce the errors in the Noah-MP LSM with dynamic vegetation, we first assessed the impacts of uncertainties in the PSs of the land surface subprocesses on model performance and then selected the statistically optimal PS combination and the combination with significant differences from the optimal combination, that is, combinations A and B, respectively. Finally, we calibrated the key parameters in combinations A and B with the SCE-UA method and compared the model improvement made by tuning the PSs and PVs. The main findings can be summarized as follows:

1. Uncertainties in five subprocesses, SFC, BTR, RAD, RUN, and SRE, had the most significant impacts on the performances of the simulated SH, LH, Rnet, and GPP by the Noah-MP LSM with dynamic vegetation;
2. Generally, the interactions of the key subprocesses were low, with values less than 10%. However, for some sites, the interactions can be quite significant with a value of greater than 80%;
3. Combining the adjustments of both PSs and PVs provides the most effective way to reduce modeling errors in the Noah-MP LSM with dynamic vegetation, particularly at the seasonal scale. Compared to tuning the PVs, tuning the PSs made more robust model improvements. More detailed findings obtained from jointly exploring the uncertainties in the PSs and PVs can be found in the discussions.

In contrast to Li et al. (2019) focusing on the reasons for model differences, this study stressed the errors in the Noah-MP LSM with dynamic vegetation, but generally, the results of the parameterization sensitivities were consistent with previous studies. For example, both this study and Li et al. (2019) agreed that the uncertainty in SFC greatly impacted the energy and water balance. The same conclusion can be found in Niu et al. (2011), Yang et al. (2011), Zhang et al. (2016), and Zheng et al. (2019), although the quantitative sensitivity of SFC varied with region due to diverse climatic conditions. Hong et al. (2014) and Gan et al. (2019) noted that evapotranspiration simulated by the Noah-MP LSM with dynamic vegetation is sensitive to RAD, SFC, RUN, and BTR, which is consistent with our results. Similar to the fact that the key parameters are generally the same for different LSMs (Li et al., 2016), the results of the parameterization sensitivities are not model dependent because, to a certain extent, the Noah-MP LSM with a specific combination of PSs can be regarded as a surrogate of another LSM. Therefore, finding the sensitive subprocesses of the Noah-MP LSM can provide scientific references for other LSMs with dynamic vegetation.

However, caution must be taken for the results of PS sensitivities. Gupta and Razavi (2018) argued that performance metric-based methods for global SA should be recognized as actually being identifiability analyses but not SA. In this sense, this study told more about the PSs for which subprocesses can be perturbed to improve model fit to the observed data than finding out the most important land subprocesses. Gupta and Razavi (2018) also showed that, due to temporal variations in the strengths of the climatic factors controlling the system drivers, use of a strictly ranked order of importance approach to characterizing sensitivity of model response to parameters is suspect, which was consistent with Prihodko et al. (2008) and Li et al. (2018). Although it was also shown that the rank assignments of land subprocesses were different among seasons, it should be admitted that the RSR used in this study integrated the daily response in a whole year onto a single metric, which unavoidably distorted important information of the Noah-MP LSM about PS sensitivities. Razavi and Gupta (2019) developed a multimethod generalized global sensitivity matrix approach that enabled a comprehensive multivariate investigation of the time-varying influence of parameters and forcings on different modeled state variables and responses, which can be an ideal choice for further SA of PSs and PVs for the Noah-MP LSM.
The results of the optimal set of PSs may differ from previous studies. For example, Zhang et al. (2016) found that for the simulated SH and LH at a crop site in Tibet, RUN3 performed worse compared to RUN1 and RUN2, and SFC2 performed better than SFC1, which are completely contrary to the conclusions of this study. Two possible explanations are that the optimal sets of PSs vary with region (Li et al., 2019) and that the optimal sets of PSs are different for different land fluxes. Hong et al. (2014) found that over the Han River basin, SFC2, RUN3, and RAD1 comprised the optimal scheme combination for evapotranspiration, but SFC1, RUN4, and RAD2 comprised the best scheme combination for runoff. As this study focused on multiple land fluxes from FLUXNET, the obtained conclusion was an average result depending on which variables and observed sites were chosen. It should be admitted that, given that the approaches we used depended on availability of observational data, the provided analysis was necessarily incomplete (Gupta & Razavi, 2018).

To date, few studies have explored or quantified the interaction effects of land surface subprocesses (Jiang et al., 2009; Koster & Milly, 1997; Li et al., 2019). Although only 92 sites were selected, it was found that the impacts of these interactions on the performances of SH and GPP can be greater than 80% at some sites, which should be taken seriously because for a region where two land surface subprocesses are highly interacted, only modifying one of the two subprocesses is insufficient for model improvement.

One of the characteristics of this study was that it jointly took the uncertainties of the PSs and PVs into consideration. Several interesting findings were obtained here. First, the superiority of either scheme combinations A or B over another was generally preserved after the process of parameter optimization. This finding indicates that to improve the land model performance, it is reasonable to tune the PS first and then calibrate the parameters in the corresponding PS. Parameter optimization should be regarded as a tool to compensate not only the model structural differences (Bonan et al., 2011; Chen et al., 2011) but also the structural error caused by missing physical features.

Second, modifying the PSs associated with calibrating the PVs can generally reduce the errors in the Noah-MP LSM with dynamic vegetation to a satisfactory level. However, for some sites (e.g., US-Los and US-Myb), tuning PSs or PVs was insufficient to improve the model due to missing physical features; inappropriate descriptions of surface conditions, such as land cover and soil types (Li, Duan, et al., 2018); or errors in the forcing data sets (Xia et al., 2005), and observations (Wilson et al., 2002; Xin et al., 2018). Attributing the errors of global LSMS to various sources is still a challenge. This study found a significant difference in the attribution of land model errors across the selected sites so that there are no proportions to approximately estimate the attributions of land model errors. This is the reason why some previous studies obtained opposite conclusions. For example, Abramowitz et al. (2007) suggested that efforts to improve the representation of fundamental processes in LSMS, rather than parameter optimization, were key to the development of LSM capability, while Li et al. (2016) showed that 60% of the model errors can be reduced by tuning the most sensitive parameters.

