

# Comparing impacts of parameter and spatial data uncertainty for a grid-based distributed watershed model

Younggu Her and Conrad Heatwole

## ABSTRACT

Parameter uncertainty in hydrologic modeling is commonly evaluated, but assessing the impact of spatial input data uncertainty in spatially descriptive 'distributed' models is not common. This study compares the significance of uncertainty in spatial input data and model parameters on the output uncertainty of a distributed hydrology and sediment transport model, HYdrology Simulation using Time-Area method (HYSTAR). The Shuffled Complex Evolution Metropolis (SCEM-UA) algorithm was used to quantify parameter uncertainty of the model. Errors in elevation and land cover layers were simulated using the Sequential Gaussian/Indicator Simulation (SGS/SIS) techniques and then incorporated into the model to evaluate their impact on the outputs relative to those of the parameter uncertainty. This study demonstrated that parameter uncertainty had a greater impact on model output than did errors in the spatial input data. In addition, errors in elevation data had a greater impact on model output than did errors in land cover data. Thus, for the HYSTAR distributed hydrologic model, accuracy and reliability can be improved more effectively by refining parameters rather than further improving the accuracy of spatial input data and by emphasizing the topographic data over the land cover data.

**Key words** | distributed model, HYSTAR, parameter uncertainty, sequential Gaussian simulation, sequential indicator simulation, spatial data

**Younggu Her** (corresponding author)

**Conrad Heatwole**

Department of Agricultural and Biological  
Engineering & Tropical Research and Education  
Center,

University of Florida,  
Homestead,

FL 33031

USA

E-mail: [yher@ufl.edu](mailto:yher@ufl.edu)

## INTRODUCTION

Mathematical models simulate watershed (catchment) hydrology based on the assumptions, relationships, parameters, boundary conditions, and descriptive spatial data that define the physical characteristics of the watershed. The spatial features of the landscape are commonly represented by averaging conditions or parameters to a single 'lumped' value. A 'distributed' watershed model considers a spatially descriptive representation of the landscape and is commonly implemented as a grid structure with separate parameters and data values for each cell. For any model, input data and parameters contain uncertainty because they represent characteristics of a complex and dynamic system that the model depicts at specific temporal and spatial scales. Errors are inherent in the system description

when single deterministic values are used to describe a system that is highly nonlinear, heterogeneous, and stochastic. Thus, input data, model algorithms, equations, data used for calibration, and temporal and spatial scale are all contributing sources of model output error in watershed modeling (Shirmohammadi *et al.* 2006).

A distributed model requires many parameters and extensive input data to effectively reflect landscape heterogeneity of a watershed in simulating hydrologic processes. Thus, a distributed model tends to be over-parameterized with the potential for propagating more error to model outputs due to the spatially detailed input data. A number of tools have been developed and used in uncertainty analysis of distributed models. The Generalized Likelihood

Uncertainty Estimator (GLUE) provides a framework and software tools that have been used in assessing uncertainty in distributed parameter models (Beven 2006). Another tool that has been applied to watershed models is the Shuffled Complex Evolution Metropolis (SCEM-UA) algorithm which is a Markov Chain Monte Carlo (MCMC) sampler (Vrugt *et al.* 2003). Feyen *et al.* (2007) assessed uncertainty in parameters and outputs of a distributed hydrologic model, LISFLOOD using the SCEM-UA algorithm. Zhang *et al.* (2012) demonstrated that distributed hydrologic simulations of SACramento Soil Moisture Accounting (SAC-SMA) were sensitive to source data types and methods used for deriving spatially varied parameters. There have been a few studies that compare different methods for distributed hydrologic models (Yang *et al.* 2008).

Use of geospatial data such as the National Elevation Dataset (NED) (USGS 2009b) and National Land Cover Dataset (NLCD) (USGS 2009c) is common in hydrologic simulation, especially with distributed models. Geospatial data, as a digital representation or model of some characteristic of the landscape, implicitly incorporates error due to the limitations of spatial, temporal, and spectral resolutions of source data, and errors in sampling, analysis and processing of the data. Identifying the best spatial scale for raster-based watershed modeling has been a question of weighing tradeoffs in spatial resolution on the accuracy of geospatial data and improvements in representing hydrologic processes. Studies have used sensitivity analysis to examine the influence of resolution and data sources on watershed modeling (Shrestha *et al.* 2006; Wu *et al.* 2007; Wang *et al.* 2014).

Effects of errors in spatial input data on model results have been rarely reported even though errors in some publicly available geographic information system (GIS) data (or geospatial layers) are explicitly defined. In particular, the vertical error of digital elevation models (DEMs) and classification error in land use data may add a great amount of uncertainty in modeling results (Miller *et al.* 2007; Wu *et al.* 2008). The impacts of the errors have not been compared to each other, and their relative significance compared to that of parameter uncertainty is unknown. The scarcity of studies can be attributed to the limited number of techniques available for modeling errors in spatially

correlated variables and the computational challenges of multi-dimensional uncertainty analysis. Although sequential simulation including sequential Gaussian/indicator simulation (SGS/SIS) techniques has been widely used to generate realizations of a spatially distributed variable in geostatistical simulation, its application to uncertainty analysis of a watershed model is seldom found in the literature (Hengl *et al.* 2010).

The objective of this study was to examine and compare for a distributed hydrologic model the relative contributions of spatial data error and parameter uncertainty on model output uncertainty. We used the HYdrology Simulation using Time-Area method (HYSTAR) model (Her & Heatwole 2016a, 2016b) as the platform for the analysis, and examined spatial data error in elevation and land cover data inputs to the model. The overall goal is to demonstrate and evaluate techniques for uncertainty analysis with distributed models and to provide specific direction for improving the efficiency and accuracy of implementation of the HYSTAR model, which may also provide inference for application to other distributed hydrologic models.

---

## METHODOLOGY

### The HYSTAR model and base application

The distributed, continuous hydrology and sediment transport model, HYSTAR, was applied in predicting runoff and sediment load for the 329 ha ORD subwatershed of Owl Run in Fauquier County, Virginia (Mostaghimi *et al.* 1989). The NLCD 1992 (USEPA 2010) shows the land cover of the ORD catchment as predominately pasture/hay (59%) and forest (26%) (Figure 1). Runoff and sediment load measurements (Mostaghimi *et al.* 1989) made at the outlet of ORD were used in calibrating the model.

The HYSTAR model is capable of simulating direct runoff, base flow, soil moisture, groundwater recharge, and sediment load within a grid-based spatial data model in a distributed manner and in an hourly time step using readily available GIS layers such as NED and NLCD. The model employs a newly developed two-dimensional overland and channel 'time-area routing' scheme to redistribute direct runoff volume along flow paths over a watershed in every

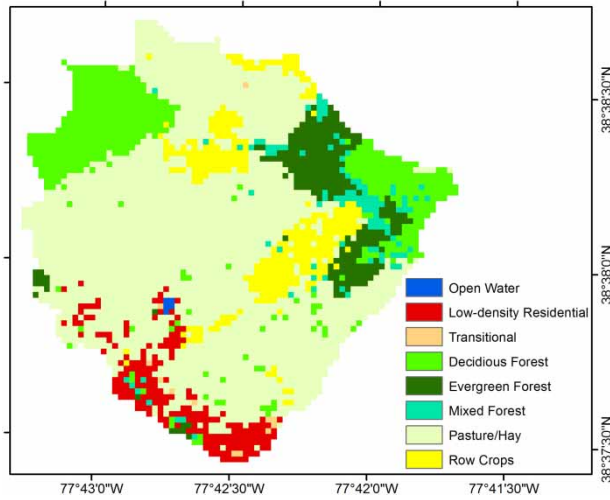


Figure 1 | Land cover and use (NLCD 1992) of the study watershed (ORD).

time interval based on flow hydraulics. In the model, a soil water accounting model is integrated with the routing scheme to continuously track soil water variations considering spatially distributed infiltration of surface flow so as to enable seamless predictions of runoff and sediment load for a long simulation period. Such features allow the model to effectively consider spatiotemporal changes in watershed processes in hydrology and sediment transport simulations. Thus, model predictions are responsive to

variations of spatial input data layers, providing the basis for investigating the impact of GIS layer uncertainty on model outputs. A detailed description of the model and its calibration can be found in Her & Heatwole (2016a, 2016b), and only a brief description is included here.

