AGU PUBLICATIONS

Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE

10.1002/2014MS000406

Key Points:
- Presented a multiobjective two-stage sensitivity analysis approach, which combines the merits of qualitative and quantitative global sensitivity analysis methods
- Described the mechanisms of the terrestrial hydrological processes of a Conjunctive Surface-Subsurface Process (CSSP) land surface model
- Identified dominant parameters and reduced parameter dimensionality for parameter estimation

Supporting Information:
- Supporting Information S1

Correspondence to:
Q. Duan,
qyduan@bnu.edu.cn

Citation:
Gan, Y., X.-Z. Liang, Q. Duan, H. I. Choi, Y. Dai, and H. Wu (2015), Stepwise sensitivity analysis from qualitative to quantitative: Application to the terrestrial hydrological modeling of a Conjunctive Surface-Subsurface Process (CSSP) land surface model, *J. Adv. Model. Earth Syst.*, *7*, 648–669, doi:10.1002/2014MS000406.

Received 14 NOV 2014
Accepted 13 APR 2015
Accepted article online 16 APR 2015
Published online 8 MAY 2015
Corrected 26 JUN 2015

This article was corrected on 26 JUN 2015. See the end of the full text for details.

© 2015. The Authors.
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Stepwise sensitivity analysis from qualitative to quantitative: Application to the terrestrial hydrological modeling of a Conjunctive Surface-Subsurface Process (CSSP) land surface model

Yanjun Gan[1,2,3], Xin-Zhong Liang[2,4], Qingyun Duan[3], Hyun Il Choi[5], Yongjiu Dai[3], and Huan Wu[2]

[1]State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing, China, [2]Earth System Science Interdisciplinary Center, University of Maryland, College Park, Maryland, USA, [3]Joint Center for Global Change Studies, College of Global Change and Earth System Science, Beijing Normal University, Beijing, China, [4]Department of Atmospheric and Oceanic Science, University of Maryland, College Park, Maryland, USA, [5]Department of Civil Engineering, Yeungnam University, Gyeongsan, South Korea

Abstract An uncertainty quantification framework was employed to examine the sensitivities of 24 model parameters from a newly developed Conjunctive Surface-Subsurface Process (CSSP) land surface model (LSM). The sensitivity analysis (SA) was performed over 18 representative watersheds in the contiguous United States to examine the influence of model parameters in the simulation of terrestrial hydrological processes. Two normalized metrics, relative bias (RB) and Nash-Sutcliffe efficiency (NSE), were adopted to assess the fit between simulated and observed streamflow discharge (SD) and evapotranspiration (ET) for a 14 year period. SA was conducted using a multiobjective two-stage approach, in which the first stage was a qualitative SA using the Latin Hypercube-based One-At-a-Time (LH-OAT) screening, and the second stage was a quantitative SA using the Multivariate Adaptive Regression Splines (MARS)-based Sobol' sensitivity indices. This approach combines the merits of qualitative and quantitative global SA methods, and is effective and efficient for understanding and simplifying large, complex system models. Ten of the 24 parameters were identified as important across different watersheds. The contribution of each parameter to the total response variance was then quantified by Sobol' sensitivity indices. Generally, parameter interactions contribute the most to the response variance of the CSSP, and only 5 out of 24 parameters dominate model behavior. Four photosynthetic and respiratory parameters are shown to be influential to ET, whereas reference depth for saturated hydraulic conductivity is the most influential parameter for SD in most watersheds. Parameter sensitivity patterns mainly depend on hydroclimatic regime, as well as vegetation type and soil texture.

1. Introduction

Land surface models (LSMs) have become indispensable, not only to general circulation models (GCMs) or regional climate models (RCMs), but also to numerical weather prediction (NWP) models, since they represent the land-atmosphere exchanges of momentum, water, energy and other constituents of the earth system cycle [*Sellers et al.*, 1997; *Pitman*, 2003; *Dickinson et al.*, 2006]. Over the past few decades, LSMs have evolved from simple parameterizations representing the surface water and energy balances [*Budyko*, 1956; *Manabe*, 1969] to complex models that represent more physical processes and include multiple soil layers [e.g., *Sellers et al.*, 1986; *Dickinson et al.*, 1986; *Liang et al.*, 1994; *Chen and Dudhia*, 2001; *Dai et al.*, 2003]. The Project for Intercomparison of Land-surface Parameterization Schemes (PILPS) [*Henderson-Sellers et al.*, 1993, 1995] has shown that different LSMs produce considerably different results in water and energy budgets, even when they are forced by identical data [*Shao and Henderson-Sellers*, 1996; *Pitman and Henderson-Sellers*, 1998; *Pitman et al.*, 1999]. The uncertainties associated with LSMs mainly come from model structure and parameters, as errors in forcing and validation data should be similar among all models [*Xia et al.*, 2005]. Inaccurate parameter values may lead to poor LSM performances, even if the model structures are perfect. Thus, the proper specification of model parameters is not a trivial issue [*Sorooshian and Gupta*, 1983; *Duan et al.*, 1992; *Tolson and Shoemaker*, 2007].

A common means of parameter estimation is to calibrate models to historical observations by tuning the model parameters [*Gupta and Sorooshian*, 1985; *Duan et al*., 2006]. However, the calibration of LSMs faces several complications: (1) LSMs become increasingly complex as they include more and more sub-physics, despite being calibrated with rather limited observations, which leads to overparameterized LSMs [*Jakeman and Hornberger*, 1993; *Renard et al*., 2010]; (2) overparameterization, along with parameter nonlinear effects, causes model parameters to be not uniquely identifiable, a condition known as equifinality by *Beven* [1993, 2006]; and (3) some parameters may be spatially and temporally variable due to the heterogeneous effects of land surfaces [*Giorgi and Avissar*, 1997; *Prihodko et al*., 2008].

To calibrate LSM parameters effectively and efficiently, we must identify the parameters that dominate model behavior, thus reducing parameter dimensionality while maintaining the reliability of model simulation. For such a purpose, sensitivity analysis (SA) has been widely used to understand and simplify complex system models [*Hornberger and Spear*, 1981; *Morris*, 1991; *Saltelli et al*., 2004; *Bastidas et al*., 2006; *Rosero et al*., 2010; *Li et al.,* 2013, *Gan et al*., 2014]. Generally, SA methods can be divided into local and global methods. Unlike the former, the latter explore the whole parameter space, allowing them to provide robust measures in the presence of nonlinearity and interactions among the parameters [*Wainwright et al*., 2014], and thus are generally preferred [*Saltelli et al*., 2008]. Global SA methods can be further classified into qualitative and quantitative methods. Qualitative SA methods aim to screen out a subset of noninfluential parameters using a small number of model evaluations, whereas quantitative SA methods aim to measure each parameter's contribution to the response variance, a process that requires a large number of model evaluations [*Campolongo et al*., 2011]. Since today's LSMs are becoming more detailed and realistic, and thus have many parameters, high computational costs can be prohibitive to use quantitative SA. This has prompted many researchers to investigate the possibility of replacing the original simulation models with computationally cheaper surrogate models that perform a similar function [*Shahsavani and Grimvall*, 2011; *Stephens et al*., 2011; *Borgonovo et al*., 2012]. Approximation or meta-modeling techniques, such as Multivariate Adaptive Regression Splines (MARS), Support Vector Machines (SVMs), Artificial Neural Networks (ANNs), and Gaussian Process Regression (GPR) models, are often used to construct surrogate models to improve overall computational efficiency [*Wang and Shan*, 2007; *Storlie et al*., 2009; *Razavi et al*., 2012]. The strengths and limitations of several qualitative and quantitative SA methods, as well as the choice of appropriate SA methods for complex system models, have been discussed by *Gan et al*. [2014].

Previous sensitivity analyses of LSMs have used either qualitative or quantitative global methods, or have used both of them simultaneously. For example, *Bastidas et al*. [1999] introduced a robust multiobjective generalized sensitivity analysis (MOGSA) method to evaluate parameter sensitivities of the Biosphere-Atmosphere Transfer Scheme (BATS) model [*Dickinson et al*., 1993]. *Rosero et al*. [2010] employed Sobol' total and first-order sensitivity indices [*Sobol'*, 1993, 2001] to compare the performance and physical realism of three versions of the Noah LSM [*Ek et al*., 2003]. *Hou et al*. [2012] investigated the sensitivity of simulated surface fluxes to selected hydrologic parameters in the Community Land Model (CLM4) [*Lawrence et al*., 2011] using multivariate generalized linear model analyses and statistical significance tests. By adopting the same methods, *Huang et al*. [2013] studied the sensitivity of various simulated fluxes and their derived variables to selected hydrologic parameters in CLM4. *Rosolem et al*. [2012] presented a rank-based multicriteria implementation of the Sobol' method and applied it to the problem of identifying the most sensitive parameters in the SiB3 model [*Sellers et al*., 1986; *Baker et al*., 2008]. *Li et al*. [2013] conducted parameter screening for six output fluxes from the Common Land Model (CoLM) [*Dai et al*., 2003] using five qualitative SA methods, employing a quantitative Sobol' method to confirm the screening results. However, few of the previous works have focused on the issue of combining the merits of both qualitative and quantitative methods for stepwise SA of high-dimensional LSMs.

The purpose of this study is to explore the parameter sensitivities of a large, complex LSM, moving seamlessly from qualitative to quantitative SA methods using an uncertainty quantification framework. A multiobjective two-stage SA approach was applied to assess parameter sensitivities of a newly developed Conjunctive Surface-Subsurface Process (CSSP) LSM [*Choi and Liang*, 2010; *Choi et al*., 2013] for terrestrial hydrological modeling over 18 representative watersheds with varying hydroclimatic regimes. We used two SA methods: Latin Hypercube-based One-At-a-Time (LH-OAT) screening [*Morris*, 1991; *van Griensven et al*., 2006] for qualitative processes and MARS-based Sobol' sensitivity indices for quantitative processes. Both methods are capable of measuring the overall effects of parameters and their interactions, and have been

applied successfully in other environmental modeling cases [*Holvoet et al*., 2005; *van Griensven et al*., 2006; *Tang et al*., 2007; *Gan et al*., 2014]. The LH-OAT method was used to screen out insensitive parameters, and was followed by the MARS-based Sobol' sensitivity indices, which were used to quantify each parameter's contribution to the response variance. By evaluating model parameter sensitivities stepwise, we aimed to efficiently understand model behavior and improve its ability to perform hydrologic simulations across watersheds.

This paper is organized as follows: section 2 gives a description of the CSSP and the experimental data; section 3 presents a framework for uncertainty quantification; results are described and discussed in section 4; and finally, conclusions are summarized in section 5.

2. Model and Data

2.1. Brief Description of the CSSP

The CSSP has been developed as a core LSM of the mesoscale Climate-Weather Research and Forecasting (CWRF) model [*Liang et al*., 2012] to predict land-atmosphere water and heat fluxes exchanges. It is rooted in the CoLM [*Dai et al*., 2003, 2004] and includes a series of crucial improvements and updates: (1) an improved land-surface albedo parameterization [*Liang et al*., 2005a]; (2) a scalable parameterization of subgrid topography for soil moisture transportation [*Choi et al*., 2007]; (3) an improved parameterization for surface-subsurface flow interaction [*Choi and Liang*, 2010]; and (4) a parameterization for adding an unconfined aquifer below bedrock [*Yuan and Liang*, 2011]. Although *Yuan and Liang* [2011] demonstrated that CSSP performance is overall better than both CoLM and CLM3.5 at regional-local scales over the contiguous United States (CONUS), we found that CSSP still has relatively large errors in terrestrial hydrological modeling.

In this study, we replaced the original canopy photosynthesis-conductance parameterization [*Dai et al*., 2004] with a newly developed scheme proposed by *Bonan et al*. [2011]. A newly developed conjunctive surface-subsurface flow scheme [*Choi et al*., 2013] was also incorporated into the CSSP for the computation of streamflow discharge. We implemented the CSSP over the CONUS domain at the 30 km resolution following *Liang et al*. [2012]. Specific streamflow discharge (SD) and evapotranspiration (ET) are two target variables analyzed in this study, the parameterizations of which are briefly described in Appendix A.

2.2. Experimental Data

2.2.1. Stations, Watersheds, and Evaluation Data Sets

We selected 18 representative watersheds across the CONUS, each of these watersheds containing one U.S. Geological Survey (USGS) streamflow gauging station. In selecting these watersheds, we taking into consideration the minimal effects of upstream anthropogenic activities such as irrigation diversion and reservoir operation. Figure 1 shows the geographic locations of these stations and the corresponding watersheds, which span different hydroclimatic, soil, and vegetation conditions over the research domain. Also shown in Figure 1 are the runoff coefficients (RCs) of 18 watersheds, where RC is the dimensionless ratio of multiyear mean monthly SD to precipitation (P). These coefficients are indicators of relative watershed wetness, with larger RC correspond to wetter watershed [*Merz et al*., 2006]. Following *Li et al*. [2010], the watersheds are therefore referred to as humid ($0.6 < \mathrm{RC} \leq 1$), semihumid ($0.4 < \mathrm{RC} \leq 0.6$), semiarid ($0.2 < \mathrm{RC} \leq 0.4$), and arid ($0 < \mathrm{RC} \leq 0.2$).

Observed monthly streamflow discharge data from 1982 to 1995 were acquired directly from the USGS website (available at http://waterdata.usgs.gov/nwis/sw) for the 18 gauging stations. Additionally, the global $0.5\mathrm{deg} \times 0.5\mathrm{deg}$ observation-based monthly evapotranspiration data sets [*Jung et al*., 2010] for the same period were remapped onto the 30 km resolution research domain by bilinear interpolation. This gridded data set was derived from FLUXNET [*Baldocchi et al*., 2001], meteorological, and remote sensing observations using model tree ensemble approach, and has recently been used to evaluate and improve canopy and hydrological processes in LSMs [*Bonan et al*., 2011; *Luo et al*., 2012; *Lawrence et al*., 2012], among other applications for the evaluation of climate change projections [*Jung et al*., 2011; *Mueller et al*., 2011; *Zeng et al*., 2012]. Station and watershed characteristics, as well as multiyear mean monthly P, ET, and SD for the observation period are summarized in Table 1, where mean P was calculated based on the forcing data while mean ET and SD were calculated based on the observations.

