Environmental Modelling & Software 54 (2014) 211–221

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

A framework for propagation of uncertainty contributed by parameterization, input data, model structure, and calibration/validation data in watershed modeling

Haw Yen [a,b,*], Xiuying Wang [b], Darrell G. Fontane [a], R. Daren Harmel [c], Mazdak Arabi [a]

[a] Department of Civil and Environmental Engineering, Colorado State University, Fort Collins, CO 80523, USA
[b] Blackland Research & Extension Center, Texas A&M Agrilife Research, 720 East Blackland Road, Temple, TX 76502, USA
[c] Grassland, Soil & Water Research Laboratory, USDA-ARS, 808 East Blackland Road, Temple, TX 76502, USA

ARTICLE INFO

Article history:
Received 12 November 2013
Received in revised form
19 December 2013
Accepted 1 January 2014
Available online xxx

Keywords:
Model calibration
Error propagation
SWAT model
Bayesian Model Averaging
Uncertainty analysis

ABSTRACT

Failure to consider major sources of uncertainty may bias model predictions in simulating watershed behavior. A framework entitled the Integrated Parameter Estimation and Uncertainty Analysis Tool (IPEAT), was developed utilizing Bayesian inferences, an input error model and modified goodness-of-fit statistics to incorporate uncertainty in parameter, model structure, input data, and calibration/validation data in watershed modeling. Applications of the framework at the Eagle Creek Watershed in Indiana shows that watershed behavior was more realistically represented when the four uncertainty sources were considered jointly without having to embed watershed behavior constraints in auto-calibration. Accounting for the major sources of uncertainty associated with watershed modeling produces more realistic predictions, improves the quality of calibrated solutions, and consequently reduces predictive uncertainty. IPEAT is an innovative tool to investigate and explore the significance of uncertainty sources, which enhances watershed modeling by improved characterization and assessment of predictive uncertainty.

© 2014 Elsevier Ltd. All rights reserved.

Software availability

Name of software: IPEAT
Program language: MATLAB
Developer: Haw Yen, research associate of the Texas A&M University, developed and completed the code of IPEAT in August, 2012.
The source code: Available from Haw Yen, 720 East Blackland Road, Temple, Texas 76502, USA.
Tel: +1 254 774 6004, hyen@brc.tamus.edu

1. Introduction

The ability of watershed models to simulate and predict real world phenomena has been considerably advanced in recent years. Simultaneously, the number of model parameters in empirically or physically based functions has also increased (Yang et al., 2008; Bai et al., 2009), which increases the difficulty of manual calibration. Fortunately, through the progressive improvement of computer science and development of auto-calibration techniques, computational expense of model calibration is no longer a major challenge (Duan et al., 1992; Tolson and Shoemaker, 2007; Vrugt et al., 2009a; Yen, 2012). Thus, modelers can now focus more attention on appropriate representation of watershed phenomena and improved modeling methodology.

However, models are only simplified representations of natural systems. Actual watershed processes are more complex and variable than what can be generally represented in even most sophisticated models (Haan et al., 1995). Uncertainty due to model parameterization, input data, model structure, and observations used for model calibration can significantly impact the accuracy of model outputs. For clarification in this study the uncertainty from forcing inputs (input data) should not be regarded as a part of error contributed from model parameterization. As pointed by Ajami et al. (2007) that it is not appropriate to assume all prediction uncertainty is contributed by model parameterization as it is typical in watershed calibration/validation. On the other hand, even when uncertainty from parameterization, input data, and/or model structure is included, the commonly used goodness-of-fit statistics

* Corresponding author. Blackland Research & Extension Center, Texas A&M Agrilife Research, Texas A&M University System, 720 East Blackland Road, Temple, TX 76502, USA. Tel.: +1 254 774 6004.
E-mail addresses: hyen@brc.tamus.edu, haw.yen@gmail.com (H. Yen).

1364-8152/$ – see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.envsoft.2014.01.004

which are calculated relative to observational data need to be modified considering measurement uncertainty. Because of imperfect observation, Harmel and Smith (2007) and Harmel et al. (2010) have proposed to consider measurement uncertainty in the evaluation of goodness-of-fit statistics in hydrologic and water quality modeling. In fact, failing to consider one or more sources of uncertainty may cause biased model calibration results and corresponding model predictions.

In review of literature, parameterization uncertainty has received the most attention in previous studies (e.g., Kuczera and Parent, 1998; Osidele et al., 2006; Gallagher and Doherty, 2007; Hassan et al., 2009; Loosvelt et al., 2011; Rasmussen and Hamilton, 2012; Joseph and Guillaume, 2013). The parameter-calibration approach indicates that parameter errors are the ultimate attribution of all possible sources (Ajami et al., 2007). Predictive uncertainty contributed by model input data has been explored and proven to have significant impact on model calibration (Kavetski et al., 2002; Ajami et al., 2007; Strauch et al., 2012). Wang (2008) applied the Bayes inferences to stochastically conduct data generation with non-concurrent, missing input data. In addition, other studies incorporate input uncertainty explicitly during the calibration processes (Kavetski et al., 2002; Ajami et al., 2007). A further source of uncertainty is model structure. The importance of structural uncertainty was demonstrated by Refsgaard et al. (2006) and Clark et al. (2008). One of the most frequently cited approaches for exploring the structural uncertainty is to aggregate different models through the Bayesian Model Averaging (BMA) technique, where the significance of simulation performance for each implemented model can be stated by BMA weights (Kavetski et al., 2006a, 2006b; Ajami et al., 2007; Duan et al., 2007). The contribution of uncertainty in calibration/validation data has only recently been incorporated into evaluation of watershed models. For example, Ullrich and Volk (2010) investigated the influence of uncertainty in NO_3–N monitoring data on model calibration and evaluation. Prior to work by (Harmel et al., 2006, 2009), no comprehensive framework for estimation of uncertainty in measured discharge and water quality data was available. Following the development of this framework, methods were further developed by Harmel and Smith (2007) and Harmel et al. (2010) to modify goodness-of-fit indicators to consider uncertainty in measured data used for model calibration/validation evaluation. Uncertainty has also been explored in associated with other techniques. A Bayesian-based framework assisted by the Morris global sensitivity analysis method was developed to evaluate performance of two models (Minunno et al., 2013), and a set-membership approach was implemented to identify parameter and prediction uncertainty while conducting sediment yield simulations (Keesman et al., 2013).

The main contribution of this study is the development of a framework that facilitates simultaneous evaluation of parameterization, input data, model structure, and calibration/validation data uncertainty and their contribution to predictive uncertainty, entitled Integrated Parameter Estimation and Uncertainty Analysis Tool (IPEAT). The specific objectives of this study were to: (i) quantify predictive uncertainty while propagating different sources of uncertainty and (ii) calibrate the Soil and Water Assessment Tool (SWAT) model with the consideration of four sources of uncertainty (as opposed to typical model calibration of considering only parameter uncertainty) to understand the role and importance of uncertainty source on model prediction (best solution) and predictive uncertainty. In this study, the model structural uncertainty analysis was limited to the two modified SCS, now the Natural Resource Conservation Service (NRCS), curve number method (USDA Soil Conservation Service, 1972; USDA–NRCS, 2004) for calculating surface runoff within the SWAT2009 model. A set of other models can also be included in the analysis, which is not the focus of this study.

2. Methods and materials

2.1. Framework of incorporating different sources of uncertainty

The proposed framework to incorporate uncertainty from parameterization, input data, model structure, and observation data used for calibration is provided in Fig. 1 where it is compared with typical watershed modeling.

2.1.1. Input data uncertainty

Input data such as rainfall, temperature, soils, and land use/cover are critical drivers for watershed simulation. For simplicity in this study only the uncertainty from rainfall was considered. The integrated Bayesian uncertainty estimator (IBUNE) (Ajami et al., 2007) was used to account for uncertainty contributed by rainfall data through an input error model as shown in Eq. (1), which assumed a random Gaussian error as a multiplier for every input observation (Ajami et al., 2007).

$$R_{a_t} = kR_t; \quad k \sim N\left(m, \sigma_m^2\right) \tag{1}$$

where R_{a_t} and R_t are the adjusted and observed rainfall depth at time step t (e.g., the given day t), respectively, k is the normally distributed random noise with mean m, $m \in [0.9, 1.1]$ and variance σ_m^2, $\sigma_m^2 \in [1e-5, 1e-3]$ as defined by (Ajami et al., 2007).

For each SWAT simulation run, the two variables (m and σ_m^2) from this input error model were added as two unknown parameters to the system and a random multiplier (k) to each time step was drawn from the normal distribution $N(m, \sigma_m^2)$. A parameter estimation technique, dynamically dimensioned search (DDS) (Tolson and Shoemaker, 2007) (see Section 2.1.3.) was used to search the SWAT model parameters and the input error model parameters (m and σ_m^2) simultaneously. Through SWAT simulation, the uncertainty associated with the input error model parameters and the SWAT model parameters were propagated through the system.

2.1.2. Model structure uncertainty

Uncertainty is also contributed by the inability of the model structure to perfectly mimic watershed processes. Different models have different degrees of complexity and different algorithms to mimic natural processes. Even within a given modeling system, alternative methods may be offered. Uncertainty due to model structure can significantly impact the accuracy of model outputs. Although researchers have included a set of mutually exclusive models in analyzing model structural uncertainty (e.g., Abrahart and See, 2002; Georgakakos et al., 2004; Ajami et al., 2007), there are potentially many alternative ways to formulate the analysis by combining different models considering there are many hydrologic and water quality models out there. In this study, we only considered the widely used SWAT model. SWAT offers two options to calculate the curve number retention parameter, s. The first one is the traditional method which allows s to vary with soil profile water content (SCSI). An alternative method (SCSII) allows s to vary with accumulated plant evapotranspiration. These two methods were considered in this study for model structural uncertainty.

