UC Irvine
UC Irvine Previously Published Works

Title

An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction

Permalink

Journal

ISSN

Authors

Ajami, Newsha K
Duan, Qingyun
Sorooshian, Soroosh

Publication Date

2007

DOI

Copyright Information

Peer reviewed

An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction

Newsha K. Ajami,[1] Qingyun Duan,[2] and Soroosh Sorooshian[3]

[1] The conventional treatment of uncertainty in rainfall-runoff modeling primarily attributes uncertainty in the input-output representation of the model to uncertainty in the model parameters without explicitly addressing the input, output, and model structural uncertainties. This paper presents a new framework, the Integrated Bayesian Uncertainty Estimator (IBUNE), to account for the major uncertainties of hydrologic rainfall-runoff predictions explicitly. IBUNE distinguishes between the various sources of uncertainty including parameter, input, and model structural uncertainty. An input error model in the form of a Gaussian multiplier has been introduced within IBUNE. These multipliers are assumed to be drawn from an identical distribution with an unknown mean and variance which were estimated along with other hydrological model parameters by a Monte Carlo Markov Chain (MCMC) scheme. IBUNE also includes the Bayesian model averaging (BMA) scheme which is employed to further improve the prediction skill and address model structural uncertainty using multiple model outputs. A series of case studies using three rainfall-runoff models to predict the streamflow in the Leaf River basin, Mississippi, are used to examine the necessity and usefulness of this technique. The results suggest that ignoring either input forcings error or model structural uncertainty will lead to unrealistic model simulations and incorrect uncertainty bounds.

1. Introduction

[2] Various hydrologic rainfall-runoff models have been used to represent the watershed physical processes which control the conversion of precipitation into streamflow and water storage changes. These models include many parameters describing the properties of the watershed that need to be estimated through calibration against historical observation data. For many years, research effort has been devoted to develop techniques to find the proper estimates of the parameter values that enable the model predictions to match the watershed observations. One major weakness of this parameter-calibration approach is that the objective function used to calibrate the model parameters implicitly assumes that all sources of uncertainties in the modeling process can be attributed to parameter errors. In fact, in addition to parameter uncertainty, model predictions are affected by many other uncertainties from various sources, among them the errors in model input (forcing) data such as

the precipitation observation data, the description of boundary and initial conditions, and the model structural deficiencies. Because of the highly nonlinear nature of the hydrologic system, it is not feasible to account for all these uncertainties from different sources through model parameter adjustments.

[3] Recently, hydrologic research [*Beven and Binley*, 1992; *Kuczera and Parent*, 1998; *Vrugt et al.*, 2003; *Marshall et al.*, 2003; *Montanari and Brath*, 2004] began to analyze various uncertainty sources in hydrological modeling. New techniques have made significant progress in estimating the propagation of confidence bounds from different uncertainty sources to the model output. Among them include the use of data assimilation techniques to tackle uncertainty in boundary and initial conditions [*Kitanidis and Bras*, 1980a, 1980b; *Beck*, 1987; *Evenson*, 1992; *Miller et al.*, 1994]; simultaneous data assimilation and parameter estimation [*Moradkhani et al.*, 2005]; and simultaneous uncertainty estimation of input (forcing) data and parameter estimation [*Kavetski et al.*, 2003]. Most of these studies focus on addressing one or two uncertainty sources based on a selected hydrologic model. However, by using a single model, those techniques (which do not change the model structures) are unable to account for the errors in model output resulting from the structural deficiencies of the specific model.

[1]Berkeley Water Center, Department of Civil and Environmental Engineering, University of California, Berkeley, Berkeley, California, USA.
[2]Lawrence Livermore National Laboratory, Livermore, California, USA.
[3]Department of Civil and Environmental Engineering, University of California, Irvine, Irvine, California, USA.

[4] Lately a new scheme has emerged which seeks to obtain a consensus from a combination of multiple model predictions so that one model's output errors can be compensated by others'. The combination techniques can be categorized into two groups. The first group [e.g., *Shamseldin et al.*, 1997; *Abrahart and See*, 2002; *Georgakakos et al.*, 2004; *Ajami et al.*, 2005, 2006] uses a set of deterministic weights to combine multiple model outputs. Methods of simple model average (equal weights), linear regression, or artificial neural network (ANN) belong to this category. The consensus prediction from these methods is an alternative deterministic prediction without uncertainty estimates. In addition, the weights in such combination can take any arbitrary real (positive or negative) values that lack physical interpretations.

[5] The second group such as Bayesian model averaging (BMA) [*Madigan et al.*, 1996; *Hoeting et al.*, 1999] uses probabilistic techniques which derive the consensus prediction from competing predictions using likelihood measures as model weights. The likelihood measure (weight) for each member model is based on the success frequency of the predictions that an individual model has made within the observations. For this reason, BMA weights are tied directly to individual model performance. BMA has been applied in a variety of fields including statistics, management science, medicine, and meteorology [e.g., *Viallefont et al.*, 2001; *Fernandez et al.*, 2001; *Raftery et al.*, 2003, 2005; *Wintle et al.*, 2003]. In many case studies, the BMA has shown to produce more accurate and reliable predictions than other multimodel techniques [*George and McCulloch*, 1993; *Raftery et al.*, 1997; *Clyde*, 1999; *Viallefont et al.*, 2001; *Raftery and Zheng*, 2003; *Ellison*, 2004]. Very recently, the BMA method was applied to hydrologic groundwater modeling [*Neuman and Wierenga*, 2003; *Neuman*, 2003].

[6] The intend of this study is to build a hybrid framework, Integrated Bayesian Uncertainty Estimator (IBUNE), to confront the uncertainties in rainfall-runoff predictions associated with input errors, model parameters estimates, and model structural deficiencies. To accomplish this objective, the paper is divided into three major parts. First, the Shuffled Complex Evolution Metropolis (SCEM) algorithm [*Vrugt et al.*, 2003], which was developed for probabilistic parameter estimation, will be studied. We will demonstrate that not accounting for existing error in the input and model structure could lead to corrupted parameter estimations, as well as unreliable uncertainty bounds on the model predictions. The second part of the paper presents a simple approach to extend SCEM to simultaneously account for the uncertainties originating from both input precipitation data and the model parameters. This is the first step toward building IBUNE. We will demonstrate that the error incorporated within the input (forcing) data is one of the major uncertainty sources in the rainfall-runoff modeling system, and by accounting for it within our uncertainty assessment procedure, we will improve the uncertainty bounds in model prediction. We will also show that not assessing model structural uncertainty is still an important limitation of this part of the study.

[7] Finally, the intent of the third part of this paper is to consider model structural uncertainty in addition to input and parameter uncertainty. We present a hybrid approach where we merge the strengths of the Bayesian model averaging scheme with the extended SCEM. This is the final step in building the new framework, called IBUNE. IBUNE further reduces the uncertainties caused by the deficiencies in individual models by using Bayesian model averaging, while also accounting for input and parameter uncertainty within individual models by applying extended SCEM. Finally, the IBUNE scheme will be applied to a real case study in the Leaf River basin.

2. Study Basin and Hydrological Models

[8] We have selected the Leaf River basin to demonstrate the performance of presented frameworks in this study. This 1949-km^2 basin is located north of Collins, Mississippi. Five years of daily historical data (1953–1957), including precipitation (millimeters per 6 hours), potential evapotranspiration (mm/d), and streamflow (m^3/s) were used for calibration and uncertainty assessment. Since many other studies were conducted over the period of 1953–1957 [*Yapo et al.*, 1998; *Gupta et al.*, 1998; *Hogue et al.*, 2003; *Vrugt et al.*, 2003], for comparison purposes we selected the same period for this study. To reduce the sensitivity to initial state variables, a 365-day (through water year 1952) warm-up period was used, during which no calibration and uncertainty estimation was performed for any of the under study hydrologic models. Three hydrologic models were selected for this study including the Sacramento soil moisture accounting (SAC-SMA) model [*Burnash et al.*, 1973], the hydrologic model (HYMOD) [*Boyle*, 2001], and the simple water balance (SWB) [*Schaake et al.*, 1996] model.

[9] SAC-SMA is a nonlinear, time-continuous, and conceptual rainfall-runoff model [*Burnash et al.*, 1973] and is being used operationally by many of the U.S. National Weather Service River Forecast Centers (NWS-RFC) for flood forecasting. The model includes two soil moisture layers, an upper and lower zone (Figure 1). This model includes 16 parameters, three of which were fixed at specified values; the remaining 13 parameters need to be determined through some kind of search process.

[10] Because the Leaf River basin has been studied extensively for optimization purposes [e.g., *Yapo et al.*, 1998; *Gupta et al.*, 1998; *Thiemann et al.*, 2001; *Hogue et al.*, 2003], we have gained a very good knowledge of what SAC-SMA parameter values should be for this basin. In order to minimize the interaction between various parameters in the SAC-SMA model and hence reduce the complications due to the nonidentifiably problem, which could cast shadow over the main objectives of this work, the SAC-SMA model was simplified. First we fixed five percolation parameters to prespecified values (Table 1). Further, we maintained the relative values of the parameters associated with the lower zone and the upper zone. Consequently the number of parameters in the SAC-SMA model that need to be identified was decreased to five: upper zone tension water maximum storage (UZTWM); upper zone free-water maximum storage (UZFWM); upper zone free-water lateral depletion rate (UZK); lower zone total maximum storage (LZTM); and lower zone supplementary free-water depletion rate (LZSK). LZTM represents the summation of all lower zone storages. The lower zone primary free-water depletion rate (LZPK) is estimated to be 3% of the lower zone supplemental free-water depletion rate (LZSK).
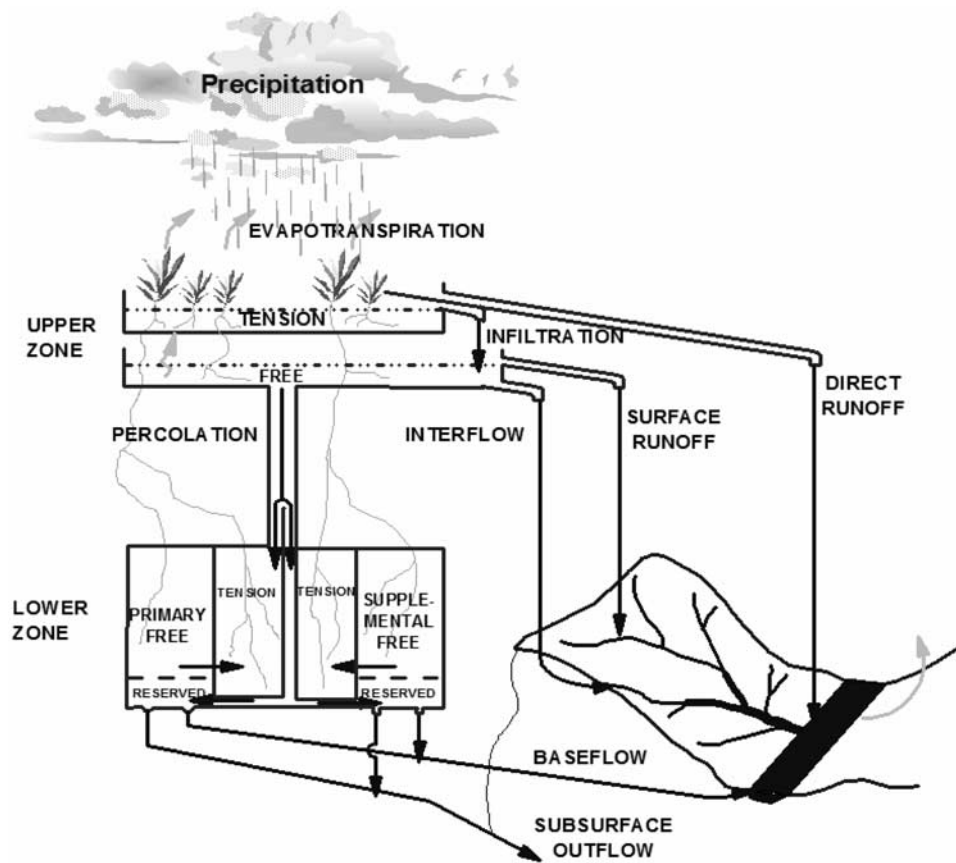
Figure 1. Schematic of the Sacramento soil moisture accounting (SAC-SMA) model [*Brazil*, 1988].