Fourteen parameters for hydrology and sediment simulation (Table 1) were calibrated utilizing root mean squared error (RMSE) and relative error as objective functions, respectively. The calibrated model provided good monthly runoff and sediment load predictions for the entire simulation period of 6 years from 1990 to 1995 with Nash–Sutcliffe efficiency coefficients of 0.60 and 0.58 and coefficient of determination of 0.83 and 0.79, respectively. The parameters are described and calibrated values shown in Table 1.

Of the fourteen calibration parameters, eight are spatially varied: curve number, Manning’s roughness coefficient for overland and channel flow, standard basal crop coefficient and fraction of the evaporable soil surface, root zone depth, and soil detachability and cohesion ratios. In order to reduce the number of parameters to be calibrated, spatial scale factors (CNF, MNO, MNC, BCC, EFS, RZD, SDR, and SCR) were introduced for the spatially varied parameters (Table 1). Values are adjusted by changing the corresponding scale factors while fixing their spatial variability in parameter calibration. The values for these parameters were determined

Table 1 | Parameters of Hystar and their calibrated values

Process	Parameter	Description	Feature	Range	Calibrated value
Direct runoff	CNF	Curve number scale factor	Spatial scale factor	0.5–1.5	1.094
	MNO	Manning’s roughness coefficient scale factor for overland areas	Spatial scale factor	0.5–1.5	1.054
	MNC	Manning’s roughness coefficient scale factor for channel	Spatial scale factor	0.5–1.5	1.171
	THA	Threshold area that defines initiation points of channel networks	Unit: ha	1–100	17.91
Soil moisture	GCL	Coefficient L of the van Genuchten equation	–	0.25–0.75	0.662
	GCM	Coefficient M of the van Genuchten equation	–	0.25–0.75	0.325
	BCC	Basal crop coefficient scale factor	Spatial scale factor	0.5–1.5	0.512
	EFS	Effective fraction of soil surface covered by vegetation scale factor	Spatial scale factor	0.5–1.5	1.187
	RZD	Root zone depth scale factor	Spatial scale factor	0.1–2.0	0.339
	SAR	Soil anisotropy ratio	–	0.1–2.0	0.164
Sediment transport	CSO	Critical Shields parameter for overland flow	–	0.1–10.0	0.744
	CSC	Critical Shields parameter for channel flow	–	0.035–2.0	0.057
	SDR	Soil detachability ratio scale factor	Spatial scale factor	0.1–2.0	0.891
	SCR	Soil cohesion ratio scale factor	Spatial scale factor	0.1–2.0	1.066

from the spatially descriptive elevation (NED), land cover (NLCD), and soil (SSURGO: Soil Survey Geographic) data.

On the other hand, the other six parameters are spatially uniform in the watershed. The threshold area (THA) is used to define the stream network of a watershed by identifying its initiation points. A smaller THA makes the stream network more dense thus the overall travel time of a watershed shorter and vice versa. The two coefficients (GCL and GCM) that define the relationship between saturated and unsaturated hydraulic conductivity come from the van Genuchten equation, and the soil anisotropy ratio (SAR) represents the ratio of vertical to horizontal hydraulic conductivity. For the sediment transport simulation, CSO and CSC were introduced to control critical Shields parameters for overland and channel conditions, respectively.

### Parameter uncertainty analysis – SCEM-UA

The MCMC technique is one of the Bayesian samplers commonly used for uncertainty analysis. The goal of MCMC is to sample parameter values from the posterior distribution by simulating a random process using a Markov chain that has the posterior distribution as its stationary distribution. MCMC is known to provide a solution to the difficult problem of sampling from a high dimensional posterior distribution (Kanso *et al.* 2006). The required number of iterations for convergence of the Markov chain is determined by diagnostic criterion such as the Gelman–Rubin statistic, while the speed of convergence is affected by the efficiency of the sampling strategy and the proposal distribution. When the Gelman–Rubin statistic becomes less than 1.2, the distribution of samples is regarded as stationary (Vrugt *et al.* 2003).

Vrugt *et al.* (2003) introduced an advanced MCMC method, called SCEM-UA algorithm, to improve efficiency in updating a proposal distribution in the MCMC. This algorithm operates by merging the strengths of the Metropolis algorithm and Shuffled Complex Evolution (SCE-UA) algorithm of Duan *et al.* (1992). In order to prevent the collapse of the SCE-UA algorithm into the relatively small region of a single best parameter set, the SCEM-UA algorithm replaced the Simplex search method with the Metropolis method. It is known as one of the most efficient and robust algorithms for identifying uncertainty of model parameters and output (Vrugt *et al.* 2003; Feyen *et al.*

2007). In this study, the SCEM-UA algorithm was employed for examining uncertainty of parameters and reliability of the modeling output. The detailed procedures of SCEM-UA are described below and summarized in Figure 2.

The formal likelihood function for the SCEM-UA algorithm was utilized to assess the posterior probability distribution of parameters, which are regarded as their uncertainty. The likelihood is calculated as:

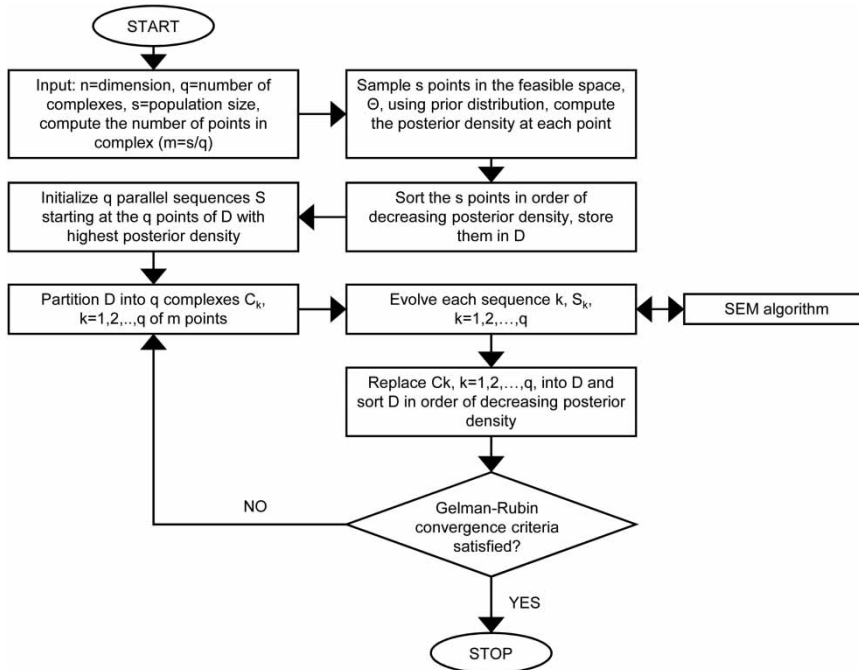
$$L(\Theta|O) = P(O|\Theta) = \exp\left(-\frac{1}{2} \sum_{i=1}^N \left|\frac{e}{\sigma_e}\right|^2\right) \quad (1)$$

where  $\Theta$  is a set of parameters,  $O$  is a set of observations,  $N$  is the length of the data and  $e$  is a vector of errors with zero expectation and constant variance ( $\sigma_e$ ) (Vrugt *et al.* 2003). It is assumed that the errors are distributed normally and are independent (Feyen *et al.* 2007).