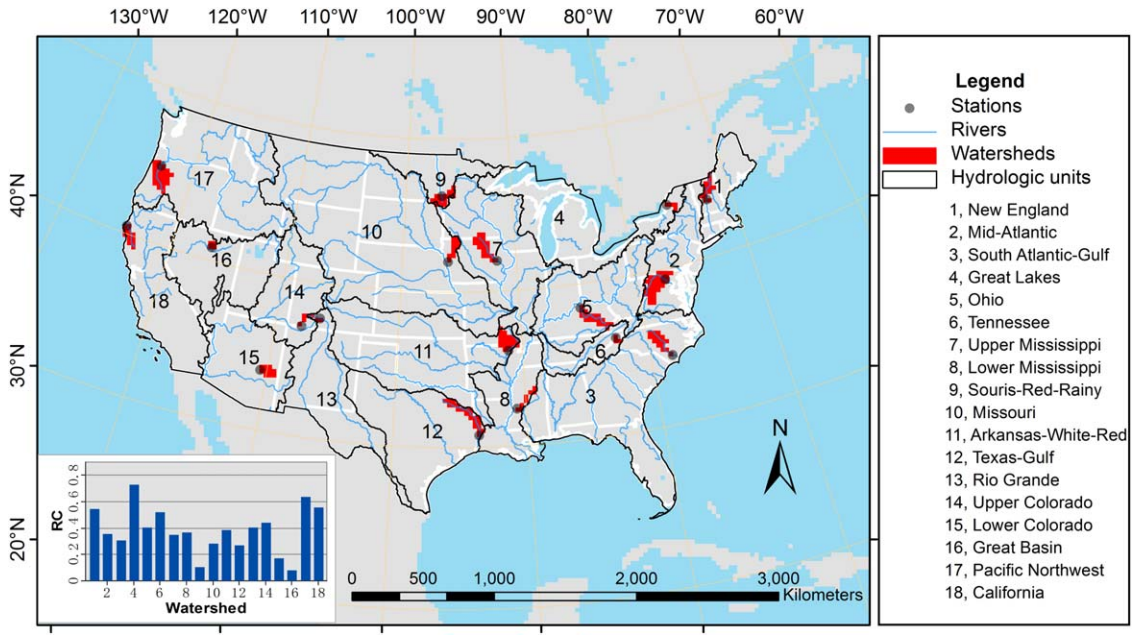

Figure 1. Geographic locations of selected USGS streamflow gauging stations and corresponding watersheds over 18 hydrologic units. Subfigure in the lower left corner shows the runoff coefficients of all 18 watersheds.

2.2.2. Surface Boundary Conditions

The surface boundary conditions incorporated in the CSSP are based on the best observational data over North America. They have been remapped onto the 30 km resolution research domain either by bilinear interpolation, in cases where the raw data resolution is coarser than 30 km, or by using a mass conservative approach which weights the fractional area of each pixel contributing to the grid [*Liang et al.*, 2005b, 2005c]. A detailed description of the surface boundary conditions is given in Table 2.

2.2.3. Initial and Forcing Conditions

Required initial and forcing fields are listed in Table 3. All fields date from 1 January 1979 to 31 December 1995 were obtained directly from the North America Regional Reanalysis (NARR) 32 km 3 h data [*Mesinger*

Table 1. Characteristics of Selected USGS Streamflow Gauging Stations and Corresponding Watersheds

Watershed	Station ID	Lat (°N)	Lon (°W)	Area (km^2)	P (mm/month)	ET (mm/month)	SD (mm/month)	Soil Texture	Vegetation Type
1	01144500	43.65	72.31	10,598	89.1	39.9	48.6	Loam	84.6% deciduous broadleaf forest; 15.4% mixed forest
2	01638500	39.27	77.54	24,996	81.6	52.5	28.8	Silt loam	70% deciduous broadleaf forest; 20% cropland/woodland mosaic
3	02105769	34.40	78.29	13,610	96.9	62.1	29.7	Sandy loam	56.3% cropland/woodland mosaic; 18.8% mixed forest
4	04260500	43.99	75.92	4,828	94.2	43.5	68.4	Loam	75% deciduous broadleaf forest; 25% cropland/grassland mosaic
5	03290500	38.44	84.96	16,006	94.5	56.4	38.1	Silt loam	61.1 deciduous broadleaf forest; 22.2% cropland/woodland mosaic
6	03465500	36.18	82.46	2,085	97.8	57.3	50.7	Loam	66.7% deciduous broadleaf forest; 33.3% dryland cropland and pasture
7	05464500	41.97	91.67	16,861	72.0	45.3	25.2	Loam	100% dryland cropland and pasture
8	07290000	32.35	90.70	7,283	118.2	65.1	43.5	Silt loam	62.5% cropland/woodland mosaic; 37.5% mixed forest
9	05051500	46.27	96.60	10,386	49.2	40.5	5.1	Loam	86.7% dryland cropland and pasture; 13.3% cropland/grassland mosaic
10	06607500	41.96	95.97	9,132	67.2	45.3	18.9	Silt loam	100% dryland cropland and pasture
11	07072500	36.10	91.10	19,088	100.2	57.3	38.4	Silt loam	47.4% deciduous broadleaf forest; 31.6% cropland/woodland mosaic
12	08028500	30.75	93.61	21,313	106.2	74.1	28.5	Loam	45.5% mixed forest; 22.7% savanna; 22.7% grassland
13	08220000	37.69	106.46	3,419	51.3	31.5	20.7	Loam	75% evergreen needleleaf forest; 25% grassland
14	09363500	37.04	107.88	2,823	60.9	32.7	27.0	Loam	66.7% evergreen needleleaf forest; 33.3% grassland
15	09497500	33.80	110.50	7,379	53.4	32.7	9.0	Loam	80% mixed shrubland/grassland; 20% evergreen needleleaf forest
16	10324500	40.83	116.59	2,235	22.5	21.9	1.8	Loam	100% grassland
17	14191000	44.94	123.04	18,855	134.1	45.6	85.5	Loam	72% evergreen needleleaf forest; 28% dryland cropland and pasture
18	11477000	40.49	124.10	8,063	120.0	52.2	66.9	Loam	91.7 % evergreen needleleaf forest

Table 2. Surface Boundary Conditions

Name	Units	Level	Time	Source
Land cover category		1	Static	Convert MODIS IGBP land cover classification (MCD12Q1) to USGS 24-category classification following *Liang et al.* [2005b]
Surface characteristic identification		1	Static	Derived from land cover category
Fractional vegetation cover		1	Static	Derived from MODIS vegetation indices (MOD13A1) following *Zeng et al.* [2000] and *Liang et al.* [2005b]
Leaf area index	$m^2\ m^{-2}$	1	Monthly	Derived from MODIS LAI (MCD15A2) following *Zeng et al.* [2002]
Stem area index	$m^2\ m^{-2}$	1	Monthly	Derived from MODIS LAI (MCD15A2) following *Zeng et al.* [2002]
Sink filled DEM	m	1	Static	Derived from SRTM30 [*Farr et al.*, 2007]
Flow direction		1	Static	Derived from sink filled DEM and corrected by DRT algorithm [*Wu et al.*, 2011]
Flow accumulation		1	Static	Derived from flow direction
X direction surface slope std. dev.	m/m	1	Static	Derived from sink filled DEM
Y direction surface slope std. dev.	m/m	1	Static	Derived from sink filled DEM
Bedrock depth	m	1	Static	CONUS-SOIL [*Miller and White*, 1998] and HWSD [*FAO/IIASA/ISRIC/ISS-CAS/JRC*, 2012] data
Soil sand fraction profile	%	11	Static	CONUS-SOIL [*Miller and White*, 1998] and HWSD [*FAO/IIASA/ISRIC/ISS-CAS/JRC*, 2012] data
Soil clay fraction profile	%	11	Static	CONUS-SOIL [*Miller and White*, 1998] and HWSD [*FAO/IIASA/ISRIC/ISS-CAS/JRC*, 2012] data
Soil albedo localization factor		4	Static	Derived from MODIS albedo (MCD43C3) following *Liang et al.* [2005a]

et al., 2006], and remapped onto the 30 km resolution domain through bilinear interpolation. After spatially interpolated, all the above NARR forcing fields except convective precipitation and total precipitation were then temporally interpolated to the hourly time step using linear interpolation. The 3 h accumulative convective precipitation and total precipitation data were evenly divided into hourly segments. However, since downward short-wave radiation, convective precipitation, and total precipitation of NARR data have been reported to have significant biases [*Becker et al.*, 2009; *Mo et al.*, 2010; *Yuan and Liang*, 2011; *Kennedy et al.*, 2011], the three fields were then replaced for all possible domains with the North American Land Data Assimilation System project phase 2 (NLDAS-2) 1/8deg 1 h data [*Xia et al.*, 2012] over the same time period, which were also remapped onto the 30 km resolution domain using bilinear interpolation.

2.3. Model Initialization and Spin-Up

The control version of the CSSP, which uses default parameter values, was spun-up by cycling the 17 year (1 January 1979 to 31 December 1995) forcing data twice, until all state variables reached an equilibrium. The final state variables of the last run were taken as the initial conditions for different uncertainty quantification experiments. For each experiment, a continuous 17 year simulation was carried out over the research domain on an hourly time step. To reduce the influence of initial conditions, the outputs of the first 3 years were excluded for all experiments in the subsequent analysis.

Table 3. Initial and Forcing Fields

	Units	Source
Initial Fields		
Temperature at the surface (TSFC)	K	NARR
Temperature at 2m (T2m)	K	NARR
Snow depth (SNOD)	m	NARR
Snow cover (SNOWC)	%	NARR
Soil moisture (SM)	fraction	NARR
Soil temperature (ST)	K	NARR
Forcing Fields		
Pressure at the lowest atmospheric level (P)	Pa	NARR
Temperature at the lowest atmospheric level (T)	K	NARR
Specific humidity at the lowest atmospheric level (Q)	kg/kg	NARR
U-component of wind at the lowest atmospheric level (U)	m/s	NARR
V-component of wind at the lowest atmospheric level (V)	m/s	NARR
Geopotential height at the lowest atmospheric level (H)	gpm	NARR
Pressure at the surface (PSFC)	Pa	NARR
Temperature at 2m (T2m)	K	NARR
Planetary boundary layer height (HPBL)	m	NARR
Downward long-wave radiation fluxes (DLWRF)	W/m^2	NARR
Downward short-wave radiation fluxes (DSWRF)	W/m^2	NLDAS-2
Convective precipitation (ACPCP)	mm	NLDAS-2
Total precipitation (APCP)	mm	NLDAS-2

3. Uncertainty Quantification Framework

Figure 2 depicts the uncertainty quantification framework implemented in this study. The methods we used are briefly described in this section. Further details can be found in Appendices B, C, and D.

3.1. Problem Specification

3.1.1. Parameter Ranges and Distributions

To quantify model parameter uncertainties for terrestrial hydrological modeling, we selected from the CSSP 24 related tunable parameters, whose names, physical meanings, and uncertainty ranges are presented in Table 4. Feasible ranges of these parameters were determined according to their physical meanings and previous literature [*Johnson*, 1967; *Judson and Doesken*, 2000; *Lefebre et al.*, 2003; *Kumar*, 2004; *Niu and Yang*, 2006; *Oleson et al.*, 2010; *Li et al.*, 2013]. Other model parameters were kept constant at the recommended values. All parameters considered have uniform prior probability distributions.

3.1.2. Evaluation Metrics

Studies have shown that sensitive parameters may vary between different evaluation metrics [*Tang et al.*, 2007; *van Werkhoven et al.*, 2008]. Therefore, SA or calibration of the model parameters based on a single metric is often inadequate [*Yapo et al.*, 1998; *Madsen*, 2000; *Vrugt et al.*, 2003; *Wagener et al.*, 2009], even in the case of a model simulating a single output flux [*Gupta et al.*, 1998]. Thus, model uncertainty quantification should simultaneously consider multiple relatively unrelated statistics that measure different characteristics of model simulation capability [*Bastidas et al.*, 1999, 2006; *van Werkhoven et al.*, 2009; *Neelin et al.*, 2010]. To avoid using metrics with different units, we adopted two normalized statistics, relative bias (RB) and Nash-Sutcliffe efficiency (NSE) [*Nash and Sutcliffe*, 1970], for evaluating SD and ET. The former metric was bias normalized with respect to the mean of the observations, while the latter was root mean square error normalized with respect to the variance of the observations. *McCuen et al.* [2006] suggested that these metrics be reported together when assessing the goodness of fit. The definitions of RB and NSE are as follows:

$$RB=\frac{\frac{1}{n}\sum_{i=1}^{n}(S_i-O_i)}{\overline{O}} \tag{1}$$

$$NSE=1-\frac{\sum_{i=1}^{n}(O_i-S_i)^2}{\sum_{i=1}^{n}(O_i-\overline{O})^2} \tag{2}$$

where S_i and O_i are simulated and observed values at time i, respectively; $\overline{O}$ is the mean of observations; and n is the total number of observations (or simulations).

Each metric for each output variable can be taken as an objective function, providing the four objective functions: RB of SD, NSE of SD, RB of ET, and NSE of ET. To solve the multiobjective problem, different objective functions were transformed into a Euclidian distance function following *Madsen* [2000]:

$$F_{agg}(\mathbf{X})=\sqrt{\sum_{j=1}^{m}\left[F_j(\mathbf{X})+A_j\right]^2} \tag{3}$$

where $F_j(\mathbf{X})$ is the jth of the m objective functions. Specifically, $F_1(\mathbf{X})=|RB_{SD}|$ and $F_2(\mathbf{X})=1-NSE_{SD}$ are objective functions for SD, and $F_3(\mathbf{X})=|RB_{ET}|$ and $F_4(\mathbf{X})=1-NSE_{ET}$ are objective functions for ET. A_j is the transformation constant reflecting prior information of the jth objective function as:

$$A_j=\max_{k=1,2,\cdots,m}\{F_{k,\min}\}-F_{j,\min}, \quad j=1,2,\cdots,m \tag{4}$$

where $F_{j,\min}$ (or $F_{k,\min}$) is the minimum value of $F_j(\mathbf{X})$ (or $F_k(\mathbf{X})$), which can be determined based on the initial experiments.

