JAMES Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2018MS001573

Key Points:
- Perform sensitivity analysis of spatially distributed model without significant information loss by running model at subset of grids
- Rank parameters by importance group identified via a grouping-based sensitivity analysis approach
- High importance parameters can be identified by running model on 5% of total grids for 5 times the number of parameters

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

Citation:
Huo, X., Gupta, H., Niu, G.-Y., Gong, W., & Duan, Q. (2019). Parameter sensitivity analysis for computationally intensive spatially distributed dynamical environmental systems models. *Journal of Advances in Modeling Earth Systems*, *11*. https://doi.org/10.1029/2018MS001573

Received 26 NOV 2018
Accepted 18 AUG 2019
Accepted article online 21 AUG 2019

Parameter Sensitivity Analysis for Computationally Intensive Spatially Distributed Dynamical Environmental Systems Models

Xueli Huo[1,2], Hoshin Gupta[3], Guo-Yue Niu[3], Wei Gong[1,2], and Qingyun Duan[4]

[1]State Key Laboratory of Earth Surface Processes and Resource Ecology, Faculty of Geographical Science, Beijing Normal University, Beijing, China, [2]Institute of Land Surface System and Sustainable Development, Faculty of Geographical Science, Beijing Normal University, Beijing, China, [3]Department of Hydrology and Atmospheric Sciences, The University of Arizona, Tucson, AZ, USA, [4]State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, College of Hydrology and Water Resources, Hohai University, Nanjing, China

Abstract Dynamical environmental systems models are highly parameterized, having large numbers of parameters whose values are uncertain. For spatially distributed continental-scale applications, such models must be run for very large numbers of grid locations. To calibrate such models, it is useful to be able to perform parameter screening, via sensitivity analysis, to identify the most important parameters. However, since this typically requires the models to be run for a large number of sampled parameter combinations, the computational burden can be huge. To make such an investigation computationally feasible, we propose a novel approach to combining spatial sampling with parameter sampling and test it for the Noah-MP land surface model applied across the continental United States, focusing on gross primary production and flux of latent heat simulations for two vegetation types. Our approach uses (a) progressive Latin hypercube sampling to sample at four grid levels and four parameter levels, (b) a recently developed grouping-based sensitivity analysis approach that ranks parameters by importance group rather than individually, and (c) a measure of robustness to grid and parameter sampling variability. The results show that a relatively small grid sample size (i.e., 5% of the total grids) and small parameter sample size (i.e., 5 times the number of parameters) are sufficient to identify the most important parameters, with very high robustness to grid sampling variability and a medium level of robustness to parameter sampling variability. The results ensure a dramatic reduction in computational costs for such studies.

1. Introduction

To properly represent the intricate feedback mechanisms between components of a complex system, dynamical environmental systems models (DESMs) are highly parameterized with empirical equations, resulting in many uncertain parameters. The difficulty of specifying optimal values for such parameters is one of the important causes for differences between DESM simulations and observed data. While parameter optimization can help improve DESM performance, a preliminary screening investigation is generally necessary to select the most important parameters for further calibration, that is, the ones that exert the strongest influence on the model behavior. Further, such an investigation can inform research on how to devote resources toward information gathering aiming at reduction of parameter uncertainties.

Various global sensitivity analysis (GSA) methods to perform parameter screening have been proposed (Campolongo et al., 2007; Friedman, 1991; Murphy et al., 2004; Oakley & O'Hagan, 2004; Ratto et al., 2007; Sacks et al., 1989; Saltelli et al., 1999; Saltelli et al., 2004; Saltelli et al., 2008; Sobol'', 1993). However, application of GSA methods to large-scale spatially distributed applications of DESMs is complicated by the high computational demands—for each sampled parameter combination, the model must be run for each of the large number of grids across the spatial domain. For example, Herman et al. (2013) applied the Hydrology Laboratory Research Distributed Hydrologic Model to only 78 grids covering the 1,248-km^2 Blue River basin in southern Oklahoma, USA, and reported that use of the Sobol' GSA method to obtain reliable estimates of parameter importance would require over six million model runs. This implies that proper application of GSA over even larger areas (e.g., continental scale) would be effectively impossible. Zhan et al.

(2013) showed that computational time greater than 1 month would be necessary to make the 100,000 model runs for use of Sobol' GSA to identify the first-order "sensitive" (i.e., most important) parameters of the distributed time-variant gain model (Xia et al., 2005) applied to the 441 subbasins of the Huaihe river basin. Similarly, Lee et al. (2012) reported that use of extend-FAST GSA with Monte Carlo sampling to produce a global map of the sensitivities of absolute and relative cloud condensation nuclei to eight parameters of a complex global aerosol microphysics model would be computationally prohibitive.

A number of approaches to reducing the costs associated with parameter GSA of computationally expensive DESMs have been discussed in the literature. In particular, a major effort has been devoted to strategies for decreasing the number of required model runs by improving the efficiency of the sensitivity analysis methods. Examples include use of the Morris screening method (Shin et al., 2013; Tian, 2013), meta-model approaches (Gan et al., 2014; Li et al., 2013; Razavi et al., 2012; Song et al., 2012; Zhan et al., 2013), and integration of qualitative and quantitative sensitivity analysis methods (Francos et al., 2003; Song et al., 2012; Zhan et al., 2013). As an example of the latter, Gan et al. (2015) examined the importance of 24 parameters of a computationally intensive land surface model to help understand and simplify the model. Campolongo et al. (2007) reduced computational costs by revising the Morris sensitivity measure and developing a more effective sampling strategy. Razavi and Gupta (2015a, 2015b) demonstrated the ability of the "Variogram Analysis of Response Surfaces" (VARS) GSA approach based on STAR-VARS sampling to provide reliable parameter importance estimates for high-dimensional DESMs using relatively small computational budgets. Ryan et al. (2018) demonstrated that a dimension reduction approach based on principle component analysis combined with an emulator-free approach enabled computation of variance-based sensitivity indices 37 times and 24 times faster on average compared to extended FAST or Sobol methods.

In this paper, we explore a novel approach to the problem of GSA for spatially distributed applications of DESMs. Our approach is based on the insight that computational costs associated with GSA and parameter optimization for such applications can be potentially reduced, without significant information loss, by running the model at only a representative subset of grids across the spatial domain. We select the grids via a process of stratified sampling and use a combination of Latin hypercube grid sampling and parameter sampling to inform the GSA procedure. Further, we implement a version of the recently developed grouping-based sensitivity analysis approach to rank parameters by importance *group* rather than individually (Sheikholeslami et al., 2019). As we demonstrate later, this combination of strategies can provide meaningful estimates of parameter importance at considerably reduced computational costs, making feasible the application of GSA to spatially distributed applications of DESMs. To verify our method, we compare the results of grid subset sampling to those obtained using the more expensive strategy of running Noah-MP model over the entire spatial domain.

This paper is structured as follows: Section 2 briefly describes the Noah-MP model and the data used. Section 3 discusses the experimental design strategy used to perform GSA for this model, via stratified spatial sampling on selected vegetation types combined with parameter sampling. Section 4 describes technical details related to sampling, sensitivity analysis, importance ranking, and assessment used in our approach. Section 5 presents and discusses our results regarding the sensitivity of gross primary production (GPP) and flux of latent heat (FLH) simulations to parameters for two vegetation types. We also discuss the robustness of the GSA results to spatial and parameter sampling variability. Section 6 concludes with the main findings of this paper.

2. Model and Data

2.1. A Brief Introduction to Noah-MP

Noah-MP models the states of terrestrial energy, water, carbon, and associated flux exchanges between the land surface and the atmosphere that are controlled by terrestrial hydrometeorological and ecohydrological processes (Niu et al., 2011). Based on the Noah model (Chen & Dudhia, 2001; Ek et al., 2003), Noah-MP was first augmented with vegetation and groundwater dynamics (Niu et al., 2007) and then equipped with multiple schemes for each of the ecohydrological processes. Structurally, Noah-MP has a single canopy layer, three snow layers, and four soil layers (having thicknesses of 0.1, 0.3, 0.6, and 1.0 m from the surface to the bottom of the layer). To compute changes in the vertical soil moisture distribution, Noah-MP solves the one-dimensional Richards' equation using the Clapp-Hornberger water retention relationship (Clapp

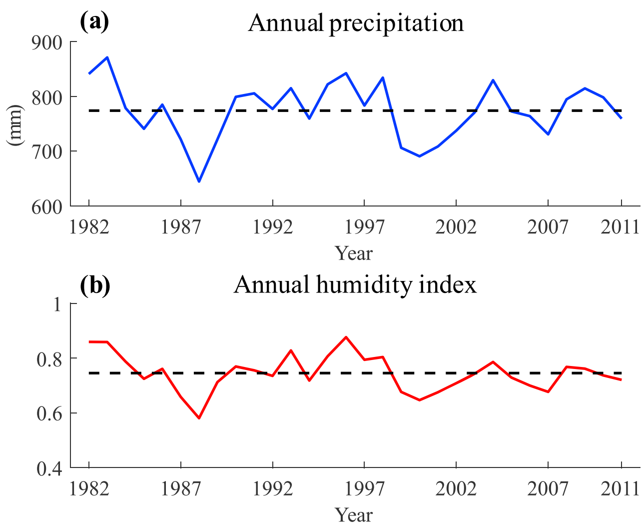

Figure 1. (a) Annual precipitation (mm) and (b) annual humidity index averaged over the continental United States. Humidity index is defined as the ratio of annual precipitation to annual potential evapotranspiration, which is calculated with the Penman (1948) approach, using the North American Land Data Assimilation System Phase 2 forcing data and the Noah-MP simulated R_n. Wet/dry years are those with precipitation more/less than mean precipitation (dotted line in (a)) and humidity index more/less than mean humidity index (dotted line in (b)), and medium years are those with values close to the mean.