A value vector of the fourteen parameters for hydrology and sediment transport simulations of HYSTAR was found through calibrating the model to observed data (Table 1). The same rainfall events, simulation conditions, and objective function as those used in the calibration were applied in the uncertainty analysis (Her & Heatwole 2016a, 2016b). The Gelman–Rubin statistic was used to assess convergence of the Markov chain into a stationary distribution (Vrugt *et al.* 2003). The SCEM-UA algorithm was run with the same numbers of complexes, population size for each complex, and total sample population as those of the SCE-UA algorithm used for calibration. For instance, the number of complexes, a population size for each complex, and the total sample population were set to 25, 13, and 325, respectively, in the uncertainty analysis for the hydrology simulation. The sampling stopped with 1,000 iterations, and each parameter had a set of 1,000 values sampled by the Markov chain. Then, the posterior distributions of the parameters were developed based on the 900 sets sampled after the Markov chain converged to its stationary distribution (in the first 100 sampling).

### GIS input data uncertainty analysis – sequential simulation

HYSTAR employs three GIS layers as input data: elevation (DEM), land cover (NLCD), and soil properties (SSURGO).



**Figure 2** | General procedure of the SCEM-UA algorithm (Vrugt *et al.* 2003).

The impact of uncertainty in the DEM and NLCD input data on model output was examined through sensitivity analysis, assuming vertical measurement error and classification error are major sources of uncertainty in the DEM and NLCD, respectively.

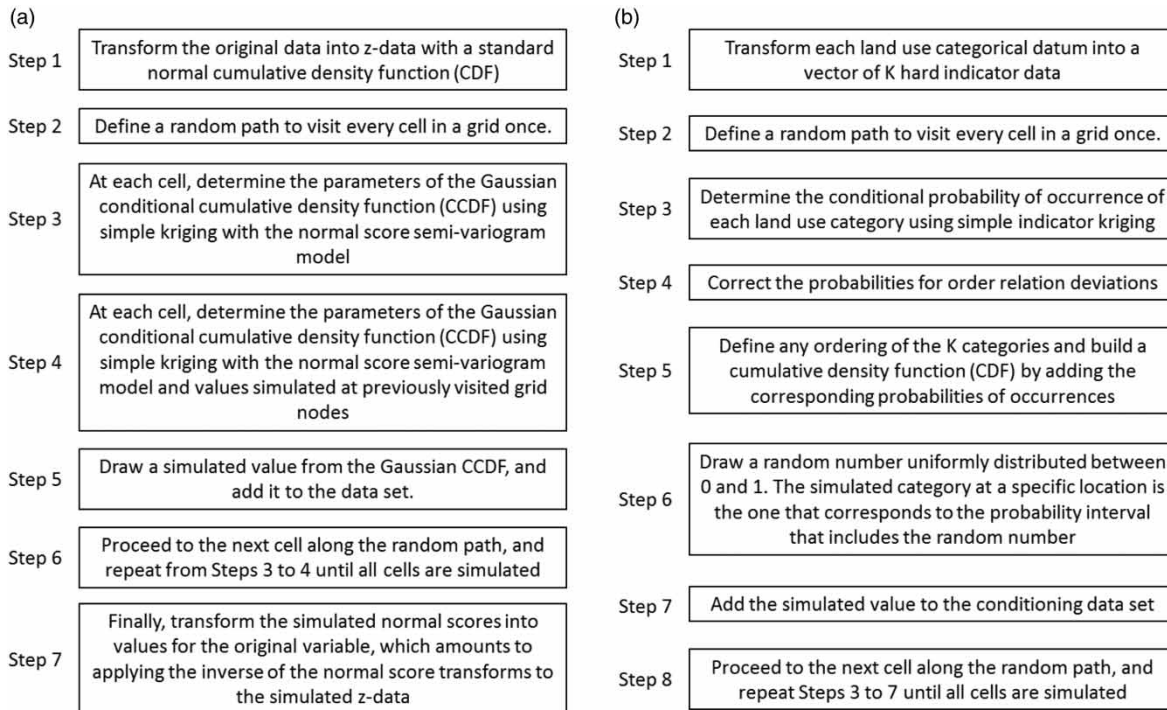
The common source of elevation data in the United States is the NED, which is developed and distributed by the US Geological Survey (USGS). The reported overall absolute vertical RMSE of the NED is 2.4 m, and the errors range from  $-42.64$  to  $18.74$  m (USGS 2009a). If error at any point occurs independently of that at any other point, RMSE can be a way to represent the variance of the overall error in DEM appropriately. When errors are spatially variable and correlated, however, a single scalar index like RMSE cannot distinguish areas with more or less uncertainty or assess spatial autocorrelation.

The NLCD provides 21 different land cover classes at a 30-meter grid resolution derived by classification of Landsat satellite imagery (USEPA 2010). For this study, we used the NLCD-1992 dataset which has an overall classification accuracy that ranges between 70 and 85% (Wickham *et al.* 2004). The error matrix is commonly used to quantify classification errors. However, Steele *et al.* (1998) pointed out that the error

matrix does not provide information on the spatial structure of error in a classification and does not have the ability to describe the variation of accuracy across a classified map.

In this study, the errors were simulated using SGS and SIS techniques (Figure 3), which consider autocorrelation structure in the spatial data and their errors (Goovaerts 1997). The techniques generate a realization of the variable of interest while sequentially visiting each cell along a predefined random path in a grid-based spatial representation of the variable. They use values simulated at previously visited grid cells as well as neighboring original data to develop conditioning information defining the probability of occurrences so as to reproduce (or 'honor') the overall covariance structure of the random field (Pebesma & Wesseling 1998). SGS assumes conditional density functions as a normal distribution and then simulates a Gaussian random field based on a procedure of sequential simulation algorithm (Goovaerts 1997). However, if attributes of data or observations do not satisfy the normal assumption, the original data can be transformed into the Gaussian space or SIS can be an alternative. SIS does not require any specific distribution model of the data so that a non-parametric method can be applied regardless of the type of data and can thus be used with the categorical land cover map.





**Figure 3** | General procedure of the sequential simulation (Goovaerts 1997). (a) SGS, (b) SIS.

SGS and SIS were utilized to simulate spatially correlated errors and produce multiple realizations of the DEM and NLCD, respectively. Elevation errors were simulated using SGS for randomly selected cells, with the total area corresponding to 30% of the watershed. For the NLCD dataset, classification errors were simulated using SIS for randomly selected cells. The total target area was initially set to 50% of the watershed. In cases where different land cover classes were simulated for a cell, the class with higher probability was assigned to the cell. Thus, in the induced error grids, the land cover class was altered in around 30% of the area. A total of 100 ‘disturbed’ datasets with induced errors were generated for the DEM and NLCD.