3.2. Qualitative Sensitivity Analysis

Qualitative SA is often employed in preliminary uncertainty quantification experiments to reduce parameter dimensionality by screening out parameters that have negligible influences on model responses from all parameters being investigated. Morris One-At-a-Time (MOAT) screening [*Morris*, 1991], which performs Monte Carlo (MC) sampling [*Meteopolis and Ulam*, 1949] followed by One-At-a-Time (OAT) sampling and

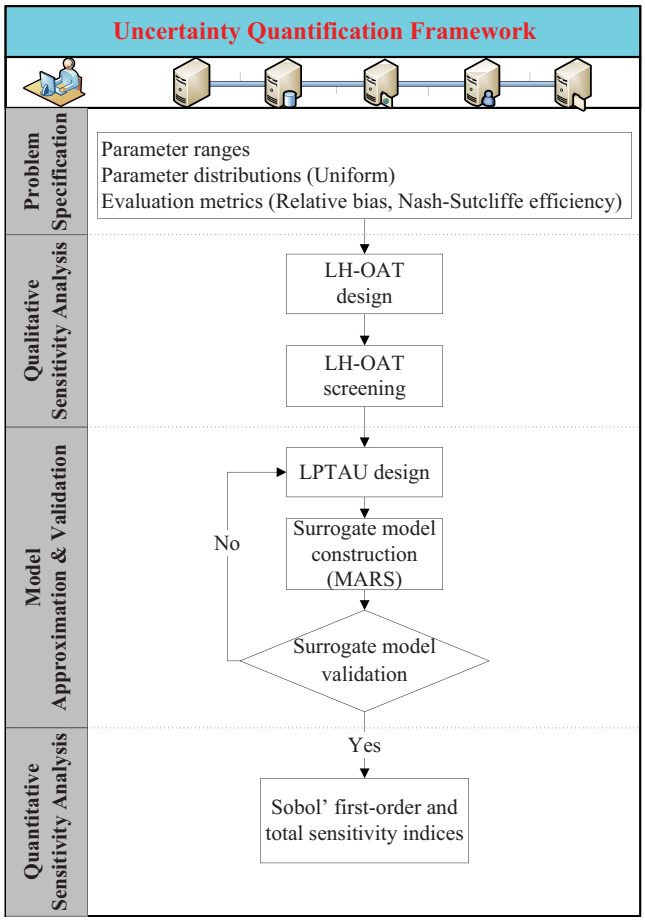

Figure 2. Uncertainty quantification framework.

evaluates parameter sensitivities according to elementary effects, has been successfully used for preliminary parameter screening [*Saltelli et al.*, 2004; *Confalonieri et al.*, 2010]. However, *Gan et al.* [2014] showed that the screening result of the MOAT method for multi-parameter complex system models may not be reliable when the number of parameter levels p is too small (e.g., $p = 4$ as used by *Morris* [1991] and *Campolongo et al.* [2007]). On the other hand, more sample points are needed to cover the whole parameter space when p is too large, because the initial MC sample points are not uniformly distributed.

Following *van Griensven et al.* [2006], we used Latin Hypercube (LH) sampling [*McKay et al.*, 1979] to replace MC sampling to generate initial sample points for the OAT screening method, since LH combines the merits of stratified and random sampling, which is known as LH-OAT screening. Briefly, p LH sample points were first generated from a n-dimension p-level orthogonal parameter space; then for each LH sample point, an OAT loop was performed by varying each of the n parameters one at a time. Therefore, the total number of

Table 4. Selected Parameters and Their Feasible Ranges

No.	Parameter	Description	Min.	Max.
P1	tcrit	Critical temperature to determine rain or snow (degC)	−2.0	4.0
P2	rous	Specific yield	0.02	0.30
P3	wtfact	Maximum saturated fraction	0.10	0.50
P4	trsmx0	Maximum transpiration rate (mm/s)	0.0001	0.01
P5	dewmx	Maximum dew allowed (mm)	0.05	0.30
P6	wssi	Irreducible water saturation of snow	0.02	0.10
P7	wimp	Water impermeable if porosity less than wimp	0.01	0.10
P8	vegmin	Minimum vegetation fraction to be counted	0.001	0.01
P9	pondmx	Ponding depth (mm)	5	30
P10	drainf	Drainage factor for bedrock flow	0.001	1.0
P11	zashc	Reference depth for saturated hydraulic conductivity (m)	0.1	3.0
P12	aniso	Anisotropic factor for lateral to vertical	1	2000
P13	csfcov	Coefficient for saturation fraction	0.2	0.8
P14	chrough	Manning's roughness coefficient	0.01	0.3
P15	z0m	Aerodynamic roughness length (m)	0.0002	2.4
P16	displa	Displacement height (zero plane distance) (m)	0.0013	15.6
P17	vmax25	Maximum carboxylation rate at 25 degC at canopy top (mol CO_2 m^{-2} s^{-1})	1.0E-05	1.2E-04
P18	effcon	Quantum efficiency of vegetation photosynthesis (mol CO_2 mol^{-1} photon)	0.03	0.1
P19	gradm	Stomatal slope factor	3	10
P20	binter	Minimum stomatal conductance (mol CO_2 m^{-2} s^{-1})	0.001	0.05
P21	d50	50% roots located above this depth (m)	0.05	0.28
P22	beta	Shape-parameter in the logistic dose-response curve for root profile	−2.6	−1.0
P23	smp_open	Soil water potential at stomata fully open for PFTs (mm)	−8.5E + 04	−3.0E + 04
P24	smp_close	Soil water potential at stomata fully close for PFTs (mm)	−5.0E + 05	−1.5E + 05

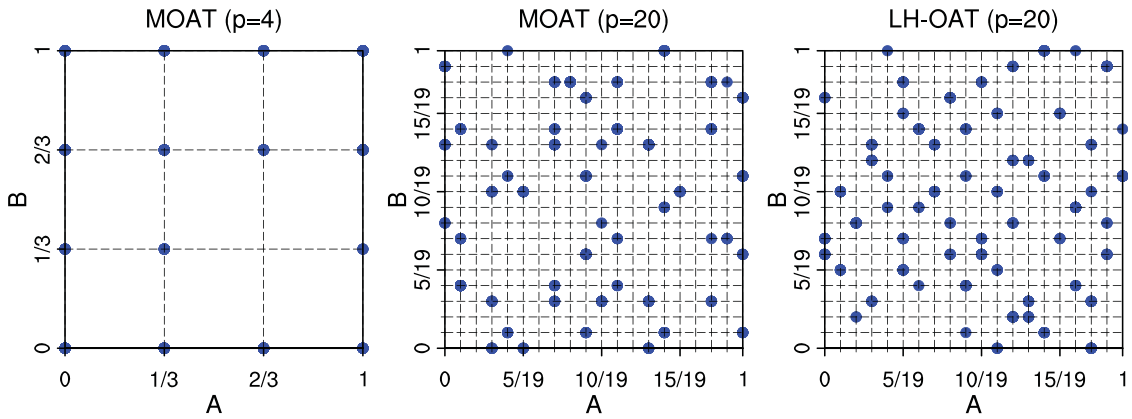

Figure 3. Projections of 500 sample points over two-dimensional parameter space, which are generated from (left) 4-level MOAT, (middle) 20-level MOAT, and (right) 20-level LH-OAT sampling. Each subfigure has some overlapping points.

sample points for p OAT loops is $(n+1)p$. When all the experiments were finished, we calculated p elementary effects for each parameter represented by partial derivatives. Overall and interaction effects of each parameter were then approximated using the mean and standard deviation of the p elementary effects.

Levy and Steinberg [2010] have suggested as a rough rule of thumb that at least $10 \times n$ sample points are needed to identify key parameters. *Gan et al.* [2014] have demonstrated that 20 OAT loops (i.e., replications) are sufficient to get a reliable result for MOAT screening, when the number of parameter levels p is set to between 16 and 32. Here for LH-OAT screening, we set p to 20, which is also the number of replications. Sample size for this 24-dimensional problem is then $(24+1) \times 20 = 500$, which is more than twice of the required least sample size 240 ($=24 \times 10$). Figure 3 compares the projections of 500 sample points over two (of twenty-four) parameters that were generated from 4-level MOAT, 20-level MOAT, and 20-level LH-OAT sampling, respectively. The advantage of LH-OAT sampling is that there is at least one sample point for each of the parameter levels, ensuring that the parameter space is better explored than in MOAT sampling.

3.3. Surrogate-Based Quantitative Sensitivity Analysis

Quantitative SA is often performed in an effort to apportion total response variance to the contribution of each parameter; it is well suited for accurate SA of low-dimensional problems due to its robustness and sophistication. The Sobol' sensitivity indices [*Sobol'*, 1993, 2001] perform ANOVA (i.e., Analysis of Variance)-like decomposition of model response variance, and measure the specific-order sensitivity index of one parameter by quantifying the attribution of response variance due to the variances of specific-order terms related to this parameter. One parameter's total sensitivity index can then be estimated by adding up all orders of sensitivity indices of this parameter. Theoretically, this method can compute sensitivity indices of any order. For a high-dimensional problem, however, only the first-order and second-order indices may be estimated practically, since the number of higher-order terms that must be computed is so large. *Homma and Saltelli* [1996] introduced an improved total sensitivity index, $S_{Ti}=S_i+S_{i,ci}=1-S_{ci}$, to solve the "curse of dimensionality" problem. Here S_i and $S_{i,ci}$ are representations of first-order and high-order effects of the ith parameter, respectively; S_{ci} is the sum of all other terms, excluding terms related to the ith parameter. In this study, we took Sobol' first-order (denoted as Sobol'-1) and total (denoted as Sobol'-t) sensitivity indices as quantitative SA measures.

A large number of model runs are required to produce reliable Sobol' sensitivity indices. Thus, computationally cheaper surrogate models are often adopted to approximate the simulation model for computing Sobol' sensitivity indices. We employed the MARS method [*Friedman*, 1991] to construct the surrogate model, as it has proven to be effective and efficient [*Storlie et al.*, 2009; *Chen et al.*, 2013; *Gan et al.*, 2014]. Quasi-random LP_τ(LPTAU) sequence sampling [*Sobol'*, 1967] was adopted to generate space-filling sample points, because it not only generates deterministic and uniformly distributed samples, but also allows the addition of more points with the same uniformity characteristics to the initial samples. Other space-filling sampling methods such as LH, Orthogonal Array (OA) [*Owen*, 1992], and OA-based LH (OALH) [*Tang*, 1993] could also be used to generate uniformly distributed sample points for constructing surrogate models.

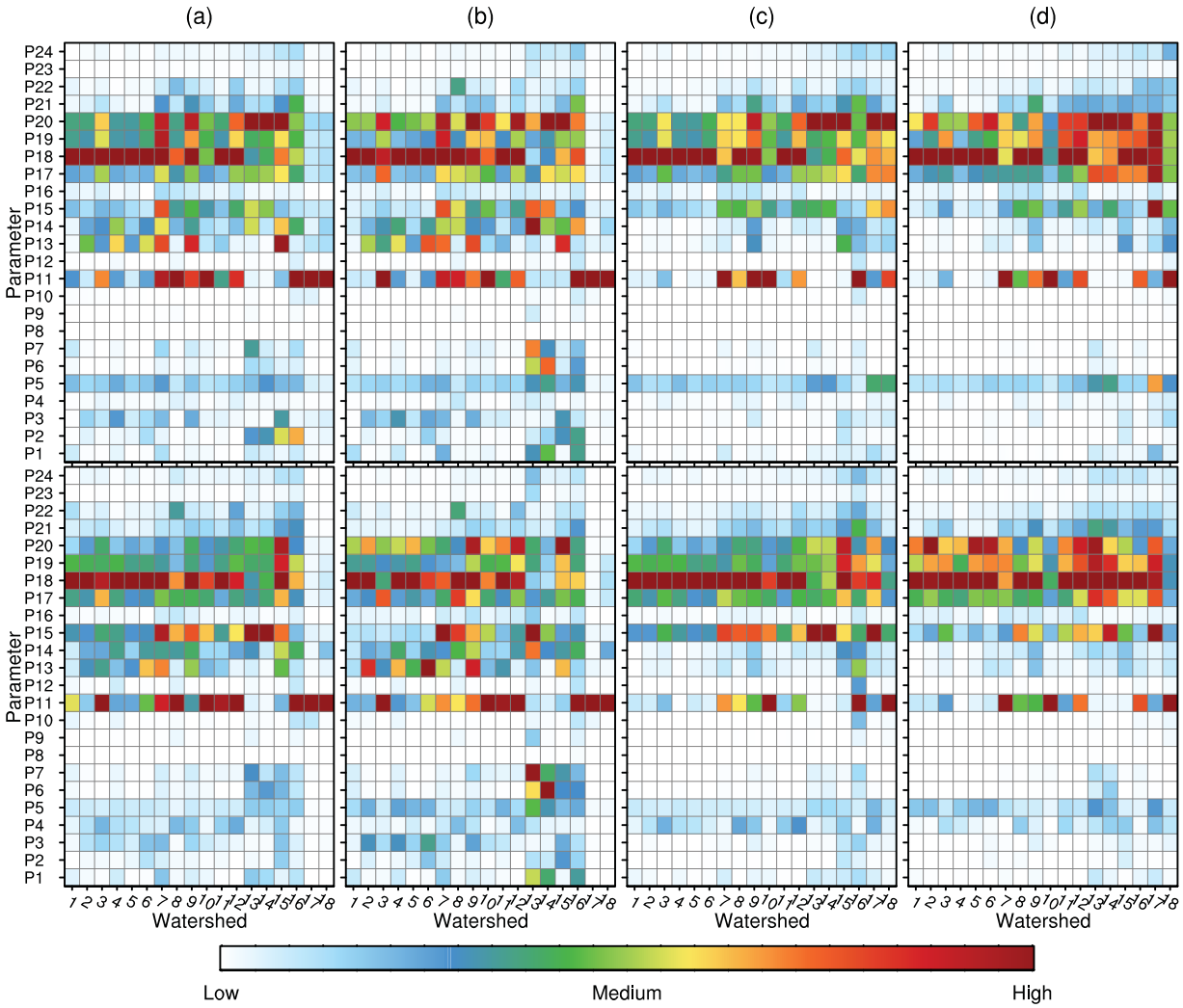

Figure 4. Relative parameter sensitivities to (top row) overall effects and (bottom row) interaction effects of 18 watersheds for (a) RB of SD, (b) NSE of SD, (c) RB of ET, and (d) NSE of ET.