Third, although parameter optimization is an effective approach to reduce the errors in the Noah-MP LSM with dynamic vegetation, the optimal values of the parameters, especially the soil parameters, are still quite uncertain. It was found that for almost each STT, the obtained optimal soil PVs for the studied sites were significantly different, which is consistent with Kishné et al. (2017), who examined the Noah default soil hydraulic parameters and compared them to soil measurements across Texas. They noted that 95% of the default soil parameters were significantly different from the region-specific measured values and that it was vital to account for the spatial variability in soil parameters rather than combine these parameters into texture types alone. Moreover, it is a signal of overfitting by optimization that the optimal soil PVs differed significantly for different combinations of PSs. This can also be inferred from the finding that tuning PVs resulted in less robust model improvements than tuning PSs in this study.

Brynjarsdóttir and Hagan (2014) demonstrated that if the model structural errors were not sufficiently considered, a science-based model may be overfitted by optimization, but they also noted that only realistic priors on the model structural errors revealed the true PVs. This study inferred that the structural error in combination A was generally smaller than that in combination B, but because missing physical features made the Noah-MP LSM with combination A not a “truth” model, we did not adopt the methods by Zhang et al. (2015) to calibrate the LSMS with biased physical features. Some previous studies assumed that the structural error was normally distributed and then ruled out input parameter settings (McNeall
et al., 2016; Williamson et al., 2013) or calibrated the parameters (Hutton & Kapelan, 2015; Xu et al., 2017) based on the Bayesian theory. These novel methods provide a solution to real-world modeling applications for which the model structural error exists but cannot be quantified or corrected, but to the best of our knowledge, their superiority to traditional optimization methods has not been fully demonstrated yet, and acceptable prior information of model structural errors is still necessary.

**Acronyms and Abbreviations**

- BTR soil moisture threshold.
- C clay.
- CL clay loam.
- CRL cropland.
- CSH closed shrubland.
- DBF deciduous broadleaf forest.
- EBF evergreen broadleaf forest.
- ENF evergreen needleleaf forest.
- FRO frozen soil permeability.
- GL grassland.
- GPP gross primary productivity.
- JULES joint UK land environment simulator.
- L loam.
- LAI leaf area index.
- LH latent heat flux.
- LSD loamy sand.
- LSM land surface model.
- MF mixed forest.
- Noah-MP Noah with multiparameterization.
- OSH open shrubland.
- PFT plant functional type.
- PS parameterization scheme.
- PV parameter value.
- PWL permanent wetland.
- RAD radiation transfer.
- RMSE root mean square error.
- Rnet net absorbed radiation.
- RSR RMSE–observations standard deviation ratio.
- RUN runoff and groundwater.
- SA sensitivity analysis.
- SAV savanna.
- SCE-UA shuffled complex evolution.
- SD sand.
- SDCL sandy clay loam.
- SDL sandy loam.
- SFC surface exchange coefficient.
- SH sensible heat flux.
- SNO snow albedo.
- SRE surface resistance to evaporation.
- STC silty clay.
- STCL silty clay loam.
- STL silty loam.
- STT soil texture type.
- SUP supercooled liquid water in frozen soil.
- WS woody savanna.
Data Availability Statement

The FLUXNET2015 data set is available at this site (http://fluxnet.fluxdata.org/data/fluxnet2015-Dataset). The global soil particle-size data set is available at this site (http://globalchange.bnu.edu.cn/research/soilw).

Appendix A: Supporting Table

In this appendix, we provide information on the FLUXNET sites used in this study (Table A1).