[11] Two simple conceptual rainfall-runoff models were also used in this study: HYMOD and SWB models. The HYMOD [*Boyle*, 2001] consists of a simple rainfall excess model, which is connected to two series of linear reservoirs to route surface and subsurface flow (three quick-flow reservoirs and a single slow-flow reservoir). This model includes five parameters: $C_{\max}$ (L) is the maximum storage capacity in the catchment; $b_{\exp}$ (·) is the shape factor of the main soil-water storage tank that represents the degree of spatial variability of the soil-moisture capacity within the catchment; *Alpha* (·) is the factor distributing flow between

two series of reservoirs; and R_q (T) and R_s (T) are the residence times of linear quick- and slow-flow reservoirs, respectively. The schematic of this model is illustrated in Figure 2. The parameters and their initial uncertainty bounds are presented in Table 2.

[12] The simple water balance (SWB) model [*Schaake et al.*, 1996] is a conceptual, parametric water balance model which is being used as an operational model in the Nile River forecast center. This model includes two soil layers. A thin upper layer represents the vegetation canopy and the soil surface, while a lower layer represents the vegetation

Table 1. Parameters of the Modified SAC-SMA Model

Parameter	Description	Prior Range
UZTWM	upper zone tension-water capacity, mm	1.00–150.0
UZFWM	upper zone free-water capacity, mm	1.00–150.0
UZK	upper zone recession coefficient, day^{-1}	0.10–0.5
LZTM	total lower zone water capacity, mm	1.00–1000.0
LZSK	lower zone supplementary recession coefficient, day^{-1}	0.01–0.25
Percolation and other parameters (not optimized)		
ADIMP	additional impervious area	0.15
PCTIM	impervious fraction of the watershed area	0.025
ZPERC	minimum percolation rate coefficient	200.0
PFREE	percentage percolating from upper to lower zone free water storage	0.1
REXP	exponent of the percolation equation	3.3
RIVA	riparian vegetation area	0.01
SIDE	ratio of deep recharge to channel base flow	0.0
RSERV	fraction of lower-zone free water not transferable to tension water	0.3
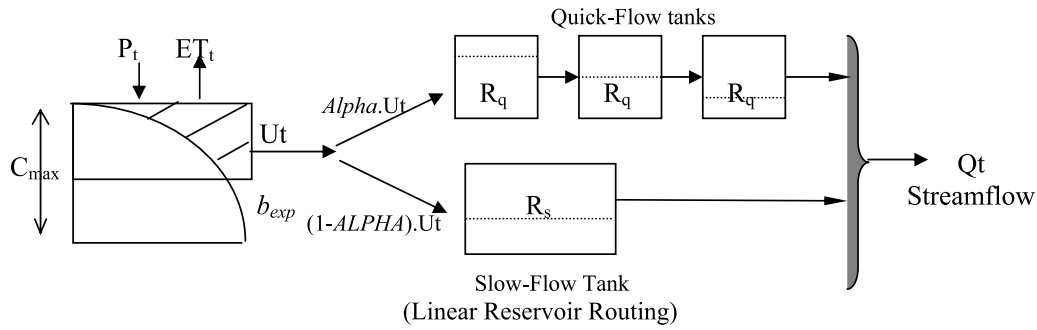
Figure 2. Schematic of the **hydrologic model** (HYMOD) [*Wagener et al.*, 2001].

root zone and groundwater system. Five parameters controlling the SWB model processes are $D_{b,\max}$, the maximum soil-moisture deficit of the bottom layer of the soil; $Q_{\max}$, the potential subsurface runoff; $\beta = Q_{\max}/S_{\max}$, the ratio of the lower level posture that produces subsurface flow ($S_{\max}$ is the minimum threshold that guarantees subsurface flow); $\alpha = D_{u,\max}/D_{b,\max}$, the upper layer deficit proportion ($D_{u,\max}$ is the maximum soil-moisture deficit of the upper layer); and K_{dt}, the timescale factor that controls infiltration into the bottom layer and the surface runoff amount. The schematic of this model is illustrated in Figure 3. The SWB model parameters and their initial uncertainty bounds are listed in Table 3.

3. Traditional Uncertainty Assessment in Hydrological Modeling

3.1. Derivation of Likelihood Function for Assessment of Parameter Uncertainty

[13] A typical hydrologic model, M, can be represented as follows:

$$y = M(\tilde{X}, \theta),\qquad(1)$$

where y represents the response matrix of the catchment (e.g., streamflow), $M(\cdot)$ denotes the nonlinear hydrologic model, θ is a set of model parameters, and $\tilde{X}$ stands for the observed forcing input matrix (e.g., precipitation). In the traditional approach, the uncertainty in the catchment response is attributed to parameter estimation uncertainty, while input and model structural uncertainty is not addressed explicitly. Assuming that the residuals are additive,

$$\tilde{y} = M(\tilde{X}, \theta) + e(\theta).\qquad(2)$$

[14] The Bayesian statistics treats hydrologic model parameter, θ, as probabilistic variables, with the joint posterior probability distribution $P(\theta \mid \tilde{X}, \tilde{y})$, which presents the probabilistic characteristic of the θ conditioned on the observed data, $\tilde{X}$ and $\tilde{y}$. Under Bayes statistics, $P(\theta|\tilde{X}, \tilde{y})$ is proportional to the product of likelihood function and the prior distribution function, $P(\theta)$. The prior probability density function explains the information about the θ, before any data are collected. Here we use a noninformative (uniform) prior over the feasible parameter space (which consists of realistic upper and lower bound for each of the parameters), $\theta \in \Theta \subset \Re^n$.

[15] Assuming that the residuals are additive, independent (uncorrelated), and normally distributed noise with mean equal to zero and constant unknown variance, σ_y, *Box and Tiao* [1973] described the likelihood of parameter set describing the observed data over the number of time steps (T) can be estimated as follows:

$$L(\theta, \sigma_y|\tilde{X}, \tilde{y}) = \frac{1}{\sigma_y^T} \exp\left(-\frac{1}{2\sigma_y^2}\left(\sum_{t=1}^{T}(e(\theta)_t)^2\right)\right).\qquad(3)$$

[16] Further assuming noninformative prior, then $P(\sigma_y) \propto \sigma_y^{-1}$, σ_y can be integrated out of the posterior density yielding the following expression [*Box and Tiao*, 1973]:

$$p(\theta|\tilde{X}, \tilde{y}) \propto \left[\sum_{t=1}^{T}(e(\theta)_t)^2\right]^{-\left(\frac{T}{2}\right)}.\qquad(4)$$

[17] In practice, it is easier to maximize the logarithm of the likelihood function. It will identify a set of plausible parameter values given the available observed data. There are several Bayesian approaches tailored for hydrologic modeling, including the Generalized Likelihood Uncertainty

Table 2. Parameters of the HYMOD Model and Their Initial Uncertainty Ranges

Parameter	Description	Prior Range
$C_{\max}$	maximum storage capacity in catchment, mm	1.0–500.0
$b_{\exp}$	factor distributing flow between two series of reservoirs (·)	0.1–2.0
$ALPHA$	shape factor for the main soil water storage tank (·)	0.1–0.990
R_s	residence time of linear slow flow reservoirs, days	0.0–0.1
R_q	residence time of linear quick flow reservoirs, days	0.1–0.99
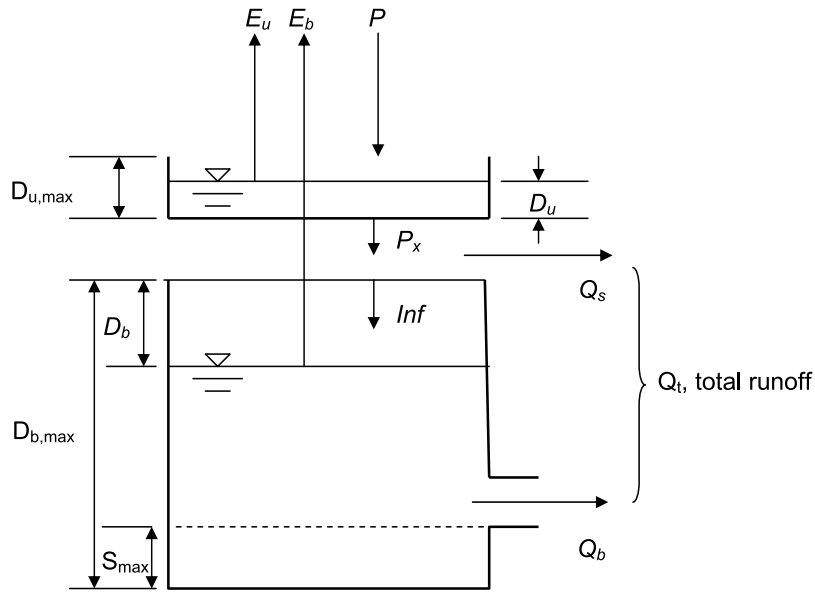
Figure 3. Schematic of the simple water balance (SWB) model.

Estimation (GLUE) framework [*Beven and Binley*, 1992] and the Shuffled Complex Evolution Metropolis (SCEM-UA) algorithm [*Vrugt et al.*, 2003] that consider model parameters in equation (1) as probabilistic variables and estimate their uncertainty bound based on the posterior pdf. In this study, we will further explore the SCEM-UA algorithm for estimating model parameters and their associated uncertainty bounds.

