Integration of soil hydraulic characteristics derived from pedotransfer functions into hydrological models: evaluation of its effects on simulation uncertainty

Wenchao Sun, Xiaolei Yao, Na Cao, Zongxue Xu and Jingshan Yu

ABSTRACT

Aimed at reducing simulation uncertainty of hydrological models in data-sparse basins where soil hydraulic data are unavailable, a method of estimating soil water parameters of soil and water assessment tool (SWAT) from readily available soil information using pedotransfer functions was introduced. The method was evaluated through a case study of Jinjiang Basin, China and was performed based on comparison between two model calibrations: (1) soil parameters estimated from pedotransfer functions and other parameters obtained from calibration; and (2) all parameters derived from calibration. The generalized likelihood uncertainty estimation (GLUE) was used as a model calibration and uncertainty analysis tool. The results show that information contained in streamflow data is insufficient to derive physically reasonable soil parameter values via calibration. The proposed method can reduce simulation uncertainty, resulting from greater average performance of behavioral parameter sets identified by GLUE. Exploring the parameter space reveals that the means of estimating soil parameters has little influence on other parameters. These facts indicate the decrease in uncertainty most likely results from a more realistic description of soil water characteristics than calibration. Thus, the proposed method is superior to calibration for estimating soil parameters of SWAT model when basin data are sparse.

Key words | generalized likelihood uncertainty estimation, hydrological modeling, pedotransfer function, simulation uncertainty, soil water characteristics

Wenchao Sun (corresponding author)
Xiaolei Yao
Zongxue Xu
Jingshan Yu
College of Water Sciences,
Beijing Normal University,
Xinjiekouwai Street 19,
Beijing 100875,
China
E-mail: *sunny@bnu.edu.cn*

Wenchao Sun
Zongxue Xu
Joint Center for Global Change Studies (JCGCS),
Beijing 100875,
China

Na Cao
Appraisal Center for Environment & Engineering,
Ministry of Environmental Protection,
Beijing 100012,
China

INTRODUCTION

Hydrological models that simulate water balance dynamics at river basin scale are indispensable for solving many engineering and environmental problems related to water (Merz & Blöschl 2004). With recent progress in geographic information system (GIS) and remote sensing technology, distributed models, which are more heavily parameterized for detailed spatial and temporal heterogeneity, are commonly used to provide decision support information to integrated water management (Sivapalan 2003). One challenge to use distributed hydrological models is minimizing prediction uncertainty, which depends on the complexity of model structure, the degree to which processes are abstracted or detailed, and the randomness of natural processes (Melching 1995). As a key component of the terrestrial water cycle, the state of soil water is a dominant control on runoff generation (Pietroniro *et al.* 2004). Correspondingly, soil information is considered critical input data to hydrological modeling (Mukundan *et al.* 2010;

doi: 10.2166/nh.2016.150

Bossa *et al.* 2012). In this context, the derivation of soil parameter values that reasonably reflect basin properties is crucial for reducing uncertainty in model simulation.

The soil and water assessment tool (SWAT; Arnold *et al.* 1998) is widely used to assess the impact of climate change and land cover variation on hydrology and the environment (e.g., Guo *et al.* 2008; Tu 2009; Singh *et al.* 2015). The physically based description of soil water balance in SWAT model requires detailed soil characteristics (e.g., available water capacity and saturated hydraulic conductivity) to describe each soil in a basin. Several studies (e.g., Muttiah & Wurbs 2002; Romanowicz *et al.* 2005) have reached a consensus that model simulation is sensitive to the deviation of soil parameters. The ideal way to obtain reasonable soil information is combining field survey with laboratory analysis. However, owing to limitations of cost and time, it is impractical to conduct field surveys in most cases. For modeling work in the USA, Wang & Melesse (2006) suggested that national databases, such as STATSGO (USDA-SCS (US Department of Agriculture-Soil Conservation Service) 1993) and SSURGO (USDA-NRCS (US Department of Agriculture-Natural Resources Conservation Service) 1995), are well suited to cases in which site-specific soil data are unavailable, because the two databases may provide sufficient information for general basin-scale modeling efforts. Furthermore, the STATSGO database has been incorporated into SWAT as the default dataset of soil information (Di Luzio *et al.* 2004).

Soil parameters are identified as sensitive parameters to runoff generation and streamflow simulation of SWAT by many studies (e.g., White & Chaubey 2005; Schuol *et al.* 2008). For model applications outside of the USA, where soil databases are unavailable, soil water characteristics, are usually identified from calibration against observed streamflow data (e.g. Akhavan *et al.* 2010; Ghaffari *et al.* 2010; Zang *et al.* 2012; Güngör & Göncü 2013). This has become easier with the popularization of the automatic calibration tool SWAT-CUP (Yang *et al.* 2008; available at: http://www.neprashtechnology.ca). This approach facilitates the model application in basins without detailed soil information. However, there are some limitations. Firstly, most hydrological models are overparameterized with respect to the limited soil information content in the streamflow records that are used for calibration (Beven & Binley 1992). Such an approach increases the number of parameters being calibrated, which could make the problem even worse, as the same information is used to calibrate more parameters. Secondly, the soil parameters in the SWAT model have explicit physical meaning. Nevertheless, automatic calibration procedures are purely numerical processes that seek to optimize the value of an objective function (Duan *et al.* 1992). The question that needs to be addressed here is whether the numerically optimized parameter sets can reflect the soil hydraulic properties in the target basin.

Extensive knowledge of soil water and its variability with soil characteristics has been gained in soil science research (Van Genuchten & Leij 1992). Several methods estimate soil water hydraulic characteristics from readily available physical parameters using pedotransfer functions (e.g., Gupta & Larson 1979; Vereecken *et al.* 1989; Rawls *et al.* 1992; Gijsman *et al.* 2002; Saxton & Rawls 2006). In this study, soil water characteristics required by SWAT were estimated from soil texture and organic matter by the method proposed in Saxton & Rawls (2006), which is expected to improve simulation, compared to a set derived from serial model calibration against streamflow data. The main objective of the present work was to evaluate the feasibility of incorporating this soil parameter estimation method into SWAT model simulation. The method may be useful in data-sparse basins where detailed information on soil hydraulic properties is unavailable or difficult to collect.

Although this scheme may yield more physically sound estimates of soil parameters than the numerically optimal values obtained from automatic calibration, spatial variability in hydraulic characteristics and model errors from pedotransfer functions may introduce their own uncertainty to model simulation. In many cases, field survey data of soil hydraulic characteristics are unavailable, which makes it impossible to evaluate error associated with pedotransfer functions. When applying hydrological models to decision making in water resource management and planning, hydrological simulation uncertainty is always a major concern (Sellami *et al.* 2013). Considering the aforementioned issues, in this study the evaluation is focused on analyzing changes of such uncertainty when using the pedotransfer functions.

A case study was carried out for Jinjiang Basin, China. The evaluation was performed through a comparison, using soil parameter values obtained from an automatic calibration based on hydrological data. We conducted two calibrations for SWAT modeling of Jinjiang Basin. The first was

considered a benchmark calibration. All model parameters, including soil parameters, were calibrated against streamflow data using an automatic calibration method. In the second, the soil parameters were specified by the proposed method, whereas the other model parameters were calibrated against streamflow data using the same automatic calibration method. Such comparison is difficult, because the calibration result is not only determined by calibration data but by settings of the automatic optimization scheme. For example, selection of the objective function may affect the identification of parameters (e.g., Freer & Beven 1996) and thereby performance of the calibrated model (e.g., Krause *et al.* 2005). In this context, an important issue is how to ensure the environments of the two calibrations are identical, except for the condition as to whether the soil parameters are derived from pedotransfer functions or from calibration against streamflow data as for other model parameters. To satisfy such a requirement, the evaluation was accomplished using generalized likelihood uncertainty estimation (GLUE) (Beven & Binley 1992; Freer & Beven 1996), the automatic calibration and uncertainty analysis tool for the two aforementioned calibrations. The paper is organized as follows. The SWAT model and study basin are introduced in the next section, followed by the approach to estimate soil water characteristics from pedotransfer functions. Then, evaluation strategies are described. Finally, feasibility of the proposed approach is discussed and conclusions drawn.