SWAT is a continuous-time and semi-distributed parameter model, which is developed to simulate/predict hydrologic and water quality processes at the large watershed scale (Arnold et al., 1993, 1998) and it is widely applied for assessing water resource and nonpoint-source pollution problems (e.g., Du et al., 2005; Jayakrishnan et al., 2005; Green et al., 2006; Moriasi et al., 2009; Arnold et al., 2010; Chiang et al., 2010; Douglas-Mankin et al., 2010; Ghebremichael et al., 2010; Kim et al., 2010; Meng et al., 2010; Srinivasan et al., 2010). Comprehensive descriptions of SWAT are presented in Gassman et al. (2007) and Arnold et al. (2012).

The Bayesian Model Averaging (BMA) was used in this study for account for model structural uncertainty. BMA is a probabilistic scheme for model combination (Raftery et al., 2005; Ajami et al., 2007; Wöhling and Vrugt, 2008). The posterior distribution of the BMA prediction, y_{BMA}, under the two SWAT model options of M_1 = SCSI and M_2 = SCSII is given as:

$$p\left(y_{BMA}\middle|M_1, M_2, \tilde{X}, \tilde{y}\right) = \sum_{k=1}^{2} p\left(M_k\middle|\tilde{X}, \tilde{y}\right) \times p_k\left(y_k\middle|M_k, \tilde{X}, \tilde{y}\right) \tag{2}$$

where $\tilde{X}$ is the SWAT model input forcing data, $\tilde{y}$ is SWAT output variables of interests (here streamflow and water quality NO_3–N), $p(M_k|.\tilde{X}, \tilde{y})$ is the posterior probability of model M_k, $p_k(y_k|.M_k, \tilde{X}, \tilde{y})$ is the forecast posterior distribution of y_k given prediction quantities from model M_k with input data $\tilde{X}$ and corresponding prediction $\tilde{y}$. The$p_k(y_k|.M_k, \tilde{X}, \tilde{y})$ is represented by the normal distribution with mean equal to the output of model M_k and standard deviation σ_k. The term $p(M_k|.\tilde{X}, \tilde{y})$ is also known as the likelihood of model M_k being the correct model, or BMA weight, which should be summed to one.

$$\sum_{k=1}^{2} p\left(M_k\middle|\tilde{X}, \tilde{y}\right) = 1 \tag{3}$$

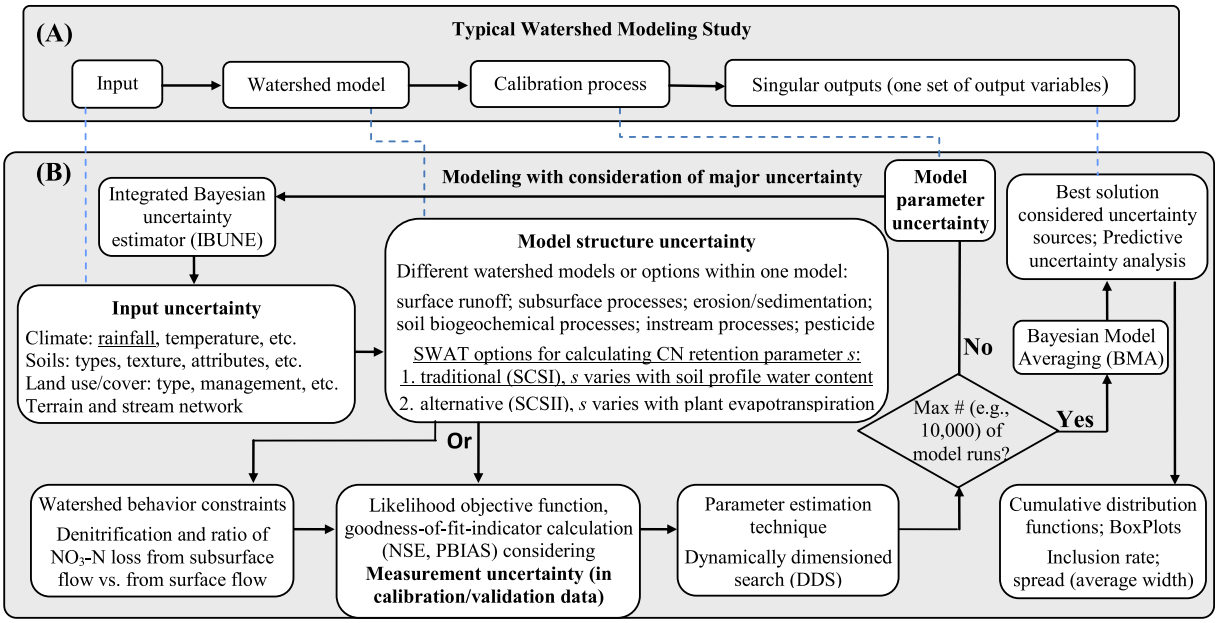

Fig. 1. Watershed modeling flowchart: (A) typical modeling study vs. (B) framework of incorporating major uncertainty in modeling.

The weights can be estimated by different optimization techniques. For example, Expectation-Maximization (EM) algorithm and the Shuffled Complex Evolution Metropolis algorithm (Raftery et al., 2005; Duan et al., 2007; Wöhling and Vrugt, 2008). In this study, the EM algorithm was used. Better simulation results produce higher BMA weights.

2.1.3. Parameter estimation

Parameterization uncertainty results from various physically or empirically based model parameters which are not known with certainty or are difficult to attain. Therefore model calibration procedure typically deals with model parameter uncertainty. In typical model calibration processes, only one set of calibrated parameters is accepted. Correspondingly, singular outputs are reported (Fig. 1A). The singular parameter set values are either identified through automatic optimization procedure or through manual calibrations where the modeler subjectively stops further adjusting of parameter values duo to insignificant improvement in terms of goodness-of-fit statistic measures.

In this study, a dynamically dimensioned search (DDS) algorithm (Tolson and Shoemaker, 2007) was used for auto-calibration and all runs are used for uncertainty analysis. The DDS is a simple stochastic neighborhood search algorithm focused on finding preferred parameter combinations fast within the user specified maximum number of model runs (here 10,000). It searches globally at the beginning and then becomes more focused on local search as the number of iterations approaches the specified maximum number of model runs, which is achieved by dynamically and probabilistically reducing the number of dimensions in the neighborhood. New search avoids poor local optima and parameter values are updated by perturbing the current solution values in the randomly selected dimensions with perturbation magnitudes randomly sampled from a normal distribution with a mean of zero. More detail can be found in Tolson and Shoemaker (2007).

Any goodness-of-fit statistics can be used as objective function in the DDS for optimization. In this study, the likelihood function developed by Ahmadi (2012) was used, which is an aggregated multisite and multi-objective (e.g., considering streamflow and water quality) function. Because the formulation/derivation of the function is long, we only cover the essence of the likelihood function as below. It represents the log-likelihood of model residuals for a given set of model parameters. Assuming that residuals are normally and independently distributed with mean equal to zero and unknown but constant standard deviation σ_e^2, the likelihood function can be written in below as stated in van Griensven and Meixner (2007):

$$l(\varepsilon|\theta) = \prod_{j=1}^{m}\prod_{i=1}^{n_j}\frac{1}{\sqrt{2\pi\sigma_{e,j}^2}}\exp\left(-\frac{\left(\widehat{P}_{i,j}(\theta)-O_{i,j}\right)^2}{2\sigma_{e,j}^2}\right) \quad (4)$$

where l is the likelihood function representing the likelihood of model residuals ε for a given set of model parameters θ, m is the total number of objective functions considering all hydrologic and water quality output variables of interest (e.g., streamflow, sediment, NO_3–N) at all sampling sites, n_j is the number (or length) of observed data for output variable j, $\widehat{P}$ is the simulated response vector of the watershed, O is the vector of observed (measured) output response.

Because in most cases the errors in hydrological and water quality modeling are not normally distributed and independent, a first-order autoregressive (AR-1) transformation of the residuals is used to account for correlated errors. The final objective function is the logarithm of Eq. (4) under AR-1 transformation (Sorooshian and Dracup, 1980). The log-likelihood of model residual ranges from 0 to positive infinity, however, the true minimum value is not known. Because the DDS algorithm is not designed to converge to the precise global optimum, the finial minimum likelihood value cannot guarantee the final parameter set truly represent the watershed system. In fact, even a precise global optimum, which in term of auto-calibration is the minimized objective function, can be achieved through a global optimization algorithm (rather than DDS), the corresponding parameters are not always the guarantees of adequate realization. That is because model calibration is only conditioned on what we feed (actual data and/or range of uncertainty) into the model and what we assigned for calibration comparisons (observed data and objective function). And any objective function (aggregated or not) has limitations because different statistical measures place emphasis on different systematic and/or dynamic behavioral errors. Therefore, although the log-likelihood function was used in the auto-calibration process to determine the best parameter set in this study, the Nash Sutcliffe efficiency (NSE, Nash and Sutcliffe, 1970) and percent bias (PBIAS) are reported and used to evaluate model performance according to the general performance ratings proposed by Moriasi et al. (2007) as shown in Table 1.