| Table A1 | FLUXNET Sites Used in This Study, as Labeled by a Country Code (First Two Letters) and Site Name (Last Three Letters) |
|----------|------------------------------------------------------------------------------------------------------------------|
| Site     | Optimization period Validation period Soil type Latitude Longitude Elevation (m)                                      |
| Evergreen needleleaf forests              |                                                                                                                  |
| CA-NS2  | 2002                                                       2003–2004 clay 55.9058 −98.5247 269                        |
| CA-NS5  | 2003                                                       2004 clay 55.8631 −98.485 266                                |
| CA-SF1  | 2004                                                       2005–2006 sandy clay loam 54.485 −105.818 573                |
| CA-SF2  | 2003                                                       2004 loam 54.2539 −105.878 518                               |
| CN-Qia  | 2003                                                       2004–2005 loam 26.7414 115.058 112                           |
| DE-Obe  | 2010                                                       2011–2014 loam 50.7867 13.7213 741                           |
| DE-Tha  | 2000                                                       2001–2014 silt loam 50.9624 13.5652 386                      |
| FI-Hyy  | 1999                                                       2000–2004 sandy loam 61.8474 24.2948 183                     |
| FI-Sod  | 2008                                                       2009–2014 sandy clay loam 67.3619 26.6378 179                |
| IT-Lav  | 1997                                                       2009 sand 44.7171 −76.93 62                                  |
| IT-SR2  | 2013                                                       2014 sandy loam 43.732 10.2909 11                            |
| US-Me2  | 2002                                                       2003–2005 sandy loam 44.4523 121.557 1,253                   |
| US-Me6  | 2011                                                       2012–2013 sandy loam 44.3233 121.608 1,000                   |
| US-NR1  | 2000                                                       2001–2014 sandy loam 40.0329 10.2909 11                      |
| CA-Qio  | 2005                                                       2006–2010 sandy loam 49.6925 10.2909 11                       |
| IT-Rem  | 2010                                                       2011–2012 loamy sand 46.5869 11.4337 1,737                   |
| IT-SRo  | 2003                                                       2004–2009 sandy loam 43.7279 10.2844 0.5                    |
| US-GLE  | 2009                                                       2010–2014 sandy loam 41.3665 −106.24 3,193                   |
| Evergreen broadleaf forests               |                                                                                                                  |
| AU-Cum  | 2013                                                       2014 loam −33.6133 150.722 27                                |
| CN-Din  | 2003                                                       2004–2005 clay loam 23.1733 112.536 298                      |
| FR-Pue  | 2001                                                       2002–2011 clay loam 43.7414 112.536 298                      |
| GF-Guy  | 2004                                                       2005–2014 silt loam 5.27877 −52.9249 48                       |
| IT-Cpz  | 2005                                                       2006–2008 loam 41.7053 12.3761 21                             |
| AU-Tum  | 2002                                                       2003–2011 sandy loam −35.6566 148.152 1,259                  |
| Deciduous broadleaf forests               |                                                                                                                  |
| DE-Hai  | 2000                                                       2001–2011 clay 51.0792 10.453 464                             |
| DK-Sor  | 2003                                                       2004–2010 sandy loam 55.4859 11.6446 46                       |
| FR-Fon  | 2005                                                       2006–2013 clay 48.4764 2.78014 109                            |
| IT-CA1  | 2012                                                       2013 sandy loam 42.3804 12.0266 204                           |
| IT-PT1  | 2003                                                       2004 loam 45.2009 9.06104 109                                |
| IT-Ro1  | 2003                                                       2004–2006 loam 42.3903 11.9209 173                           |
| IT-Ro2  | 2005                                                       2006 sandy loam 42.5378 −72.1715 344                          |
| US-Ha1  | 2010                                                       2011–2012 sandy loam 45.5598 −84.7138 239                     |
| US-UMB  | 2000                                                       2001–2007 sand 45.8126 8.63358 220                           |
| IT-Col  | 2007                                                       2008 loam 45.8059 9.0799 519                                  |
| US-WCr  | 2000                                                       2001–2006 loam 45.8059 9.0799 519                            |
| US-MMS  | 2001                                                       2002–2014 silt loam 39.3232 −86.4131 286                     |
| Mixed forests                            |                                                                                                                  |
| BE-Vie  | 1997                                                       1998–2003 silt loam 50.3051 5.99805 432                      |
| CN-Cha  | 2003                                                       2004–2005 silt loam 42.4025 128.096 767                      |
| US-Syy  | 2002                                                       2003–2007 sandy loam 46.242 −89.3477 541                      |
| BE-Bra  | 1997                                                       1998–2002 sandy loam 51.3092 4.52056 32                       |
| Closed shrublands                        |                                                                                                                  |
| IT-Noe  | 2006                                                       2007 silty clay loam 40.6062 8.1517 25                         |