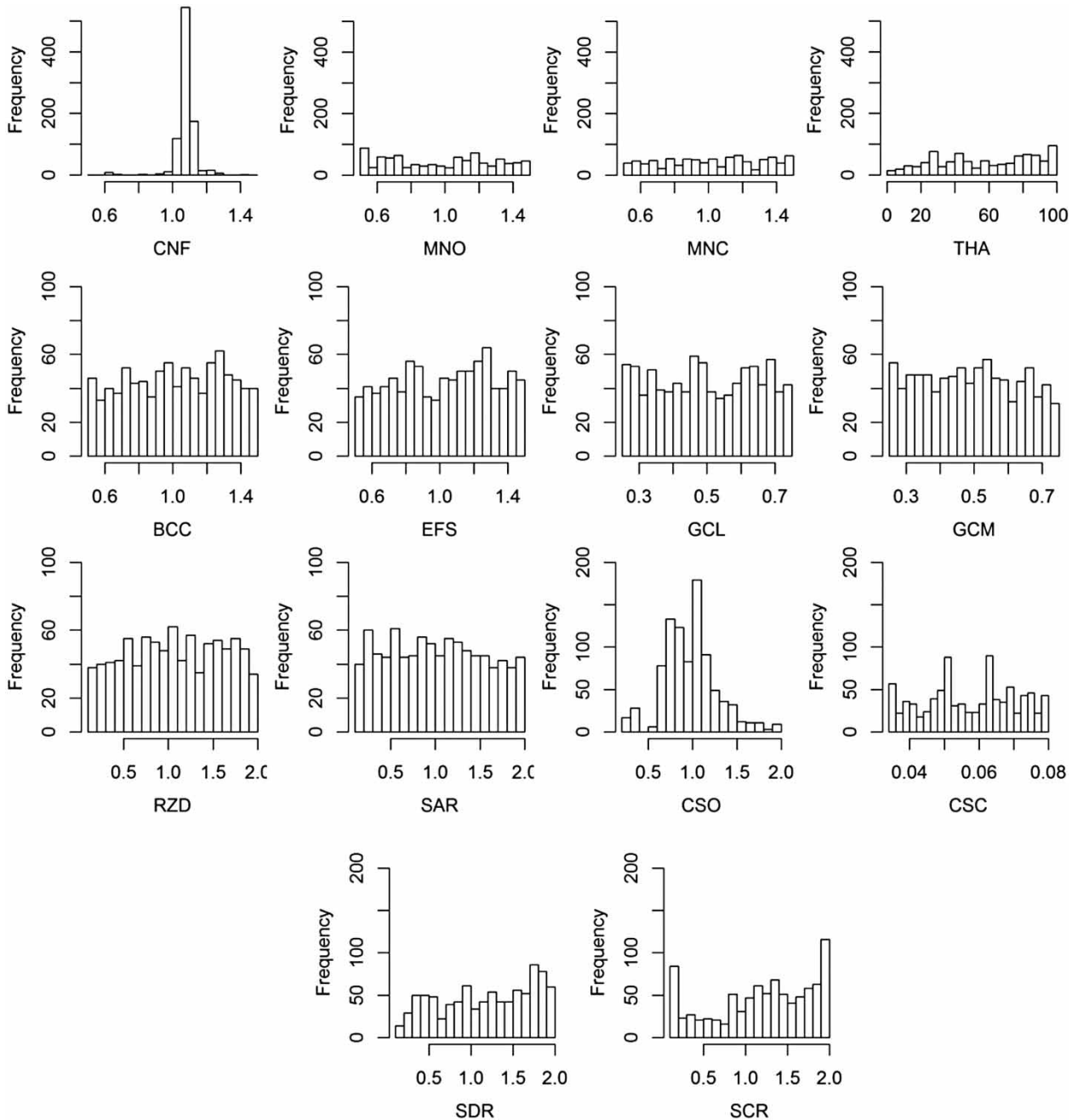
## RESULTS

### Parameter uncertainty

The resulting posterior distributions of the parameters are shown in Figure 4. The posterior distributions of the

parameters were developed using the last 900 samples drawn by the algorithm after stationary distributions were achieved during the first 100 evolutions of the Markov chain. The posterior distributions of CNF and CSO show clear peaks in contrast to the other parameters (Figure 4). CNF and CSO are parameters that influence the most runoff and sediment transport simulations of the model while simulated runoff and sediment loads are relatively insensitive to the other parameters. Thus, the degree of spread in the posterior distribution must be associated with sensitivity of a model to a parameter.

As seen in Figure 4, the posterior distributions of the parameters for soil moisture simulation have multiple modes. The calibrated values of the parameters are located at vicinities of the modes in their own posterior distributions, implying the results did not always correspond to the calibration results. Such outcomes are possible because the calibration and uncertainty analysis employed unique algorithms for different purposes, identifying a single set of the parameter that gives the best performance statistics and the posterior distributions of the parameters, respectively. However, it is worth noting that the most sensitive



**Figure 4** | Posterior distributions of the parameter values sampled by the SCEM-UA algorithm.

parameters for hydrology and sediment transport simulations, CNF and CSO, have calibrated values that corresponded well with the posterior distributions, implying the soundness of sampling for both algorithms.

Uncertainty of runoff and sediment modeling was examined through quantifying variation in the model outputs

simulated using parameter sets sampled from the posterior distributions. Uncertainty is defined as a range and confidence interval (CI) of the output hydrograph time series (Table 2). CIs of the simulated runoff and sediment load were determined at a significance level of 0.05. The estimated uncertainty in the simulated runoff and sediment load hydrographs of the

**Table 2** | Statistics of uncertainty of the model output for ORD

Item	Uncertainty measure	Runoff		Sediment load	
		m <sup>3</sup> /s	% <sup>a</sup>	Mg	%
Max <sup>b</sup>	Range	291.1	798	935.5	2,475
	CI	9.2	25	21.2	56
Average <sup>c</sup>	Range	86.6	237	150.5	398
	CI	2.2	6	3.5	9
Average calibrated <sup>d</sup>		36.5	100	37.8	100

<sup>a</sup>Percentage of the uncertainty measures to the average of the calibrated monthly runoff and sediment load.

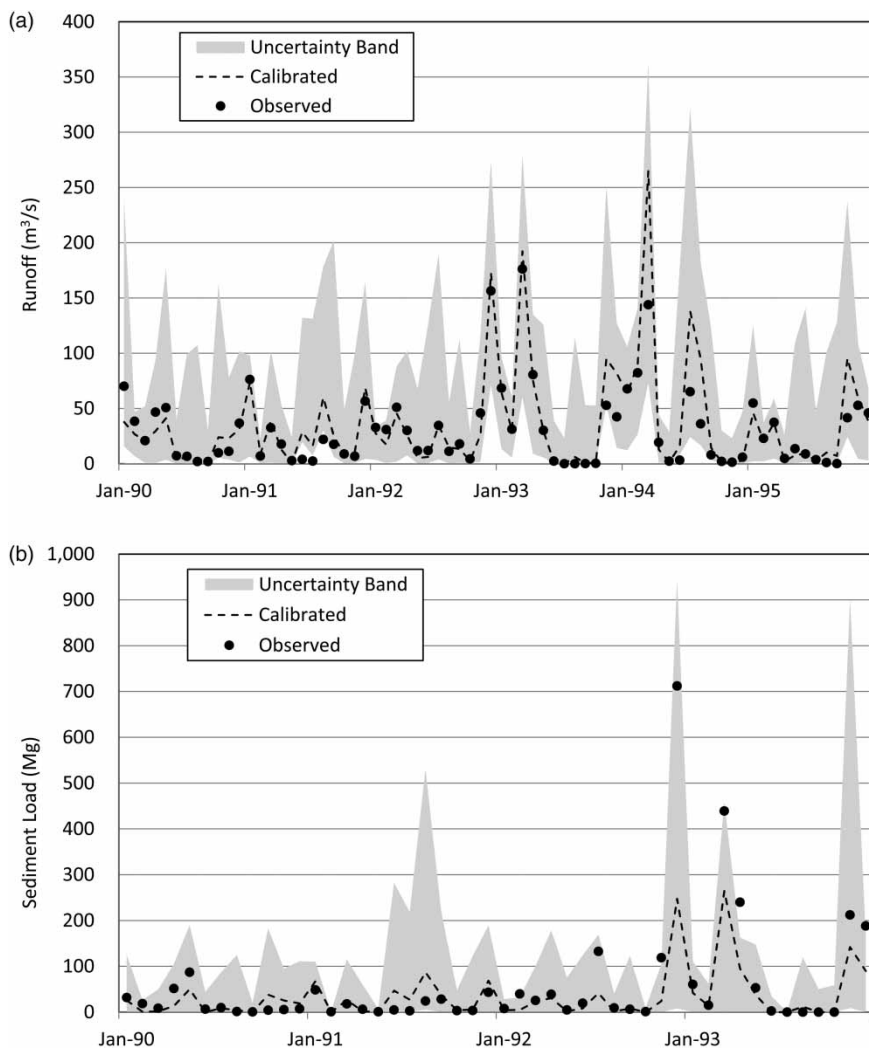
<sup>b</sup>The maximum width of the band of the uncertainty measure.

