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

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;
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. 2013). 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. 2003) 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 (US Department of Agriculture-Soil Conservation Service) 1993 and SSURGO (US Department of Agriculture-Natural Resources Conservation Service) 1993, 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 data-set 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; Schuel 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 & Gönçü 2013). This has become easier with the popularization of the automatic calibration tool SWAT-CUP (Yang et al. 2008; available at: http://www.neprashtech.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. 2015). 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 NO₃ 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 km², 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 (50 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), 1:108 scale soil map derived from the Chinese Soil Scientific Database (CSSD) (http://www.soil.cssdb.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
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. \( \text{SOL}_{\text{AWC}}, \text{SOL}_{\text{K}} \) and \( \text{SOL}_{\text{BD}} \) of SWAT model for each soil type in Jinjiang Basin were estimated as follows.

\[ \text{SOL}_{\text{AWC}} \]

\( \text{SOL}_{\text{AWC}} \) (mm H\(_2\)O/mm soil) is the fraction of water between the field capacity and permanent wilting point, which is soil moisture at tension 33 kPa \( \theta_{33} \) (volumetric...
percentage) and 1,500 kPa $\theta_{1500}$ (volumetric percentage), respectively:

$$SOLAWC = \theta_{33} - \theta_{1500}$$  \hspace{1cm} (1)$$

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

$$\theta_{33t} = -0.251S + 0.195C + 0.011OM + 0.006(S \times OM)$$
$$- 0.027(C \times OM) + 0.452(S \times C) + 0.299$$  \hspace{1cm} (2)$$

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

$$\theta_{1500t} = -0.024S + 0.487C + 0.006OM$$
$$+ 0.005(S \times OM) - 0.015(C \times OM)$$
$$+ 0.068(S \times C) + 0.031$$  \hspace{1cm} (4)$$

$$\theta_{1500} = \theta_{1500t} + (0.14 \times \theta_{1500t} - 0.02)$$  \hspace{1cm} (5)$$

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, $\theta_{33}$ is corrected based on a relationship derived from a second regression analysis:

$$\theta_{33} = \theta_{33t} + [1.283(\theta_{33t})^2 - 0.374\theta_{33t} - 0.015]$$  \hspace{1cm} (3)$$

Figure 2 | Spatial distributions of three soil group types.
**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:

\[
SOL.K = 1930(\theta_s - \theta_{33})^{(3-\lambda)}
\]

(6)

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

\[
\theta_s = \theta_{33} + \theta_{(S-33)} - 0.097S + 0.043
\]

(7)

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

\[
\begin{align*}
\theta_{S-33I} &= 0.278S + 0.034C + 0.022OM \\
&- 0.018(S \times OM) - 0.027(C \times OM) \\
&- 0.564(S \times C) + 0.078
\end{align*}
\]

(8)

\[
\theta_{S-33} = \theta_{S-33I} + \left(0.656 \times \theta_{S-33I} - 0.107\right)
\]

(9)

The slope of logarithmic tension – moisture curve \(\lambda\) is computed as

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

(10)

**SOL.BD**

**SOL.BD** is estimated from \(\theta_s\), assuming particle density 2.65 (g/cm\(^3\)):

\[
SOL.BD = (1 - \theta_s) \times 2.65
\]

(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

<table>
<thead>
<tr>
<th>Soil group name</th>
<th>SOL_AWC (mm/mm)</th>
<th>SOL_BD (g/cm²)</th>
<th>SOL_K (mm/hr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red soil</td>
<td>0.151</td>
<td>1.58</td>
<td>2.2</td>
</tr>
<tr>
<td>Paddy soil</td>
<td>0.059</td>
<td>1.54</td>
<td>45.5</td>
</tr>
<tr>
<td>Yellow soil</td>
<td>0.150</td>
<td>1.14</td>
<td>17.4</td>
</tr>
</tbody>
</table>

**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 \(\times\) 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 parameters) were calibrated. In the second calibration, the nine soil parameters estimated by pedotransfer functions were fixed. The SWAT model was calibrated and validated for each soil type, and the final SWAT model was applied to the basin to map soil-water-hydrological characteristics.