2.1.4. Calibration/validation data uncertainty

Uncertainty is also contributed by using uncertain measured discharge and constituent flux data to calibrate model and evaluate model performance. In this study, the impact of measurement uncertainty was investigated based on research by Harmel et al. (2006, 2009, 2010). Harmel et al. (2006) compiled information from available literature on uncertainty in measured streamflow and water quality data. The uncertainty is inherent in the four procedural categories: discharge measurement, sample collection, sample preservation/storage, and laboratory analysis.

Table 1
General performance ratings from Moriasi et al. (2007).

Performance rating	NSE	PBIAS (%)	
		Streamflow	Nitrogen
Very Good	$0.75 < \text{NSE} \le 1.00$	$\text{PBIAS} < \pm 10$	$\text{PBIAS} < \pm 25$
Good	$0.65 < \text{NSE} \le 0.75$	$\pm 10 \le \text{PBIAS} < \pm 15$	$\pm 25 \le \text{PBIAS} < \pm 40$
Satisfactory	$0.50 < \text{NSE} \le 0.65$	$\pm 15 \le \text{PBIAS} < \pm 25$	$\pm 40 \le \text{PBIAS} < \pm 70$
Unsatisfactory	$\text{NSE} \le 0.50$	$\text{PBIAS} \ge \pm 25$	$\text{PBIAS} \ge \pm 70$

NSE: Nash-Sutcliffe efficiency coefficient.
PBIAS: Percent bias.

Harmel et al. (2009) added another procedure category, data processing and management, to the 2006 framework. The root mean square error propagation method of Topping (1972) was used with literature reported errors to estimate the cumulative probable uncertainty for each procedural category and for the overall resulting streamflow and water quality data. Scientific estimates of uncertainty for best and worst case and a range of typical data quality scenarios are reported in Table 5 in Harmel et al. (2006). Harmel et al. (2009) reported that the typical uncertainty for measured streamflow is about ±7–23% and NO_3–N load is about ±14–31%. The range (probable error range, *PER*) of ±10% and ±30% was used for streamflow and NO_3–N, respectively, in this study.

Because of the inherent measurement uncertainty, modeling performance should be evaluated against such uncertainty. Therefore, goodness-of-fit indicators require modification to appropriately compare model predictions to observations (Harmel and Smith, 2007). In this study, the probability distribution (PD) method from Harmel and Smith (2007) was adopted to calculate NSE and PBIAS. The core of the PD method is to assign a correction factor, *CF*, to each paired error. Taking NSE calculation as an example, the error term (O_i–P_i) in the Eq. (5) is modified/replaced with the error term eu_i in Eq. (5).

$$NSE = 1 - \left[\sum_{i=1}^{n}(O_i - P_i)^2 / \sum_{i=1}^{n}\left(O_i - \overline{O}\right)^2\right] \tag{5}$$

$$eu_i = \frac{CF_i}{0.5} \times (UO_i - P_i); \quad (1 - abs(PER_i)/100) \times O_i \le UO_i \le (1 + abs(PER_i)/100) \times O_i \tag{6}$$

where O_i and P_i are measured and predicted values at each comparison point *i*, respectively, eu_i is the modified error term at comparison time *i*, CF_i is the correction factor at time *i*, UO_i is adjusted measurement at data point *i* considering the uncertainty associated with the data point, and PER_i is the probable error range for each measured data point *i*, which is from the literature values as provided above.

The underlying assumption for this modification is that the probability distribution for each observed value is symmetric with the observed value as mean and median of the distribution. The CF_i is divided by 0.5 because that is the maximum probability of one half of the pdf. The *CF* value for the symmetric distribution can range from 0 (when $O_i = P_i$) to 0.5. In this study, the observed data point is assumed normally distributed with the mean of O_i, and the variance σ^2 is calculated as below:

$$\sigma^2 = \left(\frac{O_i - UO_i(l)}{3.9}\right)^2 \tag{7}$$

where $UO_i(l)$ is the lower uncertainty boundary for each measured data point *i*. Because the normal distribution is infinite in both directions, the use of lower uncertainty boundary in the calculation is not strictly valid; however, it is appropriate by using standard deviation of 3.9 which contains >99.99% of the normal probability distribution (Haan, 2002). More detail can be found in Harmel and Smith (2007). In implementation to determine the *CF* value for the normal distribution, the normal distribution can be transformed to the standard normal distribution *N(0,1)*. And *CF* is calculated as the area under the standard normal distribution (Harmel and Smith, 2007).

2.1.5. Watershed behavior constraints

For complex watershed simulation models such as SWAT, it is possible to achieve acceptable goodness-of-fit statistic values for sequential responses (e.g., daily streamflow, monthly mean nitrate concentration), but inaccurately predict summative outputs (e.g., denitrification and NO_3–N yield) with typical optimization (or auto-calibration) methods. To avoid this situation, watershed behavior constraints are applied to ensure calibrated results are realistic. In other words, calibration that does not accurately represent watershed behavior may produce parameter sets that generate unrealistic outputs but "good" goodness-of-fit statistic values. The watershed behavior responses, such as accumulated mass of biodegradation, water and nutrient runoff from agricultural fields, amount of nutrient leaching, etc. (Sui and Frankenberger, 2008), may not reasonably simulated as a result of calibration criteria focusing only on in-stream responses (where measurements are available) of the system and associated model tuning rather than the intra-watershed response (Green et al., 2006; Amatya and Jha, 2011).

In this study, the importance of reasonably representing watershed behavior in calibration was considered by implementing watershed behavior constraints on denitrification and the ratio of NO_3–N losses contributed from subsurface flow (SSQ) verse the total loss from SSQ and surface flow (SQ) due to the case study watershed (Section 2.2) having tile drained agriculture land. In the Midwest region of the United States, the annual denitrification rate has been reported no more than 50 kg/ha (David et al., 2009). Ale et al. (2012) found that denitrification ranged from 20 to 26 kg/ha. In this study, denitrification was restricted to 0–50 kg/ha based on David et al. (2009). Although measured data of NO_3 loadings from surface runoff, lateral flow, tile drainage, and groundwater are not available for the ECW (Sui and Frankenberger, 2008), studies conducted in cultivated watersheds in Indiana and Iowa reported that tile-drain effluent is a major contributor to NO_3 in streams (e.g., Nelson et al., 1979; Lucey and Goolsby, 1993). Schilling (2002) reported that baseflow and tile drainage accounted for 61–68% of annual NO_3 loss for two watersheds in central Iowa. In this study, the ratio of NO_3–N losses contributed from SSQ vs. from both SSQ and SQ to be larger than 0.6. If results violate these constraints, then the corresponding model run is penalized by assigning an extreme value to the likelihood objective function used in DDS minimization procedure. Therefore, new search can avoid poor local optima.

2.1.6. Analysis of prediction uncertainty

In brief to summarize uncertain propagation (Fig. 1B), the input and parameter uncertainty are propagated through SWAT simulation of hydrologic processes using either the SCSI or SCSII surface runoff estimation and all other processes simulated by SWAT. Model outputs are then compared with measurements with the consideration of inherent measurement uncertainty using modified goodness-of-fit indicators and likelihood objective function. The procedure is repeated through parameter DDS optimization.

Then predictive uncertainty can be analyzed using all runs for probability distribution and 95% confidence interval. In this study SWAT outputs of streamflow and NO_3–N loss are quantified using inclusion rate and spread. The inclusion rate represents the percentage of observed data points within the 95% confidence interval, and the spread is the average width of the corresponding uncertainty band along the prediction time series. The uncertainty band was derived based on the difference between the 95% and 5% quantiles.

2.2. Case study watershed and calibration design

To illustrate this framework, a case study was developed for the 248 km^2 Eagle Creek Watershed (ECW). The ECW is located in central Indiana within the Upper White River watershed (Fig. 2). The average annual precipitation is 965 mm, and the average annual temperature is 11 °C (Newman, 1997). As in much of the Midwestern U.S., agriculture is the major land use (Fig. 2) with drainage tile. A majority of the soils in the ECW are in hydrologic soil group B (51%) and C (48%) (Fig. 2). NRCS Soil Survey Staff (2010) defines a hydrologic group as a group of soils having similar runoff potential under similar storm and cover conditions. The B and C group of soils have a moderate and slow infiltration rate, respectively. For the period from 1997 to 2003, measured discharge data are available for one station (number 35 in Fig. 2, USGS Gauge #03353200), and measured water quality data (NO_3–N) are available for four stations (numbers 20, 22, 27, 32 in Fig. 2). More detailed description of the ECW can be found in Yen (2012).

The monthly streamflow and NO_3–N loss available from 1997 to 2000 were used to calibrate SWAT2009 with the consideration of four sources of uncertainty (as opposed to typical model calibration of considering only parameter uncertainty) to examine the impact of different calibration conditions/setups on the best solution. The validation periods were from 2001 to 2003 for the 5 gauge stations (Fig. 2). Four preliminary calibration cases (Table 2) were included in this study in order to demonstrate the dilemma that model parameter sets identified through parameter optimization procedure that have reasonable comparison statistics (e.g., goodness-of-fit indicators proposed by Moriasi et al. (2007)) for watershed-scale response may not be realistic in terms of intra-watershed responses (see Section 3.2). That is what led to our consideration of using watershed behavior constraints during parameter optimization (Table 3). We have also examined the situation where the watershed constraints may not be able to clearly define due to data unavailability; therefore, no constraints may be setup during parameter optimization yet with the consideration of four sources of uncertainty (Table 3). All the calibration cases were implemented to calibrate monthly streamflow and NO_3–N loss using the same default SWAT model parameter setup and a maximum number of 10,000 runs were specified for each calibration case for the DDS optimization procedure to find optimal solutions by minimizing the aggregated multisite and multi-objective (considering streamflow and NO_3–N) objective function described in Section 2.1.3. For each of SCSI vs. SCSII paired with and without watershed behavior constraints (4 calibration cases listed in Table 2), after the specified maximum number of model runs was met under each calibration case, BMA weights were calculated for paired calibration cases (V vs. VI without constraints and VII vs. VIII with constraints).