Scaled root mean square error (SRMSE) between the simulation model responses and the surrogate model responses was estimated using the *k*-fold cross-validation method [*Picard and Cook*, 1984]. The surrogate model is applicable only when the final error of *k*-fold cross validation is less than a predefined value; if it is not, more simulation experiments must be added to achieve a reliable surrogate model.

4. Results and Discussion

4.1. LH-OAT Screening

The range for each of the 24 parameters was evenly divided into 20 levels. Thus, 500 ($=20 \times (24+1)$) LH-OAT experiments were designed to screen out the least sensitive parameters. Figure 4 plots relative parameter sensitivities to overall effects (upper row) and interaction effects (lower row) of the 18 watersheds for four objective functions (RB of SD, NSE of SD, RB of ET, and NSE of ET). Although parameter rankings vary among watersheds as well as among objective functions, some dominant patterns are notable in this figure. Overall, P6, P7, P11, P13, P14, P15, P17, P18, P19, and P20 are regarded as sensitive parameters, whereas the others are classified as insensitive ones. Among the above 10 sensitive parameters, P6, P7, P11, P13, and P14 are directly related to terrestrial hydrological processes, which mainly control runoff generation and routing; P15, the aerodynamic roughness length, influences the turbulent exchange between land surface and the atmosphere; and P17, P18, P19, and P20 are associated with leaf photosynthesis and stomatal

conductance processes, which impact canopy energy exchange, carbon exchange, and water vapor variation. Specifically, P18, the quantum efficiency of vegetation photosynthesis, is consistently sensitive across most watersheds for different objective functions.

Comparing parameter sensitivities across watersheds for SD (Figures 4a and 4b) and ET (Figures 4c and 4d), parameters P6, P7, P13, and P14 are found to be sensitive in some watersheds for SD, but are regarded as insensitive for ET, reflecting the fact that SD is sensitive to more parameter than ET. In addition, we can see that the patterns of overall or interaction effects are generally similar for RB and NSE of the same variable (e.g., top row or bottom row of Figures 4a and 4b). However, there are also differences, the most obvious of which is that P19 and/or P20 are identified as parameters with low interaction effects by most watersheds for RB of ET (lower row of Figure 4c), but are identified as those with high interaction effects by the same watersheds for NSE of ET (lower row of Figure 4d). Distinct differences are also detectable between RB of SD and NSE of SD for watersheds 13 and 14. All these observations suggest that parameter sensitivities vary to some extent with metrics as shown in *Tang et al*. [2007] and *van Werkhoven et al*. [2008].

Considering the same objective function, the number of sensitive parameters across watersheds varies from 1 (e.g., watershed 5) to 6 (e.g., watershed 7) for SD, and from 1 (e.g., watershed 4) to 5 (e.g., watershed 17) for ET, which reveals the combined influences of different hydroclimatic regimes [*Bastidas et al*., 2006; *Demaria et al*., 2007], as well as soil textures and vegetation types [*Hou et al*., 2012]. For example, in watersheds with the vegetation type of broadleaf forest (i.e., watersheds 1, 2, 4, 5, 6, and 11), the quantum efficiency of vegetation photosynthesis (P18) is identified as the most sensitive parameter, as it controls the canopy photosynthesis and hence the ET and SD. For vegetation types of cropland (watershed 7) and grassland (watershed 16), where land surfaces are sparsely vegetated and photosynthesis is weak, reference depths for saturated hydraulic conductivity (P11) become as or even more important than the quantum efficiency of vegetation photosynthesis (P18). The parameter sensitivity patterns for watersheds 3 and 8 are different even though they have similar vegetation types, which is a good illustration of the influences of hydroclimatic regime and soil texture. In fact, hydroclimatic regime plays a more important role than vegetation type and soil texture in the performance of LSM parameters [*van Werkhoven et al*., 2008; *Rosero et al*., 2010]. The sensitivity patterns of watersheds 13 and 14 are different for NSE of SD, even though they have the same vegetation types and soil textures and share the same hydroclimatic regimes. Other studies have also noted inconsistent parameter behavior between watersheds with the same physical properties [*Demaria et al*., 2007; *Rosero et al*., 2010]. The implication of our observations is that although the behaviors of LSM parameters mainly depend on hydroclimatic regime, vegetation type and soil texture, other uncontrollable factors such as nonlinear effect may also be influential. Transfer of these physically based model parameters between watersheds that share the same physical properties is sometimes infeasible, and observational data should be used to analyze more related physical processes in order to strengthen our understanding of parameter behavior before conducting the transfer as suggested by *Rosero et al*. [2010].

The importance of parameter interactions is further illustrated by the fact that parameters with low overall effects may have significant interaction effects, which should also be considered as sensitive parameters. For example, P15 is identified by watersheds 9, 10, 13, and 14 as a parameter with high interaction effects but low overall effects, as it is shown in Figure 4a.

4.2. Validation of Surrogate Model and Output Uncertainty

After qualitatively evaluating model parameter sensitivities and appropriately reducing parameter dimensionalities from 24 to 10, we designed 200, 400, 600, and 800 LPTAU experiments, and then used corresponding parameter sets and model responses (i.e., aggregate objective function values) to construct MARS models for surrogate-based quantitative SA. The prediction accuracies of the MARS models with different sample sizes were evaluated using fivefold, tenfold, fifteenfold, and twentyfold cross-validation schemes. Figure 5 compares the SRMSE values obtained by MARS with different sample sizes and cross-validation schemes. Theoretically, the performances of different sample sizes (or cross-validation schemes) reflect the tradeoff between prediction accuracy and time. As can be seen from Figure 5 that prediction accuracy increases with sample size, but shows no obvious trends for different cross-validation schemes because of the stochastic training of MARS with a limited number of iterations [*Zhang et al*., 2009]. Generally, SRMSE values are acceptable and similar for different cross-validation schemes when the sample size is equal to

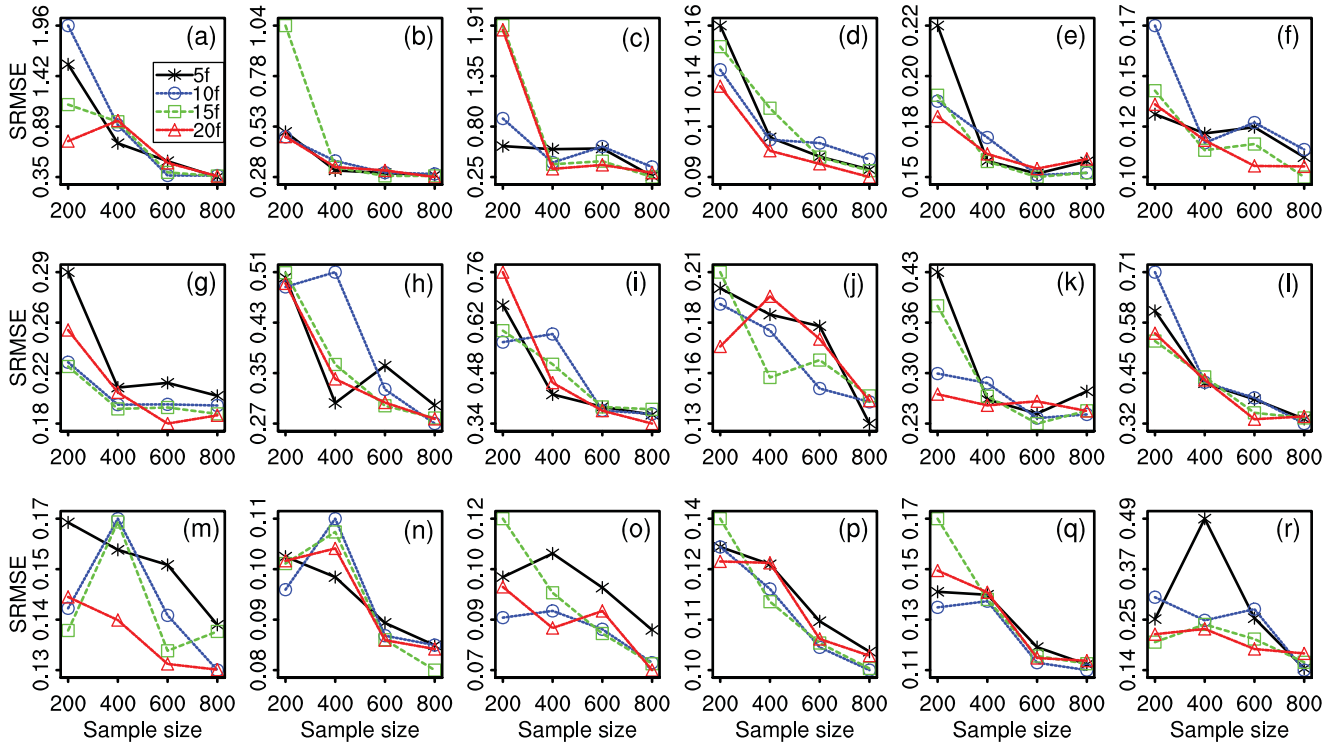

Figure 5. Comparison of scaled RMSE values obtained by MARS with different sample sizes (200, 400, 600, and 800) and cross-validation schemes (fivefold, tenfold, fifteenfold, and twentyfold). for watershed 1 to 18 (a–r), respectively.

800. Moreover, the performance of MARS varies across watersheds due to differences in their physical characteristics.

Before calculating the Sobol' sensitivity indices, we further analyzed the output uncertainty ranges of the 800 LPTAU experiments as suggested by *Huang et al*. [2013], to check whether the observations could be captured by the possible output ranges. Figure 6 shows the boxplots of SD and ET for the ensemble simulations across 18 watersheds. The output uncertainties are summarized by the minimum, first quartile (25%), median (50%), third quartile (75%), and the maximum of the ensemble simulations. The figure also presents the observations (denoted by red circles), and the outputs of the control simulation (represented by blue squares).

Compared to the observations, ET of the control simulation is generally underestimated in the summer across all watersheds except watershed 15, which overestimates ET for the whole year. The situation of SD in the control simulation is more complicated. When comparing its outputs with the observations, most watersheds overestimate SD for the whole year, whereas watersheds 4, 13, 14, and 17 underestimate SD for 3 or 4 months. These deviations between the observations and the outputs of the control simulation were the motivation behind our attempt to create a systematic SA and optimization of the model parameters.

As for the ensemble simulations, the observations overall lie within their output ranges with a few exceptions. For example, the observed SD is not well bounded in spring for watersheds 2, 5, 6, 16, 17, and 18, and is not well bounded in May or June for watersheds 13 and 14. We inferred that the physical bounds of the inputs are reasonable but the ensemble experiments are not enough to capture the whole output uncertainty ranges, because the observations fall within output bounds in June for watersheds 13 and 14 when the control simulation is included in the ensemble simulations. Note that this does not mean that 800 LPTAU experiments are not sufficient for quantitative SA or the construction of surrogate models, but rather reflects the strong interactions among the selected model parameters. Another possible reason for the observations outside of the output ranges is that the observed SD and/or ET are not highly accurate in some watersheds. Similar to other studies [*Hou et al*., 2012; *Huang et al*., 2013], the uncertainty ranges of ET are wider for the summer than for the cold seasons. This can be explained by the fact that summer ET is

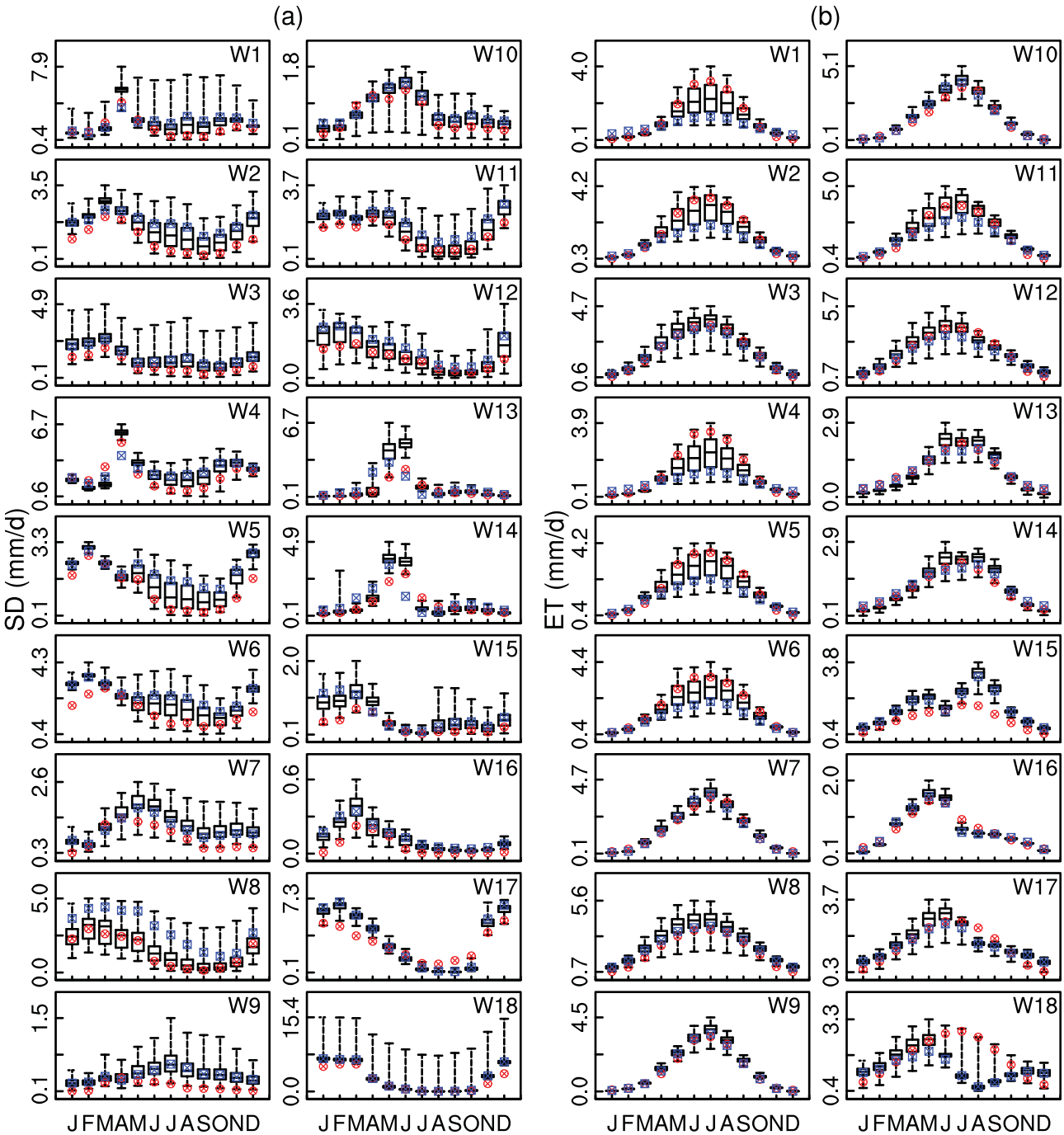

Figure 6. Monthly variations of (a) SD and (b) ET of 18 watersheds, due to parameter uncertainties. The red cycles are observations and blue squares are outputs of the control simulation.

much higher, and therefore their variations for the ensemble simulations are larger, which also implies that summer ET is more sensitive to selected parameters. In addition to addressing the poor performance of the control simulation for summer ET, as previously stated, it is certain that summer ET could be improved by parameter optimization.