<sup>c</sup>Average width of the band of the uncertainty measure.

<sup>d</sup>Average of the calibrated monthly runoff and sediment load.

watershed is presented in Figure 5, and model output uncertainty measures are summarized in Table 2.

In Figure 5, the uncertainty band (width of the range) derived using the SCEM-UA algorithm generally covers the calibrated and observed runoff and sediment loads. When the range was used as an uncertainty measure, the average widths of the uncertainty band of 86.6 m<sup>3</sup>/s and 150.5 Mg were estimated for the simulated monthly runoff and sediment load respectively in the entire simulation period (Table 2). They correspond to 237 and 398% of the average of the calibrated monthly runoff and sediment load, respectively. CI provided narrower uncertainty bands; average CIs were 2.2 m<sup>3</sup>/s (6% of the calibrated monthly runoff) and

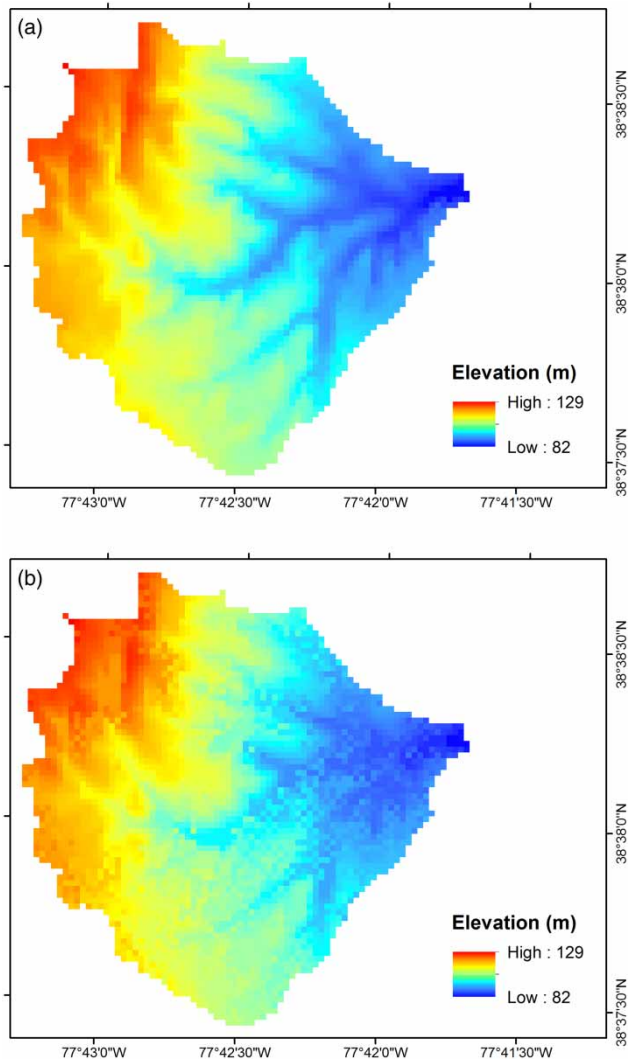
**Figure 5** | Runoff and sediment load uncertainty estimated using the SCEM-UA algorithm: (a) monthly runoff and (b) monthly sediment load.



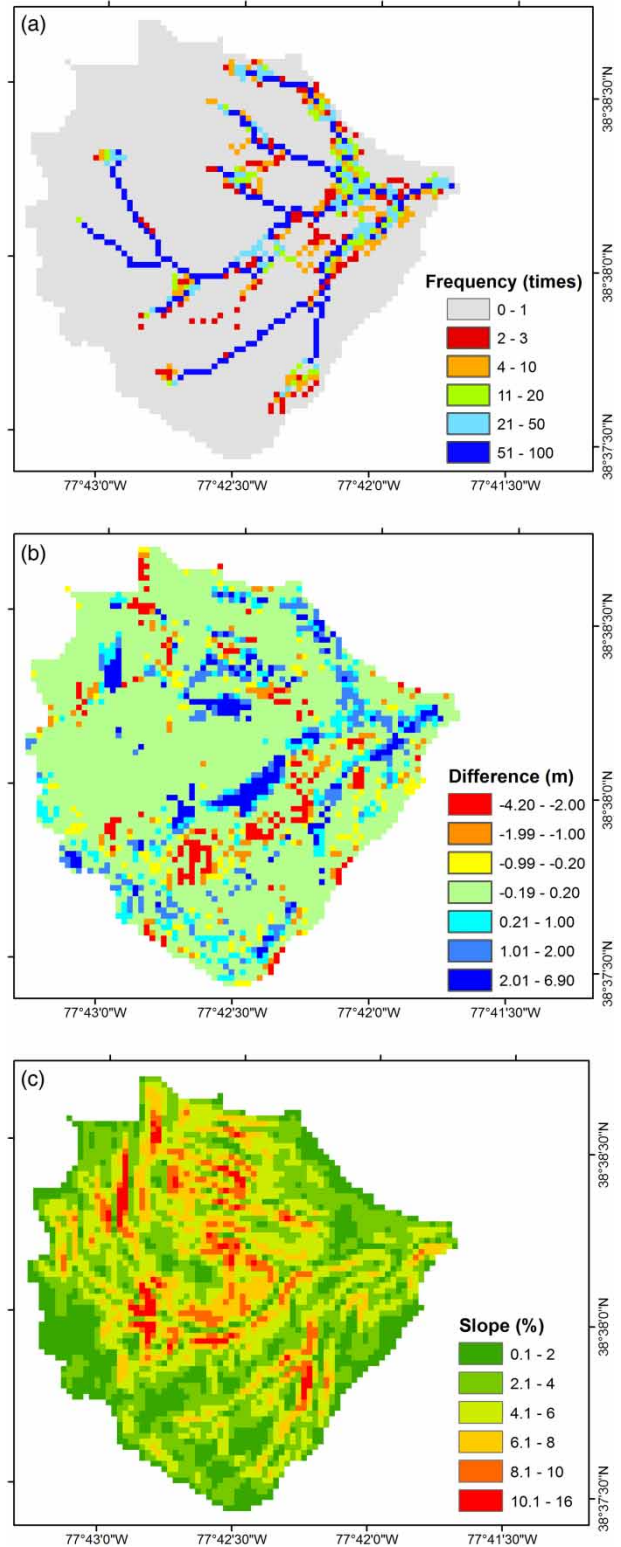
3.5 Mg (9% of the calibrated sediment load) for monthly runoff and sediment loads, respectively. When a smaller significance level is applied, the width of uncertainty band will be increased. For example, an average width of the CI for the monthly runoff increased to 2.9 m<sup>3</sup>/s (7.9%) at a level of significance of 0.01.