10.1029/2019MS001914
### Table A1

**Continued**

| Site          | Optimization period | Validation period | Soil type     | Latitude  | Longitude | Elevation (m) |
|---------------|---------------------|-------------------|---------------|-----------|-----------|---------------|
| US-KS2        | 2005                | 2006              | sandy loam    | 28.6086   | −80.6715  | 2             |
| **Open shrublands** |          |                   |               |           |           |               |
| CA-NS6        | 2002                | 2003–2004         | clay          | 55.9167   | −98.9644  | 277           |
| CA-NS7        | 2003                | 2004              | clay          | 56.6358   | −99.9483  | 273           |
| CA-SF3        | 2004                | 2005              | loam          | 54.0916   | −106.005  | 543           |
| ES-LgS        | 2007                | 2008              | sandy loam    | 37.0979   | −2.96583  | 2,261         |
| US-Whs        | 2008                | 2009–2014         | sandy loam    | 31.7438   | −110.052  | 1,374         |
| **Woody savannas** |            |                   |               |           |           |               |
| AU-Gin        | 2013                | 2014              | loamy sand    | −31.375   | 115.65    | 49            |
| AU-How        | 2010                | 2011–2014         | sandy clay loam | −12.4952  | 131.15    | 39            |
| US-SRM        | 2004                | 2005–2014         | loamy sand    | 31.8214   | −110.866  | 1,120         |
| US-Ton        | 2002                | 2003              | loam          | 38.4316   | −120.966  | 169           |
| **Savannas**  |                     |                   |               |           |           |               |
| AU-Cpr        | 2011                | 2012–2013         | sandy loam    | −34.0021  | 140.589   | 63            |
| AU-DaS        | 2010                | 2011–2013         | sandy clay loam | −14.1593  | 131.388   | 75            |
| SD-Dem        | 2005                | 2006–2009         | sandy loam    | 13.8289   | 30.4783   | 537           |
| AU-GWW        | 2013                | 2014              | clay loam     | −30.1913  | 120.654   | 449           |
| **Grasslands**|                     |                   |               |           |           |               |
| AT-Neu        | 2004                | 2005              | loam          | 47.1167   | 11.3175   | 970           |
| AU-DaP        | 2009                | 2010–2012         | sandy loam    | −14.0633  | 131.318   | 71            |
| AU-Stp        | 2010                | 2011–2014         | sandy clay loam | −17.1507  | 133.35    | 228           |
| AU-Ync        | 2013                | 2014              | sandy clay loam | −34.9893  | 146.291   | 126           |
| CH-Cha        | 2008                | 2009–2010         | loam          | 47.2102   | 8.41044   | 394           |
| CH-Oc1        | 2002                | 2003–2008         | loam          | 47.2858   | 7.73194   | 453           |
| CN-Dan        | 2004                | 2005              | loam          | 30.4978   | 91.0664   | 4,316         |
| CN-Du2        | 2007                | 2008              | sandy loam    | 42.0467   | 116.284   | 1,320         |
| CN-HeM        | 2002                | 2003              | loam          | 37.37     | 101.18    | 4,003         |
| DE-Hri        | 2004                | 2005–2014         | loam          | 50.95     | 13.5126   | 378           |
| IT-MBo        | 2003                | 2004–2008         | loam          | 46.0147   | 11.0458   | 1,554         |
| US-AR2        | 2010                | 2011              | loam          | 36.6358   | −99.5975  | 647           |
| US-SRG        | 2009                | 2010–2014         | loamy sand    | 31.7894   | −110.828  | 1,289         |
| US-Var        | 2001                | 2002–2003         | loam          | 38.4133   | −120.951  | 159           |
| US-WkG        | 2005                | 2006–2014         | sandy loam    | 31.7365   | −109.942  | 1,528         |
| AU-Emr        | 2012                | 2013              | clay          | −23.8587  | 148.475   | 175           |
| CH-Fru        | 2007                | 2008–2010         | loam          | 47.1158   | 8.53778   | 982           |
| CN-Cng        | 2008                | 2009              | silty clay    | 44.5934   | 123.509   | 142           |
| US-AR3        | 2010                | 2011              | loam          | 36.4267   | −99.42    | 613           |
| US-Cop        | 2002                | 2003              | sandy loam    | 38.09     | −109.39   | 1905          |
| **Permanent wetlands** |      |                   |               |           |           |               |
| CN-Ha2        | 2003                | 2004–2005         | loam          | 37.6086   | 101.327   | 3,200         |
| CZ-wet        | 2007                | 2008–2014         | loam          | 49.0247   | 14.7704   | 426           |
| FI-Lom        | 2007                | 2008–2009         | sandy loam    | 67.9972   | 24.2092   | 270           |
| US-Los        | 2000                | 2001–2008         | loamy sand    | 46.0827   | −89.9792  | 482           |
| US-MyB        | 2011                | 2012–2014         | sandy loam    | 38.0498   | −121.765  | −4            |
| **Croplands** |                     |                   |               |           |           |               |
| BE-Lom        | 2005                | 2006              | silt loam     | 50.5516   | 4.74613   | 168           |
| DE-Geb        | 2001                | 2002–2014         | silt loam     | 51.1001   | 10.9143   | 162           |
| DE-Kle        | 2006                | 2007–2014         | loam          | 50.8931   | 13.5224   | 481           |
| DE-Seh        | 2008                | 2009              | silt loam     | 50.8706   | 6.44965   | 102           |
| IT-BCI        | 2006                | 2007              | sandy loam    | 40.5238   | 14.9574   | 10            |
| US-ARM        | 2003                | 2004–2007         | loam          | 36.6058   | −97.4888  | 316           |
| US-Ne1        | 2002                | 2003–2012         | silty clay loam | 41.1651  | −96.4766  | 358           |
| US-Twt        | 2010                | 2011–2014         | sandy loam    | 38.1087   | −121.653  | −6.5          |
| US-Ne2        | 2002                | 2003–2012         | silty clay loam | 41.1649  | −96.4701  | 356           |
| US-Ne3        | 2004                | 2005–2012         | silty clay loam | 41.1797  | −96.4397  | 357           |
Acknowledgments
We thank Guoyue Niu from the University of Arizona for his help with our questions on model simulations and parametric uncertainty. This research was jointly supported by the National Natural Science Foundation for Young Scientists of China (Grants 41605073), the National Key R&D Program of China (Grants 2018YFC1507003 and 2017YFC1404000), and the State Key Program of the National Natural Science Foundation of China (Grant 91537211).