3.2. The Shuffled Complex Evolution Metropolis

[18] The Shuffled Complex Evolution Metropolis (SCEM) was built upon the principles of the effective and efficient global optimization technique, the Shuffled Complex Evolution (SCE-UA) developed by *Duan et al.* [1992]. *Vrugt et al.* [2003] combined the strengths of the Monte Carlo Markov Chain (MCMC) sampler with the concept of complex shuffling from SCE-UA to form an algorithm that not only provides the most probable parameter set, but also estimates the uncertainty associated with estimated parameters. The main difference between SCEM and SCE is that the downhill simplex method in SCE was replaced by the Metropolis-Hastings search algorithm [*Metropolis et al.*, 1953; *Hastings*, 1970]. Thus SCEM in every model run is able to simultaneously identify both the most likely parameter set and its associated posterior probability distribution. SCEM-UA is explained in detail by *Vrugt et al.* [2003]. The convergence of the algorithm was monitored using the Gelman-Rubin criterion [*Gelman and Rubin*, 1992], which

is a scale reduction score that quantitatively diagnoses if each parameter converges to a stationary distribution.

3.3. Case Study: Use of SCEM for Calibration and Uncertainty Assessment of Hydrologic Model Parameters

[19] In this section we demonstrate the performance and applicability of SCEM-UA to identify and estimate model parameters and their associated uncertainty bounds, by application to three above mentioned hydrologic models: SAC-SMA [*Burnash et al.*, 1973], HYMOD [*Boyle*, 2001], and SWB [*Schaake et al.*, 1996].

[20] Input-forcing data and model structures were assumed perfect in this section, and all of the uncertainty in the streamflow simulation was attributed to parameter estimation uncertainty. Uniform prior distributions were assumed on the parameter ranges of all three models. The marginal posterior probability distribution for the estimated SAC-SMA model parameters is given in Figure 4. As we mentioned earlier, the number of unknown parameters in this model was reduced to five major parameters. The distributions are generated using 20,000 samples after the algorithm converged to the final posterior distribution. Figure 4 illustrates two points. The first point is that the posterior distributions for three of the five parameters (UZTWM, LZTM, and LZSK) are approximately normal; however, the posterior distribution of UZFWM depicts the existence of two modes (multimodality). The posterior

**Table 3.** Parameters of the SWB Model and Their Initial Uncertainty Ranges

Parameters	Description	Prior Ranges
$D_{b,max}$	maximum soil moisture deficit of bottom layer of the soil, mm	10.0–800.0
Q_{max}	potential subsurface runoff, mm/d	5.0–100.0
β	ratio of the lower level posture that produces subsurface flow (·)	0.1–0.90
α	upper layer deficit proportion (·)	0.01–0.5
K_{dt}	timescale factor, days	1.0–20.0
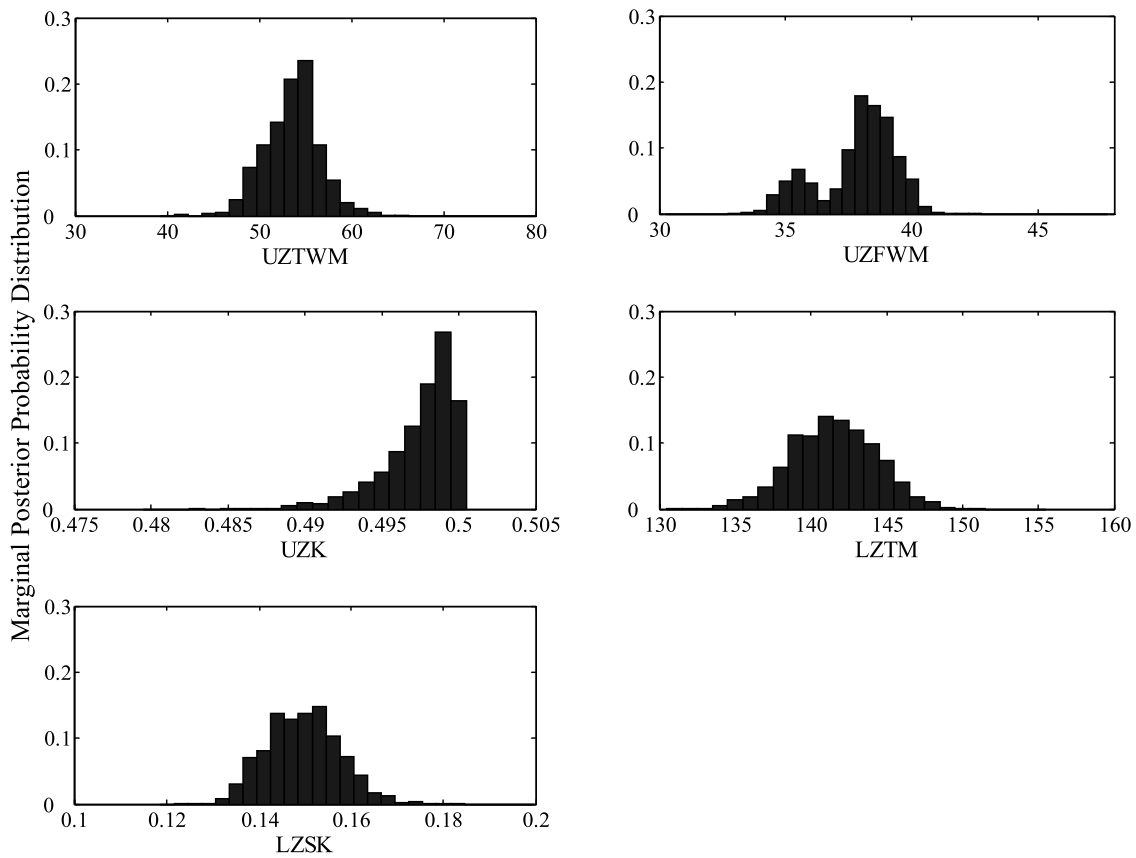
Figure 4. Marginal posterior probability distribution of the SAC-SMA parameters, using 20,000 samples generated after convergence of the SCEM-UA algorithm.
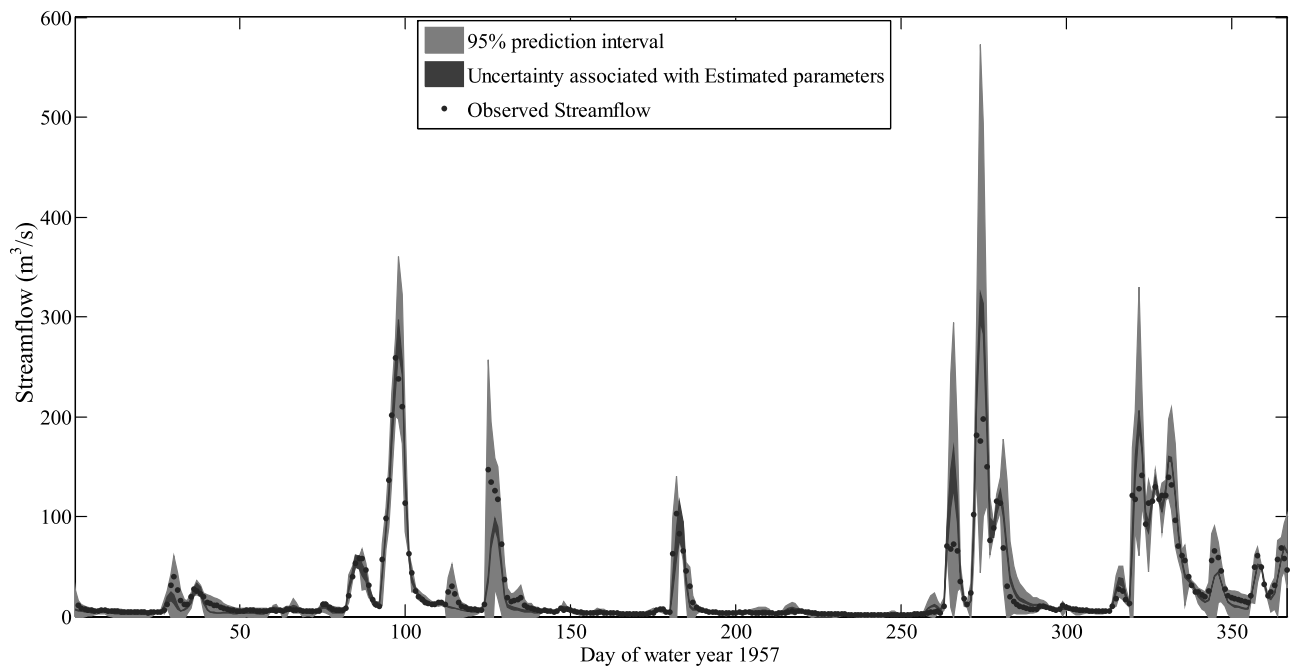
Figure 5. Streamflow hydrograph prediction uncertainty associated with estimated parameters (shown in darker gray) for the SAC-SMA model and 95% confidence interval for prediction of observed streamflow (shown in lighter gray) for water year 1957.
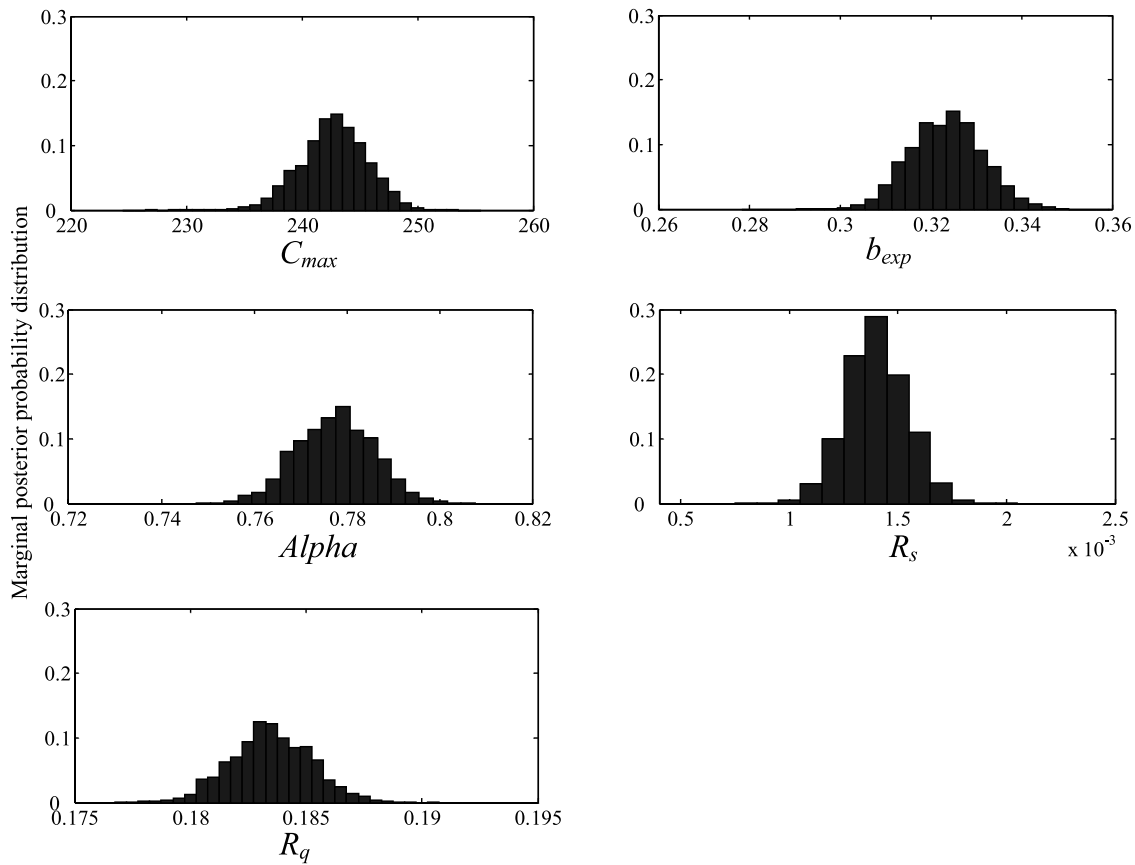
Figure 6. Marginal posterior probability distribution of the HYMOD parameters, using 20,000 samples generated after convergence of the SCEM-UA algorithm.