MATERIALS AND METHODS

Model description

The SWAT model is a continuous (daily-step) distributed model that simulates hydrological processes, fate and transport of sediment and pollutants within a basin. Based on a digital elevation model (DEM), the basin is discretized into a number of sub-basins. Then each sub-basin is further divided into several unique hydrological response units, according to differences in soil and land use. For simulation of hydrological processes in the land phase, the SCS curve number method is used to compute generated runoff volume, and channel flow is routed using the Muskingum or variable storage methods. Soil information needed by SWAT can be separated into two groups. First are physical properties such as soil particle size distribution and soil hydraulic characteristics. The second are chemical properties such as initial NO_3 concentration and soluble phosphorus, P. We focused on the estimation of three key soil hydraulic parameters: available water capacity (*SOL_AWC*), saturated hydraulic conductivity (*SOL_K*), and bulk density (*SOL_BD*), which are indexed to soil texture and organic matter.

The basin and data availability

Jinjiang is a coastal basin on the west side of the Taiwan Strait. The entire basin area is within the city of Quanzhou in Fujian Province, China. The river has two major tributaries, the Xixi and Dongxi, which join at Shuangxikou. Basin area is 5,629 km^2, which embraces a mountainous area in the northwest and a low plain in the southeast. Elevation varies from 50 to 1,366 m. The dominant land use types are forest and cropland. The basin has a subtropical monsoon climate characterized by a dry winter and rainy summer. Annual precipitation ranges from 1,000 to 1,800 mm, of which 80% occurs from March through September. A water intake infrastructure at Jinji sluice, several kilometers downstream of the Shilong gauging station, contributes greatly to the water supply of Quanzhou. The hydrological simulation was conducted for the area upstream of the Shilong station, for the period 2005–2009.

Input rainfall data and streamflow data were provided by the Quanzhou City Water Authority. Locations of gauging stations are shown in Figure 1. The input GIS data include a DEM derived from ASTER GDEM (30 meter resolution; http://gdem.ersdac.jspacesystems.or.jp), land cover data derived from Global Land Cover Characteristics Data Base Version 2.0 (1 km resolution; http://edc2.usgs.gov/glcc/tabgeo_globe.php), $1{:}10^8$ scale soil map derived from the Chinese Soil Scientific Database (CSSD) (http://www.soil.csdb.cn). Soil property data were also acquired from CSSD and contain soil particle distribution and chemical properties for soil profiles all across China.

Soil data preprocessing

Preprocessing for soil data is necessary for both SWAT modeling and soil hydraulic characteristic estimation. The soil map and soil properties dataset derived from the CSSD are

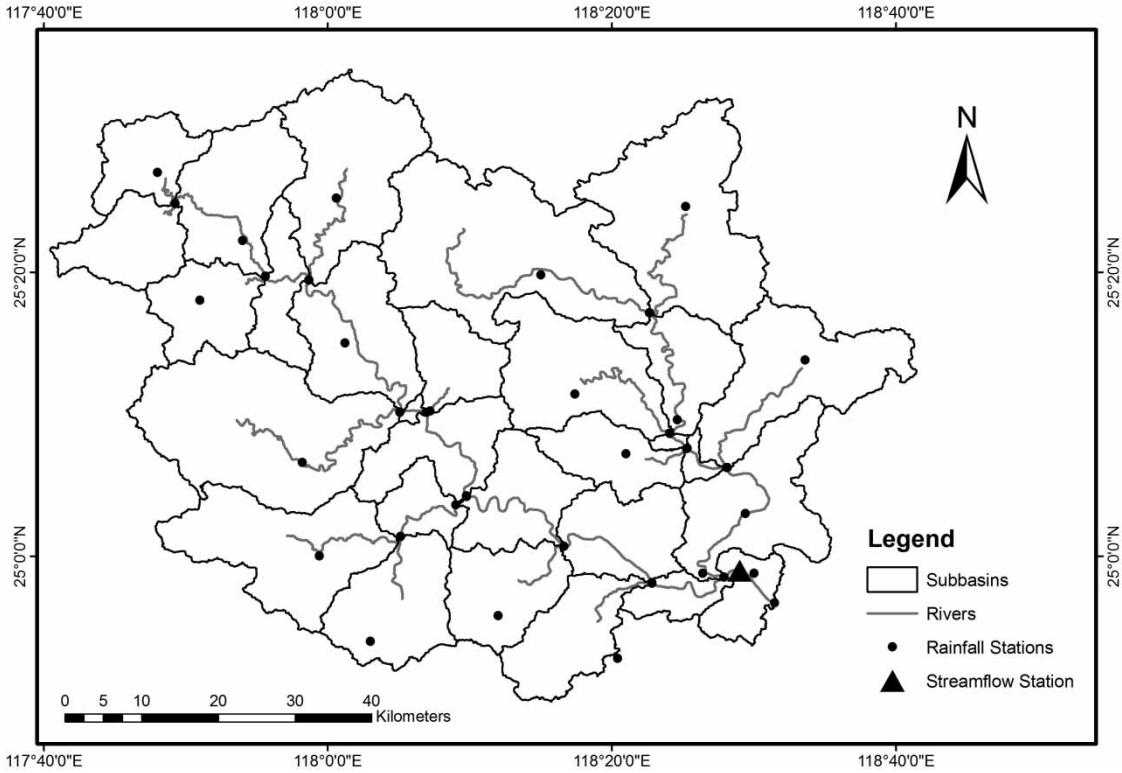

Figure 1 | Sub-basins being defined for hydrological modeling, streamflow and rainfall gauging stations.

based on a soil genetic classification system. From high to low level, soil group, soil subgroup and soil species are three soil categories involved. The physical properties data are at soil species level, whereas the soil map only contains soil spatial distribution at the soil subgroup level. The properties of each map unit (at soil subgroup level) are extracted from the dataset by finding the record (at soil species level) in the dataset at which the location description best matches the spatial distribution of the map unit. Each record usually contains observations from more than one soil profile. Arithmetic means of each type of observation in these soil profiles are used as properties for the corresponding map unit. To keep the number of soil parameters in SWAT model at a manageable level, all map units belonging to the same soil group are merged. In other words, map units are aggregated into the soil group level. Then, soil properties of the soil subgroup of largest area in a soil group are treated as the properties of the soil group. The soil particle distribution is converted from the international classification system used in CSSD into the US Department of Agriculture (USDA) classification system, as required by SWAT model. The spatial distribution and properties of the three soil types in Jinjiang Basin are shown in Figure 2 and Table 1.

Soil water characteristics estimation method

Soil water characteristic equations developed by Saxton & Rawls (2006) were selected for our study. These represent an update of the method proposed by Saxton *et al.* (1986), which has been successfully applied in agricultural hydrology and water management. The equations are derived from a USDA soil database based on regression analysis between soil water retention data and readily available soil information (soil texture and organic matter), for improved evaluation of soil water movement. *SOL_AWC*, *SOL_K* and *SOL_BD* of SWAT model for each soil type in Jinjiang Basin were estimated as follows.