References
Abramowicz, G., Pitman, A., Gupta, H., Kowalczyk, E., & Wang, Y. (2007). Systematic bias in land surface models. *Journal of Hydrometeorology*, 8(5), 989–1001. https://doi.org/10.1175/jhm628.1
Alo, C. A., & Wang, G. (2010). Role of dynamic vegetation in regional climate predictions over western Africa. *Climate Dynamics*, 35(5), 907–922. https://doi.org/10.1007/s00382-010-0744-z
Anar, A., Murray-Tortoroli, G., Friedlingstein, P., Sitch, S., Piao, S., & Zhu, Z. (2013). Evaluation of land surface models in reproducing satellite derived leaf area index over the high-latitude Northern Hemisphere. Part II: Earth system models. *Remote Sensing*, 5(3), 3637–3666. https://doi.org/10.3390/RS5036367
Ang, R., & Deuring, C. (2018). Simulating streamflow in an ungauged catchment of Tonlesap Lake Basin in Cambodia using soil and water assessment tool (SWAT) model. *Water Science*, 32(1), 89–101. https://doi.org/10.1016/j.wasj.2017.12.002
Aorca, V. (2002). Modeling vegetation as a dynamic component in soil-vegetation-atmosphere transfer schemes and hydrological models. *Reviews of Geophysics*, 40(2), 1006. https://doi.org/10.1029/2001RG000103
Arsenault, K. R., Nearing, G. S., Wang, S., Yathreendra, S., & Peters-Lidard, C. D. (2018). Parameter sensitivity of the Noah-MP land surface model with dynamic vegetation. *Journal of Hydro meteorology*, 19(5), 815–830. https://doi.org/10.1175/jhm-d-17-0265.1
Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., et al. (2001). FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bulletin of the American Meteorological Society*, 82(11), 2415–2434. https://doi.org/10.1175/1520-0477(2001)0822415:FANTTS2.3.CO;2
Ball, J. T., Woodrow, I. E., & Berry, J. A. (1987). A model predicting stomatal conductance and its contribution to the control of photosynthesis under different environmental conditions. In *Progress in photosynthesis research* (pp. 221–224). Nijhoff, Dordrecht: Springer.
Barlage, M., Tewari, M., Chen, F., Miguez-Macho, G., Yang, Z.-L., & Niu, G.-Y. (2015). The effect of groundwater interaction in North American regional climate simulations with WRF/Noah-MP. *Climatic Change*, 129(3–4), 485–498. https://doi.org/10.1007/s10584-014-1308-8
Bonan, G. B., Lawrence, P. J., Oleson, K. W., Levis, S., Jung, M., Reichstein, M., et al. (2011). Improving canopy processes in the community land model version 4 (CLM4) using global flux fields empirically inferred from FLUXNET data. *Journal of Geophysical Research*, 116, G02014. https://doi.org/10.1029/2010JG001593
Bonan, G. B., & Levis, S. (2006). Evaluating aspects of the community land and atmosphere models (CLM3 and CAM3) using a dynamic global vegetation model. *Journal of Climate*, 19(11), 2290–2301. https://doi.org/10.1175/JCLI3741.1
Bruaet, W. A. (1982). *Evaporation into the atmosphere* (p. 299). Dordrecht, Netherlands: D. Reidel.
Brynjardóttir, J., & O’Hagan, A. (2014). Learning about physical parameters: The importance of model discrepancy. *Inverse Problems*, 30(11), 144007. https://doi.org/10.1088/0266-5611/30/11/144007
Cadou, P., Friedlingstein, P., Bopp, L., Sitch, S., Jones, C. D., Claes, P., et al. (2010). Benchmarking coupled climate-monsoon models against long-term atmospheric CO2 measurements. *Global Biogeochemical Cycles*, 24, GB2016. https://doi.org/10.1029/2009GB003556
Cai, X., Yang, Z. L., David, C. H., Niu, G. Y., & Rodell, M. (2014). Hydrological evaluation of the Noah-MP land surface model for the Mississippi River basin. *Journal of Geophysical Research: Atmospheres*, 119, 23–38. https://doi.org/10.1002/2013JD020792
Campolongo, F., Cariboni, J., & Sattelli, A. (2007). An effective screening design for sensitivity analysis of large models. *Environmental Modelling & Software*, 22(10), 1509–1518. https://doi.org/10.1016/j.envsoft.2006.10.004
Chen, F., & Dudhia, J. (2001). Coupling an advanced land surface-hydrology model with the Penn State-NCAR M55 modeling system Part I: Model implementation and sensitivity. *Monthly Weather Review*, 129(4), 569–585. https://doi.org/10.1175/1520-0493(2001)1292569.4 cuckhls2.0.co;2
Chen, F., Janjić, Z., & Mitchell, K. (1997). Impact of atmospheric surface-layer parameterizations in the new land-surface scheme of the NCEP mesoscale eta model. *Boundary-Layer Meteorology*, 85(3), 391–421. https://doi.org/10.1023/A:100531001463
Chen, F., Liu, C., Dudhia, J., & Chen, M. (2014). A sensitivity study of high-resolution regional climate simulations to three land surface models over the western United States. *Journal of Geophysical Research: Atmospheres*, 119, 7271–7291. https://doi.org/10.1002/2014JD021827
Chen, F., Mitchell, K., Schaake, J., Xue, Y., Pan, H.-L., Koren, V., et al. (1996). Modeling of land surface evaporation by four schemes and comparison with FIFE observations. *Journal of Geophysical Research: Atmospheres*, 101(D3), 7251–7268. https://doi.org/10.1029/95JD02165
Chen, H., Dickinson, R. E., Dai, Y., & Zhou, L. (2011). Sensitivity of simulated terrestrial carbon assimilation and canopy transpiration to different stomatal conductance and carbon assimilation schemes. *Climate Dynamics*, 36(5–6), 1037–1054. https://doi.org/10.1007/s00382-010-0741-2
Clark, M. P., Kavetski, D., & Fenicia, F. (2011). Pursuing the method of multiple working hypotheses for hydrological modeling. *Water Resources Research*, 47, W09930. https://doi.org/10.1029/2010WR009827
Collatz, G. J., Ball, J. T., Grivet, C., & Berry, J. A. (1991). Physiological and environmental regulation of stomatal conductance, photosynthesis and transpiration: A model that includes a laminar boundary layer. *Agricultural and Forest Meteorology*, 54(2–4), 107–136. https://doi.org/10.1016/0168-1923(91)90002-8
Cosby, B., Hornberger, G., Clapp, R., & Ginn, T. (1984). A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils. *Water Resources Research*, 20, 682–690. https://doi.org/10.1029/WR020i006p06862
Cuntz, M., Mai, J., Samaniego, L., Clark, M., Wulfmeyer, V., Branch, O., et al. (2016). The impact of standard and hard-coded parameters on the hydrologic fluxes in the Noah-MP land surface model. *Journal of Geophysical Research: Atmospheres*, 121, 10,676–10,700. https://doi.org/10.1002/2016JD025097
Dee, D. P., Uppala, S. M., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656), 553–597. https://doi.org/10.1002/qj.828
Dickinson, R. E., Henderson-Sellers, A., & Kennedy, P. J. (1993). Biosphere-atmosphere transfer scheme (BATS) version 1e as coupled to the NCAR Community Climate Model. NCAR Technical Note. Boulder, CO: National Center for Atmospheric Research. Climate and Global Dynamics Division.
Dickinson, R. E., Saitoh, M., Bryant, R., & Graulich, L. (1998). Interactive canopies for a climate model. *Journal of Climate*, 11(11), 2823–2836. https://doi.org/10.1175/1520-0442(1998011283:icfcm2.0.co;2
Duan, Q., Schaake, J., Andreaeissian, V., Franks, S., Goetzii, G., Gupta, H., et al. (2006). Model parameter estimation experiment (MOPEX): An overview of science strategy and major results from the second and third workshops. *Journal of Hydrology*, 320(1–2), 3–17. https://doi.org/10.1016/j.jhydrol.2005.07.031
Duan, Q., Soroshian, S., & Gupta, V. K. (1994). Optimal use of the SCE-UA global optimization method for calibrating watershed models. *Journal of Hydrology, 158*(3-4), 265–284. https://doi.org/10.1016/0022-1694(94)90057-4