distribution of UZK is very close to the upper boundary of the National Weather Service predefined probable parameter range. This can be an indication of an inherent model structural uncertainty and/or other sources of uncertainty within the system which are not being considered here. The second observation is that the final converged samples for all the parameters capture only a small space of the predefined range for the parameters (Table 1). However, the hydrograph uncertainty bounds (Figure 5) associated with these parameter ranges do not cover the expected number of observed streamflow values (dark gray region in Figure 5). This can be argued as a problem of over-conditioning the selected relationships between observed and simulated (modeled) output. The light gray region in Figure 5 shows the 95% hydrograph prediction uncertainty associated with the total error in the hydrologic system in terms of model residuals (calculated based on predictive variance of SCEM). Even though the 95% total prediction uncertainty range captures all the observations, it is very wide compared with uncertainty bounds associated with parameter uncertainty, revealing a considerable amount of uncertainty in both the structure of the model under study and the data used to condition the model.

[21] To further demonstrate the applicability of SCEM, we used this algorithm to estimate optimal parameter sets and assess their associated uncertainty boundaries for two other hydrologic models, HYMOD [*Boyle*, 2001] and SWB [*Schaake et al.*, 1996].

[22] The final estimated marginal posterior distributions of the HYMOD model parameters, after 20,000 samples, are given in Figure 6. The results reveal that the distributions for all HYMOD parameters are approximately normal. These parameter distributions cover a very small range of predefined parameter ranges. However, in Figure 7 we can see that even though the algorithm shows high probability for these parameter sets, the estimated hydrograph prediction uncertainty bounds (dark gray) does not include many of the observed streamflow values. Similar results are presented in Figures 8 and 9 for the SWB model.

[23] The examples presented above reveal that attributing all uncertainties in hydrologic models to model parameters and ignoring input and model structural uncertainties leads to an inaccurate, biased, and inconsistent simulation of the system processes and their associated uncertainty bounds.

4. Extended SCEM-UA to Include the Input Error Model: Simultaneous Parameter and Input Uncertainty Estimation

[24] Results from the previous section indicate that dealing only with model parameter uncertainty is not enough to accurately estimate the true uncertainty in hydrologic simulation. Uncertainties from other sources must be dealt with more directly. There have been a few studies in hydrological modeling that explicitly account for input uncertainty within the system through input error models. One such approach is the Bayesian total error analysis (BATEA) by *Kavetski et*
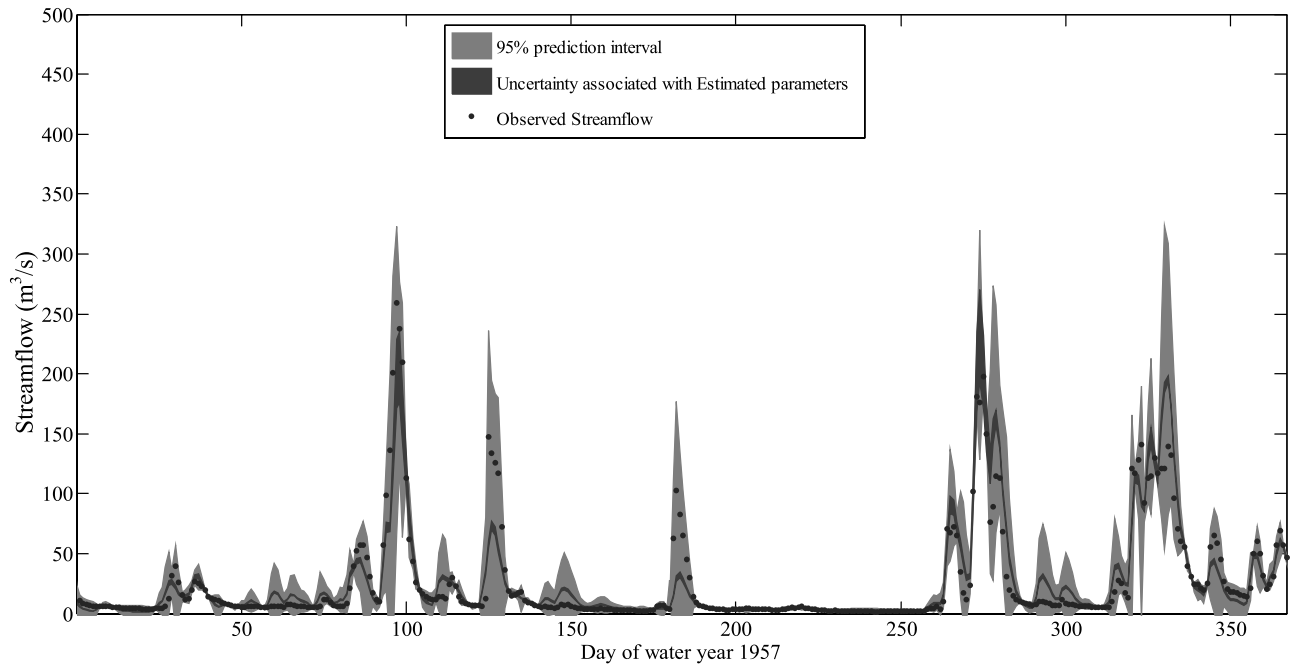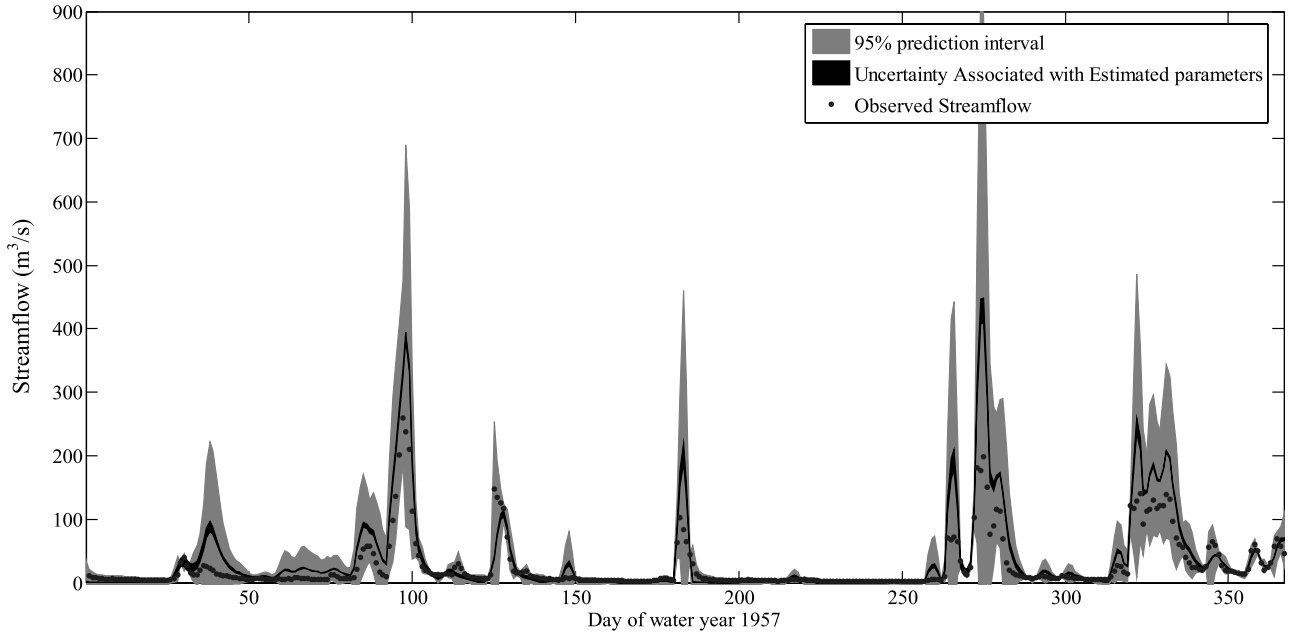
Figure 7. Streamflow hydrograph prediction uncertainty associated with estimated parameters (shown in darker gray) for the HYMOD model and 95% confidence interval for prediction of observed streamflow (shown in lighter gray) for water year 1957 (calibration period).
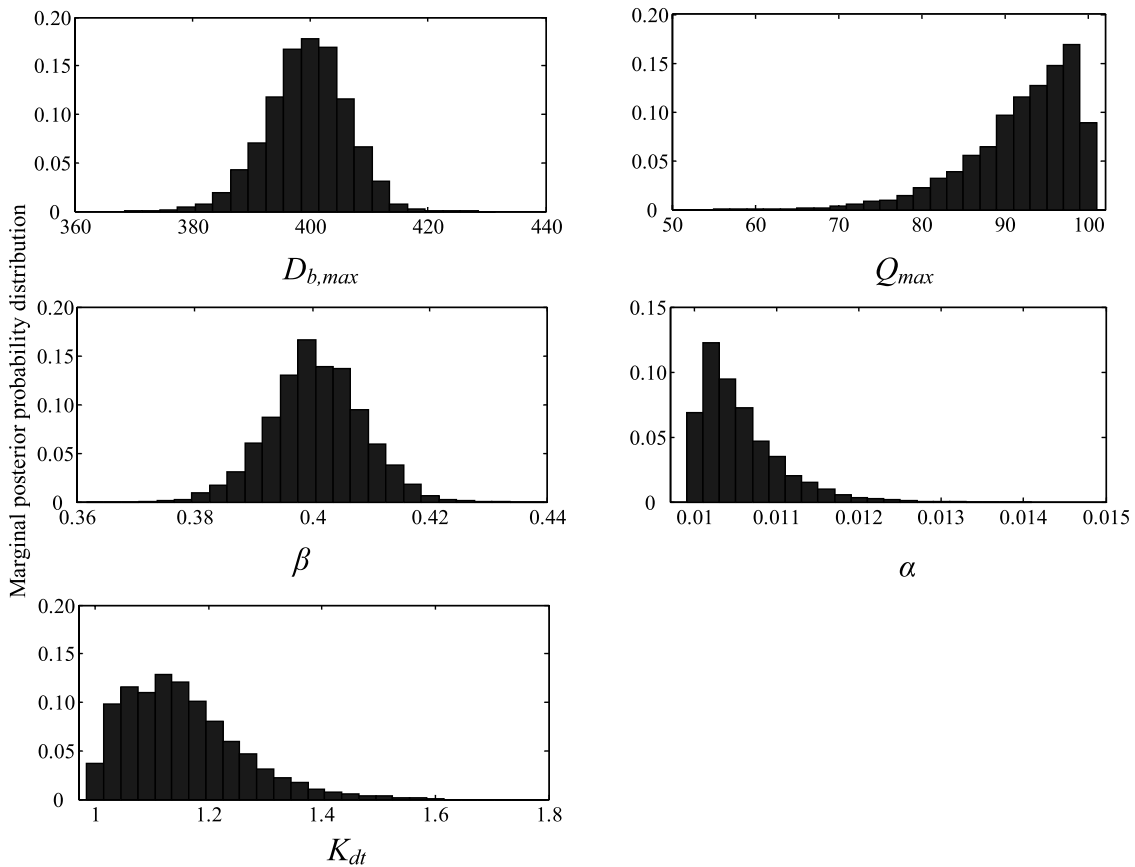
Figure 8. Marginal posterior probability distribution of the SWB parameters, using 20,000 samples generated after convergence of the SCEM-UA algorithm.