3. Results and discussion

3.1. Auto-calibration relative to computational requirements

Computational time for each calibration case, which involved 10,000 runs, ranged from 450 to 500 h (Intel® Core™ 2 Duo CPU E8400 @ 3.00 GHz, 32-bit operating system, Microsoft Windows XP). For all of the calibration cases, objective function values converged (such that changes in the objective function value for the current best parameter set showed improvement <1%) within 5000 model runs (Fig. 3). And patterns of convergence were similar where objective function value dropped significantly in the beginning of DDS iterations and after algorithm convergence

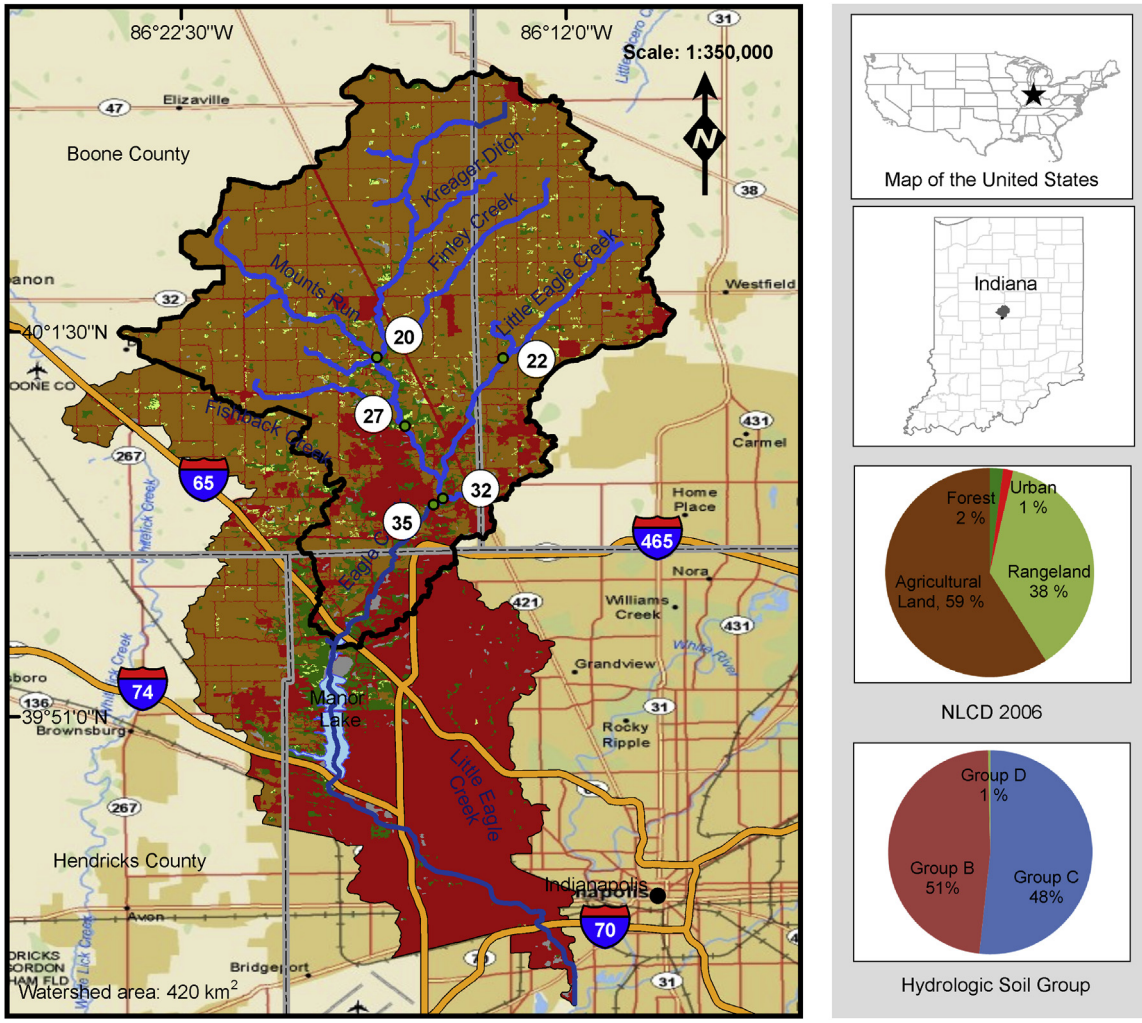

Fig. 2. Case study area: Eagle Creek Watershed, Indiana. USGS gauge stations are represented by the numbered circles. Soil hydrologic group is defined by NRCS Soil Survey Staff (2010) as a group of soils having similar runoff potential under similar storm and cover conditions.

objective function values were stable with little changes (Fig. 3). It is expected that auto-calibration with watershed behavior constraints would slow down the DDS convergence (cases VII and VIII) because the optimization would avoid poor neighborhood which violate constraints and the corresponding objective function was assigned an extreme value (penalized); therefore DDS would restart search for minimization. Based on Fig. 3, the calibration case II performed the worst in maximizing objective function values, and the SCSI without watershed behavior constraints (calibration case V) produced the best values.

3.2. Best solutions and goodness-of-fit indicators

As described earlier in Section 2.1.3 that model calibration (manual or automatic) is conditioned on what we feed (actual data and/or range of uncertainty) into the model and what we assigned for calibration comparisons (observed data and objective function), therefore, different calibration conditions/cases would result in different best solutions. Results of best objective function values and the corresponding predicted output for denitrification and the NO_3–N loss ratio (loss from SSQ vs. total loss from SSQ and SQ) are shown in Table 4.

Although the best solutions based on the objective function values for the first four calibration cases (I–IV) are comparable with calibration cases V–VIII as listed in Table 4, the annual denitrification rates were ranged between 121.4 and 243.3 (cases I to IV) which were not reasonable based on reported literature values as reviewed in Section 2.1.5. Although the streamflow comparison statistics at the USGS gauge 03353200 (#35 in Fig. 2) were satisfactory for both the calibration and validation period for all the calibration cases (Fig. 4), Fig. 5 illustrates the weakness of the four calibration cases I–IV in simulating NO3–N loss. Although the calibrated models have reasonable comparison statistics for NO_3–N loss during calibration period (Fig. 5) with NSE values ranging from 0.54 to 0.92 and PBIAS ranging from 16.1% to 26.6% (satisfactory performance based on Moriasi et al. (2007)), most of the corresponding NSE values for the validation period are unsatisfactory. Among the total of 16 comparisons for NO_3–N loss (4 stations $\times$ 4 calibrated cases $=$ 16) during the validation period, eleven comparison NSE values ranged from -0.09 to 0.41 and only five (see data labels in Fig. 5) ranged from 0.51 to 0.69 (satisfactory).

However, for calibration cases V to VIII, BMA(V vs. VI) and BMA(VII vs. VIII) (see Table 3), 23 out of 24 comparisons for NO_3–N loss had satisfactory performance rating during calibration period, and the satisfactory ratio was 21 out of 24 for the validation period

Table 2
Preliminary calibration cases for Eagle Creek Watershed SWAT modeling study without consideration of watershed behavior constraints during optimization.

Uncertainty sources	SCSI	SCSII
SWAT parameters	I	III
SWAT parameters, input data	II	–
SWAT parameters, calibration data	IV	–

Table 3
Calibration cases for Eagle Creek Watershed SWAT modeling study (uncertainty from input data, SWAT parameters, model structure, and calibration data are considered).

Behavior constraints	Runoff method		
	SCSI	SCSII	
No	V	VI	BMA (V vs. VI)
Yes	VII	VIII	BMA (VII vs. VIII)

Behavior constraints: annual denitrification rate <50 kg/ha; and ratio of NO_3–N loss contributed from subsurface flow vs. from subsurface flow and surface flow>0.6.

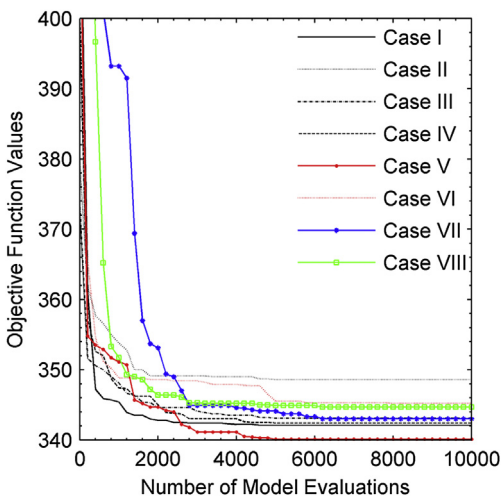

Fig. 3. Overall performance of objective function values versus model evaluations for calibration cases I–VIII.