### Table 3 | SWAT model parameters

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Initial range</th>
</tr>
</thead>
<tbody>
<tr>
<td>CN2</td>
<td>SCS runoff curve number</td>
<td>20–90</td>
</tr>
<tr>
<td>ALPHA_BF</td>
<td>Baseflow recession coefficient</td>
<td>0–1</td>
</tr>
<tr>
<td>GW_DELAY</td>
<td>Groundwater delay time (days)</td>
<td>30–450</td>
</tr>
<tr>
<td>GWQMN</td>
<td>Threshold water level in shallow aquifer for base flow</td>
<td>0–2</td>
</tr>
<tr>
<td>GW_REVAP</td>
<td>Groundwater evaporation coefficient</td>
<td>0–0.2</td>
</tr>
<tr>
<td>ESCO</td>
<td>Soil evaporation compensation coefficient</td>
<td>0.8–1</td>
</tr>
<tr>
<td>CH_N2</td>
<td>Manning coefficient for the main channel</td>
<td>0–0.3</td>
</tr>
<tr>
<td>CH_K2</td>
<td>Hydraulic conductivity in main channel (mm/hr)</td>
<td>5–150</td>
</tr>
<tr>
<td>ALPHA_BNK</td>
<td>Bank flow recession coefficient</td>
<td>0–1</td>
</tr>
<tr>
<td>SFTMP</td>
<td>Snowfall temperature (°C)</td>
<td>−5–5</td>
</tr>
<tr>
<td>SOL_AWC</td>
<td>Available soil water capacity (mm H₂O/mm soil)</td>
<td>0–1</td>
</tr>
<tr>
<td>SOL_BD</td>
<td>Soil moist bulk density (g/cm³)</td>
<td>1.1–2.5</td>
</tr>
<tr>
<td>SOL_K</td>
<td>Soil Saturated hydraulic conductivity (mm/hr)</td>
<td>0–2,000</td>
</tr>
</tbody>
</table>
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.neprashtecnhology.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 parameter sets. One advantage of GLUE is that a modeler 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 2010). 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:

\[
\text{NSE} = 1 - \frac{\sum (Q_{\text{obs},i} - Q_{\text{sim},i})}{\sum (Q_{\text{obs},i} - Q_{\text{obs},\text{avg}})}
\]

where \(Q_{\text{obs},i}\) (m\(^3\)/s) and \(Q_{\text{sim},i}\) (m\(^3\)/s) are observed and simulated streamflows at Shilong station for time step \(i\), and \(Q_{\text{obs},\text{avg}}\) (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:

\[
L_{p}[\theta|Q_{\text{obs}}] = CL[\theta|Q_{\text{obs}}]L_{a}[\theta]
\]

where \(L_{p}[\theta|Q_{\text{obs}}]\) is the prior likelihood weight for parameter set \(\theta\), which was the same for all behavioral sets, \(L[C][\theta|Q_{\text{obs}}]\) is the likelihood value calculated in step two, \(L_{p}[\theta|Q_{\text{obs}}]\) is the posterior likelihood weight conditioned by streamflow observations \(Q_{\text{obs}}\), and \(C\) is a scaling constant ensuring that the sum of \(L_{p}[\theta|Q_{\text{obs}}]\) for all behavioral sets was equal to unity.

4. Calculate uncertainty quantiles. The cumulative distribution of the predictions weighted by likelihood was calculated by:

\[
P_{t}(Q_{t} < q) = \sum_{i=1}^{m} L_{p}[\theta_{i}|Q_{t,i} < q]
\]

where \(P_{t}(Q_{t} < q)\) is the cumulative probability of predicted streamflow \(Q_{t}\) less than arbitrary value \(q\) at time step \(t\), \(L_{p}[\theta_{i}|Q_{t,i} < q]\) is the posterior likelihood of parameter set \(\theta_{i}\) 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 \( Q_{i,t} < q \). 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:

\[
R_{\text{factor}} = \frac{\sum_{i=1}^{m} (Q_{97.5\%i} - Q_{2.5\%i})}{m \times \sigma_{Q_{\text{obs}}}}
\]

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 \( \sigma_{Q_{\text{obs}}} \) 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.