SOL_AWC

SOL_AWC (mm H_2O/mm soil) is the fraction of water between the field capacity and permanent wilting point, which is soil moisture at tension 33 kPa θ_{33} (volumetric

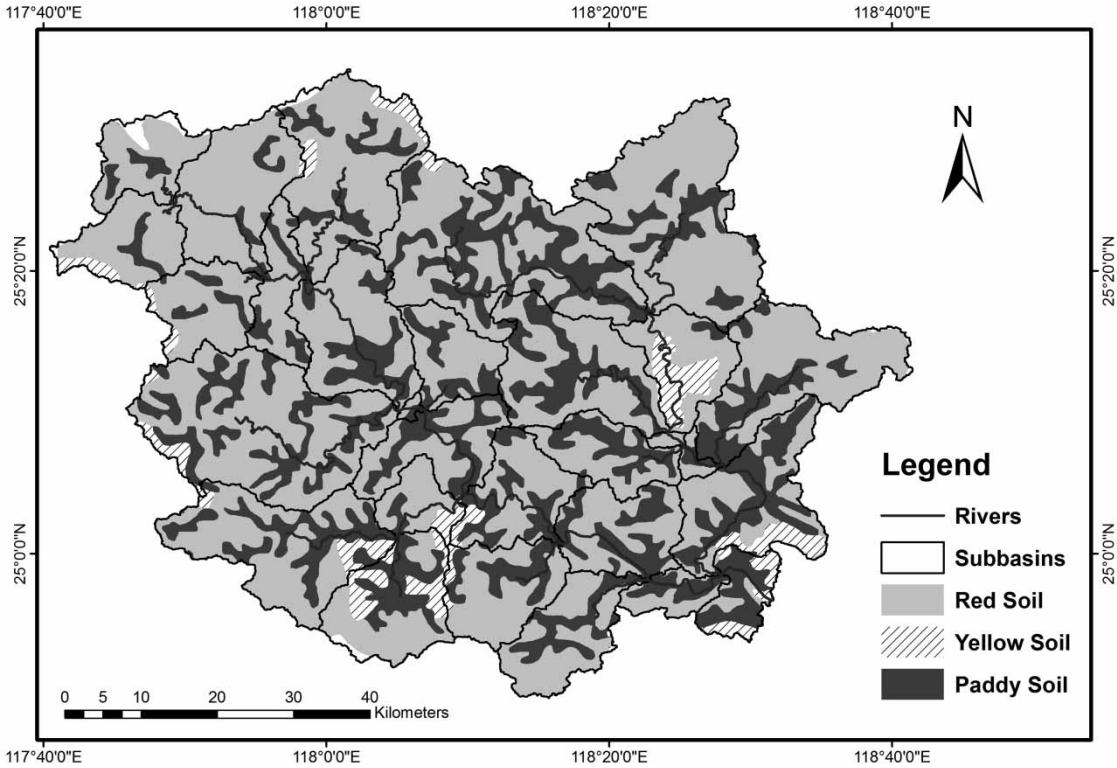

Figure 2 | Spatial distributions of three soil group types.

Table 1 | Properties of three main soil types in Jinjiang Basin

Soil group name	Percentage of area (%)	Clay (%)	Silt (%)	Sand (%)	Organic matter (%)
Red soil	67	45.7	23.8	30.5	6.05
Paddy soil	28	10.6	17.0	72.4	2.78
Yellow soil	5	27.8	24.1	48.1	12.8

percentage) and 1,500 kPa θ_{1500} (volumetric percentage), respectively:

$$SOL\_AWC = \theta_{33} - \theta_{1500} \quad (1)$$

To obtain θ_{33}, an intermediate value θ_{33t} is computed based on a relationship derived from a multivariable linear analysis:

$$\theta_{33t} = -0.251S + 0.195C + 0.011OM + 0.006(S \times OM) - 0.027(C \times OM) + 0.452(S \times C) + 0.299 \quad (2)$$

where S and C are volumetric percentages of sand and clay, and OM is the percentage of organic matter on a weight basis. To compensate the situation in which some variables may not be linearly correlated with the dependent variables, θ_{33t} is corrected based on a relationship derived from a second regression analysis:

$$\theta_{33} = \theta_{33t} + [1.283(\theta_{33t})^2 - 0.374\theta_{33t} - 0.015] \quad (3)$$

θ_{1500} is estimated via a two-step method similar to θ_{33}, with an intermediate value θ_{1500t}:

$$\theta_{1500t} = -0.024S + 0.487C + 0.006OM + 0.005(S \times OM) - 0.013(C \times OM) + 0.068(S \times C) + 0.031 \quad (4)$$

$$\theta_{1500} = \theta_{1500t} + (0.14 \times \theta_{1500t} - 0.02) \quad (5)$$

SOL_K

Hydraulic conductivity is a nonlinear function of volumetric soil water content (Rawls *et al.* 1993). In this study, *SOL_K* (mm/hr) was computed from a power function of moisture held at low tensions:

$$SOL\_K = 1930(\theta_s - \theta_{33})^{(3-\lambda)} \quad (6)$$

where θ_s is soil moisture at 0 kPa tension (saturation) and λ is slope of the logarithmic tension – moisture curve. θ_s is computed based on soil moisture at tension $0 - 33$ kPa $\theta_{(S-33)}$ (volumetric percentage), θ_{33}, and S:

$$\theta_s = \theta_{33} + \theta_{(S-33)} - 0.097S + 0.043 \quad (7)$$

$\theta_{(S-33)}$ is estimated in a two-step method similar to θ_{33}, with an intermediate value $\theta_{(S-33)t}$:

$$\theta_{(S-33)t} = 0.278S + 0.034C + 0.022OM - 0.018(S \times OM) - 0.027(C \times OM) - 0.564(S \times C) + 0.078 \quad (8)$$

$$\theta_{(S-33)} = \theta_{(S-33)t} + (0.636 \times \theta_{(S-33)t} - 0.107) \quad (9)$$

The slope of logarithmic tension – moisture curve λ is computed as

$$\lambda = [ln(\theta_{33}) - ln(\theta_{1500})]/[ln(33) - ln(1500)] \quad (10)$$

SOL_BD

SOL_BD is estimated from θ_s, assuming particle density 2.65 (g/cm^3):

$$SOL\_BD = (1 - \theta_S) \times 2.65 \quad (11)$$

In summary, to estimate the three soil parameter values, *S*, *C* and *OM* of each soil type in Jinjiang Basin are needed. These input data are acquired from the preprocessed soil data. Soil characteristics were estimated by SPAW software (http://hrsl.ba.ars.usda.gov/SPAW/Index.htm). Estimated parameter values are listed in Table 2.

Table 2 | Soil parameter values estimated by pedotransfer functions

Soil group name	SOL_AWC (mm/mm)	SOL_BD (g/cm^3)	SOL_K (mm/hr)
Red soil	0.131	1.38	2.2
Paddy soil	0.059	1.54	45.5
Yellow soil	0.150	1.14	17.4

Evaluation strategy

Effectiveness of the proposed soil parameter estimation method was assessed based on performance of SWAT hydrological simulation in the target basin. Table 3 lists the SWAT parameters considered. The first ten parameters have been regularly calibrated in studies from the literature (e.g., Yang *et al.* 2008; Li *et al.* 2010; Shen *et al.* 2012). The last three parameters in Table 3 describe soil hydraulic characteristics. As three soil types were involved in SWAT simulation, a total of nine soil parameters (three soil types × three soil parameters) were addressed. Two model calibrations were executed. In the first calibration, all the aforementioned 19 parameters (i.e., first ten parameters in Table 3 and the nine soil

Table 3 | SWAT model parameters

Name	Description	Initial range
CN2	SCS runoff curve number	20–90
ALPHA_BF	Baseflow recession coefficient	0–1
GW_DELAY	Groundwater delay time (days)	30–450
GWQMN	Threshold water level in shallow aquifer for base flow	0–2
GW_REVAP	Groundwater evaporation coefficient	0–0.2
ESCO	Soil evaporation compensation coefficient	0.8–1
CH_N2	Manning coefficient for the mail channel	0–0.3
CH_K2	Hydraulic conductivity in main channel (mm/hr)	5–130
ALPHA_BNK	Bank flow recession coefficient	0–1
SFTMP	Snowfall temperature (°C)	−5–5
SOL_AWC	Available soil water capacity (mm H_2O/mm soil)	0–1
SOL_BD	Soil moist bulk density (g/cm^3)	1.1–2.5
SOL_K	Soil Saturated hydraulic conductivity (mm/hr)	0–2,000

parameters) were obtained by calibration against streamflow data using an automatic optimization scheme. In the second calibration, the nine soil parameters were attained via the pedotransfer functions, whereas the other ten parameters were derived by model calibration using the above scheme. The aforementioned two calibrations are hereafter referred to as CAL_19 and CAL_10, respectively. The model simulation corresponding to CAL_19 was treated as a benchmark. The differences in model simulation results corresponding to the two calibrations are considered to be caused by the way of deriving soil parameters (i.e., through calibration or pedotransfer functions).