& Hornberger, 1978). Noah-MP represents an unconfined aquifer as a reservoir or bucket underlying the soil column to account for water exchanges between the soil and the bucket through gravity and capillary forces (Niu et al., 2007). For the work reported here, we used the same model version as the experiment EXP6 in Yang et al. (2011).

To run the Noah-MP model, the inputs include atmospheric forcing data and static geography data. The atmospheric forcing includes air temperature, surface pressure, wind speed, specific humidity, precipitation, downward solar radiation, and downward longwave radiation. The static geography data include the geographical location of each grid and soil category, land use category, and terrain height (to derive the topographic index) for that grid. The global 1-km hybrid State Soil Geographic Database and the United States Geological Survey 24-category vegetation data (http://www.ral.ucar.edu/research/land/technology/lsm.php) are used to determine nominal values for the soil and vegetation parameters, respectively. Both of these 1-km resolution data sets are aggregated to 0.125° with the dominant soil and vegetation types to match the spatial resolution of the North American Land Data Assimilation System Phase 2 atmospheric forcing.

2.2. Forcing Data and Evaluation Data for Model-Simulated GPP and FLH

To drive Noah-MP, we used the North American Land Data Assimilation System Phase 2 hourly, 0.125° × 0.125° atmospheric forcing fields of air temperature, specific humidity, wind speed, surface pressure, downward shortwave radiation, downward longwave radiation, and precipitation (https://ldas.gsfc.nasa.gov/nldas/NLDAS2forcing_download.php). To evaluate the model-simulated GPP and LE fields, we used the FLUXNET model tree ensembles GPP and LE data from Max Planck Institute for Biogeochemistry (https://www.bgc-jena.mpg.de/geodb/projects/Home.php). These global data sets have been widely employed to evaluate land surface models-simulated (Anav et al., 2015; Bonan et al., 2011; Xia et al., 2016) and satellite-derived (Frankenberg et al., 2011) land-atmosphere carbon, water, and energy exchanges. The GPP and LE evaluation data over the continental United States are considered to be of high quality because most of the FLUXNET sites over the "data-rich" continental United States were incorporated in this product (Jung et al., 2010). To compute the difference between model simulations and evaluation data, we used the nearest-neighbor interpolation to resample the monthly 0.5° × 0.5° gridded GPP and LE over the continental United States to 0.125° × 0.125° grids (Ma et al., 2017).

Considering the limited nature of our computational resources, we select 6 years of forcing data from the available 1982–2011 data period. To ensure that the selected years are representative, two successive wet years (1982 and 1983), two successive dry years (1987 and 1988), and two successive medium years (2002 and 2003) were selected by referring to annual precipitation and to the annual humidity index computed for 1982 to 2011 (Figure 1). The first year of each 2-year period was used to "spin up" the model, while the second year was used to study the parameter importance. Correspondingly, the GPP and LE data used for the study correspond to 1983, 1988, and 2003.

3. Spatial and Parameter Sampling Design

To minimize the computational costs associated with analysis of the sensitivity of model response to parameter specification, our strategy is to investigate the possibility of obtaining robust GSA results while running the model on only a relatively small number of (sampled) grid locations instead of on the entire relevant spatial domain with a relatively small number of parameter combinations. The flowchart in Figure 2 illustrates our strategy to achieve this objective, explained as follows:

1. Determine the Noah-MP model parameters to be investigated and select the model state/output to be considered as the response of interest;

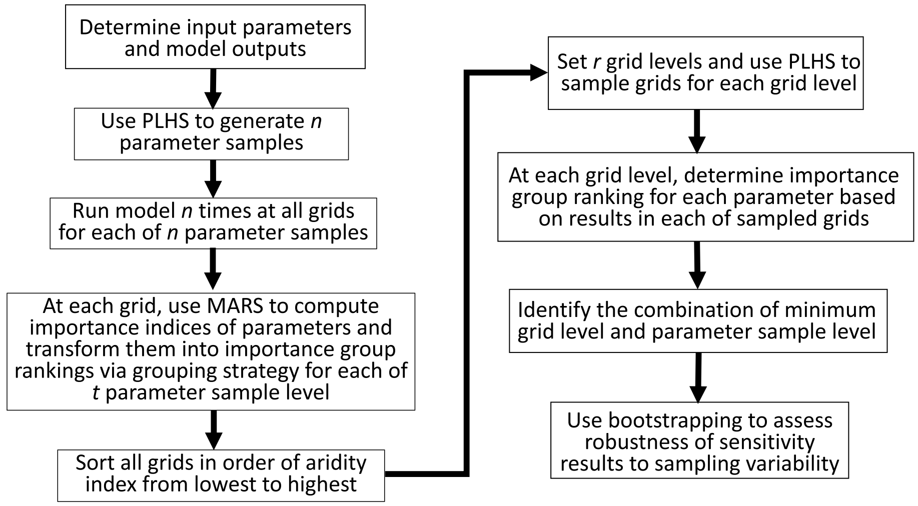

Figure 2. Flowchart of sensitivity analysis on spatial and parameter sampling. PLHS = progressive Latin hypercube sampling; MARS = multivariate adaptive regression splines.

2. Use progressive Latin hypercube sampling (PLHS; Sheikholeslami & Razavi, 2017) to generate *n* parameter samples from the feasible parameter space;
3. Run the model at *all* grids for the vegetation types selected, for each of the *n* parameter samples; this results in *n* sets of model outputs at each grid location;
4. At each grid location, set *t* parameter sample levels, each corresponding to a different number of parameter samples to include in the analysis. For each level, calculate the importance *indices* associated with each parameter via use of multivariate adaptive regression splines (MARS; Friedman, 1991), then transform the importance indices into importance group *rankings* via a grouping strategy;
5. For each vegetation type, sort all of the corresponding grids in order of humidity index from the lowest to highest;
6. Set *r* grid levels, each corresponding to a different number of grids to include in the analysis, and use PLHS randomly sample an appropriate set of grid locations from the sorted grids for each grid level;
7. At each grid level and for each parameter, select the importance group ranking to be the one that appears *most frequently* in the sampled grids;
8. Identify the combination of minimum grid level and parameter sample level for which the parameter importance group rankings stabilize; and
9. Assess robustness of the sensitivity results to sampling variability at both the grid level and parameter level, via bootstrapping.