4.3. MARS-Based Sobol' Sensitivity Analysis

The MARS model based on 800 sample points was used in lieu of the simulation model to predict output response, and 200,000 LPTAU sample points were then used to estimate Sobol' sensitivity indices for each

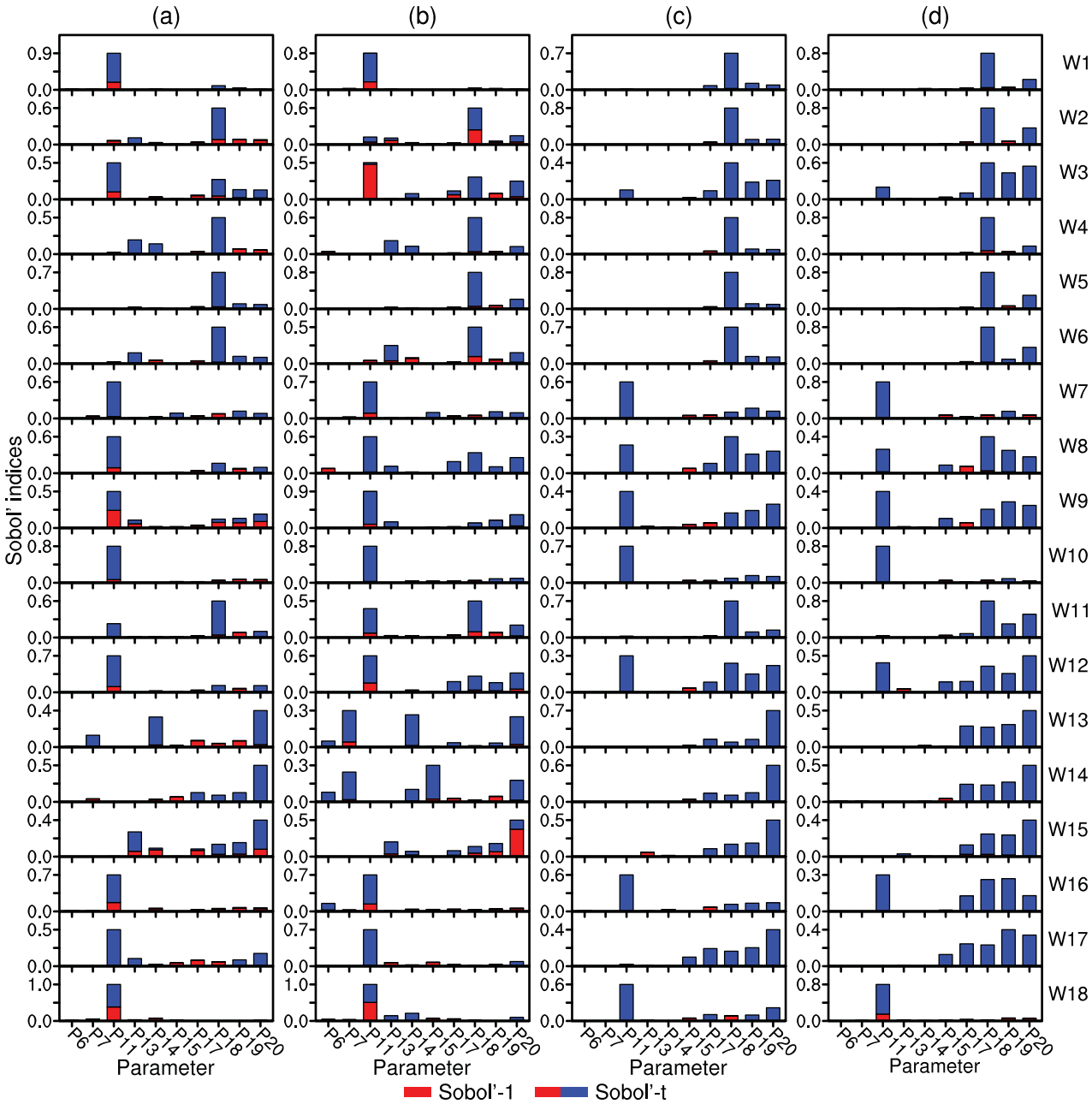

Figure 7. Sobol' sensitivity indices of 18 watersheds for (a) RB of SD, (b) NSE of SD, (c) RB of ET, and (d) NSE of ET. Each subfigure from top to bottom shows watershed 1 to watershed 18 (as represented by W1, W2, etc.), respectively.

objective function of each watershed. Figure 7 provides a detailed portrayal of the Sobol' sensitivity indices of 18 watersheds for four objective functions (RB of SD, NSE of SD, RB of ET, and NSE of ET). In each subfigure, the red bar represents the first-order sensitivity (main effect), while the blue bar indicates the higher-order sensitivity (interaction effect), and therefore the total sensitivity (total effect) is denoted by the sum of the red and blue bars. It is shown that parameter interaction is the principal mechanism for sensitivity, indicating that the model is nonadditive and that the response surface of the model is highly nonlinear.

In general, about 1–5 parameters control most of the model response variance in the 18 watersheds. The most important parameters are the reference depth for saturated hydraulic conductivity (P11), the quantum efficiency of vegetation photosynthesis (P18), and the minimum stomatal conductance (P20). The maximum carboxylation rate (P17) and the stomatal slope factor (P19) are of secondary importance, and the other

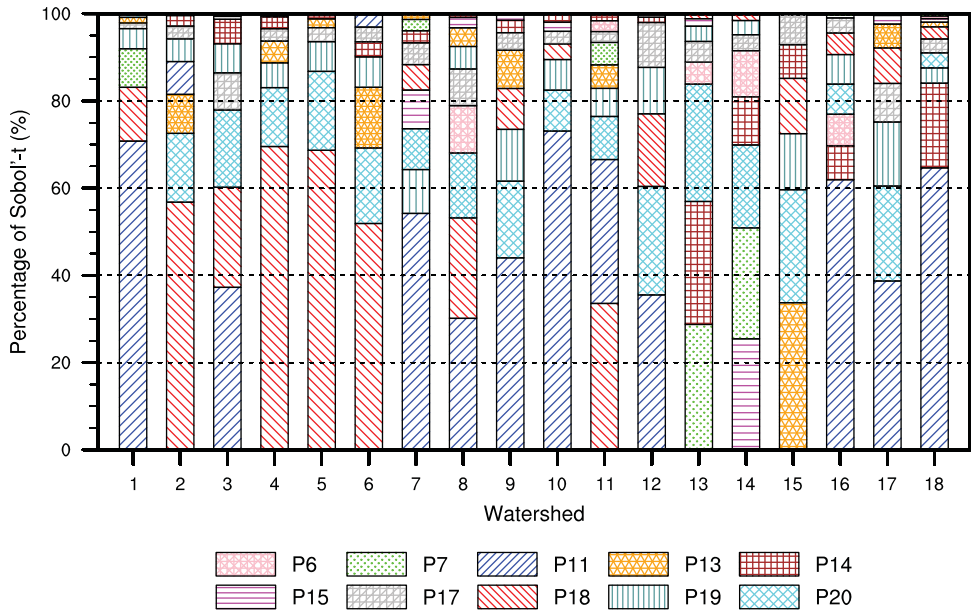

Figure 8. Ranked percentages of Sobol'-t for aggregate objective functions of 18 watersheds.

parameters have less impact on model responses. Consistent with both our expectations and the sensitivity results of LH-OAT screening, the photosynthetic and respiratory parameters P17, P18, P19, and P20, which control the variation of canopy water, exert significant influences on ET. Beyond being dominated by one of the photosynthetic and respiratory parameters such as P18 (for watersheds 2, 3, 4, 5, 6, and 11) or P20 (for watersheds 13, 14, and 15), the sensitivity pattern for SD is much more controlled by the reference depth for saturated hydraulic conductivity (P11), which directly controls the saturated hydraulic conductivity and hence the runoff generation. The variation of sensitive parameters is bigger for SD than for ET, demonstrating a shift in the dominant mechanisms for the generation of SD across watersheds.

Comparing the result of Sobol' SA (Figure 7) with that of LH-OAT screening (Figure 4), the parameter sensitivity patterns of the same watershed are generally consistent, with a few fluctuations. For example, P18 is taken as the most sensitive parameter for SD of watershed 1 by LH-OAT screening, while P11 is treated as the most sensitive parameter for SD of the same watershed by Sobol' SA. These observations demonstrate that the dominant mechanisms of SD and ET are identifiable by both LH-OAT screening and Sobol' SA methods. On the other hand, differences also exist between the results of these two methods under some circumstances, due to the influence of high nonlinearity.

To get a more objective ranking of multiobjective parameter sensitivity, we also quantified the contribution of each parameter to the variance of the aggregate objective function. Ranked percentages of Sobol' total sensitivity indices for aggregate objective function of 18 watersheds are shown in Figure 8. Parameters are represented by different fill patterns with different colors. Apparently, P11 plays the most significant role in the response variances for watersheds 1, 3, 7, 8, 9, 10, 12, 16, 17, and 18; while P18 exerts the most control for watersheds 2, 4, 5, 6, and 11. The most sensitive parameters for watersheds 13, 14, and 15 are P7, P15, and P13, respectively; these are different than the sensitive parameters in other watersheds. Overall, the most influential parameter corresponds to between 28% and 68% of the response variance across watersheds. Furthermore, the first five parameters make up 88% to 99% of the response variance across watersheds, and are sufficient to explain most of the model performance for SD and ET. Thus, the last five parameters, which have little or no influence on model response, can be fixed to default values in the parameter estimation process.

Comparing Figure 8 with Figure 7, it is notable that parameter sensitivity rankings for aggregate objective function are similar to those for objective functions of SD (i.e., RB of SD, and NSE of SD), indicating that the objective functions of SD contribute most to the aggregate objective function. This is because the best objective function values of SD are worse than those of ET, and larger weights are then given to objective functions of SD, as stated in section 3.1.2.

Rosolem et al. [2012] also presented a multiobjective implementation of the Sobol' method, which has proven to be superior to single-objective approaches while avoiding the subjectivity involved in pseudo multiobjective methods. However, their method is too time consuming to be applied in complex system models, as it not only requires large numbers of model simulations to get the Sobol' sensitivity indices, but also needs a large number of simulations to further get the linear correlation coefficients between objective function values for the samples of the rank-to-*r* sensitivity group parameter space and those for the samples of the total parameter space. On the other hand, our method is more efficient since we use the computationally cheaper surrogate model to approximate the complex LSM, and evaluate and rank the contribution of each parameter to the variance of the aggregate objective function values by using Sobol' total sensitivity indices directly. A potential shortcoming of our multiobjective implementation, as stated in *Rosolem et al.* [2013], is that a small change of the cutoff threshold may affect the number of influential parameters dramatically when the total number of considered parameters is large. Our suggestion is that a preliminary screening should be done first to reduce parameter dimensionality before adopting the multiobjective Sobol' method, and this is what we have shown in this paper.

5. Summary and Conclusions

An uncertainty quantification framework was introduced in this paper to analyze parameter sensitivities of large and complex system models. By applying it to the CSSP with 24 uncertain parameters, we aimed at identifying the parameters that dominate model behavior. Using 14 year (January 1982 to December 1995) records of the USGS observed SD and observation-based ET for 18 representative watersheds across CONUS, we gradually identified the sensitive parameters for the simulation of SD and ET across watersheds, by taking a multiobjective two-stage approach using first qualitative and then quantitative SA. We chose LH-OAT screening and MARS-based Sobol' sensitivity indices for the two stages, respectively. The former is efficient but less nuanced, while the latter is accurate and robust but computationally expensive. Although the application of the uncertainty quantification framework was used for a single offline LSM and focused on the CONUS region, we certainly believe that some conclusions from our sensitivity analyses of the CSSP have more general application.

Our sensitivity analyses showed that this multiobjective two-stage SA approach is effective and efficient for understanding and simplifying complex system models, such as the CSSP with its high-dimensional parameter space. Principal mechanisms of the terrestrial hydrological processes are identifiable and remarkably consistent with both LH-OAT screening and Sobol' sensitivity indices. In general, the dominant parameters are similar across metrics of the same output variable, with a few exceptions that vary across watersheds. Parameter interactions contribute most to the response variance of the CSSP, and only 5 out of 24 considered parameters dominate model behavior. Photosynthetic and respiratory parameters (P17, P18, P19, and P20) are more influential to ET, whereas reference depth for saturated hydraulic conductivity (P11) is the influential parameter for SD of the most watersheds. The lack of sensitivity for most of the considered parameters suggests that the CSSP may be overparameterized [*Jakeman and Hornberger*, 1993; *Bastidas et al.*, 2006; *Rosero et al.*, 2010] for terrestrial hydrological modeling, if the parameterizations are assumed to be correct.