GLUE was selected as the automatic calibration and uncertainty analysis tool. It was run in the framework of SWAT-CUP (http://www.neprashtechnology.ca/), a computer program for calibration of SWAT model. Considering equifinality in model simulation, which is the phenomenon in which very different parameters give similar model predictions, GLUE does not assume that only one optimal parameter set exists. Instead, it treats all parameter sets for which model performance exceeds a certain threshold as good ones, which are called 'behavioral' parameter sets. Then, an ensemble simulation is run using all behavioral parameter sets. One advantage of GLUE is that a modeler's subjective options are made explicit and the suitability of any one can be examined (Beven & Freer 2001). For the two calibrations in our study, all settings for GLUE implementation were made the same, except that the calibrated parameters were different. This permits the differences in model simulations results purely from the different approaches to specify soil parameters. GLUE was implemented for CAL_19 as follows.

1. Generate random samples from the entire parameter space. Random parameter sets were generated using the Latin hypercube sampling method, assuming the *a priori* parameter distribution to be uniform, which is a common assumption when parameter distribution information is unavailable (e.g., Beven & Freer 2001; Hailegeorgis & Alfredsen 2016). Initial ranges of the model parameters are specified in Table 3. For CAL_19, one parameter set includes one randomly generated value for each of the 19 parameters calibrated. In total, 10,000 parameter sets were generated.
2. Calculate likelihood values of each parameter set and select behavioral ones. Every set was input to SWAT for model simulation. The degree to which a parameter set could reflect basin reality was assessed through evaluation of streamflow simulation at basin outlet. The Nash–Sutcliffe model efficiency coefficient (NSE) was selected as the likelihood measure:

$$NSE = 1 - \frac{\sum (Q_{obs,i} - Q_{sim,i})}{\sum (Q_{obs,i} - Q_{obs,avg})} \tag{12}$$

where $Q_{obs,i}$ (m^3/s) and $Q_{sim,i}$ (m^3/s) are observed and simulated streamflows at Shilong station for time step i, and $Q_{obs,avg}$ (m^3/s) is the average observed streamflow for the simulation period. The threshold for rejecting parameter sets as non-behavioral ones was 0.7, which means that parameter sets for which NSE reached 0.7 were retained for ensemble simulation.
3. Calculate posterior likelihood for behavioral parameter sets. Conditioned by streamflow observations, the likelihood was updated based on application of the Bayes equation in the form:

$$L_p[\theta|Q_{obs}] = CL[\theta|Q_{obs}]L_o[\theta] \tag{13}$$

where $L_o[\theta]$ is the prior likelihood weight for parameter set θ, which was the same for all behavioral sets, $L[\theta|Q_{obs}]$ is the likelihood value calculated in step two, $L_p[\theta|Q_{obs}]$ is the posterior likelihood weight conditioned by streamflow observations Q_{obs}, and C is a scaling constant ensuring that the sum of $L_p[\theta|Q_{obs}]$ for all behavioral sets was equal to unity.
4. Calculate uncertainty quantiles. The cumulative distribution of the predictions weighted by likelihood was calculated by:

$$P_t(Q_t < q) = \sum_{i=1}^{m} L_p[\theta_i|Q_{t,i} < q] \tag{14}$$

where $P_t\ (Q_t < q)$ is the cumulative probability of predicted streamflow Q_t less than arbitrary value q at time step t, $L_p[\theta_i]$ is the posterior likelihood of parameter set θ_i for which the prediction at t $Q_{t,i}$ is less than q, and m is the total number of parameter sets satisfying the

condition $Q_{t,i} < q$. From this cumulative probability distribution, a lower 2.5% and upper 97.5% quantile of simulated streamflow were obtained at every t. These 95% simulation intervals for all time steps form the uncertainty band of ensemble simulation.

For CAL_10, before calibration, the nine soil parameters were estimated by pedotransfer function. Then, the same 10,000 combinations of randomly generated values for the remaining ten parameters (first ten parameters in Table 3) were used as parameter sets for the application of GLUE. Other settings of GLUE are same as CAL_19. The differences of deriving model parameters in CAL_19 and CAL_10 are also described in Table 4.

The model was calibrated using hydrological data from Shilong Station for 2005–2007. Then the behavioral parameter sets identified by GLUE were used for simulation of 2008–2009, for the purpose of model validation. Several indices were used to quantify model performance and simulation uncertainty for both calibrations. *NSE* of simulated streamflow by behavioral parameter sets at the 50% quantile was used to represent best performance of ensemble simulation. Simulation uncertainty was quantified by combination of two indices. The *P-factor* is the percentage of observations embraced by the 95% prediction intervals. The *R-factor* is a measure of the average width of the 95% prediction intervals:

$$R\_factor = \frac{\sum_{i=1}^{m} (Q_{97.5\%,i} - Q_{2.5\%,i})}{m \times \sigma_{Q_{obs}}} \quad (15)$$

Here, $Q_{97.5\%,i}$ and $Q_{2.5\%,i}$ are the 97.5 and 2.5% quantiles of simulated streamflow at time step i, m is the total time step of simulation, and σ_{Qobs} is the standard deviation of streamflow observations. A larger *P-factor* accompanied by a smaller *R-factor* indicates less simulation uncertainty. The comparison between the two calibrations, i.e., the two soil parameter estimation strategies, was done based on these indices.

Table 4 | Schemes for deriving model parameters in CAL_19 and CAL_10

	Soil parameters	Other parameters
CAL_19	Model calibration	Model calibration
CAL_10	Pedotransfer functions	Model calibration

RESULTS AND DISCUSSION

Posterior distributions of soil parameters in CAL_19

Apart from the input–output behavior of the model, exploring parameter space response to change of the soil parameter estimation method is valuable. This is because whether the identified parameters could reflect basin reality is critical if the model is expected to estimate the effects of perturbations to the structure of the hydrological system (Gupta *et al.* 2005). The number of identified behavioral parameters for CAL_19 is 4776. Figure 3 shows posterior parameter distributions of soil parameters derived from CAL_19. Deviation from the original assumed uniform distribution is deemed an indication of parameter sensitivity. Constraints of streamflow data on the three soil hydraulic parameters were negligible, because the posterior distributions were almost all uniform. This is similar to the calibration results of Shen *et al.* (2012), in which soil hydraulic parameters were also derived from model calibration against streamflow data. This implies that the amount of information in the calibration data is inadequate to identify these soil parameters effectively, and calibration is not the best choice to deriving soil parameter values. This justifies the use of new methods to estimate soil parameters, such as the pedotransfer function approach proposed herein.