Table 1
Parameters Investigated and the Feasible Range

Identifier	Parameter name	Physical meaning	Unit	Range
1	smcwlt	Wilting point of soil moisture	—	[0.5, 1.2]
2	smcmax	Porosity, saturated soil moisture	—	[0.83, 2.0]
3	bexp	Clapp and Hornberger “b” parameter	—	[0.1, 2.5]
4	dksat	Saturated soil hydraulic conductivity	m/s	[0.01, 1000]
5	dwsat	Saturated soil hydraulic diffusivity	m/s	[0.01, 1000]
6	sla	Single-side leaf area per kilogram	m^2/kg	[0, 300]
7	fragr	Fraction of carbon into growth respiration	—	[0, 0.8]
8	ltovrc	Leaf turnover	s^{-1}	[0, 10]
9	vcmx25	Maximum rate of carboxylation at 25 °C	μmol $CO_2/(m^2/s)$	[0, 200]
10	rmf25	Leaf maintenance respiration at 25 °C	μmol $CO_2/(m^2/s)$	[0, 20]
11	tdlef	Characteristic temperature for leaf freezing	K	[263, 275]
12	dilefc	Coefficient for leaf stress death	s^{-1}	[0, 5]

Table 2
Sixteen Vegetation Types in U.S. Geological Survey 24-Category Vegetation Over the Continental United States

Identifier	Name of vegetation type
1	Urban and built-up land
2	Dryland cropland and pasture
3	Irrigated cropland and pasture
5	Cropland/grassland mosaic
6	Cropland/woodland mosaic
7	Grassland
8	Shrubland
9	Mixed shrubland/grassland
10	Savanna
11	Deciduous broadleaf forest
14	Evergreen coniferous forest
15	Mixed forest
16	Water bodies
18	Wooded wetland
19	Barren or sparsely vegetated
21	Wooded tundra

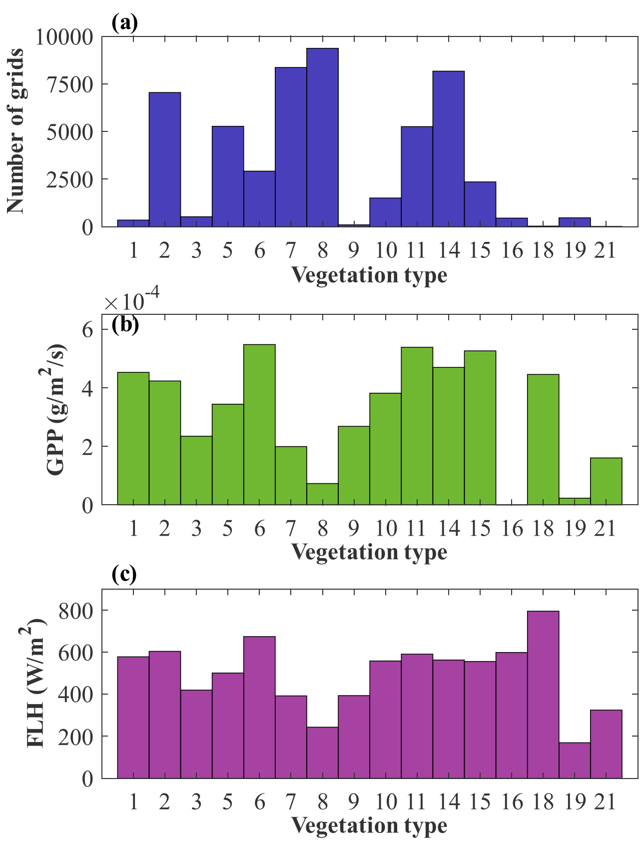

Figure 3. (a) Grid numbers of different vegetation types over the continental United States. (b) Annual average GPP (g/m^2/s) and (c) annual average FLH (W/m^2) per grid of 16 vegetation types from 1982 to 2011 computed by using evaluation data. GPP = gross primary production; FLH = flux of latent heat.

In Step 1, we selected 12 of the Noah-MP model parameters to investigate, in consultation with model experts as well as by referring to previous GSA studies related to the NOAH model, the original model on which NOAH-MP is based (Hogue et al., 2006; Rosero et al., 2010). Feasible ranges for these parameters (as suggested by model experts and based on the aforementioned papers) are indicated in Table 1. Note that the listed ranges for Parameters 1 to 5 (associated with soil type) actually refer to ranges of *multipliers* to be applied to nominal values that have been specified for these parameters; accordingly, the parameter value used in the model is obtained by multiplying the default value by the multiplier. Note also that the physical relationship requiring that SMCWLT < SMCMAX when specifying values of the first two parameters is preserved in the parameter specification process (i.e., combinations that violate this constraint are not permitted). Parameters 6 to 12 are associated with vegetation type.

The model output, selected in Step 1 to characterize changes in the behavior of the model as the parameters are perturbed, was represented by the root mean squared error between the simulations and corresponding evaluations:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (y_{\mathrm{sim}} - y_{\mathrm{eval}})^2}{N}}, \tag{1}$$

where y_{sim} is simulations, y_{eval} is evaluations, and N is the total number of values. *RMSE* is computed using simulated and evaluated values at the monthly time scale. Hereafter, *RMSE* is called the performance metric.

In Step 3, as examples for which the parameter GSA study was conducted, we chose two vegetation types—grassland and deciduous broadleaf forest—from the 16 vegetation types in the U.S. Geological Survey 24-category system (see Table 2) across the continental United States. Grassland is one of the primary vegetation types with the second largest number of grids (Figure 3a), spanning much of the transitional and arid climatological regions of the Midwest (Figure 4). Deciduous broadleaf forest, mainly distributed across the humid eastern United States, is a primary vegetation type contributing to the generation of GPP, as indicated by its large annual average GPP (Figure 3b). Further, its associated latent heat flux is very large, having an annual average value of 598.27 W/m^2 (Figure 3c).

4. Methodology

The overall procedure used in the paper includes the method of PLHS, the method of MARS, the grouping method, and the bootstrapping approach. First, PLHS was used to select representative parameter values and grid samples from their respective feasible spaces. Then MARS was used to compute the parameter sensitivity indices (precise importance rankings). Because, practically, we are more interested in the degree of importance (such as high, medium, or low importance) of a parameter rather than its precise ranking, a sensitivity index-based grouping method (Sheikholeslami et al., 2019) is implemented to group parameters into different importance groups. Finally, the bootstrapping approach was used to evaluate the robustness of model sensitivity analysis results to sampling variability. The methods are briefly described below.

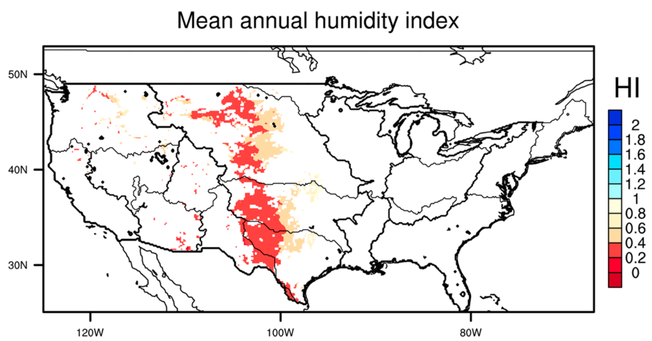

Figure 4. Geological distribution of mean annual HI of grassland from 1982 to 2011. Arid region is defined with value of mean annual HI less than 0.6, transitional region with value ranging from 0.6 to 1.2, and humid region with value ranging from 1.2 to 2. HI = humidity index.

The methods of Latin hypercube sampling and PLHS are described in detail by Sheikholeslami and Razavi (2017). The progressive Latin hypercube sample consists of T Latin hypercube sample slices whose progressive combination is a quasi-Latin hypercube sample. We implemented the strategies using the Matlab codes embedded in the VARS-Tool software package (Sheikholeslami et al., 2019).

MARS is an efficient and robust qualitative sensitivity analysis method (Gan et al., 2014; Li et al., 2013; Shahsavani et al., 2010) that uses a generalized cross-validation criterion (see equation (30) in Friedman, 1991) to measure the importance of a variable. It first constructs an "overfitted" model consisting of all the variables and then prunes the model by removing variables one at a time. The variable for which the largest increase in generalized cross validation is obtained when pruned from the overfitted model is deemed to be the most important and so on (Steinberg et al., 1999). We used the PSUADE software (Tong, 2005) to implement MARS.

The grouping method consists of a Box-Cox transformation (Box & Cox, 1964) and agglomerative hierarchical clustering (Johnson, 1967). The Box-Cox transformation, implemented using the procedure discussed in Sakia (1992), is performed to normalize the distributions of the sensitivity indices, and agglomerative hierarchical clustering is then used to group the parameters into different importance groups based on similarities between the normalized sensitivity indices. Similarity is quantified by using Ward's (1963) method.

Statistical uncertainty in the sensitivity results arising due to sampling variability can be estimated via the bootstrap method (Efron, 1979). Based on bootstrap samples via resampling with replacement, a robustness measure for the sensitivity results can be calculated via the procedure described in Razavi and Gupta (2015b). The value of the robustness measure should be in the range of (0,1].

5. Results and Discussions

5.1. Sensitivity Results Based on Grid and Parameter Sampling

In general, a randomly selected set of parameter samples with a size of approximately 10 to 20 times of the number of parameters under investigation is typically sufficient to identify the most important parameters in a model (Chapman et al., 1994; Confalonieri et al., 2010; Gan et al., 2014; Jones et al., 1998; Li et al., 2013; Loeppky et al., 2009; Sahama & Diamond, 2001). Here, due to computational considerations, the maximum number of parameter samples used to perform the sensitivity analysis was set to 240, which is 20 times the number of parameters in our investigation. These 240 parameter samples were selected using PLHS, with the size of the subsample set to 60. The Noah-MP model was then run for each sampled set of parameter values (i.e., a total of 240 times) at each grid belonging to the two selected vegetation types (grassland and deciduous broadleaf forest); note that grassland and deciduous broadleaf forest are the dominant vegetation type at 8,362 and 5,257 grids, respectively. At each grid location, we averaged the hourly model-simulated GPP and FLH for each month and computed the performance metric value for each variable using equation (1). The MARS methodology was then used to compute the sensitivity indices for each parameter, after which each parameter was ranked into one of four groups corresponding to high importance (Group 1), medium importance (Group 2), low importance (Group 3), and unimportant (Group 4). This procedure provided an importance group ranking for each parameter at each grid.