Duan, Q. Y., Gupta, V. K., & Soroshian, S. (1993). Shuffled complex evolution approach for effective and efficient global minimization. *Journal of Optimization Theory and Applications, 76*(3), 501–521. https://doi.org/10.1007/BF00939380

Ek, M., Mitchell, K., Lin, Y., Rogers, E., Grummman, P., Koren, V., et al. (2003). Implementation of Noah land surface model advances in the National Centers for environmental prediction operational mesoscale eta model. *Journal of Geophysical Research, 108*(D22). 2002JD003296. https://doi.org/10.1029/2002JD003296

Foley, J. A., Prentice, I. C., Ramankutty, N., Levis, S., Pollard, D., Sitch, S., & Haxeltine, A. (1996). An integrated biosphere model of land surface processes, terrestrial carbon balance, and vegetation dynamics. *Global Biogeochemical Cycles, 10*(4), 603–628. https://doi.org/10.1029/95GB02692

Franchini, M., Galeati, G., & Berra, S. (1998). Global optimization techniques for the calibration of conceptual rainfall-runoff models. *Hydrological Sciences Journal, 43*(3), 443–458. https://doi.org/10.1080/02626669809492137

Friedlingstein, P., Meinshausen, M., Arora, V. K., Jones, C. D., Anav, A., Liddicoat, S. K., & Knutti, R. (2013). Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks. *Journal of Climate, 27*(2), 511–526. https://doi.org/10.1175/JCLI-D-12-00579.1

Gan, Y., Duan, Q., Gong, W., Tong, C., Sun, Y., Chu, W., et al. (2014). A comprehensive evaluation of various sensitivity analysis methods: A case study with a hydrological model. *Environmental Modelling & Software, 51*, 269–285. https://doi.org/10.1016/j.envsoft.2013.09.031

Gan, Y., Liang, X.-Z., Duan, Q., Chen, F., Li, J., & Zhang, Y. (2019). 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

Gao, Y., Liang, X. Z., Duan, Q., Choi, H. I., Dai, Y., & Wu, H. (2015). Stepwise sensitivity analysis from qualitative to quantitative: Application to the terrestrial hydrological modeling of a conjunctive surface-subsurface process (CSSP) land surface model. *Journal of Advances in Modeling Earth Systems, 7*, 648–669. https://doi.org/10.1002/2014MS000406

Gao, Y., Xiao, L., Chen, D., Chen, F., Xu, J., & Xu, Y. (2016). Quantification of the relative role of land-surface processes and large-scale forcing in dynamic downscaling over the Tibetan plateau. *Climate Dynamics, 48*(5-6), 1705–1721. https://doi.org/10.1007/s00382-016-3168-6

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

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

Golmohammadi, G., Prasher, S., Madani, A., & Rudra, R. (2014). Evaluating three hydrological distributed watershed models: MIKE-SHE, APEX, SWAT. *Hydrology, 1*(1), 20–39. https://doi.org/10.3390/hydrology1010028

Good, S. P., Noone, D., & Bowen, G. (2015). Hydrologic connectivity constrains partitioning of global terrestrial water fluxes. *Science, 349*(6244), 175–177. https://doi.org/10.1126/science.aaa5931

Gupta, H. V., & Kling, H. (2011). On typical range, sensitivity, and normalization of mean squared error and Nash-Sutcliffe efficiency type metrics. *Water Resources Research, 47*, W10601. https://doi.org/10.1029/2011WR010962

Gupta, H. V., & Razavi, S. (2018). Revisiting the basis of sensitivity analysis for dynamical earth system models. *Water Resources Research, 54*, 8692–8717. https://doi.org/10.1002/2018WR022668

Gupta, H. V., Soroshian, S., & Yapo, P. O. (1998). Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resources Research, 34*(4), 751–763. https://doi.org/10.1029/97WR03495

Hogue, T. S., Soroshian, S., Gupta, H., Holz, A., & Braatz, D. (2000). A multistep automatic calibration scheme for river forecasting models. *Journal of Hydrometeorology, 1*(6), 524–542. https://doi.org/10.1175/1525-7541(2000)00100524:AMACSF2.0.CO;2

Homma, T., & Saltelli, A. (1996). Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering and System Safety, 52*(1), 1–17. https://doi.org/10.1016/0951-8320(95)00002-6

Huang, S., Yu, X., Park, S. K., Cheg, Y. S., & Mysyng, B. (2014). Assessing optimal set of implemented physical parameterization schemes in a multi-physics land surface model using genetic algorithm. *Geoscientific Model Development, 7*(5), 2517–2529. https://doi.org/10.5194/gmd-7-2517-2014

Hutton, C. J., & Kapelan, Z. (2015). A probabilistic methodology for quantifying, diagnosing and reducing model structural and predictive errors in short term water demand forecasting. *Environmental Modelling & Software, 66*, 87–97. https://doi.org/10.1016/j.envsoft.2014.12.021