Based on the comparison of parameter sensitivities across watersheds, we further conclude that parameter sensitivity patterns mainly depend on hydroclimatic regime, as well as vegetation type and soil texture. However, nonlinear effect is an unavoidable factor that governs complex system models and makes parameter estimation a tough problem. Overparameterization and high nonlinearity cause equifinality of model performance [*Beven*, 1993, 2006]. These observations remind us to carefully consider watershed physical characteristics when we attempt to use physical principles to design LSMs or explain parameter behaviors, especially when we try to transfer the optimized parameter sets from one watershed to another with so-called "similar" physical characteristics.

According to a comparison of the results of the MARS model using different sample sizes and cross-validation schemes, MARS was demonstrated to effectively approximate the computationally intensive CSSP at the cost of 800 LPTAU samples. Output uncertainty ranges for the 800 simulations generally captured the monthly observations, indicating that the model structure is warranted to predict SD and ET, and that parameter ranges and prior distributions were reasonably assigned. On the other hand, deviations between observations and outputs of control simulation suggest that the predictive ability of the CSSP can be improved by parameter optimization, which is the subject of our next study. We further assert that model skills can be better improved if more observed data, such as soil moisture, can be used to constrain the CSSP.

In light of the expensive computational resources needed for uncertainty quantification of large and complex system models, our multiobjective two-stage SA approach provides an efficient solution for identifying the principal mechanisms and reducing the parameter dimensionality of complex system models. A caveat is that we did not consider data uncertainties from the observations, forcing and boundary conditions. Since uncertainty quantification of LSM parameters are specific to the watersheds and models, in future works we will test the framework in other regions and using other LSMs. Furthermore, LSM parameters may exhibit different behaviors in coupled mode and offline mode due to land-atmosphere interactions [*Liu et al.*, 2004]. Therefore, we are also interested in conducting uncertainty quantification for the CWRF model which integrates the CSSP.

Appendix A: Streamflow and Evapotranspiration Representation of the Hydrologically Enhanced CSSP

The CSSP simulates streamflow by combining the routed surface flow with the unrouted subsurface runoff [*Choi et al.*, 2013]. To estimate specific discharge (discharge per unit drainage area) of total streamflow at a grid point (denoted as SD), the specific discharge of surface flow is added to the averaged subsurface runoff as:

$$SD = \frac{Q_s}{n_{fa}A} + \overline{R}_{sb} \tag{A1}$$

where Q_s is the discharge of surface flow, n_{fa} is the number of flow accumulation grids at the target grid point, A is the area of unit grid, and $\overline{R}_{sb}$ is the subsurface runoff averaged for the total upstream grid cells contributing to the target grid point.

The CSSP utilizes the diffusion wave equation for a surface flow routing scheme, which is an approximated solution of the full dynamic wave equations for unsteady surface flow as:

$$\frac{\partial h}{\partial t} + c_d \frac{\partial h}{\partial x_c} = D_h \frac{\partial^2 h}{\partial x_c^2} + R_s \tag{A2}$$

where h is the flow depth, t is the computational time scale, x_c is the longitudinal flow direction coordinate, c_d is the diffusion wave celerity, and D_h is the hydraulic diffusivity. R_s is the net surface runoff on the exchange of water between the surface and the subsurface, which includes both Horton and Dunne runoff generation mechanisms, as well as the influence of overland flow depth h for a time-increment Δt:

$$R_s = \underbrace{(1-F_{imp})\max\left[0, Q_w - I_{\max}\right]}_{\text{Hortonian}} + \underbrace{F_{imp}Q_w}_{\text{Dunnian}} - h/\Delta t \tag{A3}$$

where F_{imp} is the impermeable area fraction consisting of the fractional saturated area and the frozen area, and $I_{\max}$ is the maximum potential infiltration rate. Q_w is the total available water supply rate on the surface as:

$$Q_w = Q_{rain} + Q_{dew} + Q_{melt} + h/\Delta t \tag{A4}$$

where Q_{rain}, Q_{dew}, and Q_{melt} are rainfall, dewfall, and snowmelt rate at the surface, respectively. Note that the influence of overland flow depth on both infiltration rate and surface runoff is incorporated for the comprehensive surface and subsurface coupling in the CSSP.

In the CSSP, subsurface runoff consists of three components as:

$$R_{sb} = R_{sb,bas} + R_{sb,dra} + R_{sb,sat} \tag{A5}$$

where $R_{sb,bas}$, $R_{sb,dra}$, and $R_{sb,sat}$ denote subsurface runoff from baseflow, bottom drainage, and saturation excess, respectively. Note that subsurface runoff is calculated directly from the above three components in each soil column without any interacting or routing schemes for horizontal adjacent soil grids.

The evapotranspiration (ET) scheme is based on the CLM4.0 [*Oleson et al.*, 2010]. The water vapor flux ET between the surface at height of roughness length z_0 plus displacement height d and the atmosphere at reference height z_{atm} is:

$$ET = -\rho_{atm}\frac{(q_{atm} - q_s)}{r_{aw}} \tag{A6}$$

where ρ_{atm} is the air density, r_{aw} is the aerodynamic resistance to water vapor transfer, and q_{atm} and q_s are the specific humidity at the atmosphere and the surface, respectively.

The water vapor flux ET must be balanced by the sum of the water vapor flux from the vegetation E_v and the ground E_g [*Oleson et al.*, 2010] as:

$$ET = E_v + E_g \tag{A7}$$

where E_v and E_g are formulated, respectively as:

$$E_v = -\rho_{atm}\frac{(q_s - q_{sat}^{T_v})}{r_{total}} \tag{A8}$$

$$E_g = -\rho_{atm}\frac{\beta_{soi}(q_s - q_g)}{r'_{aw} + r_{litter}} \tag{A9}$$

Here $q_{sat}^{T_v}$ is the saturation specific humidity at the vegetation temperature T_v, and q_g is the specific humidity at the ground surface. r_{total} is the total resistance to water vapor transfer from the canopy to the canopy air, including contributions from the leaf boundary layer and sunlit and shaded stomatal resistances. r'_{aw} is the aerodynamic resistance to water vapor transfer between the ground and the canopy air. r_{litter} is the resistance for the plant litter layer. β_{soi} is an empirical function of soil water to represent the molecular diffusion process from the soil pore to the surface within the dry part of the soil [*Sakaguchi and Zeng*, 2009].

Equations (A6)–(A9) can be solved for the canopy specific humidity q_s:

$$q_s = \frac{c_a^w q_{atm} + c_g^w q_g + c_v^w q_{sat}^{T_v}}{c_a^w + c_g^w + c_v^w} \tag{A10}$$

where c_a^w, c_g^w, and c_v^w are the water vapor conductance terms from canopy to the atmosphere, ground to canopy air, and leaf to canopy air, respectively.

Substituting equation (A10) into equations (A8) and (A9) gives, respectively:

$$E_v = -\rho_{atm}\left[c_a^w q_{atm} + c_g^w q_g - (c_a^w + c_g^w) q_{sat}^{T_v}\right]\frac{c_v^w}{c_a^w + c_g^w + c_v^w} \tag{A11}$$

$$E_g = -\rho_{atm}\left[c_a^w q_{atm} + c_v^w q_{sat}^{T_v} - (c_a^w + c_v^w) q_g\right]\frac{c_g^w}{c_a^w + c_g^w + c_v^w} \tag{A12}$$

Finally, the water vapor flux E_v and transpiration flux E_v^t are updated for changes in leaf temperature as:

$$E_v = -\rho_{atm}\left[c_a^w q_{atm} + c_g^w q_g - (c_a^w + c_g^w)\left(q_{sat}^{T_v} + \frac{dq_{sat}^{T_v}}{dT_v}\Delta T_v\right)\right]\frac{c_v^w}{c_a^w + c_g^w + c_v^w} \tag{A13}$$

$$E_v^t = -r''_{dry}\rho_{atm}\left[c_a^w q_{atm} + c_g^w q_g - (c_a^w + c_g^w)\left(q_{sat}^{T_v} + \frac{dq_{sat}^{T_v}}{dT_v}\Delta T_v\right)\right]\frac{c_v^w}{c_a^w + c_g^w + c_v^w} \tag{A14}$$

where r''_{dry} is determined by the fraction of dry leaves, and sunlit and shaded leaf area indices.

Appendix B: LH-OAT Screening

Assume that we have n parameters $X_1, X_2, \cdots, X_n$ ($X_i \in [0,\ 1]; i = 1, 2, \cdots, n$), and the range of each parameter is partitioned into $p-1$ equal intervals. Thus, we get a n-dimension p-level orthogonal parameter space, where each parameter can take on values from p predetermined values. First, p LH sample points are generated from this orthogonal parameter space; and then, for each LH sample point, n sample points are generated by perturbing each of the n parameters one at a time with a predefined step. The total number of sample points required is thus $N = (n+1)p$.

The elementary effect of the ith parameter is defined as:

$$d_i = [f(X_1, \cdots, X_{i-1}, X_i \pm \Delta, X_{i+1}, \cdots, X_n) - f(X_1, \cdots, X_{i-1}, X_i, X_{i+1}, \cdots, X_n)]/\Delta \tag{B1}$$

where $f(\mathbf{X})$ is the response function of parameter set $\mathbf{X}$, and Δ is the space step, which is usually set to $p/[2(p-1)]$ when p is even [*Morris*, 1991]. The final sensitivity measures of the ith parameter can then be calculated as:

$$\mu_i = \sum_{j=1}^{p} |d_i(j)|/p \tag{B2}$$

and

$$\sigma_i = \sqrt{\sum_{j=1}^{p} (d_i(j)-\mu_i)^2/p} \tag{B3}$$

where μ_i and σ_i are the mean and standard deviation of d_i, and can be used to evaluate overall and interaction effects, respectively. *Campolongo et al*. [2007] suggested considering μ_i in absolute value to solve the problem of the effects of opposite signs in gradients.

Appendix C: Sobol' Sensitivity Indices

Let the function $Y=f(\mathbf{X})=f(X_1, X_2, \cdots, X_n)$, where $X_i \in [0,1]$ $(i=1,2,\cdots,n)$. Assume that the model output can be decomposed into terms of increasing dimensionality as follows:

$$Y=f(\mathbf{X})=f_0+\sum_{i=1}^{n} f_i(X_i)+\sum_{i=1}^{n}\sum_{j>i}^{n} f_{i,j}(X_i,X_j)+\cdots+f_{1,2,\cdots,n}(X_1,X_2,\cdots,X_n) \tag{C1}$$

where f_0 is a constant, $f_i(X_i)$ are the functions of one variable, and $f_{i,j}(X_i,X_j)$ are the functions of two variables, etc. The total variance of the output can be written as:

$$V(Y)=\int_0^1\cdots\int_0^1 f^2(\mathbf{X})d\mathbf{X}-f_0^2 \tag{C2}$$

while the contribution of a generic term $f_{i_1,\cdots,i_s}(1 \le i_1 < \cdots < i_s \le n)$ to the total variance can be written as:

$$V_{i_1,\cdots,i_s}=\int_0^1\cdots\int_0^1 f^2_{i_1,\cdots,i_s}(X_{i_1},\cdots,X_{i_s})dX_{i_1}\cdots dX_{i_s} \tag{C3}$$

Thus, the ANOVA-like decomposition of total variance can be expressed as:

$$V(Y)=\sum_{s=1}^{n}\sum_{i_1<\cdots<i_s}^{n} V_{i_1,\cdots,i_s}=\sum_{i=1}^{n} V_i+\sum_{i=1}^{n}\sum_{i<j}^{n} V_{i,j}+\cdots+V_{1,2,\cdots,n} \tag{C4}$$

The Sobol' sensitivity indices are defined as:

$$S_{i_1,\cdots,i_s}=V_{i_1,\cdots,i_s}/V(Y), \quad 1 \le i_1 < \cdots < i_s \le n. \tag{C5}$$

The measure proposed by *Homma and Saltelli* [1996] provides a simple way for computing the total effect of each input variable. The total effect of the ith input is defined as:

$$S_{Ti}=S_i+S_{i,ci}=1-S_{ci} \tag{C6}$$

where S_i and $S_{i,ci}$ are representations of first-order effect and high-order effect, respectively; S_{ci} is the sum of all the $S_{i_1,\cdots,i_s}$ terms that excludes the index i.

Appendix D: Multivariate Adaptive Regression Splines (MARS) and *k*-Fold Cross Validation

Let the function $Y=f(\mathbf{X})=f(X_1, X_2, \cdots, X_n)$, where parameter $X_i \in \mathbf{X}$ $(i=1,2,\cdots,n)$. MARS model can be used to approximate the relationship between parameters and function responses as:

$$\hat{Y}=\hat{f}(\mathbf{X})=a_0+\sum_{m=1}^{M} a_m \prod_{l=1}^{L_m}\left[s_{l,m}\left(X_{v(l,m)}-t_{l,m}\right)\right]_+^q \tag{D1}$$

where a_0 is a constant, a_m are fitting coefficients, M is the number of basis functions, L_m is the number of knots, $s_{l,m}$ takes on values of either 1 or -1 and indicates the right/left sense of the associated step function, $v(l,m)$ is the label of the independent variable, and $t_{l,m}$ indicates the knot location. The subscript "+" means the function is a truncated power function:

$$\left[s_{l,m}\left(X_{v(l,m)}-t_{l,m}\right)\right]_+^q=\begin{cases}\left[s_{l,m}\left(X_{v(l,m)}-t_{l,m}\right)\right]^q & s_{l,m}\left(X_{v(l,m)}-t_{l,m}\right)>0\\ 0 & s_{l,m}\left(X_{v(l,m)}-t_{l,m}\right)\le 0\end{cases} \tag{D2}$$

The k-fold cross validation can be employed to assess the accuracy of the MARS model. Namely, N parameter-response pairs $\left(\mathbf{X}^j, Y^j\right)$ for $j=1,2,\cdots,N$ were splitted into k roughly equal subsets, and the surrogate model was fitted k times, each time leaving out one of the subsets from training and using the omitted subset to compute error measures [*Queipo et al.*, 2005]. SRMSE is calculated by

$$SRMSE=\sqrt{\frac{1}{N}\sum_{p=1}^{k}\sum_{j=1}^{N/k}\left(\frac{\hat{Y}_p^j-Y_p^j}{Y_p^j}\right)^2} \tag{D3}$$

where $\hat{Y}_p^j$ and Y_p^j are the predicted and actual response values of the jth sample point of the pth subset, respectively.