Next, we selected four levels of parameter sampling, corresponding to 60, 120, 180, and 240 samples. At each grid, we determined the importance group ranking of each parameter for each of the four parameter sampling levels. Note that these parameter samples were formed by progressive union of the size-60 subsamples used in PLHS. Similarly, we selected four levels of grid sampling, corresponding to 5%, 10%, 20%, and 100% of all grids, with the size of subsample set to 5% for each vegetation type. At each parameter level, the parameter importance group rankings were calculated for each of the four grid levels for each vegetation type. For each combination of grid and parameter sampling level, we obtained the importance group ranking for each parameter (this can be 1, 2, 3, or 4) in two steps: First, we counted the number of grids having

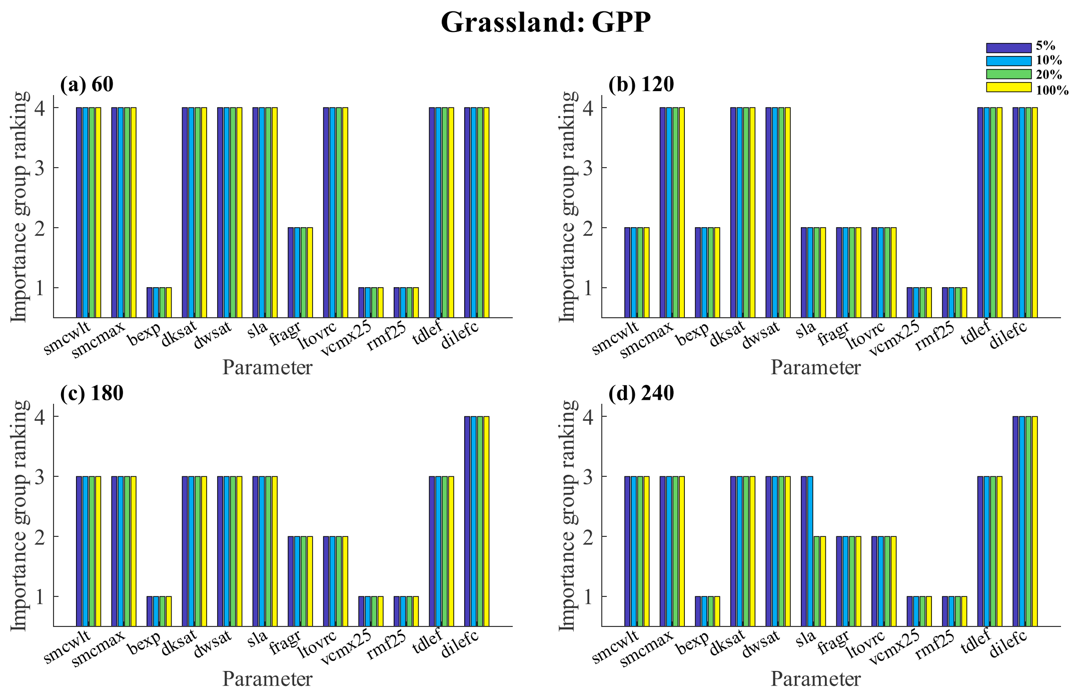

Figure 5. Importance group rankings of parameters for GPP in grassland at four grid levels consisting of 5%, 10%, 20%, and 100% of grids with (a) 60, (b) 120, (c) 180, and (d) 240 parameter samples. GPP = gross primary production.

the same group ranking and then selected the parameter group ranking to be that for which the number of grids is the largest. The results, for each vegetation type, are displayed in Figures 5–8.

From Figure 5a, we see that the parameter importance group ranking for GPP in grassland does not change significantly with grid level, indicating that (for a given number of parameter samples) the ranking results

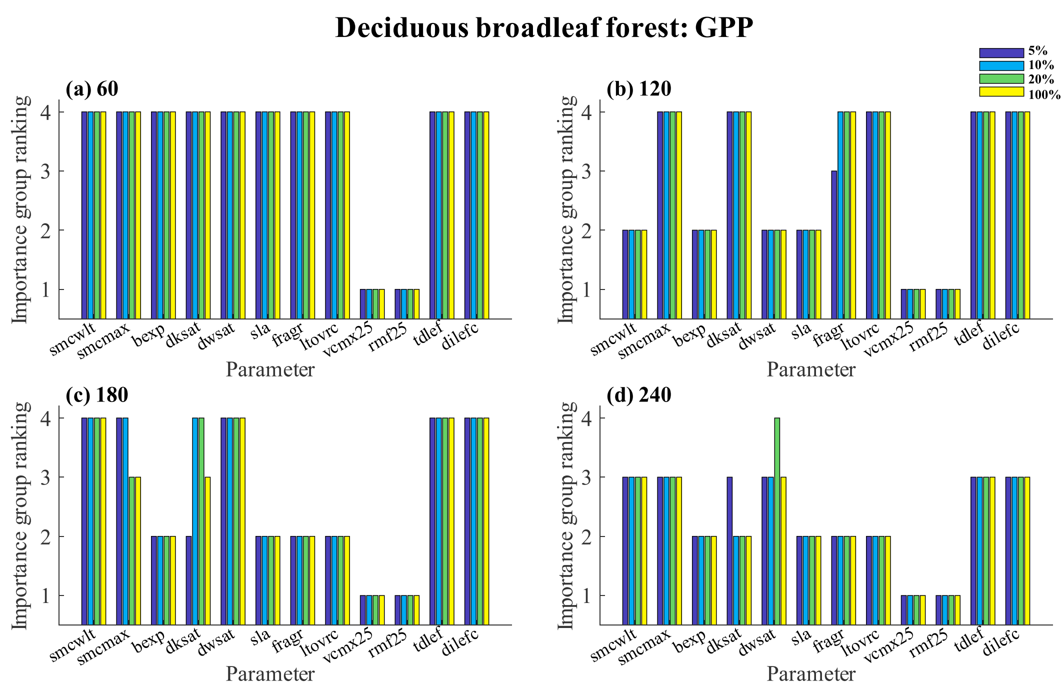

Figure 6. Importance group rankings of parameters for GPP in deciduous broadleaf forest at four grid levels consisting of 5%, 10%, 20%, and 100% of grids with (a) 60, (b) 120, (c) 180, and (d) 240 parameter samples. GPP = gross primary production.

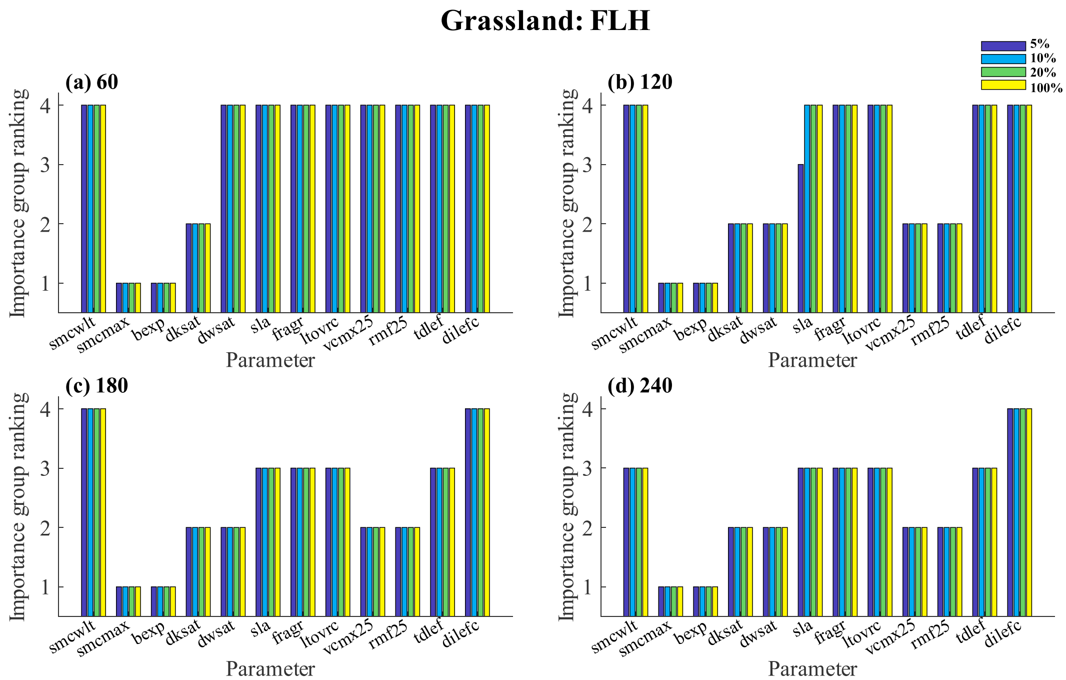

Figure 7. Importance group rankings of parameters for FLH in grassland at four grid levels consisting of 5%, 10%, 20%, and 100% of grids with (a) 60, (b) 120, (c) 180, and (d) 240 parameter samples. FLH = flux of latent heat.