Jasechko, S., Sharp, Z. D., Gibson, J. J., Birks, S. J., Yi, Y., & Fawcett, P. J. (2013). Terrestrial water fluxes dominated by transpiration. *Nature, 496*, 347. https://doi.org/10.1038/nature12183

Jiang, X., Niu, G.-Y., & Yang, Z.-L. (2009). Impacts of vegetation and groundwater dynamics on warm season precipitation over the Central United States. *Journal of Geophysical Research, 114*, D06109. https://doi.org/10.1029/2008JD010756

Kishnè, A. S., Yimam, Y. T., Morgan, C. L., & Dornblaser, B. C. (2017). Evaluation and improvement of the default soil hydraulic parameters for the Noah land surface model. *Geoderma, 285*, 247–259. https://doi.org/10.1016/j.geoderma.2016.09.022

Koren, V., Schaake, J., Mitchell, K., Duan, Q. Y., Chen, F., & Baker, J. (1999). A parameterization of snowpack and frozen ground intended for NCEP weather and climate models. *Journal of Geophysical Research, 104*(D16), 19,569–19,585. https://doi.org/10.1029/1999JD900232

Koster, R. D., & Milly, P. (1997). The interplay between transpiration and runoff formulations in land surface schemes used with atmospheric models. *Journal of Climate, 10*(7), 1579–1591. https://doi.org/10.1175/1520-0442(1997)0101578:TBATRE2.0.CO;2

Kuczera, G. (1997). Efficient subspace probabilistic parameter optimization for catchment models. *Water Resources Research, 33*(1), 177–185. https://doi.org/10.1029/96WR02671

Legates, D. R., & McCabe, G. J. Jr. (1999). Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. *Water Resources Research, 35*(1), 233–241. https://doi.org/10.1029/1998WR900018

Li, J., Chen, F., Zhang, G., Barlage, M., Gan, Y., Xin, Y., & Wang, C. (2018). Impacts of land cover and soil texture uncertainty on land model simulations over the central Tibetan plateau. *Journal of Advances in Modeling Earth Systems, 10*, 2121–2146. https://doi.org/10.1029/2018MS001377

Li, J., Duan, Q., Wang, Y.-P., Gong, W., Gan, Y., & Wang, C. (2018). Parameter optimization for carbon and water fluxes in two global land surface models based on surrogate modelling. *International Journal of Climatology, 38*(S1), e1016–e1031. https://doi.org/10.1002/joc.5428
Li, J., Duan, Q. Y., Gong, W., Ye, A., Dai, Y., Miao, C., et al. (2013). Assessing parameter importance of the common land model based on qualitative and quantitative sensitivity analysis. Hydrology and Earth System Sciences, 17(8), 3279–3293. https://doi.org/10.5194/hess-17-3279-2013

Li, J., Wang, Y.-P., Duan, Q., Lu, X., Pak, B., Wilshire, A., et al. (2016). Quantification and attribution of errors in the simulated annual gross primary production and latent heat fluxes by two global land surface models. Journal of Advances in Modeling Earth Systems, 8, 1270–1288. https://doi.org/10.1002/2015MS000583

Li, J., Zhang, G., Chen, F., Peng, X., & Gan, Y. (2019). Evaluation of land surface subprocesses and their impacts on model performance with global flux data. Journal of Advances in Modeling Earth Systems, 11, 1329–1348. https://doi.org/10.1002/2019MS001606

Lu, X., Wang, Y.-P., Ziehn, T., & Dai, Y. (2013). An efficient method for global parameter sensitivity analysis and its applications to the Australian community land surface model (CABLE). Agricultural and Forest Meteorology, 182-183, 292–303. https://doi.org/10.1016/j.agrformet.2013.04.003

Luce, C. H., & Cundy, T. W. (1994). Parameter identification for a runoff model for forest roads. Water Resources Research, 30(4), 1057–1069. https://doi.org/10.1029/93WR03348

Ma, N., Niu, G.-Y., Xia, Y., Cai, X., Zhang, Y., Ma, Y., & Fang, Y. (2017). A systematic evaluation of Noah-MP in simulating land-atmosphere energy, water, and carbon exchanges over the continental United States. Journal of Geophysical Research: Atmospheres, 122, 12,245–12,268. https://doi.org/10.1002/2017JD027597

McNeall, D., Williams, J., Booth, B., Betts, R., Challenor, P., Wilshire, A., & Sexton, D. (2016). The impact of structural error on parameter constraint in a climate model. Earth System Dynamics, 7(4), 917–935. https://doi.org/10.5194/esd-7-917-2016

Miralles, D. G., De Jeu, R. A. M., Gash, J. H., Holmes, T. R. H., & Dolman, A. J. (2011). Magnitude and variability of land evaporation and its components at the global scale. Hydrology and Earth System Sciences, 15(3), 967–981. https://doi.org/10.5194/hess-15-967-2011

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE, 50(3), 885–900. https://doi.org/10.13031/2013.23153

Morriss, M. D. (1991). Factorial sampling plans for preliminary computational experiments. Technometrics, 33(2), 161–174. https://doi.org/10.1080/00401706.1991.10484804

Murray-Tortarolo, G., Anav, A., Friedlingstein, P., Sitch, S., Piao, S., Zhu, Z., et al. (2013). Evaluation of land surface models in reproducing satellite-derived LAI over the high-latitude Northern Hemisphere Part I: Uncoupled DGVMs. Remote Sensing, 5(10), 4819–4838. https://doi.org/10.3390/rs5104819