Figure 9. Streamflow hydrograph prediction uncertainty associated with estimated parameters (shown in darker gray) for the SWB model and 95% confidence interval for prediction of observed streamflow (shown in lighter gray) for water year 1957 (calibration period).

al. [2003]. BATEA is one of the few techniques which explicitly considers input error in the development of the likelihood function in hydrological modeling. The rainfall events are predefined and each is given a unique multiplying constant. Theses multipliers allow the pattern of rainfall as well as the event magnitude to change. *Kavetski et al.* [2003] introduced rainfall depth multipliers as some "latent variables" to the system and introduced an explicit term to the likelihood function to estimate these variables. If $\tilde{r}_t$ represents the true rainfall depth $\hat{X} = [\tilde{r}_1, \tilde{r}_2, \ldots \tilde{r}_t, t = 1:T]$, and r_t is the observed rainfall depth, their input error model has the following form:

$$r_j = m_j \tilde{r}_j; \qquad m \sim N\left(1, \sigma_m^2\right), \tag{5}$$

where j indicates the storms within the rainfall series and m_j is the random noise from a normal distribution with mean equal to one and known (prespecified) variance σ_m^2 in the form of a multiplier that corrupts the true rainfall depth and yields the observed rainfall depth. *Kavetski et al.* [2003] assumed the rainfall multipliers, m_t, as latent variables and estimated both them and the model parameters through their probabilistic calibration procedure called BATEA. By considering the multipliers just for the predefined rainfall events, they decreased dimensions of the system. Considering Bayes' law, and assuming that (1) $\tilde{X}$ (observed input) and $\tilde{y}$ (observed catchment response) are statistically independent because catchment response $\tilde{y}$ depends only on the true input forcing $\hat{X}$, not necessarily on observed forcing, and (2) $\tilde{X}$ is statistically independent of θ (model parameter set), because observed input is uncorrelated to the hydrologic model parameters, *Kavetski et al.* [2003] derived the final form of their likelihood function as follows:

$$p\left(\theta, \hat{X} | \tilde{X}, \tilde{y}\right) \propto L\left(\tilde{y} | \theta, \hat{X}\right) \times L\left(X | \hat{X}\right) \times p\left(\theta, \hat{X}\right), \tag{6}$$

where $L(\tilde{y}|\theta, \hat{X})$ is the likelihood of observing $\tilde{y}$ given a parameter set θ, and the true input forcing $\hat{X}$. $L(\tilde{X}|\hat{X})$ is the likelihood based on input error model, and $p(\theta, \hat{X})$ represents the prior distribution of parameters and true input forcing.

[25] *Kavetski et al.* [2003] applied their BATEA framework to a series of synthetic case studies and demonstrated that considering an input error model explicitly and adding a new term to the likelihood function can improve the response surface and assessment of uncertainty bounds. Nonetheless, even though equation (6) allows the use of explicit input error models, it has two drawbacks. First, it is impossible to know what the true input forcing is in a real-world problem, and therefore it is impossible to assess the input error model likelihood, $L(\tilde{X}|\hat{X})$. Second, in some cases the number of these "latent variables" can increase considerably and cause some dimensionality issues. To circumvent these two problems, in this study the input error model was changed as follows:

[26] 1. Instead of introducing latent variables to the system, we considered a multiplier in the following form:

$$\tilde{r}_t = \phi_t r_t; \qquad \phi \sim N\left(m, \sigma_m^2\right), \tag{7}$$

where ϕ_t represents a random multiplier at time step t with mean equal to m, $m \in [0.9, 1.1]$ and variance equal to σ_m^2, $\sigma_m^2 \in [1e-5, 1e-3]$. In this implementation we assume true rainfall depth $\tilde{r}_t$ is corrupted at all times by random multipliers from the identical distribution with unknown mean, m, and variance, σ_m^2. Thus, instead of searching for every single multiplier as a latent variable, we introduce two new parameters to the system including mean and variance of error model multiplier (instead of additive) distribution, $\eta = \{m, \sigma_m^2\}$. Considering the error term in the form of the multiplier helps to maintain the heteroscedastic (nonhomo-
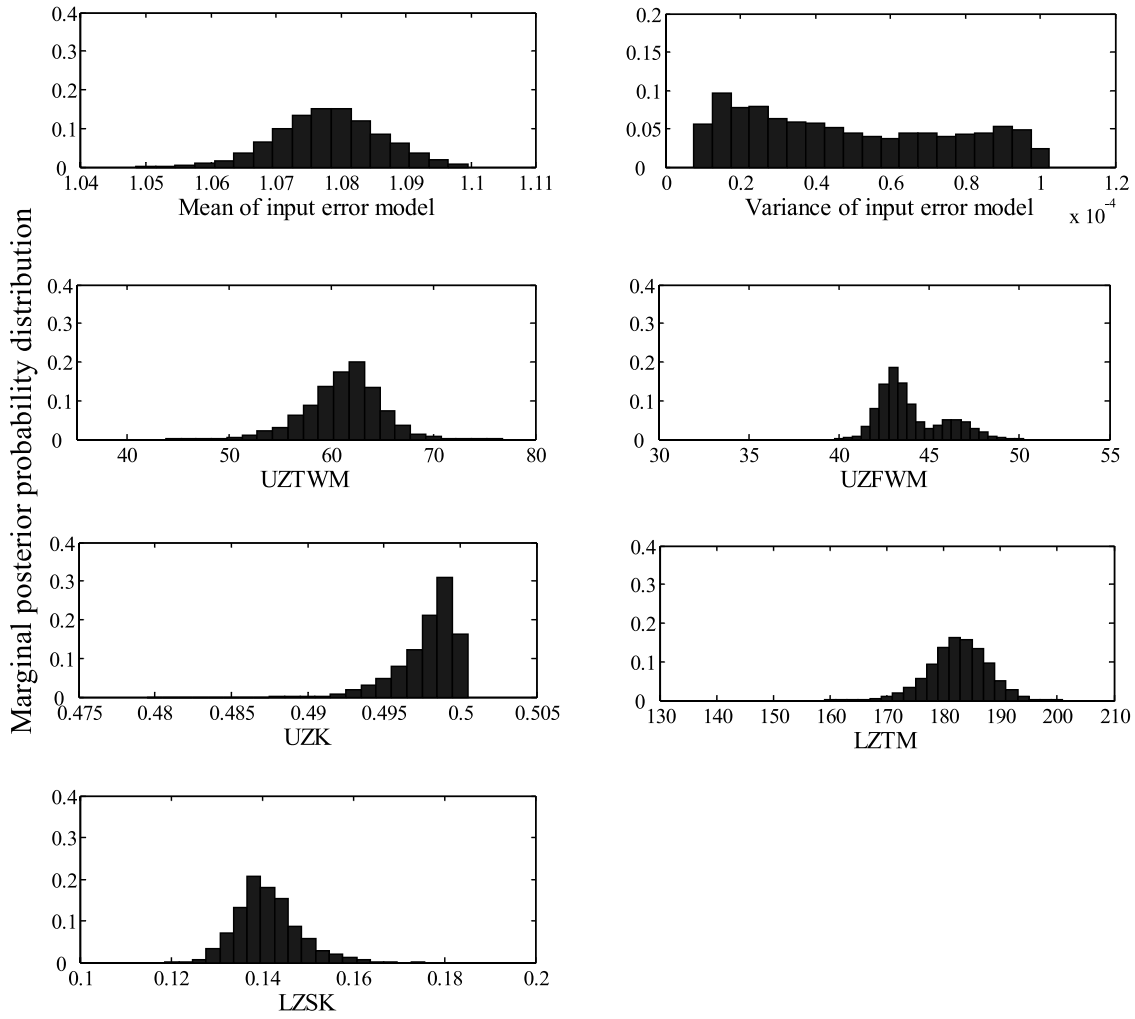
Figure 10. Marginal posterior probability distribution of the input error model parameters and the SAC-SMA model parameters using 20,000 samples generated after convergence of the SCEM-UA algorithm.

geneous) nature of the error (higher deviation in higher rainfall depths) [*Sorooshian and Dracup*, 1980].

[27] 2. To deal with the issue of not having true observations of input forcing data, it was decided to integrate the input error model into the model error term:

$$e(\theta) = y(\theta, \eta) - \tilde{y} \qquad (8)$$

Therefore the likelihood function will have the following form:

$$p(\theta, \eta | \tilde{X}, \tilde{y}) \propto L(\tilde{y} | \theta, \eta, \tilde{X}) \times p(\theta, \eta). \qquad (9)$$

[28] In brief the implemented changes into the hydrologic input-output system included introduction of a random multiplier to each time step, drawn from the same normal distribution with unknown mean and variance (m and σ_m^2). These two variables of the input error model (mean and variance of the distribution) were added as two unknown parameters to the system. The SCEM-UA was used to estimate the model parameters and input error model parameters simultaneously. Later the uncertainty associated with input error model parameters and hydrologic model

parameters were propagated through the system to estimate associated uncertainty with streamflow simulations and predictions.

4.1. Case Study: Use of Extended SCEM for Calibration and Uncertainty Assessment of Hydrologic Model Parameters and Input Error Model Parameters

[29] By means of a case study, we illustrate the performance of the SCEM-UA while considering an input error model to specify the hydrologic system. Again, we applied SCEM-UA to calibrate and assess uncertainty bounds for SAC-SMA, HYMOD, and SWB model parameters along with input error model parameters on the Leaf River basin. The idea is to compare the results from this part of the study to those from section 2.3.