**Impact of GIS data uncertainty on model outputs**

The original DEM and resulting slope map along with one of the DEMs with SGS induced error are shown in Figure 6. Figure 7 displays spatial variations in the stream networks



**Figure 6** | (a) The original DEM of the watershed and (b) one of the 100 elevation maps with added errors generated by SGS.



**Figure 7** | Characteristics of DEMs with induced error. (a) Frequency of cells included in the flow network for the 100 disturbed DEMs, (b) difference between the original DEM and average of the 100 disturbed cases, and (c) slopes of the source DEM.

derived from the disturbed DEMs. The numbers in Figure 7(a) represent the frequency of being classified as a cell on the channel networks (defined using a channel initiation THA of 10 ha) from the 100 DEM realizations. Greater variations are found at the vicinity of channel heads and junctions and in the downstream portions of the channel network where slopes are relatively shallow (Figure 7(c)). Figure 7(b) shows differences between elevations of the original and the average of disturbed DEMs range from 6.90 to -4.20 m, and are mainly distributed along stream-lines on relatively shallow areas.

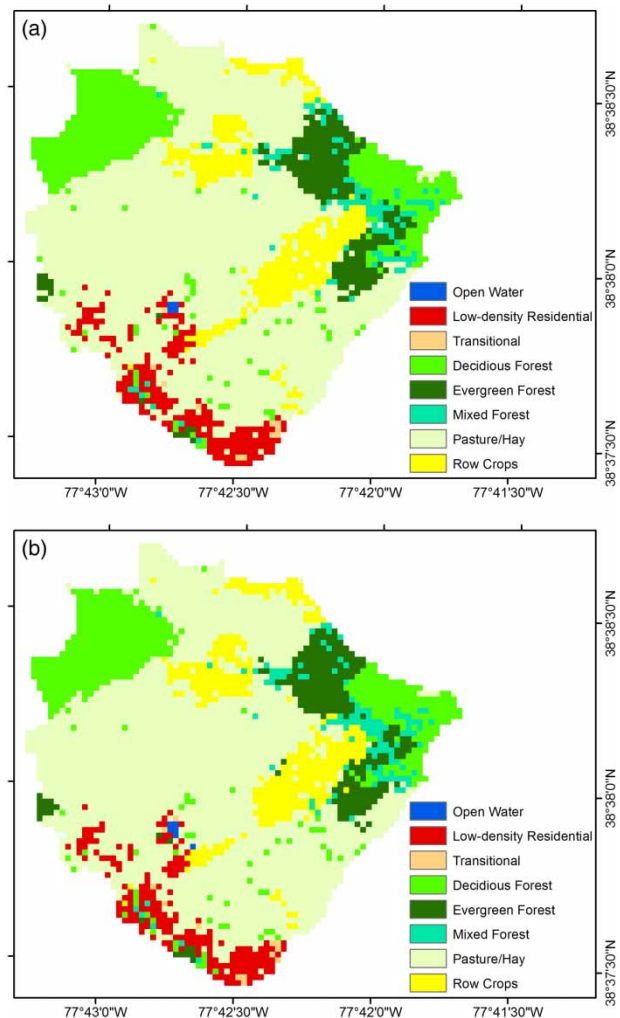
The differences between elevations of the original and the disturbed DEMs range from 9.90 to -8.37 m (Table 3), which are within the reported error ranges (USGS 2009a), and the spatially averaged standard deviation of the errors is 1.24 m. In Table 3, the ranges of errors for every elevation class show that variations of the error are symmetric and greater in the mid-classes of elevation, 90–120 m. On the other hand, the minimum, maximum, and mean statistics of the error for the classes indicate that the SGS algorithm produced biased errors for lower and higher elevations. The errors tend to be positive and negative on lower and higher elevations respectively, while the overall average error, 0.04 m (median error is 0.00), is close to zero (Figure 7(b)). In other words, the disturbed DEMs overestimated and underestimated elevation on low and high elevation areas, respectively. It implies that, as desired, the correlation between error and elevation was considered in the algorithm.

**Table 3** | Statistics of differences (disturbed – original) between the original and the disturbed DEMs by elevation classes (unit: m)

Class	Area (ha)	Min	Max	Range	Mean
80–85	4.95	-1.94	6.01	7.96	0.85
85–90	24.93	-5.94	6.66	12.61	0.29
90–95	48.24	-6.13	9.90	16.03	0.28
95–100	42.57	-7.43	9.88	17.30	0.10
100–105	52.38	-8.37	8.71	17.09	-0.05
105–110	51.66	-6.84	6.00	12.84	-0.13
110–115	35.55	-7.52	8.79	16.32	0.03
115–120	40.32	-6.63	7.31	13.94	0.00
120–125	23.13	-6.52	4.36	10.89	-0.27
125–130	5.13	-7.30	2.72	10.02	-0.40

Land cover maps disturbed by SIS are presented in Figure 8. The original land cover map (USEPA 2010) is given in Figure 1. As seen in Figure 8, land cover classes became more fragmented in the disturbed maps, and small patches surrounded by larger ones shrunk or disappeared. Area statistics for land use classes of the disturbed land use map are provided and compared with the original in Table 4. The averaged areas of forest, low intensity residential, and open water in the disturbed land cover maps are greater than those of the original whereas the others are smaller.

The maximum ranges of the monthly runoff and sediment load (29.3 m<sup>3</sup>/s and 64.6 Mg) simulated with the



**Figure 8** | Two samples of the 100 disturbed land cover data grids.

**Table 4** | Area statistics of the disturbed NLCDs to land cover classes (unit: ha)

Class	Original	Min	Max	Mean	Stdev <sup>a</sup>	cv <sup>b</sup>
Open water	0.36	0.27	1.89	0.69	0.26	0.381
Low intensity residential	18.99	19.26	22.59	21.01	0.82	0.039
Commercial/industrial	1.98	1.17	3.15	1.93	0.35	0.184
Deciduous forest	50.04	50.67	55.44	52.82	1.02	0.019
Evergreen forest	21.78	21.78	25.65	23.59	0.81	0.034
Mixed forest	14.31	10.89	14.67	12.45	0.76	0.061
Pasture/hay	192.24	184.95	193.95	189.43	1.64	0.009
Row Crops	29.16	24.30	28.89	26.94	0.90	0.033

<sup>a</sup>Standard deviation.<sup>b</sup>Coefficient of variation.

disturbed DEMs occurred when the model provided the greatest monthly direct runoff and sediment load (194 m<sup>3</sup>/s and 264 Mg), March 1994 and March 1993, respectively. The maximum ranges of the monthly runoff and sediment load (17.0 m<sup>3</sup>/s and 47.0 Mg) simulated with the disturbed NLCDs were also found in March 1994 and March 1993, respectively, which implies a proportional relationship between quantities of simulated runoff and sediment load and their uncertainty. Overall, the impact of the simulated error in the land use layer on runoff is relatively insignificant compared to that of the elevation layer.

Runoff and sediment load were simulated using the combination of disturbed DEMs and NLCDs. For the sake of efficiency in analysis, the 100 realizations for each DEM and NLCD were combined one-to-one to create 100 realizations that have errors in both DEM and NLCD. As seen in Table 5, the runoff and sediment load simulated with the disturbed DEM and NLCD are not greatly different from those of the disturbed DEM alone. The patterns of the discrepancy between the simulated runoff and sediment load and the variation statistics are also very close to each other. The implication is that the impact of the DEM error on the runoff and sediment load simulation is much greater than that of the NLCD error. In addition, as seen in Figure 9, the slopes and coefficients of determination (R<sup>2</sup>) of the linear regression equations are close to one, meaning the simulated errors of the GIS layers do not significantly influence the runoff and sediment transport simulation.