Naeini, M. R., Analui, B., Gupta, H. V., Duan, Q., & Sorooshian, S. (2019). Three decades of the shuffled complex evolution (sce-UA) optimization algorithm: Review and applications. Scientia Iranica, 26(4A), 2015–2031. https://doi.org/10.24200/sci.2019.21500

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. Journal of Hydrology, 10(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Niu, G.-Y., & Yang, Z.-L. (2008). Effects of vegetation canopy processes on snow surface energy and mass balances. Journal of Geophysical Research: Atmospheres, 109, D23111. https://doi.org/10.1029/2004JD004854

Niu, G.-Y., & Yang, Z.-L. (2006). Effects of frozen soil on snowmelt runoff and soil water storage at a continental scale. Journal of Hydro meteorology, 7(5), 937–952. https://doi.org/10.1175/JHM538.1

Niu, G.-Y., Yang, Z. L., Dickinson, R. E., Gulden, L. E., & Su, H. (2007). Development of a simple groundwater model for use in climate models and evaluation with gravity recovery and climate experiment data. Journal of Geophysical Research: Atmospheres, 112, D07103. https://doi.org/10.1029/2006JD007522

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

Oleson, K., Dai, Y., Bonan, G. B., Bosilovich, M., Dickinson, R., Dirmeyer, P., et al. (2004). Technical description of the community land model (CLM)

Prentice, I. C., Bondeau, A., Cramer, W., Harrison, S. P., Hickler, T., Lucht, W., et al. (2007). Dynamic global vegetation modeling: Quantifying terrestrial ecosystem responses to large-scale environmental change. In Terrestrial ecosystems in a changing world (pp. 175–192). Berlin: Springer.

Prentice, I. C., Liang, X., Medlyn, B. E., & Wang, Y. P. (2015). Reliable, robust and realistic: The three Rs of next-generation land-surface modelling. Atmospheric Chemistry and Physics, 15(10), 5987–6005. https://doi.org/10.5194/acp-15-5987-2015

Prihodko, L., Denning, A. S., Hanan, N. P., Baker, I., & Davis, K. (2008). Sensitivity, uncertainty and time dependence of parameters in a complex land surface model. Agricultural and Forest Meteorology, 148(2), 268–287. https://doi.org/10.1016/j.agrformet.2007.08.006

Purves, D., & Pacala, S. (2008). Predictive models of forest dynamics. Science, 320(5882), 1452–1453. https://doi.org/10.1126/science.1155359

Randerson, J. T., Hoffman, F. M., Thornton, P. E., Mahowald, N. M., Lindsay, K., Lee, Y. H., et al. (2009). Systematic assessment of terrestrial biogeochemistry in coupled climate-carbon models. Global Change Biology, 15(10), 2462–2484. https://doi.org/10.1111/j.1365-2486.2009.01912.x

Razavi, S., & Gupta, H. V. (2015). What do we mean by sensitivity analysis? The need for comprehensive characterization of “global” sensitivity in earth and environmental systems models. Water Resources Research, 51, 3070–3092. https://doi.org/10.1002/2014WR016527

Razavi, S., & Gupta, H. V. (2016a). A new framework for comprehensive, robust, and efficient global sensitivity analysis: 1 Theory. Water Resources Research, 52, 423–439. https://doi.org/10.1002/2015WR017558

Razavi, S., & Gupta, H. V. (2016b). A new framework for comprehensive, robust, and efficient global sensitivity analysis: 2 Application. Water Resources Research, 52, 440–455. https://doi.org/10.1002/2015WR017559

Razavi, S., & Gupta, H. V. (2019). A multi-method generalized global sensitivity matrix approach to accounting for the dynamical nature of earth and environmental systems models. Environmental Modelling & Software, 114, 1–11. https://doi.org/10.1016/j.enssoft.2018.12.002

Richardson, A. D., Anderson, R. S., Arain, M. A., Barr, A. G., Bohrer, G., Chen, G., et al. (2012). Terrestrial biosphere models need better representation of vegetation phenology: Results from the North American carbon program site synthesis. Global Change Biology, 18(2), 566–584. https://doi.org/10.1111/j.1365-2486.2011.02562.x

Rosolem, R., Gupta, H. V., Shuttleworth, W. J., Zeng, X., & Goncalves de Goncalves, L. G. (2012). A fully multiple-criteria implementation of the Sobol' method for parameter sensitivity analysis. Journal of Geophysical Research, 117, D07103. https://doi.org/10.1029/2011JD016355
Yang, Z.-L., Niu, G. Y., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., et al. (2011). The community Noah land surface model with multiparameterization options (Noah MP): 2. Evaluation over global river basins. Journal of Geophysical Research, 116, D12110. https://doi.org/10.1029/2010JD015140

Zaehle, S., Medlyn, B. E., De Kauwe, M. G., Walker, A. P., Dietze, M. C., Hickler, T., et al. (2014). Evaluation of 11 terrestrial carbon-nitrogen cycle models against observations from two temperate Free-air CO2 enrichment studies. New Phytologist, 202(3), 803–822. https://doi.org/10.1111/nph.12697

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

Zhang, X., Zhang, S., Liu, Z., Wu, X., & Han, G. (2015). Parameter optimization in an intermediate coupled climate model with biased physics. Journal of Climate, 28(3), 1227–1247. https://doi.org/10.1175/JCLI-D-14-00348.1

Zheng, H., Yang, Z.-L., Lin, P., Wei, J., Wu, W.-Y., Li, L., et al. (2019). 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

Zhao, L., Dai, Q., Han, D., Chen, N., & Zhao, B. (2019). Assessment of simulated soil moisture from WRF Noah, Noah-MP, and CLM land surface schemes for landslide hazard application. Hydrology and Earth System Sciences, 23(10), 4199–4218. https://doi.org/10.5194/hess-23-4199-2019