[30] Figure 10 shows the new marginal posterior distribution estimated for each parameter of the SAC-SMA model while considering an input error model's first two moments as two additional parameters in the system, using SCEM-UA. Looking at Figure 10 and comparing the results with Figure 4, two observations can be made. One is that considering input error model, the final estimated marginal distribution for the model parameters moved over the possible parameter ranges and assigned the mode of the
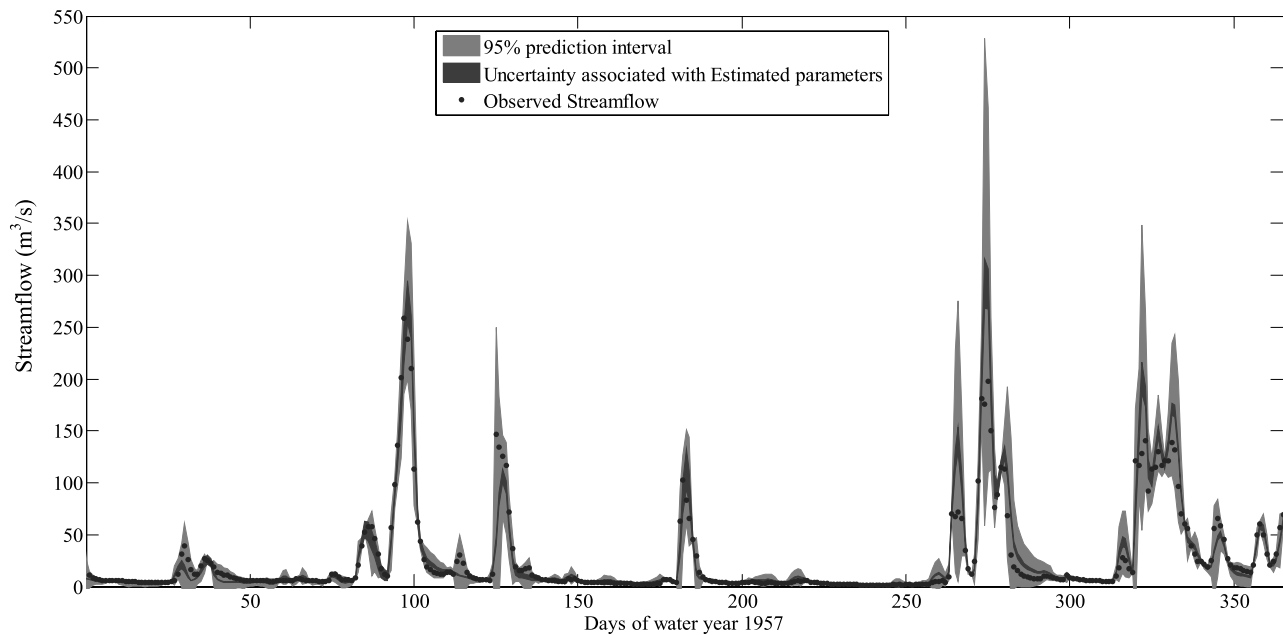
Figure 11. Streamflow hydrograph prediction uncertainty associated with estimated parameters and input error model parameters (shown in darker gray) for the SAC-SMA model and 95% confidence interval for prediction of observed streamflow (shown in lighter gray) for water year 1957 (calibration period).

probability distribution to different parameter values. The second observation is that the mean of the input error model has a mode different than one. If the input forcing was correct, the mean of the input error model would concentrate around one, and the final marginal distribution of the parameters would be the same as if we did not account for input uncertainty. This can also be the indication that the input error model is somehow compensating for the existing model structural deficiencies. The estimated uncertainty bounds for the hydrograph associated with input and model parameter uncertainty are shown in Figure 11. The 95% prediction intervals are narrower here compared with the original case (only considering uncertainty in model parameters). This reveals that the final uncertainty bounds associated with both input and model parameters are more accurate and that variance of the residuals at each point is smaller compared with the original scenario.

[31] These above mentioned results for the SAC-SMA model are confirmed in Table 4. The observation coverage for the estimated 95% uncertainty bounds for the simulation has increased by almost 70% when we account for input uncertainty. The same results are presented in the table for the HYMOD and SWB models, which reveals that accounting for input uncertainty improved the final streamflow simulation of these models as well. These results illustrate that not accounting for input uncertainty can lead to biased parameter estimates, which are compensating for other sources of uncertainty. Accounting for input uncertainty improves the daily root-mean-square (DRMS) error for all three models across all of their ensembles, as seen in Figure 12. We can also see that this improvement is more significant for the SWB and HYMOD models and less significant for the SAC-SMA model.

[32] One of the important observations from the set of experiments presented in this section was that the estimated mean and variance of input error model and their associated

uncertainty bound are different from one hydrological model to the other one. This is an inevitable result since we are still ignoring model structural uncertainty. Therefore all the model parameters as well as input error model parameter are still compensating for model structural uncertainty. The next section focuses on this important source of uncertainty in hydrologic system simulation.

5. Uncertainty Assessment in Hydrological Modeling: Simultaneous Parameter and Input and Model Structural Uncertainty Estimation

5.1. Classical Model Structural Error

[33] The dominant approach in hydrological modeling and streamflow forecasting has been the use of a single model. However, dependence on a single hydrological model, which presumably does not adequately represent all of the physical processes of the watershed well, results in unreliable, uncertain, and overconfident forecasts. This is the case even if we account for all other sources of uncertainty such as parameter estimation and input forcing uncertainty [*Georgakakos et al.*, 2004]. To date, all of the approaches set forth to identify model structural inadequacy focused on a single-model structure and how it can be improved to more adequately represent the system [e.g., *Vrugt et al.*, 2005].

[34] A new kind of approach that has recently emerged to identify model structural uncertainty is to use multimodel combination techniques, which provide a better understand-

Table 4. Percentage of Observations Being in 95% Uncertainty Bounds.

	SAC-SMA	HYMOD	SWB
SCEM (hydrologic model parameters)	15%	10%	5.6%
SCEM (hydrologic + input model parameters)	25%	17%	10%
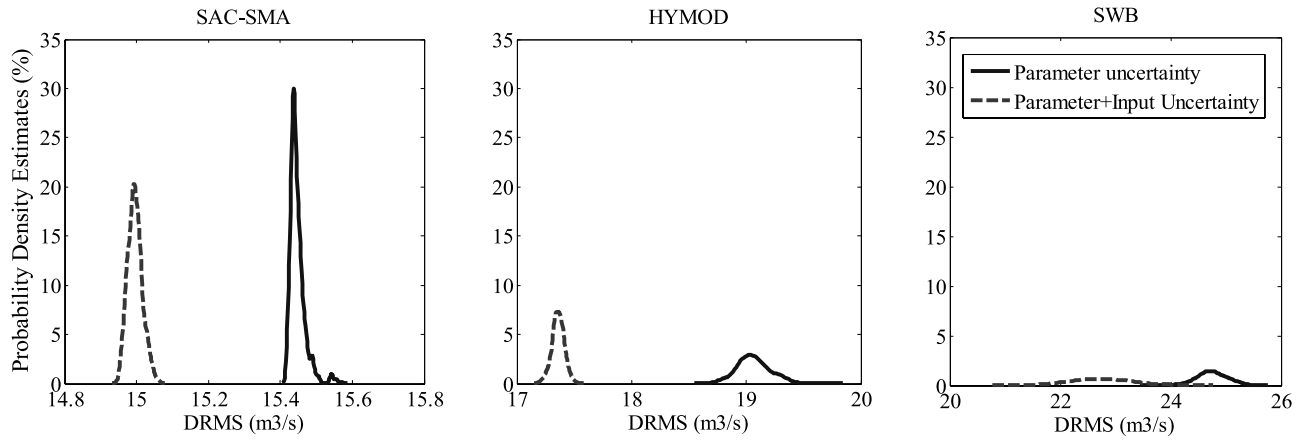
Figure 12. Distribution of daily root-mean square (DRMS) of SAC-SMA, HYMOD, and SWB considering only parameter uncertainty compared to parameter and input uncertainty.

ing of the watershed processes by investigating multiple model structures.

5.2. Bayesian Model Averaging

[35] Bayesian model averaging is a probabilistic scheme for model combination. It is a coherent technique for accounting for model structural uncertainty [*Madigan et al.*, 1996]. Below is a brief description of the essence of the BMA scheme. Consider a quantity $\tilde{y}$ as the observed output variable to be forecasted and $M = [M_1, M_2, \ldots, M_K]$ the set of all considered models. The $p_k (y_k | M_k, \tilde{X}, \tilde{y})$ is the posterior distribution of y_k which represents the quantity to be forecasted under model M_k, given a discrete data set, $\tilde{X}$ (input forcing data) and $\tilde{y}$ (observed system processes, here streamflow). The posterior distribution of the BMA prediction, y_{bma}, is thus given as

$$p\left(y_{bma}|M_1,\ldots,M_k,\tilde{X},\tilde{y}\right) = \sum_{k=1}^{K} p\left(M_k|\tilde{X},\tilde{y}\right) \cdot p_k\left(y_k|M_k,\tilde{X},\tilde{y}\right),$$

(10)

where $p(M_k | \tilde{X}, \tilde{y})$ is the posterior probability of model M_k. This term is also known as the likelihood of model M_k being the correct model. If we denote $w_k = p(M_k | \tilde{X}, \tilde{y})$, we should obtain $\sum_{k=1}^{K} w_k = 1$. 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. Suppose that y_k is a prediction made by model M_k. Weights can be estimated through the expectation-maximization algorithm [*Dempster et al.*, 1977] which will be discussed in the next section. The posterior mean and variance of the BMA prediction for variable y_{bma} are

$$E\left[y_{bma}|y_1 \ldots, y_K, \tilde{X}, \tilde{y}\right] = \sum_{k=1}^{K} w_k y_k$$

(11)

$$Var\left[y_{bma}|y_1 \ldots, y_K, \tilde{X}, \tilde{y}\right] = \sum_{k=1}^{K} w_k \left(y_k - \sum_{i=1}^{K} w_i y_i\right)^2 + \sigma^2,$$

(12)

where σ^2 is the variance of the time series shaped based on one of the model predictions (ensembles) being the best at each time step. Suppose if we build a time series that at each

time step includes the best prediction (closest to the observation) from one of the K models; σ^2 represents the variance of such time series considering observations.

[36] In essence, the BMA prediction is the average of predictions weighted by the likelihood that an individual model is correct. There are several attractive properties to the BMA predictions. First, the BMA prediction receives higher weights from better performing models, as the likelihood of a model is essentially a measure of the agreement between the model predictions and the observations. Second, the BMA variance is a measure of the uncertainty of the BMA prediction. This measure is a better description of predictive uncertainty than that in a non-BMA scheme, which estimates uncertainty based only on the model ensemble spread (i.e., only the between-model variance is considered), and consequently results in underdispersive predictions [*Raftery et al.*, 2003, 2005].

5.3. Combination of Global Optimization and Bayesian Multimodel Combination: An Integrated Bayesian Uncertainty Estimator

[37] Because the Bayesian multimodel combination framework offers an excellent statistical approach to account for model structural uncertainty, the BMA framework was combined with the SCEM-UA to form a hybrid framework to exploit the strengths of these two techniques for integrated schemes for quantification of input, parameter estimation, and model structural uncertainty. This framework should provide a more precise measure of uncertainty in system simulations. Throughout the remainder of this paper we will refer to this integrated Bayesian uncertainty estimator framework as IBUNE.