(see data labels for unsatisfactory NSE values in Fig. 6). In addition, although watershed behavior constraints were not used during optimization processes in the calibration cases V and VI where uncertainty from input data, SWAT parameters, and calibration data were considered, the optimum runs with the best objective function values had denitrification rates and NO_3–N loss ratios within the behavior ranges (Table 4).

Table 4
Best objective function values and the corresponding denitrification amounts and NO_3–N loss ratios for the calibration period.

Calibration case	Objective function value	Denitrification (kg/ha)	NO_3–N loss ratio
I	342.1 (399.6)	121.4 (16.9)	0.98 (0.94)
II	348.6 (379.4)	214.2 (30.4)	0.97 (0.99)
III	343.1 (373.1)	243.3 (7.1)	0.96 (0.98)
IV	342.4 (399.6)	211.4 (49.7)	0.98 (0.98)
V	340.1	36.1	0.96
VI	345.2	14.5	0.63
VII	343.0	36.1	0.92
VIII	344.7	49.8	0.79

The values in the parentheses are the "best" objective function values and corresponding denitrification and NO_3–N loss ratios after removing runs which violate watershed behavior constraints. Behavior constraints: annual denitrification rate <50 kg/ha; and ratio of NO_3–N loss contributed from subsurface flow vs. from subsurface flow and surface flow>0.6.

These results illustrate that not accounting for the major sources of uncertainty can lead to biased parameter estimates, which are compensating for other sources of uncertainty not considered and may not adequately represent the watershed system. For example, although the objective function values were converged within 5000 model runs (only half of the maximum number of model runs specified, Fig. 3) for each of the calibration cases, the calibrated model or parameter sets from calibration cases I–IV (considered 1 or 2 sources of uncertainty) did not realistically represent the watershed behavior in term of annul denitrification rate (Table 4) (although with reasonable comparison statistics for total NO_3–N loss during calibration period). As expected, these calibrated parameter sets did not do a good job to produce satisfactory predictions of NO_3–N loss for the validation period because they do not adequately represent the watershed process. On the other hand, the calibration cases V–VIII (considered 3 sources of uncertainty) and BMA (V vs. VI) and BMA (VII vs. VIII) (considered 4 sources of uncertainty) all produced more realistic results than cases I–IV did in terms denitrification rate and satisfactory model performance for the validation period.

Because all runs during auto-calibration were saved, a post-process procedure was conducted for calibration cases I–IV to remove runs which violate watershed behavior constraints. As expected, after this post-process the goodness-of-fit indicators (NSE and PBIAS) for the remaining "best" runs were significantly reduced (Fig. 5). If the constraints were applied during auto-calibration for I–IV rather than this post-process, the quality of calibration/validation results could be improved (see Yen, 2012). However, the question we were trying to answer here is if no constraints may be setup for parameter optimization due to data unavailability, then will the calibration cases I–IV (considered only one or two sources of uncertainty) lead to satisfactory or adequate model realization? The answer is "no", at least for this case study watershed, based on results presented in Table 4 and Fig. 5. However, by comparing the calibration cases V vs. VII (same SCSI and same 3 uncertainty sources as V but with constraints), or the calibration cases VI vs. VIII (same SCSII and same 3 uncertainty sources as VI but with constraints), less or not much differences were observed between the without and with watershed behavior constraints cases in terms of model performances (see Fig. 6). The same is true when BMA was used to consider model structural uncertainty in the calibration case BMA (V vs. VI) with on constraints used and case BMA (VII vs. VIII) with constraints applied in optimization (Fig. 6).

3.3. BMA applications

As shown in Table 5, results at the 5 USGS stations were mixed in terms of BMA weights for SWAT SCSI (soil moisture-based CN method) and SCSII (revised soil moisture index SMI CN method). The SWAT SCSII method is based on the value of water depletion or evapotranspiration (ET) results by Williams et al. (2012) indicate that this method is generally robust and produces realistic runoff estimates over a wide range of soil properties. In 7 out of 10 paired weights, the SCSII BMA weights were relatively larger than those for SCSI. For stations #32 and #27, the SCSII had larger BMA weights with or without constraints embed in calibration. The BMA's application for the VII and VIII cases (constraints embed in calibration) also illustrates that SCSII is slightly better than SCSI with all weights for SCSII > weights for SCSI except at station #22 where the two weights were very close to each other.

However, BMA weights for SCSI were larger than for SCSII under the no constraints cases V vs. VI at stations #35 and #20; yet at the

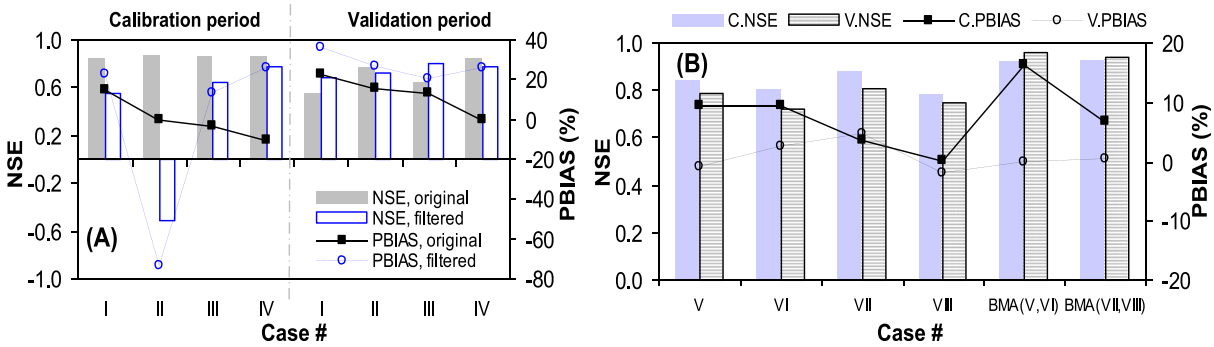

Fig. 4. Nash-Sutcliffe efficiency (NSE) and percent error (PBIAS) for calibration/validation periods at station #35 for streamflow. For (A) cases I–IV, "original": original calibration results; "filtered": post-processed results after removing runs violated behavior constraints and for (B) cases V-BMA (VII, VIII), C: calibration and V: validation.

same two stations under the constraints cases VII vs. VIII, the weights for SCSI were smaller than that for SCSII. The inconsistence can be the indication that the weights are somehow compensating for the other uncertainty sources (e.g., land use/cover, management practices, soils at spatial scale) and/or other model structural deficiencies not considered in this study. By comparing the BMA weight differences between the two groups (Table 5), the group BMA (VII vs. VIII) with constraints embed in calibration had relatively smaller weight differences than those of group BMA (V vs. VI) for all the 5 stations.

3.4. Uncertainty analysis

As described in Section 2.1.6, the uncertainty analysis was focused on the calibration cases which incorporate major uncertainty in watershed modeling as illustrated in Fig. 1, namely, cases V–VIII (parameterization, input data, and calibration/validation date uncertainty) and the two corresponding cases with BMA application.

For the input error model (Eq. (1)), during parameter optimization the random Gaussian error (multiplier) was identified with means ranged from 1.00 to 1.05 (input as from 0.9 to 1.1) and variance generally less than 0.0005 (input as from 10^{-5} to 10^{-3} as recommended by Ajami et al. (2007)). The mean had a narrower range with low variance during parameter optimization, which means the measured precipitation data was not significantly alerted before entering into the SWAT model or they are not too far from the assumed "true" rainfall depths. However, the result of mean > 1 might indicate that the precipitation input data likely underestimated the actual precipitation in the watershed.

The optimization procedure identified that the average values of correction factors (*CF*, Eq. (6)) for measured streamflow and NO_3–N data ranged from 0.90 to 0.93. This means that more than 90% of simulated outputs were within the estimated uncertainty ranges for streamflow (±10%) and NO_3–N (±30%).

Table 6 summarizes the prediction uncertainty results due to uncertainty propagation through the SWAT simulation system for the simulation period. The percentage of observed data at the 5 USGS gauge stations which are within the estimated 95% uncertainty bounds ranged from 21% to 69% and the spreads of uncertainty bands ranged from 0.6 to 31.7 (Table 6). The average width of uncertainty band (spread) in case VI is wider than for case V for all of the water quality gauges (NO_3–N prediction), which indicates that the selection of the surface runoff prediction procedure (model structure) has a large influence on predictive uncertainty. Moreover, implementing watershed behavior constraints (cases VII vs. V and cases VIII vs. VI) decreased the predictive uncertainty for all water quality gauges. This is an important result because ensuring predicted behavior realistically not only improved the quality of calibrated solutions as discussed in Section 3.2 but also reduced predictive uncertainty for the water quality stations.

When model structure uncertainty was also included by the application of BMA (BMA (V vs. VI) and BMA (VII vs. VIII)), the prediction uncertainty (as indicated by spread, which is affected primarily by the BMA weights) did not change substantially (Table 6). Although accounting for the major sources of uncertainty in watershed modeling generated more reliable results, these uncertainty results were still suffering from common limitations in hydrologic/water quality modeling. First, not all sources of uncertainty were accounted for, e.g., soil, land cover, agricultural management which were not considered in this study. Second, only two alternative methods for surface runoff simulation in SWAT (only one component and one model) were considered for model structural uncertainty. The model structural uncertainty was clearly underestimated, for example, the uncertainty in the water quality component and tile drainage simulation were missing.