Comparison of simulation uncertainty between CAL_19 and CAL_10

As observations of soil water state variables are unavailable for Jinjiang Basin, effectiveness of the proposed soil hydraulic parameter estimation method was evaluated according to the performance of ensemble streamflow simulation, which is a temporally and spatially integrated indicator of basin hydrological behavior. The ensemble simulations and corresponding model performance criteria of CAL_10 and CAL_19 for the calibration period are shown in Figure 4 and Table 5. It is evident that variations in the observed hydrograph were reasonably reproduced by ensemble

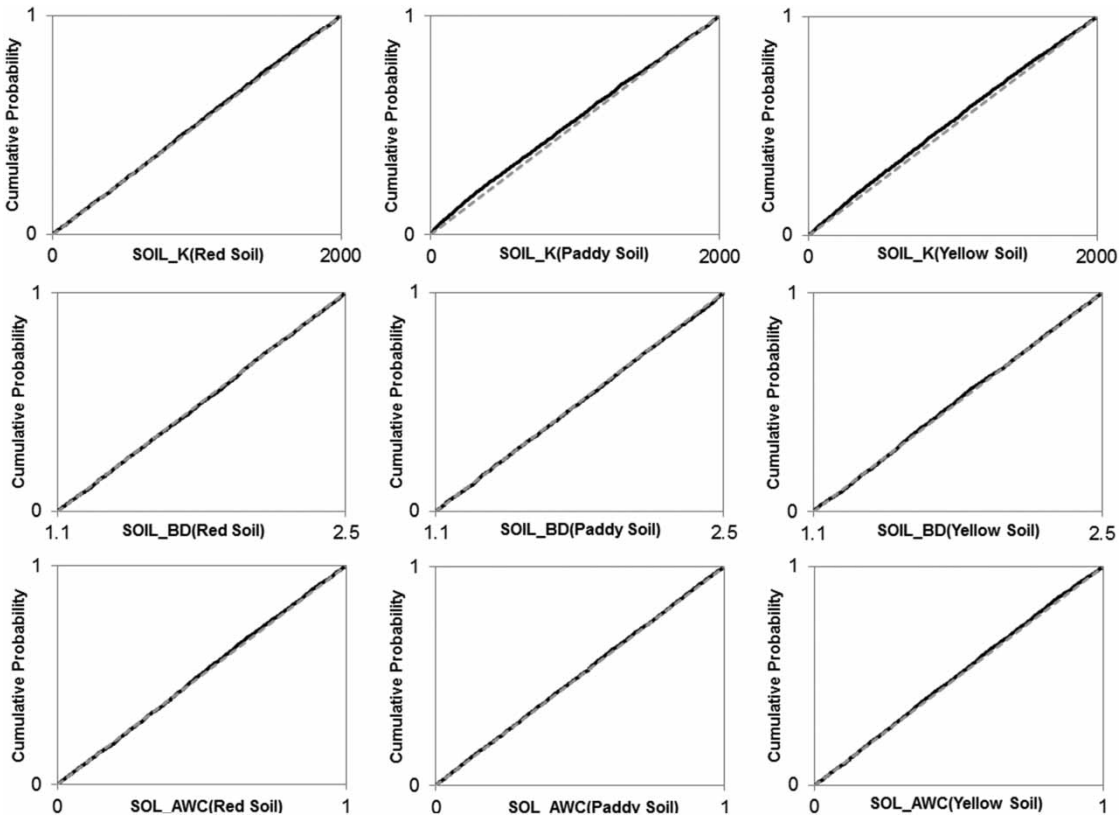

Figure 3 | Posterior distributions of SWAT soil parameters derived from CAL_19 (black lines) and prior uniform distributions (gray lines).

simulations and the best performances of ensemble simulations corresponding to the two calibrations are satisfactory, judging from *NSE*. Meanwhile, the best performance of CAL_10 was superior to CAL_19. Judging from *P-factor* and *R-factor* values, 8% more observations were included by the uncertainty band of CAL_10, and its width was narrower than that of CAL_19. Similar results were obtained for the validation period, as demonstrated in Figure 5 and Table 6; the *NSE* of best performance was greater for CAL_10. In this case, 25% more observations were covered by the uncertainty band of CAL_10, with a narrower width than CAL_19. All these findings indicate that simulation uncertainty of CAL_10 is less than that of CAL_19.

Possible reasons for reduction of simulation uncertainty

To explore the reason for reduction of simulation uncertainty when using pedotransfer functions, it is valuable to investigate *NSE* distributions of streamflow simulations produced by behavioral parameter sets identified in CAL_19 and CAL_10. Figure 6 depicts histograms of *NSE* values for the calibration period. It is revealed that compared with CAL_19, the *NSE* value corresponding to the distribution peak for CAL_10 is larger, and that number of parameter sets within the large-value range ($NSE > 0.8$) of the x-axis is greater. From this, it is clear that better average performance of behavioral parameter sets reduced simulation uncertainty for CAL_10. Applying behavioral parameter sets to the validation period and then computing the *NSE* of each set based on the difference between observed and simulated streamflow, we derived *NSE* distributions for the two calibrations (Figure 7). It is understandable that the variation in performances of behavioral sets was greater and average performance was poorer than that of the calibration period, because streamflow data for the validation period were not used in model calibration. Differences between the two distributions in