[38] IBUNE first estimates the two terms in the right-hand side of equation (10), $p_k (y_k | M_k, \tilde{X}, \tilde{y})$ and $p(M_k | \tilde{X}, \tilde{y})$, for each model. The $p_k (y_k | M_k, \tilde{X}, \tilde{y})$, which represents the posterior distribution of estimated hydrologic response (e.g., streamflow), y_k, under model M_k, is directly related to the input and parameter uncertainty under model M_k, expressed as follows:

$$p\left(y_k|M_k,\tilde{X},\tilde{y}\right) \propto p\left(\theta_k, \eta_k|M_k,\tilde{X},\tilde{y}\right).$$

(13)
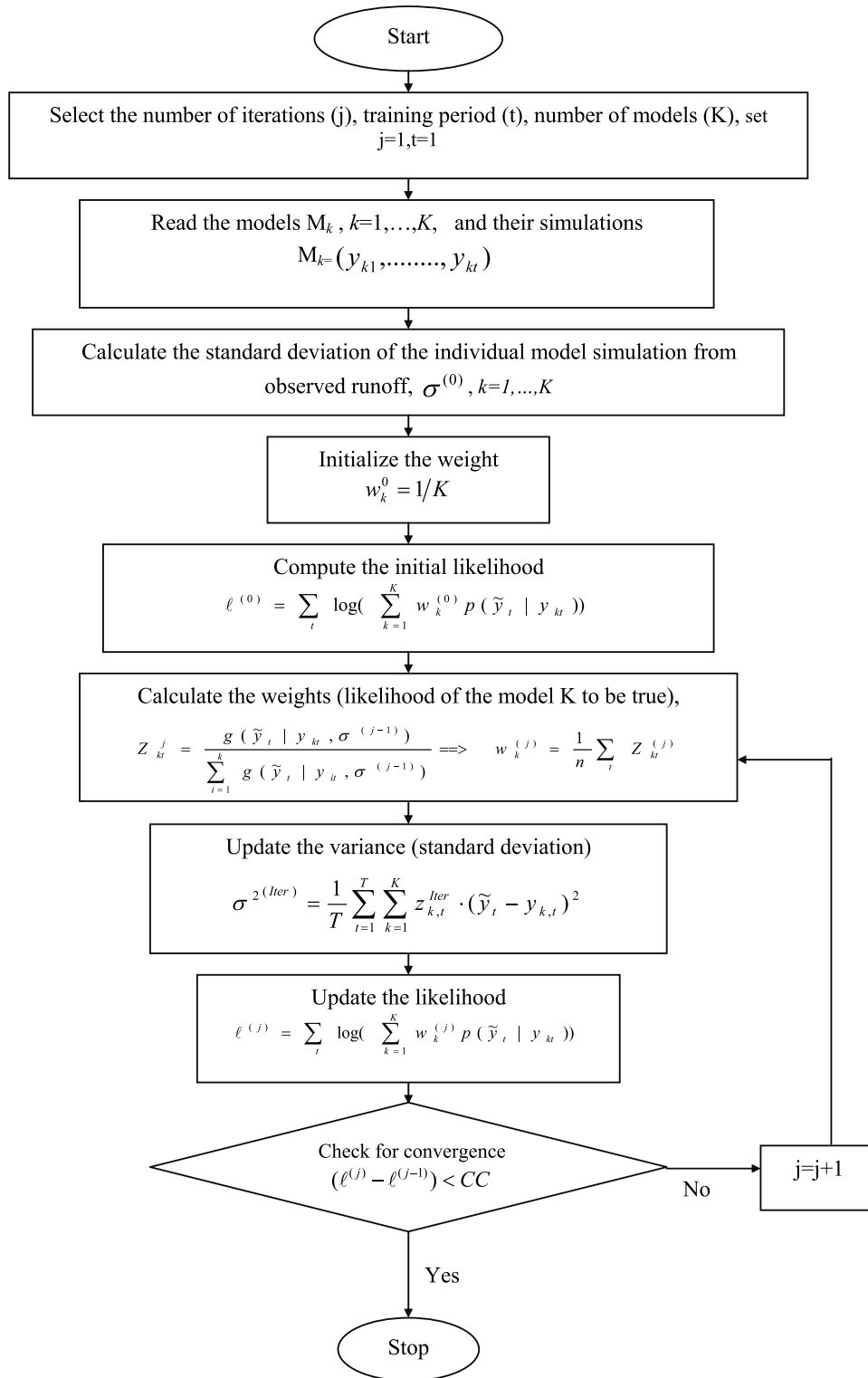
Figure 13. Expectation-maximization (EM) flowchart.

[39] We can substitute the left-hand side of equation (13), which is the outcome of SCEM-UA, into equation (10) directly. The first term of equation (10), $p(M_k|\tilde{X}, \tilde{y})$, which represents the posterior probability of the model M_k being a correct model, reflects how well model M_k matches the observed quantity of interest. To estimate $p(M_k|\tilde{X}, \tilde{y})$ or as we mentioned earlier in the previous section w_k and σ^2 (the variance of the best time series shaped based on one of the model predictions (ensembles) being the best at each time step), we used the maximum likelihood approach. The idea is to estimate w_k and σ^2 by maximizing the likelihood of occurrence of the observed data, $\tilde{y}$. As we mentioned earlier,
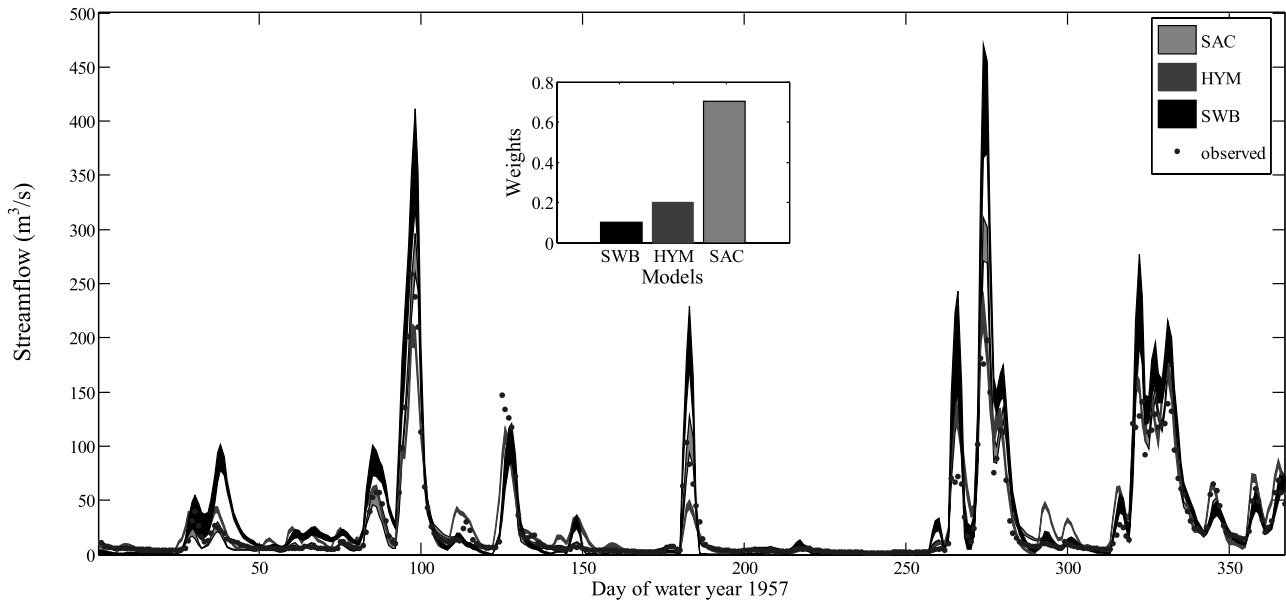
Figure 14. Streamflow hydrograph prediction uncertainty associated with estimated parameters and input error model parameters for all three models for water year 1957 and estimated combination weights for each model using Integrated Bayesian Uncertainty Estimator (IBUNE) (training/calibration period).

it is easier to maximize the logarithm of likelihood function, and therefore we define the logarithm of likelihood function as follows:

$$L\left(w_1, \ldots, w_k, \sigma^2\right) = \sum_{t=1}^{T} \log\left(\sum_{k=1}^{K} w_k . p(\tilde{y}_t | y_{kt})\right). \quad (14)$$

[40] Because of the high dimensionality of this problem, it is hard and inefficient to maximize equation (14) through direct nonlinear maximization methods such as Newton-Raphson or its variants [*Raftery et al.*, 2003]. In this study, to approximate equation (14) and estimate the model weights as *Raftery et al.* [2003] suggested, we maximized the equation (14) by performing the expectation-maximization technique. All of the conditional densities were described as Gaussian distribution for computational simplicity; however, the BMA scheme can be applied by assuming other probability distributions. The EM algorithm is applied to estimate w_k and σ^2 for each model. In brief, the expectation-maximization [*Dempster et al.*, 1977] algorithm casts the maximum likelihood problem as a "missing data" problem. The missing data here are introduced as a latent variable $Z_{k,t}$ that needs to be estimated. If the kth model ensemble is the best prediction at time t, $Z_{k,t} = 1$; otherwise $Z_{k\,t} = 0$. At any time t, there is only one $Z_{k,t}$ equal to 1 and the rest are equal to 0. The EM algorithm starts with an initial guess for w_k and σ^2 (the variance of the time series shaped based on one of the models being best at each time step) and then alternates between the E (or expectation) step, which estimates $Z_{k,t}$ based on the current value of w_k and σ^2, and the M (or maximization) step, where new values for w_k and σ^2 are estimated based on the current value of $Z_{k,t}$. The EM algorithm is described in Figure 13. For more detail description of the EM algorithm, readers are referred to *McLachlan and Krishnan* [1997].

[41] After convergence of this algorithm, we will have specified weights for each model. Therefore equation (10) can be derived and the posterior mean and variance of the forecast can be estimated through equations (11) and (12), respectively.

[42] In brief, the IBUNE framework can be implemented as follows:

[43] 1. Select the number of hydrologic models.

[44] 2. Assign prior probability to each model (we assume noninformative prior which gives uniform weights to all of the models).

[45] 3. Define an input error model.

[46] 4. Obtain posterior distribution of model parameters and input error model parameters for each model using SCEM [*Vrugt et al.*, 2003].

[47] 5. Generate a prespecified number of streamflow ensembles for each model, using probabilistic parameter estimates obtained from steps 2–4.

[48] 6. Estimate the model weight and variance of each ensemble member using the EM algorithm [*Dempster et al.*, 1977].

[49] 7. Compute the model weights by summing the weights for all ensemble members of each model.

[50] 8. Assess predictive mean and variance using equations (11) and (12).

[51] A case study on the applicability and robustness of IBUNE for reliable assessment of predictive uncertainty propagated through the system from all the important sources of uncertainty is provided in the next section.