The inclusion rates ranged from 31% to 83% (Fig. 7) and spread from 0.89 to 18.6 during the validation period (Fig. 8). Although it does has limitations, the framework is a flexible tool to enhance watershed modeling by addressing different sources of uncertainty simultaneously; it can be used to improve confidence in model application even during validation period.

3.5. Comparison with generalized likelihood uncertainty estimation (GLUE)

GLUE (Beven and Binley, 1992; Beven, 1993) is based on Monte Carlo simulations and the likelihood of a given parameter set is a function of the model performance expressed in terms of the objective function chosen. In GLUE, model responses are compared with observations and each parameter set is weighted via the likelihood measures. GLUE focuses on the generation of a large number of solutions that exceed some nominal threshold (arbitrary) which defines behavioral solutions; therefore, it claims for equifinality that there are many different parameter sets that may be behavioral/acceptable in reproducing the observed behavior of the simulation system (Beven and Freer, 2001). In IPEAT, the efficiency in both calibration and uncertainty analysis are enhanced by implementing the auto-calibration technique, DDS, which can rapidly converge to good calibration solutions and avoid poor local optima. As the results of calibration improved, it only adjusts one or a few limited parameters simultaneously to keep the current gain in calibration results. Therefore, newer parameter sets are more focused on providing better solutions without having to have each set weighted. The presented framework explicitly incorporates the

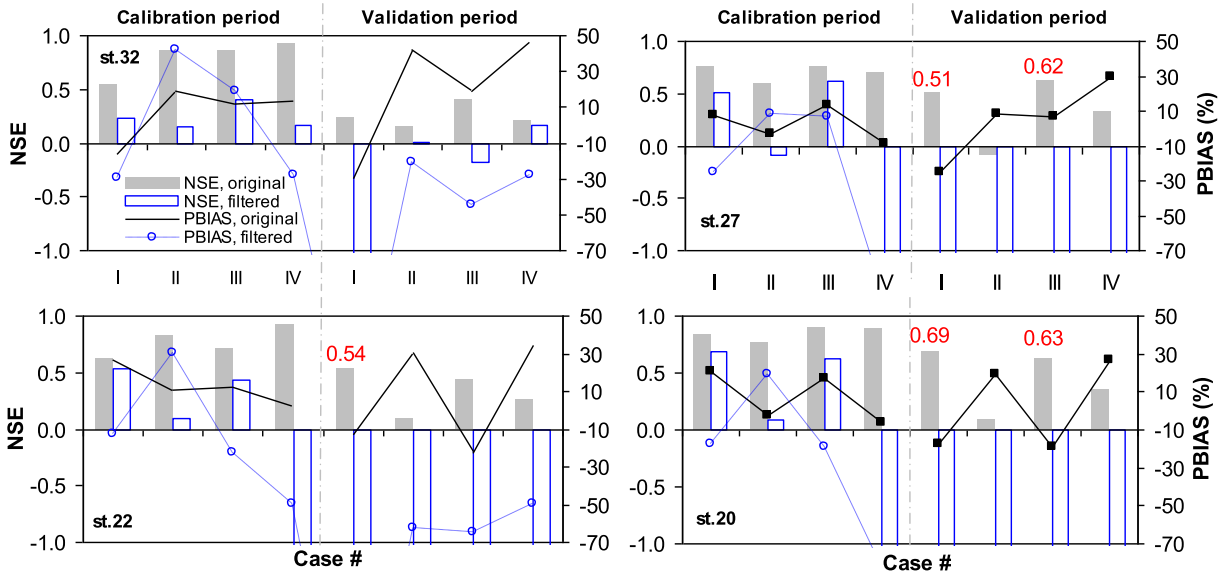

Fig. 5. Nash-Sutcliffe efficiency (NSE) and percent error (PBIAS) for calibration/validation periods (cases I–IV) at the 4 stations (st.) for NO_3–N loss. "Original": original calibration results; "filtered": post-processed results after removing runs violated behavior constraints. Due to magnitude issue, NSE < −1 and PBIAS < −70% data points were not mapped.

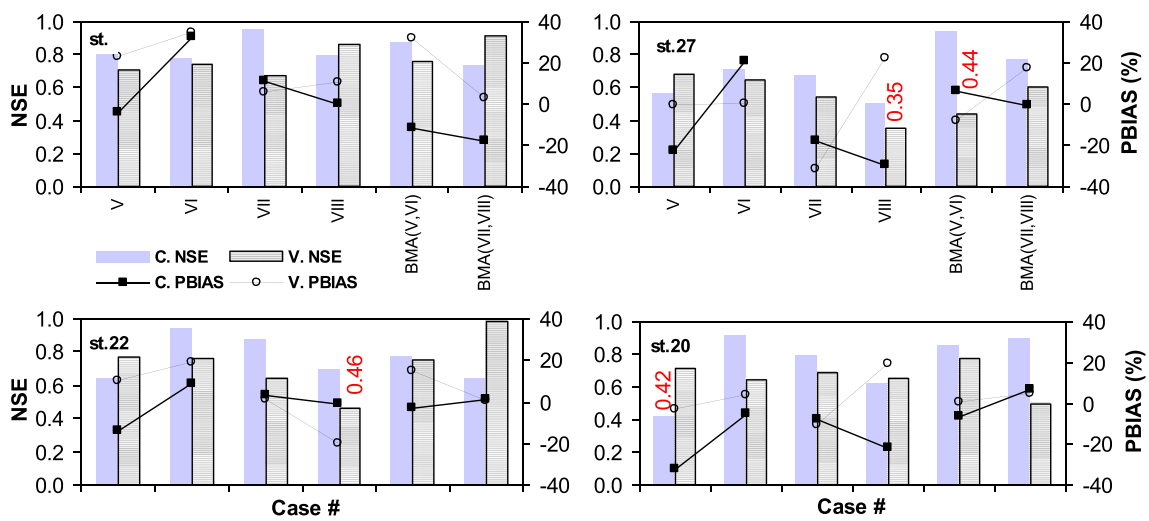

Fig. 6. Nash-Sutcliffe efficiency (NSE) and percent error (PBIAS) for calibration/validation periods for calibration cases V-BMA (VII–VIII) at the 4 USGS stations (st.) for NO_3–N loss. C: calibration and V: validation.

measurement uncertainty during calibration process using hyper-parameters, while GLUE simply uses subjective goodness-of-fit statistics. In addition, the proposed framework adds penalty terms to the objective function. These penalty terms are essentially regularization using prior knowledge (from history records in literature or it can also be experts' decision) of intra-model simulation results. The proposed framework is able to cover and improve the capacity of GLUE in both parameter estimation and uncertainty analysis.

4. Conclusions

The proposed framework incorporates parameterization, input data, model structure, and calibration/validation data uncertainty

Table 5
Bayesian Model Averaging (BMA) weights for BMA (V vs. VI) and BMA (VII vs. VIII) for the best runs.

Calibration case		ω @ st. 35 (discharge)	ω @ st. 32 (water quality)	ω @ st. 27 (water quality)	ω @ st. 22 (water quality)	ω @ st. 20 (water quality)
BMA (V vs. VI)	V (SCSI)	0.729	0.162	0.165	0.134	0.999
	VI (SCSII)	0.271	0.838	0.835	0.866	0.001
	Difference	0.457	−0.677	−0.671	−0.731	0.998
BMA (VII vs. VIII)	VII (SCSI)	0.354	0.316	0.260	0.514	0.488
	VIII (SCSII)	0.646	0.684	0.740	0.486	0.512
	Difference	−0.293	−0.368	−0.480	0.028	−0.024

ω: BMA weight.
st.: USGS gauge station number on Fig. 2.

Table 6
Inclusion rate of observed streamflow and NO_3–N loss data within the 95% confidence interval and the corresponding spread for the simulation period (1997–2003).

Scenario		st. 35 (discharge)	st. 32 (water quality)	st. 27 (water quality)	st. 22 (water quality)	st. 20 (water quality)
V	IR (%)	44.05	53.70	51.58	55.82	62.44
	Spread	1.01	2.28	2.37	2.15	1.91
VI	IR (%)	27.38	58.20	51.33	49.21	55.82
	Spread	1.06	26.03	31.73	25.94	20.97
VII	IR (%)	28.57	56.09	55.82	69.05	51.33
	Spread	0.77	1.97	2.05	1.39	1.47
VIII	IR (%)	40.47	56.35	53.97	65.34	53.97
	Spread	0.87	5.67	9.63	1.71	2.47
BMA (V, VI)	IR (%)	21.43	29.10	35.98	31.74	42.33
	Spread	0.62	27.12	29.55	24.35	2.63
BMA (V, VI)	IR (%)	25.00	33.60	40.21	27.25	29.63
	Spread	0.74	5.22	8.38	1.60	1.69

IR(%) – Inclusion rate is the % of observed data points within the 95% confidence interval.
Spread – average width of the corresponding uncertainty band along the predicted time series.