are not strongly affected by the density level of grid sampling. The results for 120 and 180 parameter samples are identical. Further, while the group rankings for parameter *sla* at the 5% and 10% grid levels are different from that at the 20% grid level and from the most densely sampled case of all grids with 240 parameter samples (Figure 5d), the results of parameter importance group rankings are quite similar across all four grid

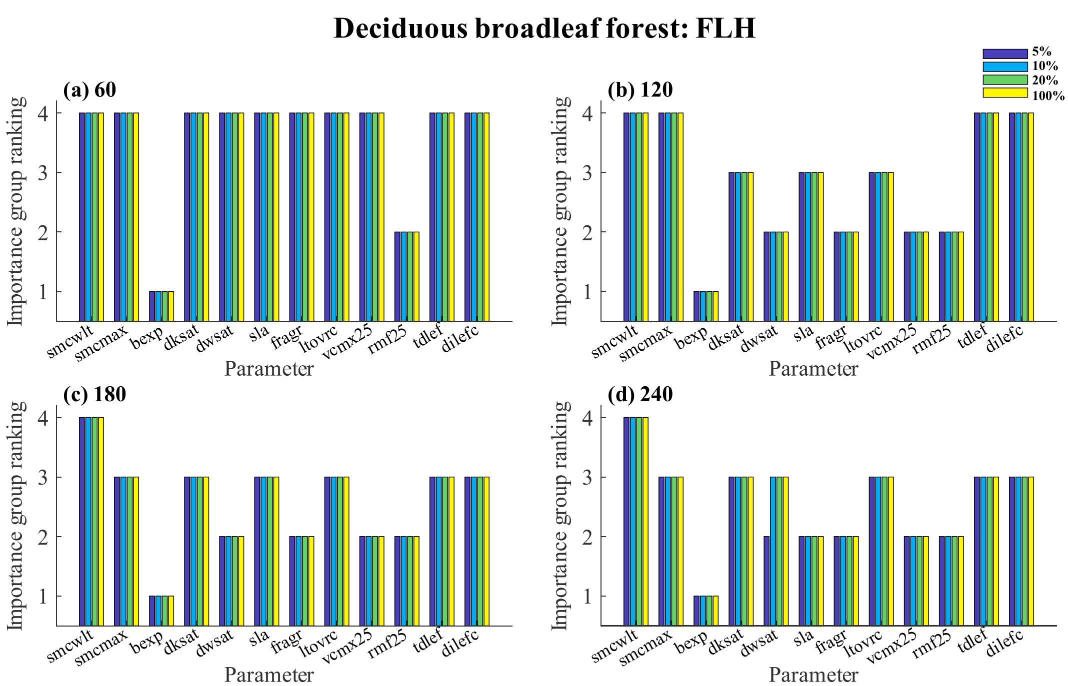

Figure 8. Importance group rankings of parameters for FLH in deciduous broadleaf forest at four grid levels consisting of 5%, 10%, 20%, and 100% of grids with (a) 60, (b) 120, (c) 180, and (d) 240 parameter samples. FLH = flux of latent heat.

Table 3
High Importance, Medium Importance, Low Importance, and Unimportant Parameters for GPP and FLH at Each Grid and Parameter Sample Level for Grassland

Percentage of grid samples	GPP				FLH			
	Number of parameter samples				Number of parameter samples			
	60	120	180	240	60	120	180	240
	High importance				High importance			
5%	3, 9, 10	9, 10	3, 9, 10	3, 9, 10	2, 3	2, 3	2, 3	2, 3
10%	3, 9, 10	9, 10	3, 9, 10	3, 9, 10	2, 3	2, 3	2, 3	2, 3
20%	3, 9, 10	9, 10	3, 9, 10	3, 9, 10	2, 3	2, 3	2, 3	2, 3
100%	3, 9, 10	9, 10	3, 9, 10	3, 9, 10	2, 3	2, 3	2, 3	2, 3
	Medium importance				Medium importance			
5%	7	1, 3, 6, 7, 8	7, 8	7, 8	4	4, 5, 9, 10	4, 5, 9, 10	4, 5, 9, 10
10%	7	1, 3, 6, 7, 8	7, 8	7, 8	4	4, 5, 9, 10	4, 5, 9, 10	4, 5, 9, 10
20%	7	1, 3, 6, 7, 8	7, 8	6, 7, 8	4	4, 5, 9, 10	4, 5, 9, 10	4, 5, 9, 10
100%	7	1, 3, 6, 7, 8	7, 8	6, 7, 8	4	4, 5, 9, 10	4, 5, 9, 10	4, 5, 9, 10
	Low importance				Low importance			
5%	—	—	1, 2, 4, 5, 6, 11	1, 2, 4, 5, 6, 11	—	6	6, 7, 8, 11	1, 6, 7, 8, 11
10%	—	—	1, 2, 4, 5, 6, 11	1, 2, 4, 5, 6, 11	—	—	6, 7, 8, 11	1, 6, 7, 8, 11
20%	—	—	1, 2, 4, 5, 6, 11	1, 2, 4, 5, 11	—	—	6, 7, 8, 11	1, 6, 7, 8, 11
100%	—	—	1, 2, 4, 5, 6, 11	1, 2, 4, 5, 11	—	—	6, 7, 8, 11	1, 6, 7, 8, 11
	Unimportant				Unimportant			
5%	1, 2, 4, 5, 6, 8, 11, 12	2, 4, 5, 11, 12	12	12	1, 5, 6, 7, 8, 9, 10, 11, 12	1, 7, 8, 11, 12	1, 12	12
10%	1, 2, 4, 5, 6, 8, 11, 12	2, 4, 5, 11, 12	12	12	1, 5, 6, 7, 8, 9, 10, 11, 12	1, 6, 7, 8, 11, 12	1, 12	12
20%	1, 2, 4, 5, 6, 8, 11, 12	2, 4, 5, 11, 12	12	12	1, 5, 6, 7, 8, 9, 10, 11, 12	1, 6, 7, 8, 11, 12	1, 12	12
100%	1, 2, 4, 5, 6, 8, 11, 12	2, 4, 5, 11, 12	12	12	1, 5, 6, 7, 8, 9, 10, 11, 12	1, 6, 7, 8, 11, 12	1, 12	12

Note. Parameter is represented by its identifier listed in Table 1. GPP = gross primary production; FLH = flux of latent heat.

levels. In contrast, the parameter importance group rankings do vary significantly as the number of parameter samples increases from 60 to 180 (Figures 5a to 5c).

Overall, these results indicate that using a subset of the total number of grids to perform the parameter sensitivity analysis is justified and that computational effort can therefore be redirected toward obtaining larger numbers of parameter samples; that is, the additional information provided by running the model at larger numbers of grids does not significantly affect the parameter importance group ranking result. This finding also applies for the parameter importance group rankings obtained for FLH in grassland (Figure 7) and for GPP (Figure 6) and FLH (Figure 8) in deciduous broadleaf forest.

Next, we note that while the importance group rankings for most of the parameters do vary with the number of parameter samples (see Figures 6–8), the classification of "high importance" parameters (those in Group 1) does not change with increasing numbers of parameter samples. The result implies that only 60 samples (i.e., 5 times the number of parameters investigated) is sufficient to identify the most important model parameters, providing results identical to when 240 parameter samples are used. The finding also applies for the parameters affecting GPP in grassland (Figure 5).

A summary of the high importance, medium importance, low importance, and unimportant parameters for GPP and FLH, at each grid and parameter sample level, for the two vegetation types (grassland and deciduous broadleaf forest) is provided in Tables 3 and 4. These tables show that the set of parameters classified as being of high importance remains stable as discussed above. However, the set of parameters classified as being unimportant is *not* stable, and the number of unimportant parameters declines as parameter sampling density increases; that is, the "unimportant" parameters shift into groups associated with higher importance as the number of parameter samples increases. This result differs from that reported by Sheikholeslami et al. (2019), where the least important parameters were found to be highly robust to the increasing number of parameter samples.