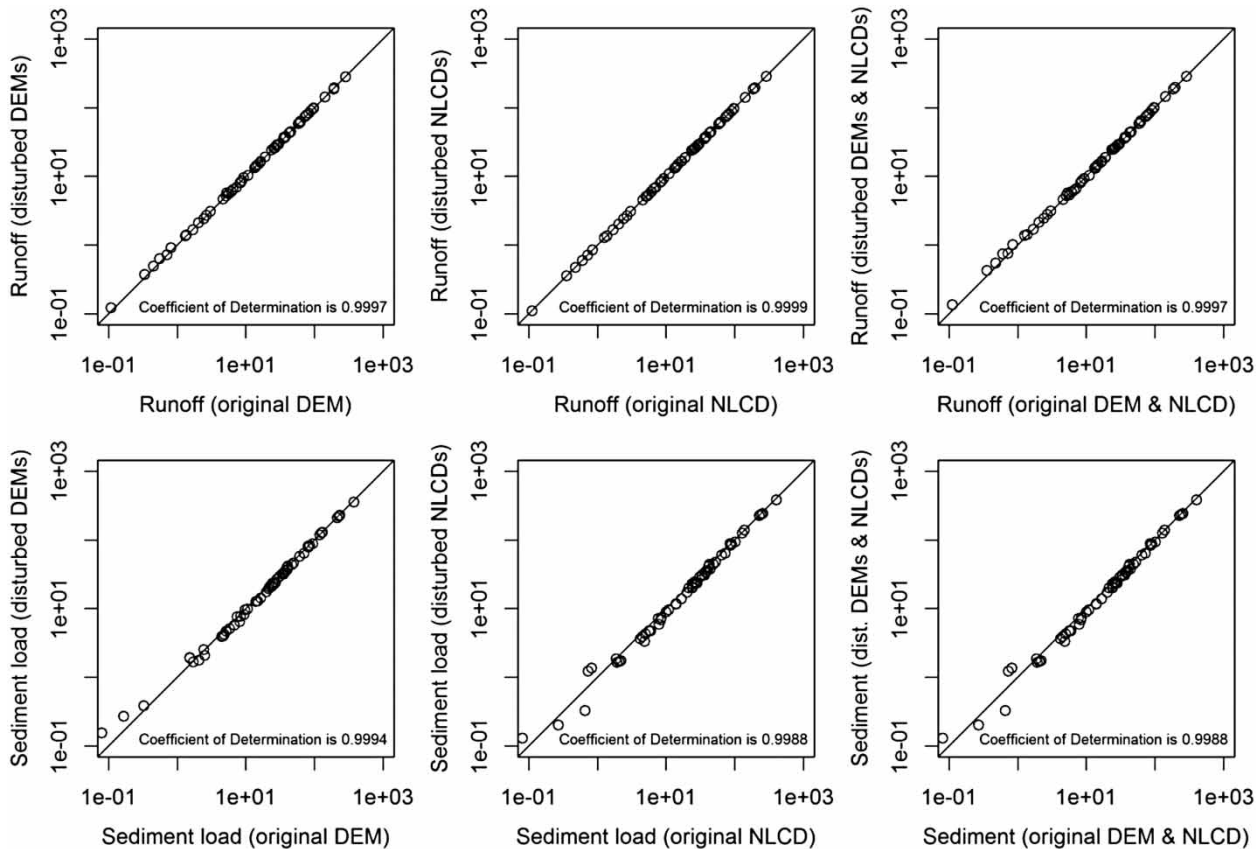
**Table 5** | Statistics of variations in the monthly runoff and sediment load simulated with the disturbed DEMs and NLCDs

Dataset with error	Statistics	Runoff		Sediment load	
		m <sup>3</sup> /s	%	Mg	%
DEM	Max range	29.3	80.3	64.6	170.0
	Ave. range	3.28	9.0	7.40	19.6
	Max CI	2.84	7.8	4.99	13.2
	Ave. CI	0.28	0.8	0.61	1.6
NLCD	Max range	17.0	46.7	47.0	124.4
	Ave. range	1.83	5.0	5.78	15.3
	Max. CI	1.03	2.8	3.96	10.5
	Ave. CI	0.14	0.4	0.45	1.2
DEM/NLCD	Max range	30.7	84.1	75.2	199.1
	Ave. range	3.49	9.6	8.60	22.8
	Max. CI	2.77	7.6	6.47	17.1
	Ave. CI	0.29	0.8	0.72	1.9
Average runoff rate and sediment load <sup>a</sup>		36.5	100.0	37.8	100.0

<sup>a</sup>These values, 36.5 m<sup>3</sup>/s and 37.7 Mg, were calculated by averaging monthly runoff rates and sediment loads simulated with the original DEM and NLCD using a calibrated HYSTAR model.

## DISCUSSION

The results of this study indicate that the runoff simulation of a spatially distributed hydrologic model is affected more by parameter uncertainty than by uncertainty in GIS layers (Tables 2 and 5). Parameters, particularly critical parameters such as CNF and CSO, have a strong influence on hydrologic components in specific directions (increasing/decreasing runoff volume or sediment load). On the other hand, elevation and land use data errors generated using



**Figure 9** | Agreement of the average monthly runoff ( $\text{m}^3/\text{s}$ ) and sediment load (Mg) for the original and the disturbed DEM and NLCDs. Monthly runoff and sediment loads were simulated with the 100 disturbed DEMs, the 100 disturbed NLCDs, and their one-to-one combinations, and then each 100 simulated runoff and sediment loads were averaged.

the sequential simulation methods were locally distributed, particularly in the vicinity of stream networks and between boundaries of different land uses. Thus, impacts of an individual error could be easily intermingled with those of other errors and canceled out while routing runoff and sediment to the outlet along flow paths.

In this study, we also found that errors in elevation data are more influential on the hydrologic modeling outputs than are errors in land use data (Table 5). Topography affects the hydrologic response of a watershed to rainfall events through its influence on flow directions, stream networks, and flow velocity. Lindsay & Evans (2008) demonstrated that ‘channel network morphometric’ of a watershed is very sensitive to elevation errors, and a small elevation error can lead to the very different presentation of the channel networks. On the other hand, an error in land use classification may not influence watershed-wide processes but only local conditions (Dong

*et al.* 2015), thus its impact on hydrologic simulation tends to be limited.

Sequential simulation using the SGS/SIS techniques was an effective tool for generating spatially correlated errors of spatial input data for uncertainty analysis of a distributed watershed model (Figures 6–8 and Tables 3 and 4). Although the application of the sequential simulation techniques was limited to elevation and land use layers in this study, they could also be useful in simulating or interpolating soil features such as textures, hydraulic conductivity, and field capacity required as input data for hydrologic modeling. It would be interesting to see how errors in soil classifications, mapping and attributes would affect distributed modeling outputs.

There are many methods and sampling algorithms developed for parameter uncertainty analysis, and the SCEM-UA algorithm was selected for this study because of its proven applicability to a distributed hydrologic model



(Vrugt *et al.* 2003; Feyen *et al.* 2007). Although this study did not intend to compare different methods of estimating parameter uncertainty, the GLUE method was also applied as another way to derive the posterior parameter distributions for the purpose of comparison. From the comparison, we observed that the GLUE method provided different uncertainty measures depending on a cut-off value used to identify 'behavioral' parameter sets (Beven 2006). While the results of the analysis with GLUE are not presented here, we found that the GLUE method led to the same conclusion that parameter uncertainty has a greater impact on model outputs than uncertainty in spatial data (Her 2011).