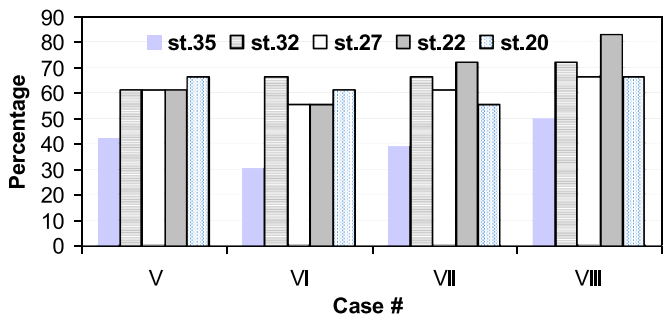

Fig. 7. Percentage of observations in 95% uncertainty bounds during the validation period at the 5 stations (st.) for cases considered uncertainty in parameter, input data and calibration/validation data.

in watershed modelling. Analysis of calibration cases I–IV (incorporated one or two uncertainty sources) demonstrated the dilemma that model parameter sets identified through parameter optimization procedure without accounting for all major sources of uncertainty produced reasonable comparison statistics for watershed-scale response during calibration period; however, they were not realistic in terms of intra-watershed responses and also performed poor for validation period. The calibration cases V and VI which considered uncertainty in parameters, input data, and calibration data in auto-calibration procedure led to more realistic/reliable model simulations such that the interior watershed behavior/response can be represented realistically (not violate watershed behavior constraints even without constraints embed in calibration) and improved model performance for the validation period.

The cases V–VIII and BMA (V vs. VI) and BMA (VII vs. VIII) involved comparison of model structure uncertainty and the effects of applying watershed behavior constraint. Important results of the comparisons include: 1) the inclusion of four uncertainty sources improved model simulations for both the calibration period and validation period; 2) the application of watershed behavior constraints improved the quality of calibration results (more solutions that represent actual watershed behavior) and reduced predictive uncertainty especially for water quality; and 3) watershed behavior was more realistically represented when the three or four major sources of uncertainty were considered without having to embed watershed behavior constraints in auto-calibration procedure, which is important especially in watersheds which little knowledge is available on actual watershed behavior.

Therefore, based on this study it is very beneficial to calibrate model with the consideration of three or four sources of uncertainty (as opposed to typical model calibration of considering only parameter uncertainty). The calibration case VII (under the consideration of three sources of uncertainty) and case BMA (VII vs. VIII) (under the consideration of four sources) provided the best results in terms of NSE and PBIAS values (Figs. 4B and 6, satisfactory performance for both calibration and validation period for all stations) and improved model prediction uncertainty (Table 6 and Fig. 8, better inclusion rate and smaller spread).

For complex large-scale watershed simulation models such as SWAT, calibration can be improved by considering the uncertainty contributed from all possible sources. The proposed framework, IPEAT, is an innovative tool to investigate the significance of uncertainty sources simultaneously. Although it has limitations, such as accounting for only the precipitation uncertainty in the input side and only two alternative methods for surface runoff simulation

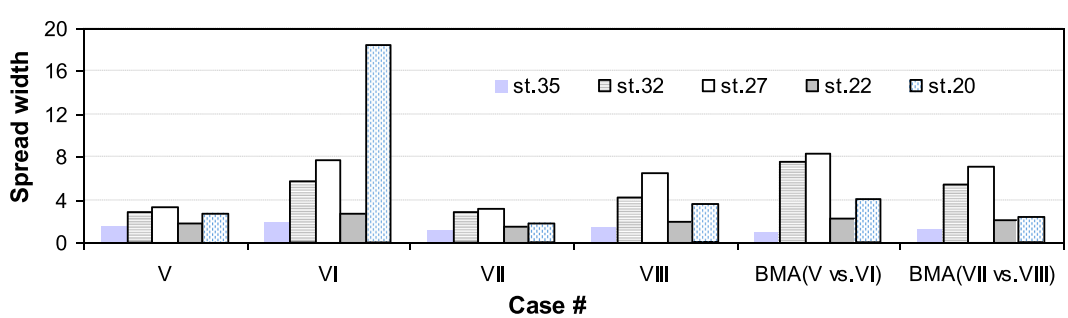

Fig. 8. Spread average width of uncertainty band along the prediction time series for the validation period at the 5 stations (st.) for cases listed in Table 3.

in SWAT in the structural uncertainty side, the framework is a flexible tool to enhance watershed modeling with improved confidence even during validation period.

References

Abrahart, R.J., See, L., 2002. Multi-model data fusion for river flow forecasting: an evaluation of six alternative methods based on two contrasting catchments. Hydrol. Earth Syst. Sci. 6 (4), 655–670.

Ahmadi, M., 2012. A Multi Criteria Decision Support System for Watershed Management under Uncertain Conditions (Ph.D. Dissertation). Colorado State University.

Ajami, N.K., Duan, Q., Sorooshian, S., 2007. An integrated hydrologic Bayesian multimodel combination framework: confronting input, parameter, and model structural uncertainty in hydrologic prediction. Water Resour. Res. 43 (1), 1–19. http://dx.doi.org/10.1029/2005WR004745.

Ale, A., Bowling, L.C., Youssef, M.A., Brouder, S.M., 2012. Evaluation of simulated strategies for reducing nitrate-nitrogen losses through subsurface drainage systems. J. Environ. Qual. 41, 217–228.

Amatya, D.M., Jha, M.K., 2011. Evaluating the SWAT model for a low-gradient forested watershed in coastal South Carolina. Trans. ASABE 54 (6), 2151–2163.

Arnold, J.G., Allen, P.M., Bernhardt, G., 1993. A comprehensive surface-groundwater flow model. J. Hydrol. 142, 47–69.

Arnold, J.G., Allen, P.M., Volk, M., Williams, J.R., Bosch, D.D., 2010. Assessment of different representations of spatial variability on SWAT model performance. Trans. ASABE 53 (5), 1433–1443.

Arnold, J.G., Srinivasan, R., Muttiah, R.S., Williams, J.R., 1998. Large area hydrologic modeling and assessment. I. Model development. J. Am. Water Resour. Assoc. 34 (1), 73–89.

Arnold, J.G., Moriasi, D.N., Gassman, P.W., Abbaspour, K.C., White, M.J., Srinivasan, R., Santhi, C., Harmel, R.D., van Griensven, A., Van Liew, M.W., Kannan, N., Jha, M.K., 2012. SWAT: model use, calibration, and validation. Trans. ASABE 55 (4), 1491–1508.

Bai, Y., Wagener, T., Reed, P., 2009. A top-down framework for watershed model evaluation and selection under uncertainty. Environ. Model. Softw. 24 (8), 901–916. http://dx.doi.org/10.1016/j.envsoft.2008.12.012.

Beven, K., 1993. Prophecy, reality and uncertainty in distributed hydrological modelling. Adv. Water Resour. 16, 41–51.

Beven, K.J., Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. J. Hydrol. 249, 11–29. http://dx.doi.org/10.1016/S0022-1694(01)00421-8.

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

Chiang, L., Chaubey, I., Gitau, M.W., Arnold, J.G., 2010. Differentiating impacts of land use changes from pasture management in a CEAP watershed using the SWAT model. Trans. ASABE 53 (2), 1569–1584.

Clark, M.P., Slater, A.G., Rupp, D.E., Woods, R.A., Vrugt, J.A., Gupta, H.V., Wagener, T., Hay, L.E., 2008. Framework for Understanding Structural Errors (FUSE): a modular framework to diagnose differences between hydrological models. Water Resour. Res. 44, 1–14. http://dx.doi.org/10.1029/2007WR006735.

David, M.B., Del Grosso, S.J., Hu, X., Marshall, E.P., McIsaac, G.F., Parton, W.J., Tonitto, C., Youssef, M., 2009. Modeling denitrification in a tile-drained, corn and soybean agroecosystem of Illinois, USA. Biogeochemistry 93, 7–30.

Douglas-Mankin, K.R., Srinivasan, R., Arnold, J.G., 2010. Soil and Water Assessment Tool (SWAT) model: current developments and applications. Trans. ASABE 53 (5), 1423–1431.

Du, B., Arnold, J.G., Saleh, A., Jaynes, D.B., 2005. Development and application of SWAT to landscapes with tiles and potholes. Trans. ASABE 48 (3).

Duan, Q., Ajami, N.K., Gao, X., Sorooshian, S., 2007. Multi-model ensemble hydrologic prediction using Bayesian model averaging. Adv. Water Resour. 30 (5), 1371–1386. http://dx.doi.org/10.1016/j.advwatres.2006.11.014.

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

Gallagher, M., Doherty, J., 2007. Parameter estimation and uncertainty analysis for a watershed model. Environ. Model. Softw. 22 (7), 1000–1020.

Gassman, P.W., Reyes, M., Green, C.H., Arnold, J.G., 2007. The soil and water assessment tool: historical development, applications, and future directions. Trans. ASABE 50 (4), 1211–1250.

Georgakakos, K.P., Seo, D.J., Gupta, H.V., Schake, J., Butts, M.B., 2004. Characterizing streamflow simulation uncertainty through multimodel ensembles. J. Hydrol. 298 (1–4), 222–241.

Ghebremichael, L.T., Veith, T.L., Watzin, M.C., 2010. Determination of critical source areas for phosphorus loss: Lake Champlain Basin, Vermont. Trans. ASABE 53 (5), 1595–1604.

Green, C.H., Tomer, M.D., Luzio, M.D., Arnold, J.G., 2006. Hydrologic evaluation of the soil and water assessment tool for a large tile-drianed watershed in Iowa. Trans. ASABE 49 (2), 413–422.

Haan, C.T., 2002. Statistical Methods in Hydrology, second ed. Iowa State Press, Ames, IA.

Haan, C.T., Allred, B., Storm, D.E., Sabbagh, G.J., Prabhu, S., 1995. Statistical procedure for evaluating hydrologic/water quality models. Trans. ASAE 38 (3), 725–733.