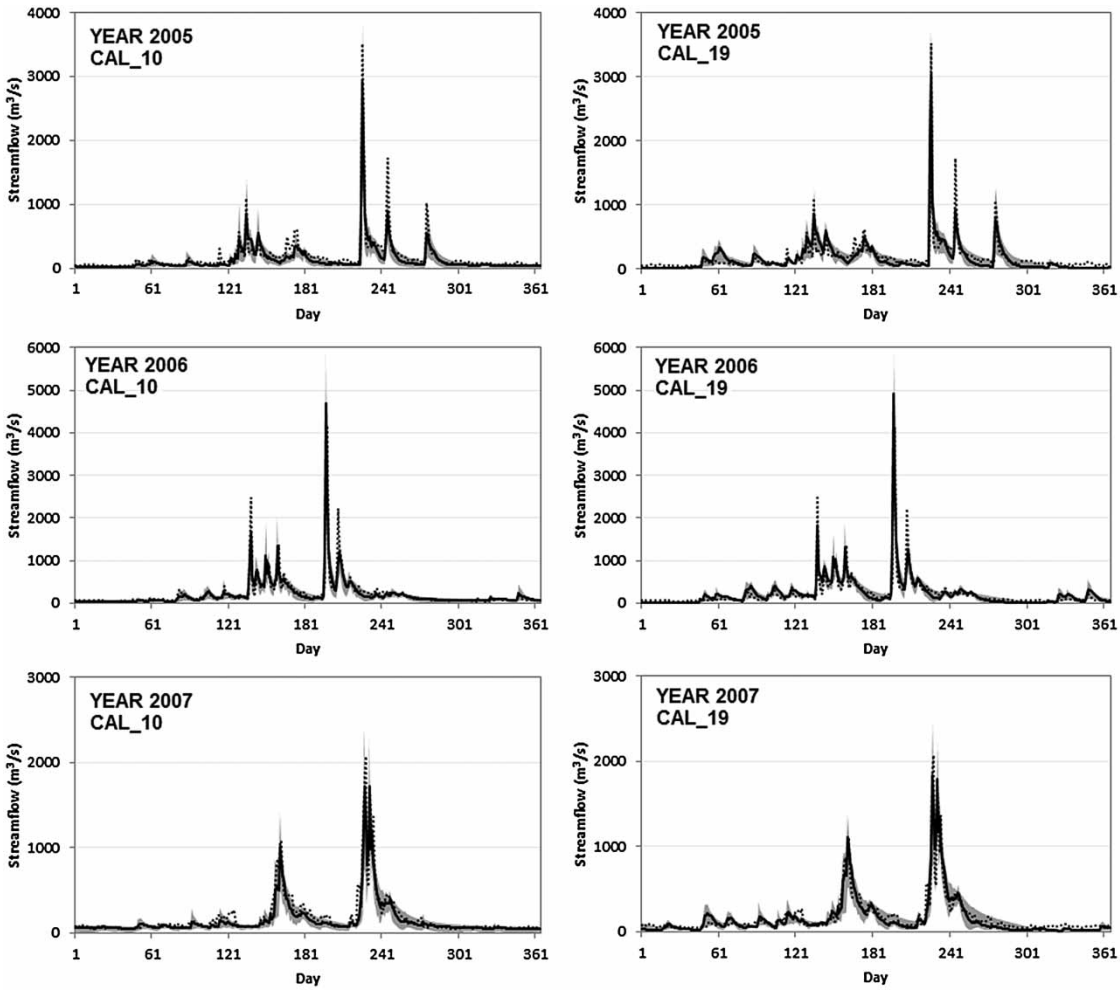

Figure 4 | Comparison of simulated streamflow by behavioral parameter sets obtained from CAL_10 and CAL 19 for calibration period (2005–2007). Dashed lines: observed streamflow; gray band: 95% uncertainty band of ensemble simulation; solid lines: best simulation of ensemble prediction.

Table 5 | Model performance criteria of calibration period (2005–2007)

	Number of behavioral parameter sets	NSE of best simulation	P-factor	R-factor
CAL_19	4,776	0.83	69%	0.59
CAL_10	6,256	0.85	77%	0.57

Figure 7 are similar to the calibration period, indicating that for a similar reason, simulation uncertainty of CAL_10 was less than CAL_10 in the validation period. Because settings of the hydrological modeling and GLUE were identical, except for the derivation of soil hydraulic parameters, the differences in ensemble simulations of the two calibrations originated solely from differences of the constrained parameter space.

Posterior distributions of the ten parameters calibrated in both CAL_10 and CAL_19 were compared. GLUE settings were exactly the same except for the number of calibrated parameters. Therefore, it is understandable that the differences in the distributions of each parameter resulted from the approach to specify soil parameters. The posterior distributions of *ALPHA_BF*, *GW_DELAY*, *GWQMN*, *GW_REVAP*, *ESCO* and *SFTMP* (Table 3) for both cases are uniform. Visually detectable differences were observed for parameters identified as sensitive in CAL_19, i.e., *CH_N2*, *CH_K2*, *ALHPA_BNK* and *CN2*

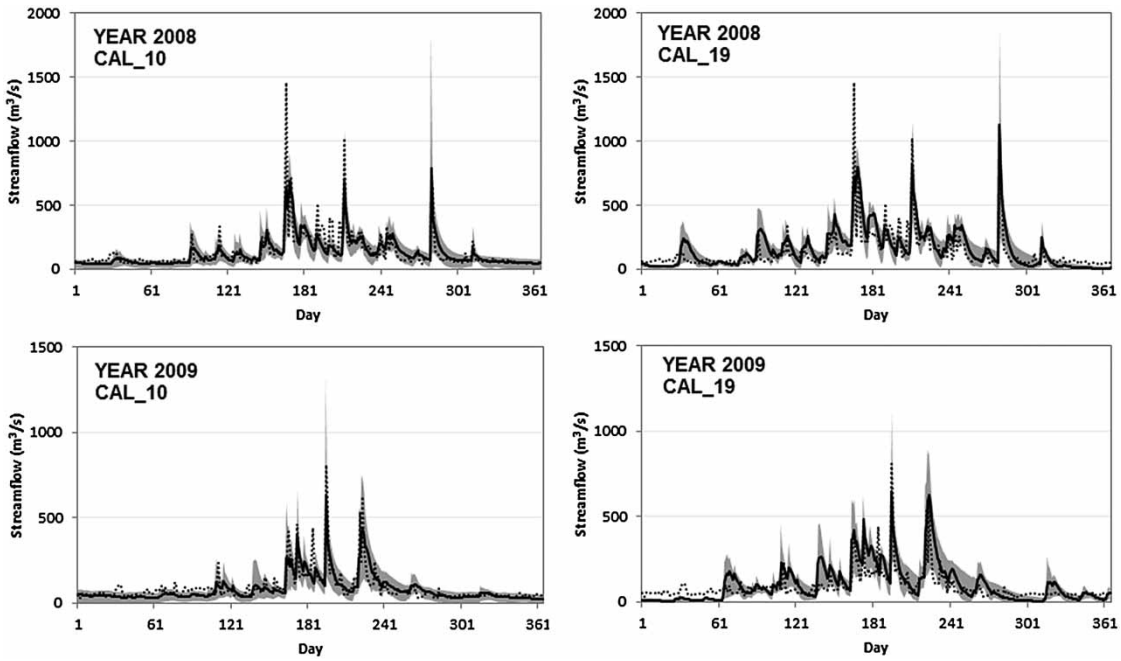

Figure 5 | Comparison of simulated streamflow by the behavioral parameter sets obtained from CAL_10 and CAL 19 for validation period (2008–2009). Dashed lines: observed streamflow; gray band: 95% uncertainty band of ensemble simulation; solid lines: best simulation of ensemble prediction.

Table 6 | Model performance criteria of validation period (2008–2009)

	NSE of best performance	*P*-factor	*R*-factor
CAL_19	0.65	59%	1.16
CAL_10	0.70	84%	1.09

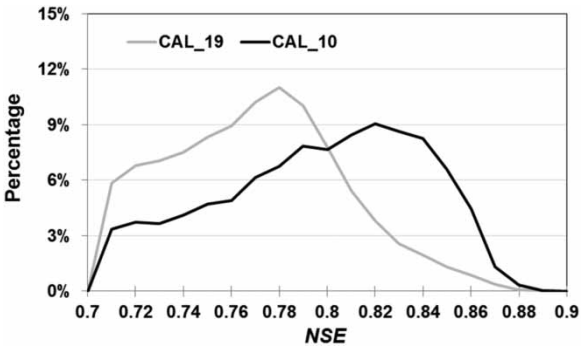

Figure 6 | NSE distributions for streamflow simulations made by behavioral parameter sets of CAL_19 and CAL_10 for calibration period (2005–2007).

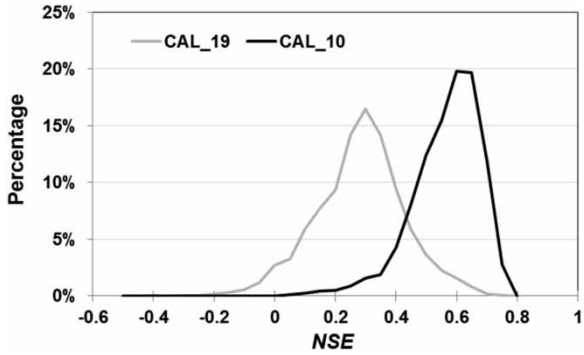

Figure 7 | NSE distributions for streamflow simulations made by behavioral parameter sets of CAL_19 and CAL_10 for validation period (2008–2009).

(Figure 8). However, the differences are not significant, because the posterior distributions of those four sensitive parameters did not change dramatically. But further interpretation is limited because GLUE only works with parameter sets and display of the distribution of individual parameter has value only in evaluating the sensitivity of that parameter (Beven & Freer 2001).