Acknowledgments
This study was supported by the National Natural Sciences Foundation of China (41375139), Chinese Ministry of Science and Technology 973 Project (2015CB452805), USEPA Science to Achieve Results award RD83418902, NOAA Education Partnership Program (EPP) COM Howard 631017, and the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MEST) (NRF-2013R1A2A2A01008881). The first author was also funded by China Scholarship Council (201206040137). We would like to acknowledge the valuable suggestions from Xing Yuan at the Chinese Academy of Sciences and Min Xu and Yuxiang He at the University of Maryland. We also want to thank the reviewers and Editors for their constructive comments. The explanation and access to the data used in this study are explained in supporting information attached to this paper.

References

Baker, I. T., L. Prihodko, A. S. Denning, M. Goulden, S. Miller, and H. R. da Rocha (2008), Seasonal drought stress in the Amazon: Reconciling models and observations, *J. Geophys. Res.*, *113*, G00B01, doi:10.1029/2007JG000644.

Baldocchi, D., 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, *Bull. Am. Meteorol. Soc.*, *82*(11), 2415–2434, doi:10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2.

Bastidas, L. A., H. V. Gupta, S. Sorooshian, W. J. Shuttleworth, and Z.-L. Yang (1999), Sensitivity analysis of a land surface scheme using multi-criteria methods, *J. Geophys. Res.*, *104*(D16), 19,481–19,490, doi:10.1029/1999JD900155.

Bastidas, L. A., T. S. Hogue, S. Sorooshian, H. V. Gupta, and W. J. Shuttleworth (2006), Parameter sensitivity analysis for different complexity land surface models using multicriteria methods, *J. Geophys. Res.*, *111*, D20101, doi:10.1029/2005JD006377.

Becker, E. J., E. H. Berbery, and R. W. Higgins (2009), Understanding the characteristics of daily precipitation over the United States using the North American Regional Reanalysis, *J. Clim.*, *22*(23), 6268–6286, doi:10.1175/2009JCLI2838.1.

Beven, K. (1993), Prophecy, reality and uncertainty in distributed hydrological modelling, *Adv. Water Resour.*, *16*(1), 41–51, doi:10.1016/0309-1708(93)90028-E.

Beven, K. (2006), A manifesto for the equifinality thesis, *J. Hydrol.*, *320*(1-2), 18–36, doi:10.1016/j.jhydrol.2005.07.007.

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

Borgonovo, E., W. Castaings, and S. Tarantola (2012), Model emulation and moment-independent sensitivity analysis: An application to environmental modelling, *Environ. Modell. Software*, *34*, 105–115, doi:10.1016/j.envsoft.2011.06.006.

Budyko, M. I. (1956), *The Heat Balance of the Earth's Surface*, 255 pp., Gidrometeoizdat, Leningrad, Russia.

Campolongo, F., J. Cariboni, and A. Saltelli (2007), An effective screening design for sensitivity analysis of large models, *Environ. Modell. Software*, *22*(10), 1509–1518, doi:10.1016/j.envsoft.2006.10.004.

Campolongo, F., A. Saltelli, and J. Cariboni (2011), From screening to quantitative sensitivity analysis. A unified approach, *Comput. Phys. Commun.*, *182*(4), 978–988, doi:10.1016/j.cpc.2010.12.039.

Chen, F., and J. Dudhia (2001), Coupling an advanced land surface-hydrology model with the Penn state-NCAR MM5 modeling system. Part I: Model implementation and sensitivity, *Mon. Weather Rev.*, *129*(4), 569–585, doi:10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2.

Chen, M., Y. Sun, P. Fu, C. R. Carrigan, Z. Lu, C. H. Tong, and T. A. Buscheck (2013), Surrogate-based optimization of hydraulic fracturing in pre-existing fracture networks, *Comput. Geosci.*, *58*, 69–79, doi:10.1016/j.cageo.2013.05.006.

Choi, H. I., and X.-Z. Liang (2010), Improved terrestrial hydrologic representation in mesoscale land surface models, *J. Hydrometeorol.*, *11*(3), 797–809, doi:10.1175/2010JHM1221.1.

Choi, H. I., P. Kumar, and X.-Z. Liang (2007), Three-dimensional volume-averaged soil moisture transport model with a scalable parameterization of subgrid topographic variability, *Water Resour. Res.*, *43*, W04414, doi:10.1029/2006WR005134.

Choi, H. I., X.-Z. Liang, and P. Kumar (2013), A Conjunctive Surface-Subsurface Flow representation for mesoscale land surface models, *J. Hydrometeorol.*, *14*, 1421–1441, doi:10.1175/JHM-D-12-0168.1.

Confalonieri, R., G. Bellocchi, S. Bregaglio, M. Donatelli, and M. Acutis (2010), Comparison of sensitivity analysis techniques: A case study with the rice model WARM, *Ecol. Modell.*, *221*(16), 1897–1906, doi:10.1016/j.ecolmodel.2010.04.021.

Dai, Y., X. Zeng, R. E. Dickinson, I. Baker, G. B. Bonan, M. G. Bosilovich, A. S. Denning, P. A. Dirmeyer, P. R. Houser, and G.-Y. Niu (2003), The Common Land Model, *Bull. Am. Meteorol. Soc.*, *84*(8), 1013–1023, doi:10.1175/BAMS-84-8-1013.

Dai, Y., R. E. Dickinson, and Y.-P. Wang (2004), A two-big-leaf model for canopy temperature, photosynthesis, and stomatal conductance, *J. Clim.*, *17*(12), 2281–2299, doi:10.1175/1520-0442(2004)017<2281:ATMFCT>2.0.CO;2.

Demaria, E. M., B. Nijssen, and T. Wagener (2007), Monte Carlo sensitivity analysis of land surface parameters using the Variable Infiltration Capacity model, *J. Geophys. Res.*, *112*, D11113, doi:10.1029/2006JD007534.

Dickinson, R. E., A. Henderson-Sellers, P. J. Kennedy, and M. F. Wilson (1986), *Biosphere-Atmosphere Transfer Scheme (BATS) for the NCAR Community Climate Model*, 69 pp., NCAR, Boulder, Colo., doi:10.5065/D6668B58.

Dickinson, R. E., A. Henderson-Sellers, and P. J. Kennedy (1993), *Biosphere-Atmosphere Transfer Scheme (BATS) Version 1e as Coupled to the NCAR Community Climate Model*, 72 pp., NCAR, Boulder, Colo., doi:10.5065/D67W6959.

Dickinson, R. E., K. W. Oleson, G. Bonan, F. Hoffman, P. Thornton, M. Vertenstein, Z.-L. Yang, and X. Zeng (2006), The Community Land Model and its climate statistics as a component of the Community Climate System Model, *J. Clim.*, *19*(11), 2302–2324, doi:10.1175/JCLI3742.1.

Duan, Q., S. Sorooshian, and V. Gupta (1992), Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resour. Res.*, *28*(4), 1015–1031, doi:10.1029/91WR02985.

Duan, Q., et al. (2006), Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops, *J. Hydrol.*, *320*(1-2), 3–17, doi:10.1016/j.jhydrol.2005.07.031.

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

FAO/IIASA/ISRIC/ISS-CAS/JRC (2012), *Harmonized World Soil Database (Version 1.2)*, FAO, Rome.

Farr, T. G., et al. (2007), The Shuttle Radar Topography Mission, *Rev. Geophys.*, *45*, RG2004, doi:10.1029/2005RG000183.

Friedman, J. H. (1991), Multivariate adaptive regression splines, *Ann. Stat.*, *19*(1), 1–67, doi:10.1214/aos/1176347963.

Gan, Y., Q. Duan, W. Gong, C. Tong, Y. Sun, W. Chu, A. Ye, C. Miao, and Z. Di (2014), A comprehensive evaluation of various sensitivity analysis methods: A case study with a hydrological model, *Environ. Modell. Software*, *51*, 269–285, doi:10.1016/j.envsoft.2013.09.031.

Giorgi, F., and R. Avissar (1997), Representation of heterogeneity effects in earth system modeling: Experience from land surface modeling, *Rev. Geophys.*, *35*(4), 413–437, doi:10.1029/97RG01754.

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

Gupta, V. K., and S. Sorooshian (1985), The relationship between data and the precision of parameter estimates of hydrologic models, *J. Hydrol.*, *81*(1–2), 57–77, doi:10.1016/0022-1694(85)90167-2.

Henderson-Sellers, A., Z.-L. Yang, and R. E. Dickinson (1993), The Project for Intercomparison of Land-surface Parameterization Schemes, *Bull. Am. Meteorol. Soc.*, *74*(7), 1335–1349, doi:10.1175/1520-0477(1993)074<1335:TPFIOL>2.0.CO;2.

Henderson-Sellers, A., A. J. Pitman, P. K. Love, P. Irannejad and T. H. Chen (1995), The Project for Intercomparison of Land-surface Parameterization Schemes (PILPS): Phases 2 and 3, *Bull. Am. Meteorol. Soc.*, *76*(4), 489–503, doi:10.1175/1520-0477(1995)076<0489:TPFIOL>2.0.CO;2.

Holvoet, K., A. van Griensven, P. Seuntjens, and P. A. Vanrolleghem (2005), Sensitivity analysis for hydrology and pesticide supply towards the river in SWAT, *Phys. Chem. Earth*, *30*(8–10), 518–526, doi:10.1016/j.pce.2005.07.006.

Homma, T., and A. Saltelli (1996), Importance measures in global sensitivity analysis of nonlinear models, *Reliab. Eng. Syst. Safety*, *52*(1), 1–17, doi:10.1016/0951-8320(96)00002-6.

Hornberger, G. M., and R. C. Spear (1981), An approach to the preliminary analysis of environmental systems, *J. Environ. Manage.*, *12*(1), 7–18.

Hou, Z., M. Huang, L. R. Leung, G. Lin, and D. M. Ricciuto (2012), Sensitivity of surface flux simulations to hydrologic parameters based on an uncertainty quantification framework applied to the Community Land Model, *J. Geophys. Res.*, *117*, D15108, doi:10.1029/2012JD017521.

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

Jakeman, A. J., and G. M. Hornberger (1993), How much complexity is warranted in a rainfall-runoff model?, *Water Resour. Res.*, *29*(8), 2637–2649, doi:10.1029/93WR00877.

Johnson, A. I. (1967), *Specific Yield: Compilation of Specific Yields for Various Materials*, 74 pp., U.S. Gov. Print. Off., Washington, D. C.

Judson, A., and N. Doesken (2000), Density of freshly fallen snow in the central Rocky Mountains, *Bull. Am. Meteorol. Soc.*, *81*(7), 1577–1588, doi:10.1175/1520-0477(2000)081<1577:DOFFSI>2.3.CO;2.

Jung, M., et al. (2010), Recent decline in the global land evapotranspiration trend due to limited moisture supply, *Nature*, *467*(7318), 951–954, doi:10.1038/nature09396.

Jung, M., et al. (2011), Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations, *J. Geophys. Res.*, *116*, G00J07, doi:10.1029/2010JG001566.

Kennedy, A. D., X. Dong, B. Xi, S. Xie, Y. Zhang, and J. Chen (2011), A comparison of MERRA and NARR reanalyses with the DOE ARM SGP data, *J. Clim.*, *24*(17), 4541–4557, doi:10.1175/2011JCLI3978.1.

Kumar, P. (2004), Layer averaged Richard's equation with lateral flow, *Adv. Water Resour.*, *27*(5), 521–531, doi:10.1016/j.advwatres.2004.02.007.

Lawrence, D. M., et al. (2011), Parameterization improvements and functional and structural advances in version 4 of the Community Land Model, *J. Adv. Model. Earth Syst.*, *3*, M03001, doi:10.1029/2011MS00045.

Lawrence, D. M., K. W. Oleson, M. G. Flanner, C. G. Fletcher, P. J. Lawrence, S. Levis, S. C. Swenson, and G. B. Bonan (2012), The CCSM4 land simulation, 1850–2005: Assessment of surface climate and new capabilities, *J. Clim.*, *25*(7), 2240–2260, doi:10.1175/JCLI-D-11-00103.1.

Lefebre, F., H. Gallée, J. van Ypersele, and W. Greuell (2003), Modeling of snow and ice melt at ETH Camp (West Greenland): A study of surface albedo, *J. Geophys. Res.*, *108*(D8), 4231, doi:10.1029/2001JD001160.

Levy, S., and D. M. Steinberg (2010), Computer experiments: A review, *AStA Adv. Stat. Anal.*, *94*(4), 311–324, doi:10.1007/s10182-010-0147-9.

Li, C. Z., H. Wang, J. Liu, D. H. Yan, F. L. Yu, and L. Zhang (2010), Effect of calibration data series length on performance and optimal parameters of hydrological model, *Water Sci. Eng.*, *3*(4), 378–393, doi:10.3882/j.issn.1674-2370.2010.04.002.

Li, J., Q. Duan, W. Gong, A. Ye, Y. Dai, C. Miao, Z. Di, C. Tong, and Y. Sun (2013), Assessing parameter importance of the Common Land Model based on qualitative and quantitative sensitivity analysis, *Hydrol. Earth Syst. Sci.*, *17*, 3279–3293, doi:10.5194/hess-17-3279-2013.

Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges (1994), A simple hydrologically based model of land surface water and energy fluxes for general circulation models, *J. Geophys. Res.*, *99*(D7), 14,415–14,428, doi:10.1029/94JD00483.

Liang, X.-Z., et al. (2005a), Development of land surface albedo parameterization based on Moderate Resolution Imaging Spectroradiometer (MODIS) data, *J. Geophys. Res.*, *110*, D11107, doi:10.1029/2004JD005579.

Liang, X.-Z., H. I. Choi, K. E. Kunkel, Y. Dai, E. Joseph, J. X. L. Wang, and P. Kumar (2005b), *Development of the Regional Climate-Weather Research and Forecasting (CWRF) Model: Surface Boundary Conditions*, 32 pp., Ill. State Water Surv., Champaign.

Liang, X.-Z., H. I. Choi, K. E. Kunkel, Y. Dai, E. Joseph, J. X. L. Wang, and P. Kumar (2005c), Surface boundary conditions for mesoscale regional climate models, *Earth Interact.*, *9*, 1–28, doi:10.1175/EI151.1.