**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
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 3](https://iwaponline.com/hr/article-pdf/47/5/964/368266/nh0470964.pdf)
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 4](https://iwaponline.com/hr/article-pdf/47/5/964/368266/nh0470964.pdf)

**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)

<table>
<thead>
<tr>
<th>Number of behavioral parameter sets</th>
<th>NSE of best simulation</th>
<th>P-factor</th>
<th>R-factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAL_19 4,776</td>
<td>0.83</td>
<td>69%</td>
<td>0.59</td>
</tr>
<tr>
<td>CAL_10 6,256</td>
<td>0.85</td>
<td>77%</td>
<td>0.57</td>
</tr>
</tbody>
</table>
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 2004). 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). 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 2004).

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
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>
<thead>
<tr>
<th>Table 7</th>
<th>Correlation matrix of posterior parameter distribution obtained from CAL_19 (bold text indicates significant at 0.01 level)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CN2</td>
<td>1</td>
</tr>
<tr>
<td>ALPHA_BF</td>
<td>0.004 1</td>
</tr>
<tr>
<td>GW_DELAY</td>
<td>0.001 −0.000 1</td>
</tr>
<tr>
<td>GWQMN</td>
<td>0.020 −0.020 0.012 1</td>
</tr>
<tr>
<td>GW_REVAP</td>
<td>−0.009 0.018 0.015 0.015 1</td>
</tr>
<tr>
<td>ESCO</td>
<td>−0.016 −0.030 −0.016 0.018 0.006 1</td>
</tr>
<tr>
<td>CH_N2</td>
<td>0.133 0.010 0.005 0.005 −0.014 −0.002 1</td>
</tr>
<tr>
<td>CH_K2</td>
<td>0.092 0.027 0.008 −0.001 0.009 0.043 −0.339 1</td>
</tr>
<tr>
<td>ALPHA_BNK</td>
<td>−0.103 0.014 0.002 −0.014 0.004 0.021 0.127 0.328 1</td>
</tr>
<tr>
<td>SFTMP</td>
<td>−0.024 0.010 −0.013 −0.004 0.015 0.013 −0.020 0.060 0.02 1</td>
</tr>
</tbody>
</table>
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


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

<table>
<thead>
<tr>
<th></th>
<th>CN2</th>
<th>ALPHA_BF</th>
<th>GW_DELAY</th>
<th>GWQMN</th>
<th>GW_REVAP</th>
<th>ESCO</th>
<th>CH_N2</th>
<th>CH_K2</th>
<th>ALPHA_BNK</th>
<th>SFTMP</th>
</tr>
</thead>
<tbody>
<tr>
<td>CN2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALPHA_BF</td>
<td>-0.021</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GW_DELAY</td>
<td>0.000</td>
<td>-0.011</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GWQMN</td>
<td>-0.007</td>
<td>0.013</td>
<td>0.005</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GW_REVAP</td>
<td>0.006</td>
<td>0.014</td>
<td>0.008</td>
<td>-0.010</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESCO</td>
<td>-0.130</td>
<td>0.013</td>
<td>0.017</td>
<td>-0.017</td>
<td>0.007</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CH_N2</td>
<td>0.213</td>
<td>0.003</td>
<td>0.004</td>
<td>-0.018</td>
<td>-0.009</td>
<td>0.021</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CH_K2</td>
<td>0.182</td>
<td>0.002</td>
<td>-0.008</td>
<td>-0.007</td>
<td>0.008</td>
<td>0.006</td>
<td>-0.174</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALPHA_BNK</td>
<td>-0.101</td>
<td>0.007</td>
<td>0.034</td>
<td>0.003</td>
<td>0.014</td>
<td>-0.023</td>
<td>0.121</td>
<td>0.288</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>SFTMP</td>
<td>0.002</td>
<td>0.003</td>
<td>-0.004</td>
<td>0.008</td>
<td>0.014</td>
<td>0.07</td>
<td>-0.009</td>
<td>-0.004</td>
<td>-0.016</td>
<td>1</td>
</tr>
</tbody>
</table>


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.


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