Constraints of the two calibrations on parameter space were further explored from the standpoint of parameter correlation. Ideally, model parameters are assumed to be mutually independent. However, parameter correlation is usually found in hydrological modeling and can be a source of modeling uncertainty (Blasone & Vrugt 2008). Therefore, examining parameter correlation can help to

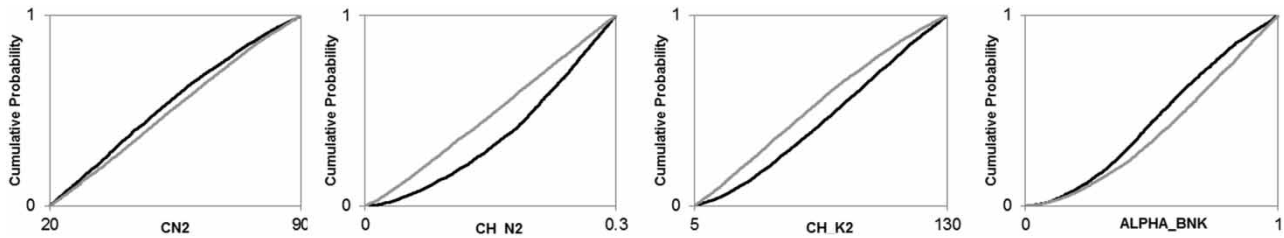

Figure 8 | Comparison of posterior parameter distributions between CAL_10 (gray lines) and CAL_19 (black lines).

explore the reason for changes in simulation uncertainty when using soil parameter values estimated from pedotransfer functions. Correlation matrices of CAL_10 and CAL_19 are shown in Tables 7 and 8, respectively. Patterns of parameter correlation for the two cases are similar, i.e., most correlations were weak and only a very few were significant at the 0.01 confidence level. A similar weak correlation pattern was detected by Yang *et al.* (2008). They concluded that this phenomenon arises from the inherent assumptions in GLUE, which tend to flatten the true response surface by removing sharp peaks and valleys. In our study, the parameter correlation patterns from CAL_10 and CAL_19 were not much different. Combining this fact with the analysis results concerning posterior parameter distribution, it is implied that the effects of incorporating soil parameter values derived from pedotransfer functions on other calibrated parameters are minor. Therefore, the differences in parameter space mainly stem from soil hydraulic parameters. More specifically, the reduction in simulation uncertainty most likely originated from the more reasonable description of soil hydraulic characteristics by the estimates obtained from pedotransfer functions than those obtained from calibration against streamflow data. This reveals the necessity of implementing the proposed method to SWAT hydrological modeling in data-sparse basins that lack field survey data of soil hydraulic characteristics.

CONCLUSIONS

This study examined the value of specifying three soil hydraulic parameters in the SWAT hydrological model through estimation from soil texture and organic matter using pedotransfer functions. Considering that the method was designed for use in basins where field survey data of soil hydraulic information are unavailable, the evaluation was accomplished through a comparison with calibrating soil parameters together with other model parameters using streamflow data. The calibrations were carried out using the GLUE scheme for avoiding the influence of the automatic calibration method itself on model simulation. In the case study, from posterior parameter distributions, it was shown that the

Table 7 | Correlation matrix of posterior parameter distribution obtained from CAL_19 (bold text indicants significant at 0.01 level)

	CN2	ALPHA_BF	GW_DELAY	GWQMN	GW_REVAP	ESCO	CH_N2	CH_K2	ALPHA_BNK	SFTMP
CN2	1									
ALPHA_BF	0.004	1								
GW_DELAY	0.001	−0.000	1							
GWQMN	0.020	−0.020	0.012	1						
GW_REVAP	−0.009	0.018	0.015	0.015	1					
ESCO	−0.016	−0.030	−0.016	0.018	0.006	1				
CH_N2	**0.133**	0.010	0.005	0.005	−0.014	−0.002	1			
CH_K2	**0.092**	0.027	0.008	−0.001	0.009	**0.043**	**−0.339**	1		
ALPHA_BNK	**−0.103**	0.014	0.002	−0.014	0.004	0.021	**0.127**	**0.328**	1	
SFTMP	−0.024	0.010	−0.013	−0.004	0.015	0.013	−0.020	0.060	0.02	1

Table 8 | Correlation matrix of the posterior parameter distribution obtained from CAL_10 (bold text indicants significant at 0.01 level)

	CN2	ALPHA_BF	GW_DELAY	GWQMN	GW_REVAP	ESCO	CH_N2	CH_K2	ALPHA_BNK	SFTMP
CN2	1									
ALPHA_BF	−0.021	1								
GW_DELAY	0.000	−0.011	1							
GWQMN	−0.007	0.013	0.005	1						
GW_REVAP	0.006	0.014	0.008	−0.010	1					
ESCO	−0.130	0.013	0.017	−0.017	0.007	1				
CH_N2	**0.213**	0.003	0.004	−0.018	−0.009	0.021	1			
CH_K2	**0.182**	0.002	−0.008	−0.007	0.008	0.006	**−0.174**	1		
ALPHA_BNK	**−0.101**	0.007	0.034	0.003	0.014	−0.023	**0.121**	**0.288**	1	
SFTMP	0.002	0.003	−0.004	0.008	0.014	0.07	−0.009	−0.004	−0.016	1

streamflow data could not sufficiently constrain the soil hydraulic parameters. This indicated that the information contained in streamflow calibration data is not adequate to derive physically reasonable soil parameter values. From the perspectives of both posterior distribution and parameter correlation, it was seen that the influence of the soil parameter estimation method on other calibrated parameters was minor. However, comparisons of streamflow simulations demonstrated that simulation uncertainty was reduced by the proposed pedotransfer method, which comes from better average performance of the behavioral parameters sets. Together, these findings indicate that the most probable reason for reduction of simulation uncertainty was a more reasonable description of soil hydraulic characteristics. From the successful application to Jinjiang Basin, it is concluded that in data-sparse basins where texture and organic matter data from the CSSD are the only available soil information, compared with using soil parameter values derived from model calibration, less simulation uncertainty of SWAT model can be achieved by incorporating estimates from the proposed method. To achieve a more general understanding of the feasibility of this method, it must be tested intensively in additional Chinese basins with various climatic and geophysical conditions.

ACKNOWLEDGEMENTS

This study was supported by the National Natural Science Foundation of China (Grant Nos 41201018, 91125015), the National Key Technology R&D Program (Grant No. 2013BAB05B04), the Non-profit Industry Financial Program of Ministry of Water Resources of China (Grant No. 201401036) and the Fundamental Research Funds for the Central Universities. The soil data were provided by the Soil Sub-Center, Institute of Soil Science, Chinese Academy of Sciences (Chinese Soil Scientific Database, http://www.soil.csdb.cn/).

REFERENCES

Akhavan, S., Abedi-Koupai, J., Mousavi, S. F., Afyuni, M., Eslamian, S. S. & Abbaspour, K. C. 2010 Application of SWAT model to investigate nitrate leaching in Hamadan-Bahar Watershed, Iran. *Agr. Ecosyst. Environ.* **139**, 675–688.

Arnold, J. G., Srinivasan, R., Muttiah, R. S. & Williams, J. R. 1998 Large area hydrologic modeling and assessment – Part 1: model development. *J. Am. Water Resour. Assoc.* **34**, 73–89.

Beven, K. & Binley, A. 1992 The future of distributed models: model calibration and uncertainty prediction. *Hydrol. Process.* **6**, 279–298.

Beven, K. & Freer, J. 2001 Equifinality, data assimilation, and uncertainty estimation in mechanistic modeling of complex environmental systems using the GLUE methodology. *J. Hydrol.* **249**, 11–29.

Blasone, R. S. & Vrugt, J. A. 2008 Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov Chain Mote Carlo sampling. *Adv. Water Resour.* **31**, 630–648.

Bossa, A. Y., Diekkrüger, B., Igué, A. M. & Gaiser, T. 2012 Analyzing the effects of different soil databases on modeling of hydrological processes and sediment yield in Benin (West Africa). *Geoderma* **173–174**, 61–74.