Table 5 lists the combination of grid and parameter level that seems to be sufficiently informative to identify the parameters to which GPP and FLH in grassland and deciduous broadleaf forest are most sensitive. It suggests that use of a 5% grid density and 5 times the number of parameters investigated can be sufficient. This results in a large decrease in the computational burden required to study the model performance.

Table 4
High Importance, Medium Importance, Low Importance, and Unimportant Parameters for GPP and FLH at Each Grid and Parameter Sample Level for Deciduous Broadleaf Forest

Percentage of grid samples	GPP				FLH			
	Number of parameter samples				Number of parameter samples			
	60	120	180	240	60	120	180	240
	High importance				High importance			
5%	9, 10	9, 10	9, 10	9, 10	3	3	3	3
10%	9, 10	9, 10	9, 10	9, 10	3	3	3	3
20%	9, 10	9, 10	9, 10	9, 10	3	3	3	3
100%	9, 10	9, 10	9, 10	9, 10	3	3	3	3
	Medium importance				Medium importance			
5%	—	1, 3, 5, 6	3, 4, 6, 7, 8	3, 6, 7, 8	10	5, 7, 9, 10	5, 7, 9, 10	5, 6, 7, 9, 10
10%	—	1, 3, 5, 6	3, 6, 7, 8	3, 4, 6, 7, 8	10	5, 7, 9, 10	5, 7, 9, 10	6, 7, 9, 10
20%	—	1, 3, 5, 6	3, 6, 7, 8	3, 4, 6, 7, 8	10	5, 7, 9, 10	5, 7, 9, 10	6, 7, 9, 10
100%	—	1, 3, 5, 6	3, 6, 7, 8	3, 4, 6, 7, 8	10	5, 7, 9, 10	5, 7, 9, 10	6, 7, 9, 10
	Low importance				Low importance			
5%	—	7	—	1, 2, 4, 5, 11, 12	—	4, 6, 8	2, 4, 6, 8, 11, 12	2, 4, 8, 11, 12
10%	—	—	—	1, 2, 5, 11, 12	—	4, 6, 8	2, 4, 6, 8, 11, 12	2, 4, 5, 8, 11, 12
20%	—	—	2	1, 2, 11, 12	—	4, 6, 8	2, 4, 6, 8, 11, 12	2, 4, 5, 8, 11, 12
100%	—	—	2, 4	1, 2, 5, 11, 12	—	4, 6, 8	2, 4, 6, 8, 11, 12	2, 4, 5, 8, 11, 12
	Unimportant				Unimportant			
5%	1, 2, 3, 4, 5, 6, 7, 8, 11, 12	2, 4, 8, 11, 12	1, 2, 5, 11, 12		1, 2, 4, 5, 6, 7, 8, 9, 11, 12	1, 2, 11, 12	1	1
10%	1, 2, 3, 4, 5, 6, 7, 8, 11, 12	2, 4, 7, 8, 11, 12	1, 2, 4, 5, 11, 12		1, 2, 4, 5, 6, 7, 8, 9, 11, 12	1, 2, 11, 12	1	1
20%	1, 2, 3, 4, 5, 6, 7, 8, 11, 12	2, 4, 7, 8, 11, 12	1, 4, 5, 11, 12	5	1, 2, 4, 5, 6, 7, 8, 9, 11, 12	1, 2, 11, 12	1	1
100%	1, 2, 3, 4, 5, 6, 7, 8, 11, 12	2, 4, 7, 8, 11, 12	1, 5, 11, 12		1, 2, 4, 5, 6, 7, 8, 9, 11, 12	1, 2, 11, 12	1	1

Note. Parameter is represented by its identifier listed in Table 1. GPP = gross primary production; FLH = flux of latent heat.

5.2. High Importance Parameters

The groups of parameters that GPP and FLH in grassland and deciduous broadleaf forest are most sensitive to are listed in Table 6; these include the soil parameters *smcmax* and *bexp* and the vegetation parameters *vcmx25* and *rmf25*.

We see that the vegetation parameters *vcmx25* and *rmf25* strongly affect GPP of both vegetation types. The result can be explained by the fact that *vcmx25* and *rmf25* are the rates of maximum carboxylation and leaf maintenance respiration at optimal temperatures (25 °C). They control the leaf carbon budgets and hence leaf area index that feeds back to affect the actual rate of photosynthesis. The actual rate of photosynthesis impacts GPP directly. For GPP, the soil parameter *bexp* is a high importance parameter for grassland but not for deciduous broadleaf forest. In general, due to the shading effects of the vegetation canopy, the modeled GPP is less sensitive to soil parameters for a denser canopy, like the deciduous broadleaf forest.

For FLH in both grassland and deciduous broadleaf forest, the soil parameter *bexp* is the high importance parameter. This is because FLH includes evaporation from the soil surface, which is sensitive to soil moisture conditions, in turn controlled by the soil parameters. Compared with the parameter *smcmax* (soil porosity), the power *bexp* in the power function plays a greater role in controlling soil water retention and hydraulic

Table 5
The Combination of Grid and Parameter Level to Identify High Importance Parameters

Vegetation type	GPP	FLH
Grassland	(5% grids, 60 samples)	(5% grids, 60 samples)
Deciduous broadleaf forest	(5% grids, 60 samples)	(5% grids, 60 samples)

Note. GPP = gross primary production; FLH = flux of latent heat.

Table 6
High Importance Parameters Affecting GPP and FLH for the Two Vegetation Types

Vegetation type	GPP	FLH
Grassland	smcmax, vcmx25, rmf25	smcmax, bexp
Deciduous broadleaf forest	vcmx25, rmf25	bexp

Note. GPP = gross primary production; FLH = flux of latent heat.

conductivity and thus the soil moisture conditions. Note that the soil parameter *smcmax* strongly affects FLH only in grassland. This is because the grasslands located in the relative dry and dry-to-wet transitional zones of the Midwest are more water stressed, and thus, the modeled FLH shows sensitivity to *smcmax* (soil porosity), while almost all of deciduous broadleaf forests are located in the humid regions of the East where soil moisture is abundant and not as sensitive to *smcmax*.

5.3. Robustness of Results to Grid and Parameter Sampling

Due to sampling variability associated with selection of the grids and parameter values, some degree of uncertainty in the result of parameter importance group is unavoidable. Therefore, we computed a measure of robustness of our results—particularly related to identification of the high importance parameters—for the case when our result is based on 5% grid sampling and 60 parameter samples. For grid sampling, we estimated robustness of the importance group ranking of each parameter by computing the number of grid bootstrap replicates that generates the same parameter importance group ranking as that obtained using the entire set of locations and dividing the number by the total number of bootstrap replicates (set to 80). For parameter sampling, we first compute a measure of robustness of parameter importance group ranking at each grid (using 300 bootstrap replicates) and then average over the grids having the same importance group ranking. Each bootstrap replicate generates a sample of the parameter importance group ranking, so for each grid for each parameter, we compute the fraction of bootstrap replicates that generate the same parameter importance group ranking as that obtained using the total set of 60 PLHS parameter samples. We then compute the robustness measure as the average over all having the same parameter importance group ranking as that obtained with the total number of grids.

After the robustness measures (of parameter importance group ranking for GPP and FLH in grassland and deciduous broadleaf forest) were computed, we focus on high importance parameter group (i.e., Group 1). With respect to grid sampling, the high importance parameters for GPP (*bexp*, *vcmx25*, and *rmf25* in grassland and *vcmx25* and *rmf25* in deciduous broadleaf forest) and for FLH (*smcmax* and *bexp* in grassland and *bexp* in deciduous broadleaf forest) have robustness values of 1.0, meaning that all of the parameters classified into this group are fully (i.e., 100%) robust to grid sampling variability. The parameter importance group ranking robustness results for parameter sampling are shown in Figure 9. From Figure 9, we can see that the