Harmel, R.D., Cooper, R.J., Slade, R.M., Haney, R., Arnold, J.G., 2006. Cumulative uncertainty in measured streamflow and water quality data for small watersheds. Trans. ASABE 49 (3), 689–701.

Harmel, R.D., Smith, P., 2007. Consideration of measurement uncertainty in the evaluation of goodness-of-fit in hydrologic and water quality modeling. J. Hydrol. 337 (3–4), 326–336.

Harmel, R.D., Smith, D.R., King, K.W., Slade, R.M., 2009. Estimating storm discharge and water quality data uncertainty: a software tool for monitoring and modeling applications. Environ. Model. Softw. 24 (7), 832–842.

Harmel, R.D., Smith, P.K., Migliaccio, K.W., 2010. Modifying goodness-of-fit indicators to incorporate both measurement and model uncertainty in model calibration and validation. Trans. ASABE 53 (1), 55–63.

Hassan, A.E., Bekhit, H.M., Chapman, J.B., 2009. Using Markov Chain Monte Carlo to quantify parameter uncertainty and its effect on predictions of a groundwater flow model. Environ. Model. Softw. 24 (6), 749–763. http://dx.doi.org/10.1016/j.envsoft.2008.11.002.

Jayakrishnan, R., Srinivasan, R., Santhi, C., Arnold, J.G., 2005. Advances in the application of the SWAT model for water resources management. Hydrol. Process. 19 (3), 749–762. http://dx.doi.org/10.1002/hyp.5624.

Joseph, J.F., Guillaume, J.H.A., 2013. Using a parallelized MCMC algorithm in R to identify appropriate likelihood functions for SWAT. Environ. Model. Softw. 46, 292–298. http://dx.doi.org/10.1016/j.envsoft.2013.03.012.

Kavetski, D., Franks, S.W., Kuczera, G., 2002. Confronting input uncertainty in environmental modeling. Water Sci. Appl. 6, 49–68. http://dx.doi.org/10.1016/j.envsoft.2005.04.010.

Kavetski, D., Kuczera, G., Franks, S.W., 2006a. Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. Water Resour. Res. 42 (3), 1–9. http://dx.doi.org/10.1029/2005WR004368.

Kavetski, D., Kuczera, G., Franks, S.W., 2006b. Bayesian analysis of input uncertainty in hydrological modeling: 2. Application. Water Resour. Res. 42 (3), 1–10. http://dx.doi.org/10.1029/2005WR004376.

Keesman, K.J., Norton, J.P., Croke, B.F.W., Newham, L.T.H., Jakeman, A.J., 2013. Set-membership approach for identification of parameter and prediction uncertainty in power-law relationships: the case of sediment yield. Environ. Model. Softw. 40, 171–180. http://dx.doi.org/10.1016/j.envsoft.2012.09.007.

Kim, N.W., Shin, A.H., Lee, J., 2010. Effects of streamflow routing schemes on water quality with SWAT. Trans. ASABE 53 (5), 1457–1468.

Kuczera, G., Parent, E., 1998. Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. J. Hydrol. 211 (1–4), 69–85. http://dx.doi.org/10.1016/S0022-1694(98)00198-X.

Loosvelt, L., Pauwels, V.R.N., Cornelis, W.M., De Lannoy, G.J.M., Verhoest, N.E.C., 2011. Impact of soil hydraulic parameter uncertainty on soil moisture modeling. Water Resour. Res. 47 (3), 1–16. http://dx.doi.org/10.1029/2010WR009204.

Lucey, K.J., Goolsby, D.A., 1993. Effects of climatic variations over 11 years on nitrate nitrogen concentrations in the Raccoon River, Iowa. J. Environ. Qual. 22, 38–46.

Meng, H., Sexton, A.M., Maddox, M.C., Sood, A., Brown, C.W., Ferraro, R.R., Murtugudde, R., 2010. Modelling Rappahannock river basin using SWAT – Pilot for Chesapeake Bay watershed. Appl. Eng. Agric. 26 (5), 795–806.

Minunno, F., van Oijen, M., Cameron, D.R., Cerasoli, S., Pereira, J.S., Tomé, M., 2013. Using a Bayesian framework and global sensitivity analysis to identify strengths and weaknesses of two process-based models differing in representation of autotrophic respiration. Environ. Model. Softw. 42, 99–115.

Moriasi, D.N., Arnold, J.G., Engel, B.A., Rossi, C.G., 2009. Incorporation of a new shallow water table depth algorithm into SWAT2005. Trans. ASABE 52 (3), 771–784.

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans. ASABE 50 (3), 885–900.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models, part I: a discussion of principles. J. Hydrol. 10 (3), 282–290.

Nelson, D.W., Monke, E.J., Bottcher, A.D., Sommers, L.E., 1979. Sediment and nutrient contributions to the Maumee River from an agricultural watershed. In: Best Management Practices for Agriculture and Silviculture. Proc. 1978 Cornell Agric. Waste Manag.Conf. Dep. of Agronomy, Purdue Univ, W. Lafayette, IN, pp. 491–505.

Newman, J.E., 1997. Our changing climate. In: Jackson, M.T. (Ed.), The Natural Heritage of Indiana.

NRCS, 2010. Keys to Soil Taxonomy, 11th ed. USDA-Natural Resources Conservation Service, Washington, DC.

Osidele, O.O., Zeng, W., Beck, M.B., 2006. A random search methodology for examining parametric uncertainty in water quality models. Water Sci. Technol. 53 (1), 33. http://dx.doi.org/10.2166/wst.2006.005.

Raftery, A.E., Gneiting, T., Balabdaoui, F., Polakowski, M., 2005. Using Bayesian Model Averaging to calibrate forecast ensembles. Am. Meteorol. Soc., 1155–1174.

Rasmussen, R., Hamilton, G., 2012. An approximate Bayesian computation approach for estimating parameters of complex environmental processes in a cellular automata. Environ. Model. Softw. 29 (1), 1–10. http://dx.doi.org/10.1016/j.envsoft.2011.10.005.

Refsgaard, J.C., van der Sluijs, J.P., Brown, J., van der Keur, P., 2006. A framework for dealing with uncertainty due to model structure error. Adv. Water Resour. 29 (11), 1586–1597. http://dx.doi.org/10.1016/j.advwatres.2005.11.013.

Schilling, K.E., 2002. Chemical transport from paired agricultural and restored prairie Watersheds. J. Environ. Qual. 31, 1846–1851.

Sorooshian, S., Dracup, J.A., 1980. Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: correlated and heteroscedastic error cases. Water Resour. Res. 16 (2), 430. http://dx.doi.org/10.1029/WR016i002p00430.

Srinivasan, R., Zhang, X., Arnold, J.G., 2010. SWAT ungauged: hydrological budget and crop yield predictions in the Upper Mississippi River Basin. Trans. ASABE 53 (5), 1533–1546.
Strauch, M., Bernhofer, C., Koide, S., Volk, M., Lorz, C., Makeschin, F., 2012. Using precipitation data ensemble for uncertainty analysis in SWAT streamflow simulation. J. Hydrol. 414–415, 413–424.
Sui, Y., Frankenberger, J.R., 2008. Nitrate loss from subsurface drains in an agricultural watershed using SWAT2005. Trans. ASABE 51 (4), 1263–1272.
Tolson, B.A., Shoemaker, C.A., 2007. Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. Water Resour. Res. 43 (1), 1–16. http://dx.doi.org/10.1029/2005WR004723.
Topping, J., 1972. Errors of Observation and Their Treatment, fourth ed. Chapman and Hall, London, U.K.
Ullrich, A., Volk, M., 2010. Influence of different nitrate-N monitoring strategies on load estimation as a base for model calibration and evaluation. Environ. Monit. Assess. 171, 513–527.
van Griensven, A., Meixner, T., 2007. A global and efficient multi-objective auto-calibration and uncertainty estimation method for water quality catchment models. J. Hydroinformatics 9 (4), 277. http://dx.doi.org/10.2166/hydro.2007.104.
Vrugt, J.A., Braak, C.J.F., Diks, C.G.H., Robinson, B.A., Hyman, J.M.., Higdon, D., 2009a. Accelerating Markov Chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. Int. J. Nonlinear Sci. Numer. Simul. 10, 271–288 (March 2009).
Wang, Q., 2008. A Bayesian method for multi-site stochastic data generation: dealing with non-concurrent and missing data, variable transformation and parameter uncertainty. Environ. Model. Softw. 23 (4), 412–421. http://dx.doi.org/10.1016/j.envsoft.2007.04.013.
Williams, J.R., Kannan, N., Wang, X., Santhi, C., Arnold, J.G., 2012. Evolution of the SCS runoff curve number method and its application to continuous runoff simulation. J. Hydrol. Eng. 17 (11), 1221–1229.
Wöhling, T., Vrugt, J.A., 2008. Combining multiobjective optimization and Bayesian model averaging to calibrate forecast ensembles of soil hydraulic models. Water Resour. Res. 44 (12), 1–18. http://dx.doi.org/10.1029/2008WR007154.
Yang, J., Reichert, P., Abbaspour, K., Xia, J., Yang, H., 2008. Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China. J. Hydrol. 358 (1–2), 1–23. http://dx.doi.org/10.1016/j.jhydrol.2008.05.012.
Yen, H., 2012. Confronting Input, Parameter, Structural, and Measurement Uncertainty in Multi-site Multiple Responses Watershed Modeling Using Bayesian Inferences (Ph.D. Dissertation). Colorado State University.