Di Luzio, M., Arnold, J. G. & Srinivasan, R. 2004 Integration of SSURGO maps and soil parameters within a geographic information system and nonpoint source pollution model system. *J. Soil Water Conserv.* **59**, 123–133.

Duan, Q., Gupta, V. K. & Sorooshian, S. 1992 Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resour. Res.* **28**, 1015–1031.

Freer, J. & Beven, K. 1996 Bayesian estimation of uncertainty in runoff prediction and the value of data: an application of the GLUE approach. *Water Resour. Res.* **32**, 2161–2173.

Ghaffari, G., Keesstra, S., Ghodousi, J. & Ahmadi, H. 2010 SWAT-simulated hydrological impact of land-use change in the Zanjanrood basin, Northwest Iran. *Hydrol. Process.* **24**, 892–903.

Gijsman, A. J., Jagtap, S. S. & Jones, J. W. 2002 Wading through a swamp of complete confusion: how to choose a method for estimating soil water retention parameters for crop models. *Eur. J. Agron.* **18**, 75–105.

Güngör, Ö. & Göncü, S. 2013 Application of the soil and water assessment tool model on the Lower Porsuk Stream Watershed. *Hydrol. Process.* **27**, 453–466.

Guo, H., Hu, Q. & Jiang, T. 2008 Annual and seasonal streamflow responses to climate and land-cover changes in the Poyang Lake basin, China. *J. Hydrol.* **355**, 106–122.

Gupta, S. C. & Larson, W. E. 1979 Estimating soil water retention characteristics from particle size distribution, organic matter content, and bulk density. *Water Resour. Res.* **15**, 1633–1635.

Gupta, H. V., Beven, K. J. & Wagener, T. 2005 Model calibration and uncertainty analysis. In: *Encyclopedia of Hydrological Science* (M. G. Anderson, ed.). John Wiley & Sons, Ltd, Chichester, UK.

Hailegeorgis, T. T. & Alfredsen, K. 2016 Multi-basin and regional calibration based identification of distributed precipitation–runoff models for hourly runoff simulation: calibration and transfer of full and partial parameters. *Hydrol. Res.* **47** (2), 239–259.

Krause, P., Boyle, D. P. & Bäse, F. 2005 Comparison of different efficiency criteria for hydrological model assessment. *Adv. Geosciences* **5**, 89–97.

Li, Z., Shao, Q., Xu, Z. & Cai, X. 2010 Analysis of parameter uncertainty in semi-distributed hydrological models using bootstrap method: a case study of SWAT model applied to Yingluoxia watershed in northwest China. *J. Hydrol.* **385**, 76–83.

Melching, C. S. 1995 Reliability estimation. In: *Computer Models of Watershed Hydrology* (V. P. Singh, ed.). Water Resources Publications, Highlands Ranch.

Merz, R. & Blöschl, G. 2004 Regionalisation of catchment model parameters. *J. Hydrol.* **287**, 95–123.

Mukundan, R., Radcliffe, D. E. & Risse, L. M. 2010 Spatial resolution of soil data and channel erosion effects on SWAT model predictions of flow and sediment. *J. Soil Water Conserv.* **65**, 92–104.

Muttiah, R. S. & Wurbs, R. A. 2002 Scale-dependent soil and climate variability effects on watershed water balance of the SWAT model. *J. Hydrol.* **256**, 264–285.

Pietroniro, A., Soulis, E. D. & Kouwen, N. 2004 Scaling soil moisture for hydrological models. *IAHS Publication* **287**, 77–95.

Rawls, W. J., Ahuja, L. R. & Brakensiek, D. L. 1992 Estimating soil hydraulic properties from soils data. In: *Indirect Methods for Estimating the Hydraulic Properties of Unsaturated Soils* (M. T. Van Genuchten, ed.). Univ. of California, Riverside.

Rawls, W. J., Ahuja, L. R., Brakensiek, D. L. & Shirmonhammadi, A. 1993 Infiltration and soil water movement. In: *Handbook of Hydrology* (D. R. Maidment, ed.). McGraw-Hill, New York.

Romanowicz, A. A., Vanclooster, M., Rounsevell, M. & La Junesse, I. 2005 Sensitivity of the SWAT model to soil and land use data parameterisation: a case study in the Thyle catchment, Belgium. *Ecol. Model.* **187**, 27–39.

Saxton, K. E. & Rawls, W. J. 2006 Soil water characteristic estimates by texture and organic matter for hydrologic solutions. *Soil Sci. Soc. Am. J.* **70**, 1569–1578.

Saxton, K. E., Rawls, W. J., Romberger, J. S. & Papendick, R. I. 1986 Estimating generalized soil water characteristics from texture. *Soil Sci. Soc. Am. J.* **50**, 1031–1036.

Schuol, J., Abbaspour, K. C., Srinivasan, R. & Yang, H. 2008 Estimation of freshwater availability in the West African sub-continent using the SWAT hydrologic model. *J. Hydrol.* **352**, 30–49.

Sellami, H., La Jeunesse, I., Benabdallah, S. & Vanclooster, M. 2013 Parameter and rating curve uncertainty propagation analysis of the SWAT model for two small Mediterranean catchments. *Hydrolog. Sci. J.* **58**, 1635–1657.

Shen, Z. Y., Chen, L. & Chen, T. 2012 Analysis of parameter uncertainty in hydrological and sediment modeling using GLUE method: a case study of SWAT model applied to Three Gorges Reservoir Region, China. *Hydrol. Earth Syst. Sci.* **16**, 121–132.

Singh, H. V., Kalin, L., Morrison, A., Srivastava, P. & Lockaby, G. 2015 Post-validation of SWAT model in a coastal watershed for predicting land use/cover change impacts. *Hydrol. Res.* **46** (6), 837–853.

Sivapalan, M. 2003 Prediction in ungauged basins: a grand challenge for theoretical hydrology. *Hydrol. Process.* **17**, 3163–3170.

Tu, J. 2009 Combined impact of climate and land use changes on streamflow and water quality in eastern Massachusetts, USA. *J. Hydrol.* **379**, 268–283.

USDA-NRCS (US Department of Agriculture-Natural Resources Conservation Service) 1995 *Soil Survey Geographic (SSURGO) Data Base: Data Use Information*. National Cartography and GIS Center, Fort Worth, TX, USA.

USDA-SCS (US Department of Agriculture-Soil Conservation Service) 1993 *State Soil Geographic Data Base (STATSGO)*. US Government Printing Office, Washington, DC, USA.

Van Genuchten, M. T. & Leij, F. J. 1992 On estimating the hydraulic properties of unsaturated soils. In: *Indirect Methods for Estimating the Hydraulic Properties of Unsaturated Soils* (M. T. Van Genuchten, ed.). Univ. of California, Riverside.

Vereecken, H., Maes, J., Feyen, J. & Darius, P. 1989 Estimating the soil moisture retention characteristic from texture, bulk density, and carbon content. *Soil Sci.* **148**, 389–403.

Wang, X. & Melesse, A. M. 2006 Effects of STATSGO and SSURGO as inputs on SWAT model's snowmelt simulation. *J. Am. Water Resour. Assoc.* **42**, 1217–1236.

White, K. L. & Chaubey, I. 2005 Sensitivity analysis, calibration, and validation for a multisite and multivariable SWAT model. *J. Am. Water Resour. Assoc.* **41**, 1077–1089.

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.

Zang, C. F., Liu, J., Velde, M. V. D. & Kraxner, F. 2012 Assessment of spatial and temporal patterns of Green and blue water flows under natural conditions in inland river basins in Northwest China. *Hydrol. Earth Syst. Sci.* **16**, 2859–2870.

First received 28 July 2015; accepted in revised form 18 November 2015. Available online 27 January 2016