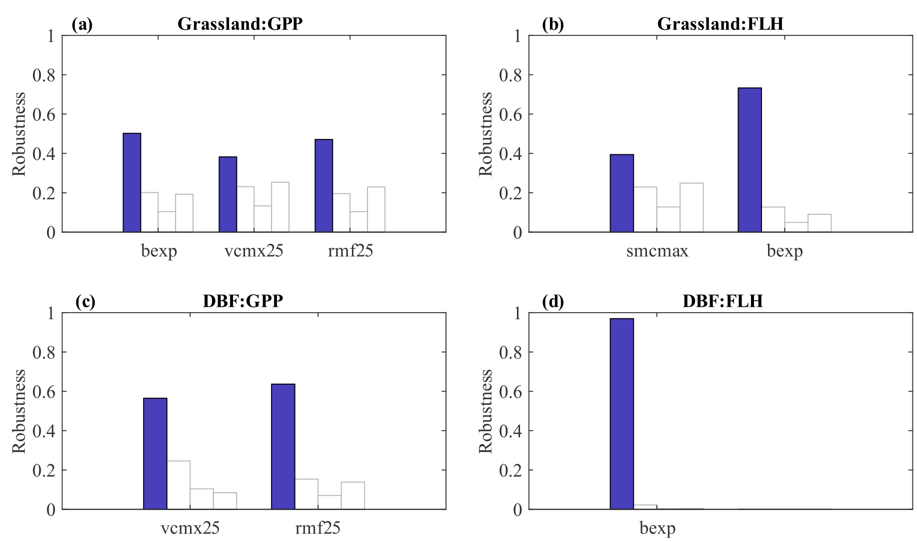

Figure 9. Robustness of importance group rankings of parameters for grassland GPP (a), grassland FLH (b), DBF (c), and DBF FLH (d) to parameter sampling variability. In each subplot, four bars located at each parameter from left to right correspond to the robustness of the parameter classified into Importance Groups 1, 2, 3, and 4, respectively. The blue bars in each subplot correspond to the robustness of the most important parameters identified. GPP = gross primary production; FLH = flux of latent heat; DBF = deciduous broadleaf forest.

robustness value of the high importance parameters for GPP and for FLH classified as Group 1 (blue bar in four subplots) is higher than these parameters classified as any other group. The result indicates that the high importance parameters for GPP and for FLH identified are robust to parameter variability.

6. Conclusion

Dynamical models of spatially distributed Earth and environmental systems are typically highly parameterized with many uncertain parameters and are usually very computationally expensive. The process of identifying which parameters the model behaviors are most sensitive to can require very large numbers of model runs. By running the model at a representative set of grid locations, a considerable amount of computational expense can be saved. Here we have explored the possibility of being able to identify the high importance parameters of such models by running the model at a relatively small fraction of locations sampled throughout the entire model domain.

Accordingly, we ran experiments for an implementation of the Noah-MP model across the continental United States, using four grid sampling levels (5%, 10%, 20%, and 100%) and four parameter sampling levels (60, 120, 180, and 240 parameter samples). Before sampling, all of the grids associated with a particular vegetation type were sorted in order of decreasing annual humidity index, with a view to ensuring representation of their climatological distributions. A maximum number of 240 parameter samples consisting of four subset samples was generated using PLHS. We focused our study on two vegetation types—grassland and deciduous broadleaf forest—and as a benchmark, the model was run on all of the corresponding grids, using the largest feasible number of parameter samples (240). We then used the MARS method, combined with a recently developed parameter importance grouping strategy, to classify the model parameters into four importance groups, and assessed the stability of the results obtained using smaller numbers of grid and parameter samples.

Computational savings were achieved in four ways: (a) by using forcing data for a representative set of wet, medium, and dry years; (b) by running the model at only a climatologically representative set of sampled grid locations; (c) by the use of PLHS to explore the sensitivity of the results to parameter sample size; and (d) by the use of grouping-based importance ranking (developed by Sheikholeslami et al., 2019) which requires far fewer parameter samples to provide relatively robust results because it does not require precise parameter ranking, only convergence of the parameter ranking results into an “importance” group.

This latter point is important because, from a practical view of point, we are often really only interested in determining whether a parameter belongs to a high importance group—for example, when targeting parameters for calibration via optimization. Our results indicated that grouping-based rankings of the Noah-MP parameters into a high importance group give stable results using only 5% of the model domain grids and a relatively small number of parameter samples (here 5 times the number of parameters). If identification of medium importance parameters is required, then a larger number of parameter samples will likely be necessary. At least for the two vegetation types studied here, the results seem robust to the strategy of selecting only a subsample of the total number of grids, thereby resulting in considerable computational savings.

Finally, an assessment of robustness of our results was carried out using bootstrapping. In our case, all of the high importance parameters were found to be fully robust to grid sampling variability, and most were found to be of medium to strong robustness to parameter sampling variability (the exception being *vcmx25* for GPP in grassland). Overall, the implication is that most of the information necessary for a robust analysis of model behavioral sensitivity to parameter specification lies in representativeness of the parameter space samples and that spatial grid-to-grid variability may not be as significant a factor. This means that we may be able to obtain suitable robust GSA results by focusing on higher parameter space sampling densities while saving computational costs by running the model at a limited number of carefully selected spatially distributed grid locations.

Note that the finding that 5% of the total grids and 5 times the number of parameters is sufficient to identify most important parameters, with strong robustness to grid sampling variability and good robustness to parameter sampling variability may be limited by two factors: (1) the way in which we determined the parameter importance group rankings for parameters at the grid and parameter levels and (2) the specific vegetation types investigated. Here we determined the importance group rankings of parameters in two steps: First, we computed the importance group rankings of the parameters at each sampled grid and then selected the group ranking for each parameter to be the one that appears most frequently in the sampled grids. If we choose another approach to determine the importance group, we might obtain different results. Second, our findings are based on results obtained for only the two main vegetation types across the continental United States, and the behavior of other vegetation types still needs to be explored. These issues will be investigated in future work.

Compared with sensitivity analysis, calibration of computationally intensive spatially distributed DESMs can require a computational burden that is typically infeasible. The approach proposed in this paper is to reduce the computational burden associated with parameter importance analysis, especially by selecting representative grids, is a potentially promising way to also reduce the computational time required for parameter optimization of such models. In ongoing work, we are investigating the applicability of the approach proposed in this paper to calibration of computationally intensive spatially distributed DESMs and will report on our findings in future work.

Acknowledgments

This research was supported by the Special Fund for Meteorological Scientific Research in Public Interest (GYHY201506002, CRA-40: the 40-year CMA global atmospheric reanalysis), the National Basic Research Program of China (2015CB953703), the State Key Laboratory of Earth Surface Processes and Resource Ecology (2017-KF-05), and the Fundamental Research Funds for the Central Universities-Beijing Normal University Research Fund (2015KJJCA04). Ms. Xueli Huo gratefully acknowledges the scholarship provided by China Scholarship Council Joint Graduate Program (201706040197), and the opportunity provided by the visiting research scholar program of the Department of Hydrology and Atmospheric Sciences (HAS), to visit and study at the University of Arizona. The second author acknowledges partial support by the Australian Centre of Excellence for Climate System Science (CE110001028). The PLHS algorithm and grouping algorithm were implemented using VARS-Tool software package, a MATLAB Toolbox developed by Sheikholeslami & Razavi, 2017, Sheikholeslami et al., 2019) and can be downloaded online (http://vars-tool.com/). The codes used to generate results and data used in the paper can be downloaded online (https://github.com/shirleyhuo/spatially-sampling-based-sensitivity-analysis-procedure).

References

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

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

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B: Methodological*, *26*, 211–252. https://www.jstor.org/stable/2984418

Campolongo, F., Cariboni, J., & Saltelli, A. (2007). An effective screening design for sensitivity analysis of large models. *Environmental Modelling & Software*, *22*, 1509–1518. https://doi.org/10.1016/j.envsoft.2006.10.004

Chapman, W. L., Welch, W. J., Bowman, K. P., Sacks, J., & Walsh, J. E. (1994). Arctic sea ice variability: Model sensitivities and a multidecadal simulation. *Journal of Geophysical Research*, *99*, 919–935. https://doi.org/10.1029/93JC02564

Chen, F., & Dudhia, J. (2001). Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 Modeling System. Part I: Model implementation and sensitivity. *Monthly Weather Review*, *129*, 569–585. https://doi.org/10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2

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

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

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. In *The Annals of Statistics* (Vol. 7, pp. 1–26). New York, NY: Springer. https://www.jstor.org/stable/2958830

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

Francos, A., Elorza, F. J., Bouraoui, F., Bidoglio, G., & Galbiati, L. (2003). Sensitivity analysis of distributed environmental simulation models: Understanding the model behaviour in hydrological studies at the catchment scale. *Reliability Engineering & System Safety*, *79*(2), 205–218. https://doi.org/10.1016/S0951-8320(02)00231-4

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

Friedman, J. H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics*, *19*(1), 1–67. https://doi.org/10.1214/aos/1176347963

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

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

Herman, J. D., Kollat, J. B., Reed, P. M., & Wagener, T. (2013). Technical note: Method of Morris effectively reduces the computational demands of global sensitivity analysis for distributed watershed models. *Hydrology and Earth System Sciences*, *17*, 2893–2903. https://doi.org/10.5194/hess-17-2893-2013

Hogue, T. S., Bastidas, L. A., Gupta, H. V., & Sorooshian, S. (2006). Evaluating model performance and parameter behavior for varying levels of land surface model complexity. *Water Resources Research*, *42*, W08430. https://doi.org/10.1029/2005WR004440