## CONCLUSIONS

This study compared the significance of parameter and spatial input data uncertainty on predictions of a two-dimensional continuous hydrology model, HYSTAR. Posterior distributions of the parameters were derived using the SCEM-UA algorithm, and errors in the elevation and land cover spatial data grids were simulated using SGS and SIS to assess the impact of these errors on hydrology and sediment transport predictions of the model. The study showed that the impact of parameter uncertainty was much greater than the impact of GIS data errors on model output. The average ranges of runoff and sediment loads simulated with parameter sets sampled from the posterior distributions are one to four times greater than the averages of the base simulation, signifying greater significance of parameter uncertainty on model predictions. On the other hand, average variations of runoff and sediment loads simulated with error-added spatial input data layers were only 10–20% of the base simulation. Therefore, the accuracy and reliability of a distributed model will be more efficiently improved by refining the parameters than by exploring the better quality of the GIS input data. This conclusion is not to say that the accuracy of the layers is not important to the model, but should be interpreted that the spatial input layers are accurate enough to provide acceptable levels of errors in this distributed model application. In addition, the impact of the topographic data error on the model output was greater than that of the land cover data error. Thus, improving the topographic data

should be emphasized over the land cover data for better accuracy in the model results.

## REFERENCES

- Beven, K. 2006 *A manifesto for the equifinality thesis*. *J. Hydrol.* **320** (1), 18–36.
- Dong, M., Bryan, B. A., Connor, J. D., Nolan, M. & Gao, L. 2015 *Land use mapping error introduces strongly-localised, scale-dependent uncertainty into land use and ecosystem services modelling*. *Ecosyst. Serv.* **15**, 63–74.
- Duan, Q., Sorooshian, S. & Gupta, V. 1992 *Effective and efficient global optimization for conceptual rainfall-runoff models*. *Water Resour. Res.* **28** (4), 1015–1031.
- Feyen, L., Vrugt, J. A., Nuallain, B. O., van der Knijff, J. & Roo, A. D. 2007 *Parameter optimization and uncertainty assessment for large-scale streamflow simulation with the lisflood model*. *J. Hydrol.* **332** (3), 276–289.
- Goovaerts, P. 1997 *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York.
- Hengl, T., Heuvelink, G. B. M. & van Loon, E. E. 2010 *On the uncertainty of stream networks derived from elevation data: the error propagation approach*. *Hydrol. Earth Syst. Sci.* **14** (7), 1153–1165.
- Her, Y. 2011 *HYSTAR: Hydrology and Sediment Transport Simulation using Time-Area Method*. Doctoral Dissertation, Virginia Polytechnic Institute and State University.
- Her, Y. & Heatwole, C. 2016a *Two-dimensional continuous simulation of watershed hydrology using the time-area method*. *Hydrol. Process.* **30** (5), 751–770.
- Her, Y. & Heatwole, C. 2016b *HYSTAR sediment model: distributed two-dimensional simulation of landscape erosion and sediment transport using the time-area routing method*. *J. Am. Water Resour. Assoc.* **52** (2), 376–396.
- Kanso, A., Chebbo, G. & Tassin, B. 2006 *Application of mcmc – GSA model calibration method to urban runoff quality modeling*. *Reliabil. Eng. & Syst. Safety* **91** (10), 1398–1405.
- Lindsay, J. B. & Evans, M. G. 2008 *The influence of elevation error on the morphometrics of channel networks extracted from dems and the implications for hydrological modelling*. *Hydrol. Process.* **22** (11), 1588–1603.
- Miller, S. N., Guertin, D. P. & Goodrich, D. C. 2007 *Hydrologic modeling uncertainty resulting from land cover misclassification*. *J. Am. Water Resour. Assoc.* **43** (4), 1065–1075.
- Mostaghimi, S., McClellan, P. W., Tim, U. S., Carr, J. C., Byler, K. K., Dillaha, T. A., Shanholtz, V. O. & Pratt, J. R. 1989 *Watershed/water quality monitoring for evaluating animal waste BMP effectiveness: Pre-BMP evaluation, Final Report: Report No. O-P1-8906*. Virginia Polytechnic Institute and State University, Blacksburg, Virginia.
- Pebesma, E. J. & Wesseling, C. G. 1998 *GSTAT: a program for geostatistical modelling, prediction and simulation*. *Comput. Geosci.* **24** (1), 17–31.



- Shirmohammadi, A., Chaubey, I., Harmel, R. D., Bosch, D. D., Munoz-Carpena, R., Dharmasri, C., Sexton, A., Arabi, M., Wolfe, M. L., Frankenberger, J., Graff, C. & Shorabi, T. M. 2006 **Uncertainty in TMDL models**. *Trans. ASABE* **49** (4), 1033–1049.
- Shrestha, R., Tachikawa, S. & Takara, K. 2006 **Input data resolution analysis for distributed hydrological modeling**. *J. Hydrol.* **319** (1), 36–50.
- Steele, B. M., Winne, J. C. & Redmond, R. L. 1998 **Estimation and mapping of misclassification and probabilities for thematic land cover maps**. *Remote Sens. Environ.* **66** (2), 192–202.
- USEPA 2010 **1992 National Land Cover Data (NLCD)**. Available from: [www.epa.gov/mrlc/nlcd.html](http://www.epa.gov/mrlc/nlcd.html) (accessed March 2010).
- USGS 2009a **Accuracy Assessment of Elevation Data**. Available from: <http://topochange.cr.usgs.gov/assessment.php> (accessed August 2009).
- USGS 2009b **The National Map Viewer**. Available from: <http://nationalmap.gov/viewer.html> (accessed 2009).
- USGS 2009c **The USGS Land Cover Institute (LCI)**. Available from: <http://landcover.usgs.gov/natl/landcover.php> (accessed 2009).
- Vrugt, J. A., Gupta, H. V., Bouten, W. & Sorooshian, S. 2003 **A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters**. *Water Resour. Res.* **39** (8), SWC 1-1-SWC 5–14.
- Wang, H., Li, X., Li, W. & Du, X. 2014 **The impact of digital elevation model and land use spatial information on hydrologic simulation program-fortran – predicted stream flow and sediment uncertainty**. *J. Hydroinform.* **16** (5), 989–1003.
- Wickham, J. D., Stehman, S. V., Smith, J. H. & Yang, L. 2004 **Thematic accuracy of the 1992 National Land-Cover Data for the western United States**. *Remote Sens. Environ.* **91** (3), 452–468.
- Wu, S., Li, J. & Huang, G. H. 2007 **Modeling the effects of elevation data resolution on the performance of topography-based watershed runoff simulation**. *Environ. Model. Softw.* **22** (9), 1250–1260.
- Wu, S., Li, J. & Huang, G. H. 2008 **Characterization and evaluation of elevation data uncertainty in water resources modeling with GIS**. *Water Resour. Manage.* **22** (8), 959–972.
- Yang, J., Reichert, P., Abbaspour, K. C., Xia, J. & Yang, H. 2008 **Comparing uncertainty analysis techniques for a swat application to the Chaohe Basin in China**. *J. Hydrol.* **358**, 1–23.
- Zhang, Z., Koren, V., Reed, S., Smith, M., Zhang, Y., Moreda, F. & Cosgrove, B. 2012 **SAC-SMA a priori parameter differences and their impact on distributed hydrologic model simulations**. *J. Hydrol.* **420**, 216–227.

First received 4 January 2016; accepted in revised form 9 June 2016. Available online 27 July 2016