Liang, X.-Z., et al. (2012), Regional Climate-Weather Research and Forecasting Model (CWRF), *Bull. Am. Meteorol. Soc.*, *93*, 1363–1387, doi:10.1175/BAMS-D-11-00180.1.

Liu, Y., H. V. Gupta, S. Sorooshian, L. A. Bastidas, and W. J. Shuttleworth (2004), Exploring parameter sensitivities of the land surface using a locally coupled land-atmosphere model, *J. Geophys. Res.*, *109*, D21101, doi:10.1029/2004JD004730.

Luo, Y. Q., J. T. Randerson, G. Abramowitz, C. Bacour, E. Blyth, N. Carvalhais, P. Ciais, D. Dalmonech, J. B. Fisher, and R. Fisher (2012), A framework for benchmarking land models, *Biogeosciences*, *9*, 3857–3874, doi:10.5194/bg-9-3857-2012.

Madsen, H. (2000), Automatic calibration of a conceptual rainfall-runoff model using multiple objectives, *J. Hydrol.*, *235*(3), 276–288, doi:10.1016/S0022-1694(00)00279-1.

Manabe, S. (1969), Climate and the ocean circulation: 1. The atmospheric circulation and the hydrology of the earth's surface, *Mon. Weather Rev.*, *97*(11), 739–774, doi:10.1175/1520-0493(1969)097<0739:CATOC>2.3.CO;2.

McCuen, R. H., Z. Knight, and A. G. Cutter (2006), Evaluation of the Nash-Sutcliffe efficiency index, *J. Hydrol. Eng.*, *11*(6), 597–602, doi:10.1061/(ASCE)1084-0699(2006)11:6(597).

McKay, M. D., R. J. Beckman, and W. J. Conover (1979), A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics*, *21*(2), 239–245, doi:10.1080/00401706.2000.10485979.

Merz, R., G. Blöschl, and J. Parajka (2006), Spatio-temporal variability of event runoff coefficients, *J. Hydrol.*, *331*(3), 591–604, doi:10.1016/j.jhydrol.2006.06.008.

Mesinger, F., G. DiMego, E. Kalnay, K. Mitchell, P. C. Shafran, W. Ebisuzaki, D. Jovic, J. Woollen, E. Rogers, and E. H. Berbery (2006), North American Regional Reanalysis, *Bull. Am. Meteorol. Soc.*, *87*(3), 343–360, doi:10.1175/BAMS-87-3-343.

Meteopolis, N., and S. Ulam (1949), The Monte Carlo method, *J. Am. Stat. Assoc.*, *44*(247), 335–341, doi:10.1080/01621459.1949.10483310.

Miller, D. A., and R. A. White (1998), A conterminous United States multilayer soil characteristics dataset for regional climate and hydrology modeling, *Earth Interact.*, *2*(2), 1–26, doi:10.1175/1087-3562(1998)002<0001:ACUSMS>2.3.CO;2.

Mo, K. C., L. N. Long, Y. Xia, S. K. Yang, J. E. Schemm, and M. Ek (2010), Drought indices based on the climate forecast system reanalysis and ensemble NLDAS, *J. Hydrometeorol.*, *12*(2), 181–205, doi:10.1175/2010JHM1310.1.

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

Mueller, B., et al. (2011), Evaluation of global observations-based evapotranspiration datasets and IPCC AR4 simulations, *Geophys. Res. Lett.*, *38*, L06402, doi:10.1029/2010GL046230.

Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models: Part I—A discussion of principles, *J. Hydrol.*, *10*(3), 282–290, doi:10.1016/0022-1694(70)90255-6.

Neelin, J. D., A. Bracco, H. Luo, J. C. McWilliams, and J. E. Meyerson (2010), Considerations for parameter optimization and sensitivity in climate models, *Proc. Natl. Acad. Sci. U. S. A.*, *107*(50), 21,349–21,354, doi:10.1073/pnas.1015473107.

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

Oleson, K. W., et al. (2010), *Technical Description of Version 4.0 of the Community Land Model (CLM)*, 257 pp., NCAR, Boulder, Colo., doi:10.5065/D6FB50WZ.

Owen, A. B. (1992), Orthogonal arrays for computer experiments, integration and visualization, *Stat. Sin.*, *2*(2), 439–452.

Picard, R. R., and R. D. Cook (1984), Cross-validation of regression models, *J. Am. Stat. Assoc.*, *79*(387), 575–583, doi:10.1080/01621459.1984.10478083.

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

Pitman, A. J., and A. Henderson-Sellers (1998), Recent progress and results from the Project for the Intercomparison of Land-surface Parameterization Schemes, *J. Hydrol.*, *212*, 128–135, doi:10.1016/S0022-1694(98)00206-6.

Pitman, A. J., et al. (1999), Key results and implications from phase 1(c) of the Project for Intercomparison of Land-surface Parameterization Schemes, *Clim. Dyn.*, *15*(9), 673–684, doi:10.1007/s003820050309.

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

Queipo, N. V., R. T. Haftka, W. Shyy, T. Goel, R. Vaidyanathan, and P. Kevin Tucker (2005), Surrogate-based analysis and optimization, *Prog. Aerosp. Sci.*, *41*(1), 1–28, doi:10.1016/j.paerosci.2005.02.001.

Razavi, S., B. A. Tolson, and D. H. Burn (2012), Review of surrogate modeling in water resources, *Water Resour. Res.*, *48*, W07401, doi:10.1029/2011WR011527.

Renard, B., D. Kavetski, G. Kuczera, M. Thyer, and S. W. Franks (2010), Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, *Water Resour. Res.*, *46*, W05521, doi:10.1029/2009WR008328.

Rosero, E., Z.-L. Yang, T. Wagener, L. E. Gulden, S. Yatheendradas, and G.-Y. Niu (2010), Quantifying parameter sensitivity, interaction, and transferability in hydrologically enhanced versions of the Noah land surface model over transition zones during the warm season, *J. Geophys. Res.*, *115*, D03106, doi:10.1029/2009JD012035.

Rosolem, R., H. V. Gupta, W. J. Shuttleworth, X. Zeng, and L. G. G. de Gonçalves (2012), A fully multiple-criteria implementation of the Sobol′ method for parameter sensitivity analysis, *J. Geophys. Res.*, *117*, D07103, doi:10.1029/2011JD016355.

Rosolem, R., H. V. Gupta, W. J. Shuttleworth, L. G. G. de Gonçalves, and X. Zeng (2013), Towards a comprehensive approach to parameter estimation in land surface parameterization schemes, *Hydrol. Processes*, *27*(14), 2075–2097, doi:10.1002/hyp.9362.

Sakaguchi, K., and X. Zeng (2009), Effects of soil wetness, plant litter, and under-canopy atmospheric stability on ground evaporation in the Community Land Model (CLM3.5), *J. Geophys. Res.*, *114*, D01107, doi:10.1029/2008JD010834.

Saltelli, A., S. Tarantola, F. Campolongo, and M. Ratto (2004), *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*, 219 pp., John Wiley, Chichester, U. K.

Saltelli, A., M. Ratto, T. Andres, and F. Campolongo (2008), *Global Sensitivity Analysis: The Primer*, 292 pp., John Wiley, Chichester, U. K.

Sellers, P. J., Y. Mintz, Y. E. A. Sud, and A. Dalcher (1986), A simple biosphere model (SiB) for use within general circulation models, *J. Atmos. Sci.*, *43*(6), 505–531, doi:10.1175/1520-0469(1986)043<0505:ASBMFU>2.0.CO;2.

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

Shahsavani, D., and A. Grimvall (2011), Variance-based sensitivity analysis of model outputs using surrogate models, *Environ. Modell. Software*, *26*(6), 723–730, doi:10.1016/j.envsoft.2011.01.002.

Shao, Y., and A. Henderson-Sellers (1996), Modeling soil moisture: A Project for Intercomparison of Land-surface Parameterization Schemes phase 2(b), *J. Geophys. Res.*, *101*(D3), 7227–7250, doi:10.1029/95JD03275.

Sobol', I. M. (1967), On the distribution of points in a cube and the approximate evaluation of integrals, *USSR Comput. Math. Math. Phys., Engl. Transl.*, *7*, 86–112, doi:10.1016/0041-5553(67)90144-9.

Sobol', I. M. (1993), Sensitivity analysis for nonlinear mathematical models, *Math. Model. Comput. Exp.*, *1*, 407–414.

Sobol', I. M. (2001), Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, *Math. Comput. Simul.*, *55*(1-3), 271–280, doi:10.1016/S0378-4754(00)00270-6.

Sorooshian, S., and V. K. Gupta (1983), Automatic calibration of conceptual rainfall-runoff models: The question of parameter observability and uniqueness, *Water Resour. Res.*, *19*(1), 260–268, doi:10.1029/WR019i001p00260.

Stephens, D. W., D. Gorissen, K. Crombecq, and T. Dhaene (2011), Surrogate based sensitivity analysis of process equipment, *Appl. Math. Model.*, *35*(4), 1676–1687, doi:10.1016/j.apm.2010.09.044.

Storlie, C. B., L. P. Swiler, J. C. Helton, and C. J. Sallaberry (2009), Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models, *Reliab. Eng. Syst. Safety*, *94*(11), 1735–1763, doi:10.1016/j.ress.2009.05.007.

Tang, B. (1993), Orthogonal array-based Latin hypercubes, *J. Am. Stat. Assoc.*, *88*(424), 1392–1397, doi:10.2307/2291282.

Tang, Y., P. Reed, T. Wagener, and K. van Werkhoven (2007), Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation, *Hydrol. Earth Syst. Sci.*, *11*(2), 793–817, doi:10.5194/hess-11-793-2007.

Tolson, B. A., and C. A. Shoemaker (2007), Dynamically dimensioned search algorithm for computationally efficient watershed model calibration, *Water Resour. Res.*, *43*, W01413, doi:10.1029/2005WR004723.

van Griensven, A., T. Meixner, S. Grunwald, T. Bishop, M. Diluzio, and R. Srinivasan (2006), A global sensitivity analysis tool for the parameters of multi-variable catchment models, *J. Hydrol.*, *324*(1–4), 10–23, doi:10.1016/j.jhydrol.2005.09.008.

van Werkhoven, K., T. Wagener, P. Reed, and Y. Tang (2008), Characterization of watershed model behavior across a hydroclimatic gradient, *Water Resour. Res.*, *44*, W01429, doi:10.1029/2007WR006271.

van Werkhoven, K., T. Wagener, P. Reed, and Y. Tang (2009), Sensitivity-guided reduction of parametric dimensionality for multi-objective calibration of watershed models, *Adv. Water Resour.*, *32*(8), 1154–1169, doi:10.1016/j.advwatres.2009.03.002.

Vrugt, J. A., H. V. Gupta, L. A. Bastidas, W. Bouten, and S. Sorooshian (2003), Effective and efficient algorithm for multiobjective optimization of hydrologic models, *Water Resour. Res.*, *39*(8), 1214, doi:10.1029/2002WR001746.

Wagener, T., K. van Werkhoven, P. Reed, and Y. Tang (2009), Multiobjective sensitivity analysis to understand the information content in streamflow observations for distributed watershed modeling, *Water Resour. Res.*, *45*, W02501, doi:10.1029/2008WR007347.

Wainwright, H. M., S. Finsterle, Y. Jung, Q. Zhou, and J. T. Birkholzer (2014), Making sense of global sensitivity analyses, *Comput. Geosci.*, *65*, 84–94, doi:10.1016/j.cageo.2013.06.006.

Wang, G. G., and S. Shan (2007), Review of metamodeling techniques in support of engineering design optimization, *J. Mech. Design*, *129*(4), 370–380, doi:10.1115/1.2429697.

Wu, H., J. S. Kimball, N. Mantua, and J. Stanford (2011), Automated upscaling of river networks for macroscale hydrological modeling, *Water Resour. Res.*, *47*, W03517, doi:10.1029/2009WR008871.

Xia, Y., Z.-L. Yang, P. Stoffa, and M. Sen (2005), Optimal parameter and uncertainty estimation of a land surface model: Sensitivity to parameter ranges and model complexities, *Adv. Atmos. Sci.*, *22*(1), 142–157, doi:10.1007/BF02930878.

Xia, Y., et al. (2012), Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products, *J. Geophys. Res.*, *117*, D03109, doi:10.1029/2011JD016048.

Yapo, P. O., H. V. Gupta, and S. Sorooshian (1998), Multi-objective global optimization for hydrologic models, *J. Hydrol.*, *204*(1–4), 83–97, doi:10.1016/S0022-1694(97)00107-8.

Yuan, X., and X.-Z. Liang (2011), Evaluation of a Conjunctive Surface–Subsurface Process model (CSSP) over the contiguous United States at regional–local scales, *J. Hydrometeorol.*, *12*(4), 579–599, doi:10.1175/2010JHM1302.1.

Zeng, X., R. E. Dickinson, A. Walker, M. Shaikh, R. S. DeFries, and J. Qi (2000), Derivation and evaluation of global 1-km fractional vegetation cover data for land modeling, *J. Appl. Meteorol.*, *39*(6), 826–839, doi:10.1175/1520-0450(2000)039<0826:DAEOGK>2.0.CO;2.

Zeng, X., M. Shaikh, Y. Dai, R. E. Dickinson, and R. Myneni (2002), Coupling of the Common Land Model to the NCAR Community Climate Model, *J. Clim.*, *15*(14), 1832–1854, doi:10.1175/1520-0442(2002)015<1832:COTCLM>2.0.CO;2.

Zeng, Z., S. Piao, X. Lin, G. Yin, S. Peng, P. Ciais, and R. B. Myneni (2012), Global evapotranspiration over the past three decades: Estimation based on the water balance equation combined with empirical models, *Environ. Res. Lett.*, *7*(1), 014026, doi:10.1088/1748-9326/7/1/014026.

Zhang, X., R. Srinivasan, and M. Van Liew (2009), Approximating SWAT model using artificial neural network and support vector machine, *J. Am. Water Resour. Assoc.*, *45*(2), 460–474, doi:10.1111/j.1752-1688.2009.00302.x.

Erratum

In the originally published version of this article, there were errors in the text of sections 1 and 2. These errors have since been corrected, and this version may be considered the authoritative version of record.