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, *32*(3), 241–254. https://doi.org/10.1007/BF02289588

Jones, D. R., Schonlau, M., & Welch, W. J. (1998). Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, *13*(4), 455–492. https://doi.org/10.1023/A:1008306431147

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

Lee, L. A., Carslaw, K. S., Pringle, K. J., & Mann, G. W. (2012). Mapping the uncertainty in global CCN using emulation. *Atmospheric Chemistry and Physics*, *12*, 9739–9751. https://doi.org/10.5194/acp-12-9739-2012

Li, J., Duan, Q. Y., Gong, W., Ye, A., Dai, Y., Miao, C., et al. (2013). Assessing parameter importance of the Common Land Model based on qualitative and quantitative sensitivity analysis. *Hydrology and Earth System Sciences*, *17*(8), 3279–3293. https://doi.org/10.5194/hess-17-3279-2013

Loeppky, J. L., Sacks, J., & Welch, W. J. (2009). Choosing the sample size of a computer experiment: A practical guide. *Technometrics*, *51*(4), 366–376. https://doi.org/10.1198/TECH.2009.08040
Ma, N., Niu, G.-Y., Xia, Y., Cai, X., Zhang, Y., Ma, Y., & Fang, Y. (2017). A systematic evaluation of Noah-MP in simulating land-atmosphere energy, water, and carbon exchanges over the continental United States. *Journal of Geophysical Research: Atmospheres*, *122*, 12,245–12,268. https://doi.org/10.1002/2017JD027597
Murphy, J. M., Sexton, D. M. H., Barnett, D. N., Jones, G. S., Webb, M. J., Collins, M., & Stainforth, D. A. (2004). Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, *430*(7001), 768–772. https://doi.org/10.1038/nature02771
Niu, G.-Y., Yang, Z.-L., Dickinson, R. E., Gulden, L. E., & Su, H. (2007). Development of a simple groundwater model for use in climate models and evaluation with Gravity Recovery and Climate Experiment data. *Journal of Geophysical Research*, *112*, D07103. https://doi.org/10.1029/2006JD007522
Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., et al. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research*, *116*, D12109. https://doi.org/10.1029/2010JD015139
Oakley, J. E., & O'Hagan, A. (2004). Probabilistic sensitivity analysis of complex models: A Bayesian approach. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, *66*(3), 751–769. https://www.jstor.org/stable/3647504, https://doi.org/10.1111/j.1467-9868.2004.05304.x
Penman, H. L. (1948). Natural evaporation from open water, bare soil and grass. *Proceedings of the Royal Society of London, Series A*, *193*(1032), 120–145. https://doi.org/10.1098/rspa.1948.0037
Ratto, M., Pagano, A., & Young, P. (2007). State dependent parameter metamodelling and sensitivity analysis. *Computer Physics Communications*, *177*, 863–876. https://doi.org/10.1016/j.cpc.2007.07.011
Razavi, S., & Gupta, H. V. (2015a). A new framework for comprehensive, robust, and efficient global sensitivity analysis: 1. Theory. *Water Resources Research*, *52*, 423–439. https://doi.org/10.1002/2015WR017558
Razavi, S., & Gupta, H. V. (2015b). A new framework for comprehensive, robust, and efficient global sensitivity analysis: 2. Application. *Water Resources Research*, *52*, 440–455. https://doi.org/10.1002/2015WR017559
Razavi, S., Tolson, B. A., & Burn, D. H. (2012). Review of surrogate modeling in water resources. *Water Resources Research*, *48*, W07401. https://doi.org/10.1029/2011WR011527
Rosero, E., Yang, Z.-L., Wagener, T., Gulden, L. E., Yatheendradas, S., & Niu, G.-Y. (2010). Quantifying parameter sensitivity, interaction, and transferability in hydrologically enhanced versions of the Noah land surface model over transition zones during the warm season. *Journal of Geophysical Research*, *115*, D03106. https://doi.org/10.1029/2009JD012035
Ryan, E., Wild, O., Voulgarakis, A., & Lee, L. (2018). Fast sensitivity analysis methods for computationally expensive models with multi-dimensional output. *Geoscientific Model Development*, *11*, 3131–3146. https://doi.org/10.5194/gmd-11-3131-2018
Sacks, J., Welch, W. J., Mitchell, T. J., & Wynn, H. P. (1989). Design and analysis of computer experiments. *Statistical Science*, *4*, 409–423. https://www.jstor.org/stable/2245858
Sahama, T. R., & Diamond, N. T. (2001). Sample size considerations and augmentation of computer experiments. *Journal of Statistical Computation and Simulation*, *68*, 307–319. https://doi.org/10.1080/00949650108812073
Sakia, R. M. (1992). The Box-Cox transformation technique: A review. *Journal of the Royal Statistical Society. Series D (The Statistician)*, *41*, 169–178. https://www.jstor.org/stable/2348250
Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., et al. (2008). *Global sensitivity analysis: The primer*. Chichester: John Wiley & Sons Ltd. https://doi.org/10.1002/9780470725184.fmatter
Saltelli, A., Tarantola, S., Campolongo, F., & Ratto, M. (2004). *Sensitivity analysis in practice: A guide to assessing scientific models*. Chichester: John Wiley & Sons Ltd. https://doi.org/10.1002/0470870958.fmatter
Saltelli, A., Tarantola, S., & Chan, K. P. S. (1999). A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics*, *41*, 39–56. https://www.jstor.org/stable/1270993
Shahsavani, D., Tarantola, S., & Ratto, M. (2010). Evaluation of MARS modeling technique for sensitivity analysis of model output. *Procedia - Social and Behavioral Sciences*, *2*, 7737–7738. https://doi.org/10.1016/j.sbspro.2010.05.204
Sheikholeslami, R., & Razavi, S. (2017). Progressive Latin hypercube sampling: An efficient approach for robust sampling-based analysis of environmental models. *Environmental Modelling & Software*, *93*, 109–126. https://doi.org/10.1016/j.envsoft.2017.03.010
Sheikholeslami, R., Razavi, S., Gupta, H. V., Becker, W., & Haghnegahdar, A. (2019). Global sensitivity analysis for high-dimensional problems: How to objectively group factors and measure robustness and convergence while reducing computational cost. *Environmental Modelling & Software*, *111*, 282–299. https://doi.org/10.1016/j.envsoft.2018.09.002
Shin, M.-J., Guillaume, J. H. A., Croke, B. F. W., & Jakeman, A. J. (2013). Addressing ten questions about conceptual rainfall–runoff models with global sensitivity analyses in R. *Journal of Hydrology*, *503*, 135–152. https://doi.org/10.1016/j.jhydrol.2013.08.047
Sobol', I. M. (1993). Sensitivity estimates for nonlinear mathematical models. *Mathematical Modeling and Computational Experiment*, *1*, 407–414.
Song, X., Zhan, C., Xia, J., & Kong, F. (2012). An efficient global sensitivity analysis approach for distributed hydrological model. *Journal of Geographical Sciences*, *22*, 209–222. https://doi.org/10.1007/s11442-012-0922-5
Steinberg, D., Colla, P. L., & Martin, K. (1999). *MARS user guide*. San Diego, CA: Salford Systems.
Tian, W. (2013). A review of sensitivity analysis methods in building energy analysis. *Renewable and Sustainable Energy Reviews*, *20*, 411–419. https://doi.org/10.1016/j.rser.2012.12.014
Tong, C. (2005). PSUADE User's Manual, Lawrence Livermore National Laboratory, LLNL-SM-407882. Livermore, CA.
Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*(301), 236–244. https://doi.org/10.1080/01621459.1963.10500845
Xia, J., Wang, G. S., Tan, G., Ye, A. Z., & Huang, G. H. (2005). Development of distributed time-variant gain model for nonlinear hydrological systems. *Science in China Series D: Earth Sciences*, *48*, 713–723.
Xia, Y., Cosgrove, B. A., Mitchell, K. E., Peters-Lidard, C. D., Ek, M. B., Brewer, M., et al. (2016). Basin-scale assessment of the land surface water budget in the National Centers for Environmental Prediction operational and research NLDAS-2 systems. *Journal of Geophysical Research: Atmospheres*, *121*, 2750–2779. https://doi.org/10.1002/2015JD023733
Yang, Z.-L., Niu, G.-Y., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., et al. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 2. Evaluation over global river basins. *Journal of Geophysical Research*, *116*, D12110. https://doi.org/10.1029/2010JD015140
Zhan, C.-S., Song, X.-M., Xia, J., & Tong, C. (2013). An efficient integrated approach for global sensitivity analysis of hydrological model parameters. *Environmental Modelling & Software*, *41*, 39–52. https://doi.org/10.1016/j.envsoft.2012.10.009