5.4. Use of IBUNE: Uncertainty Assessment of Hydrologic Model Parameters and Input Error Model Parameters and Model Structure

[52] The IBUNE scheme promises better assessment of total uncertainty because it accounts for model parameters, input, and model structural uncertainty. In this section we will present the results for IBUNE and compare it with all
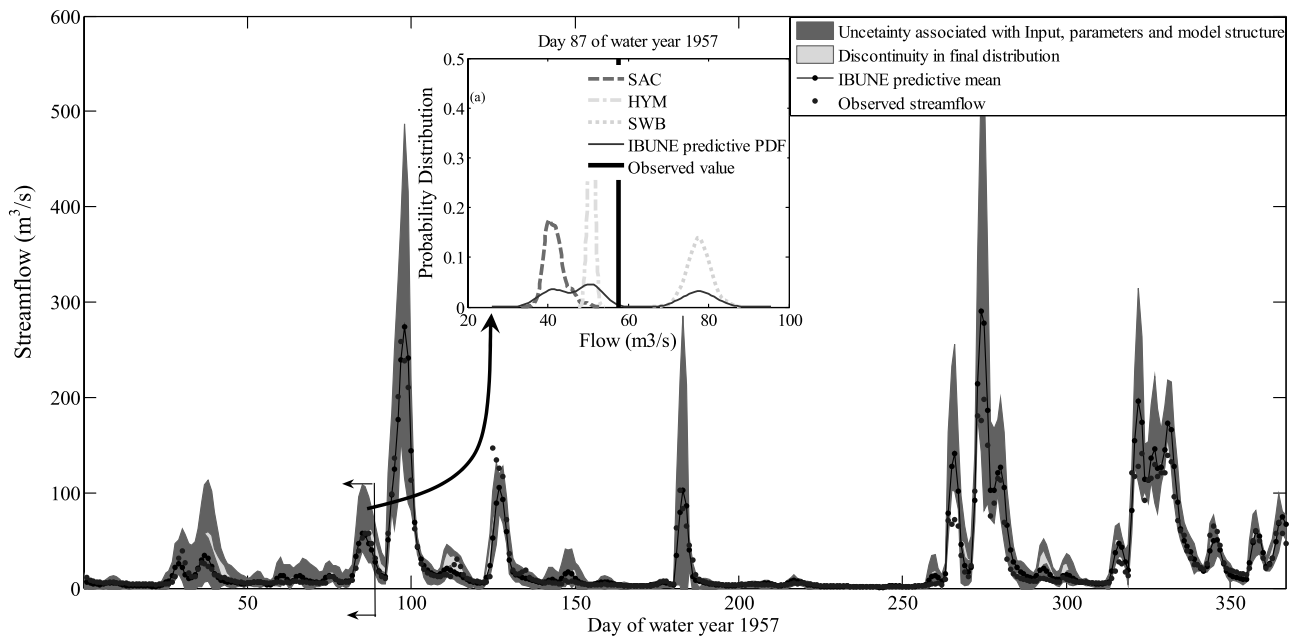
Figure 15. Streamflow hydrograph prediction uncertainty associated with estimated parameters and input error model parameters as well as model structural uncertainty (in shaded gray) for water year 1957 (calibration period). The lighter patches in the uncertainty bounds represent the discontinuity of the final model distributions. (inset) Profile of the selected cross section which includes the final distribution of each member model and the final IBUNE predictive probability density function.

other scenarios. Figure 14 illustrates the estimated uncertainty bound using SCEM, associated with input and model parameters for the three above mentioned models for the water year 1957. The solid dots in Figure 15 represent observed streamflow. Notice that different models include different observation values from various parts of the observed hydrograph, which can be interpreted as skill of the model to capture various processes within the watershed. On the basis of step 5 of IBUNE (presented in the previous section), the posterior probability distribution of each model in capturing observations (i.e., the weight) for each model was estimated. The weights are presented in Figure 14. As expected, the model with the higher skill (SAC-SMA) was assigned the highest weight, while the model with the lowest skill (SWB) was assigned the lowest weight. Both HYMOD and SWB gain very small weights. However, their contribution to the final results is considerable because they represent variety of the watershed processes which were not well represented in the SAC-SMA.

[53] The final IBUNE predictive probability which was estimated based on the probability of contributing model in the combination is given in Figure 15. The width of this final probability can be calculated through equation (12); however, the shape and intensity of the distribution can be captured through summation of the posterior probability distribution of contributing models in the combination (Figure 15a). The connected dots depict the IBUNE predictive mean which was estimated through equation (11) using the estimated weights and model simulations at each point. Another interesting observation from Figure 15 is that in some parts of the hydrograph, the final posterior probability of the three contributing model does not meet and therefore causes discontinuity in the final posterior probability distribution at these parts of the hydrograph (Figure 15). These
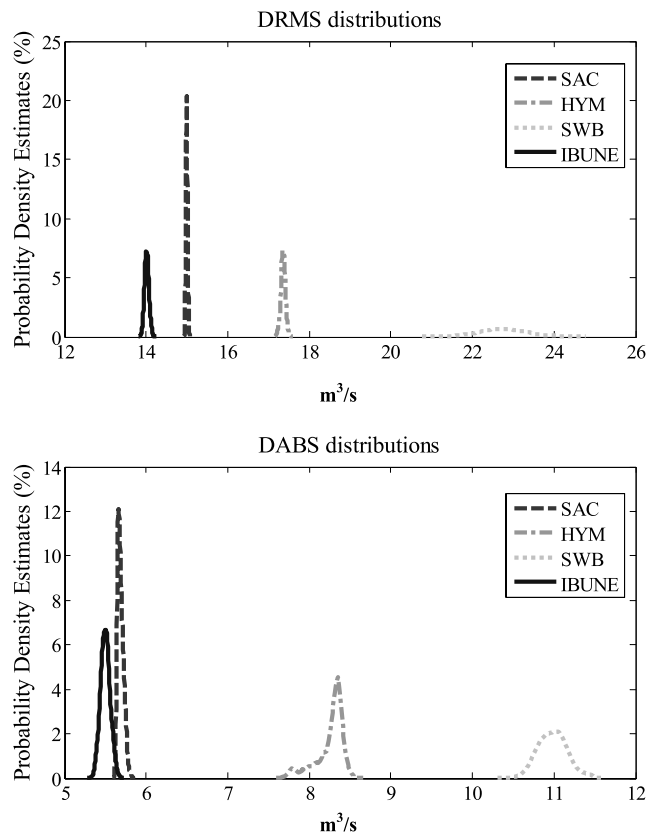
Figure 16. Distribution of DRMS and DABS for individual models and IBUNE.
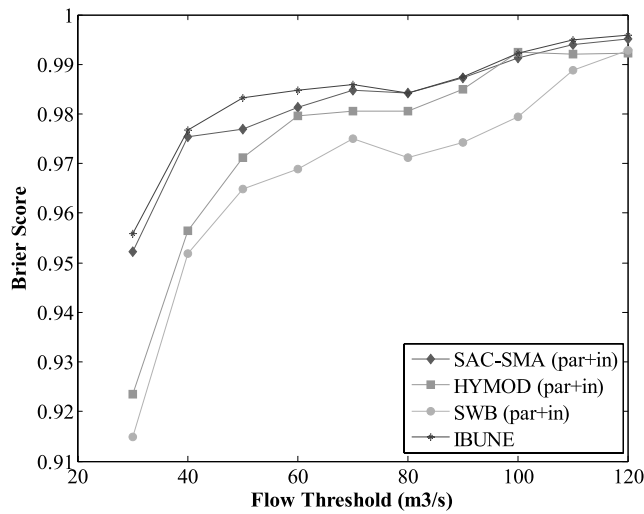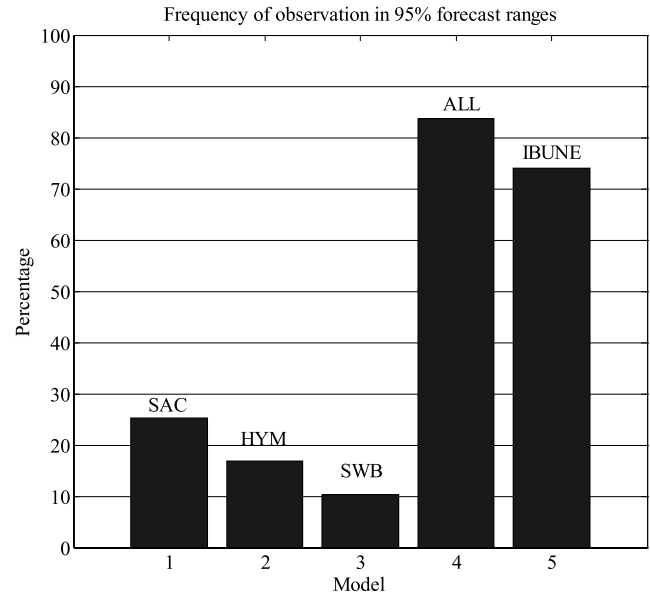
Figure 17. Brier score for IBUNE and three member models.

Figure 18. Percentage of observation in 95% uncertainty bounds of different models and IBUNE.

discontinuity areas are presented by the light gray color in Figure 15. Also shown in Figure 15a is the profile of a cross section in the hydrograph for clarification. Notice that the posterior probability distribution at this time step is discontinuous with three distinct modes. This is a clear indication that these three models do not represent the model space well and that more models are needed to avoid this problem. This also suggests that just looking at the uncertainty bounds as set of percentiles can be misleading in some cases. Figure 16 shows the distribution of daily root-mean-square error as well as the daily absolute error for all three contributing models and simulations generated through IBUNE. These distributions were estimated based on the ensemble of simulations generated by each model through their input and model parameter distributions. Figure 16 illustrates that IBUNE improved DRMS more than DABS, and these results indicate that IBUNE improved simulation of the high-flow values more than the low-flow values.

[54] The Brier score (BS) was also used to compare the skill of the individual model ensembles (considering both parameter and input uncertainty) with IBUNE. The Brier score is a scalar measure of the quality of probabilistic forecast and has been commonly used in literature. BS is defined as follows [*Georgakakos et al.*, 2004]:

$$BS = 1 - \frac{1}{N} \sum_{t=1}^{N} (f(t) - o(t))^2, \qquad (15)$$

where $f(t)$ is frequency of target event at time step t estimated by the fraction of model ensemble simulations which are larger than prespecified threshold; $o(t)$ is equal to 1 if the observation at that time step is larger than threshold and equal to zero otherwise; and N is the number of time steps in the record. Here BS is a positively oriented score, and therefore in Figure 16 the higher the BS the better. Figure 17 shows the BS for all the models and IBUNE. The findings in Figure 16 that IBUNE produces superior predictions than individual member models are confirmed in Figure 17. One can see that IBUNE gained a higher score in most of the thresholds. Another observation from this

figure is that IBUNE outperformed other models over the low-flow periods as well as the high- flow periods. This suggests that IBUNE is a promising flood forecasting framework because it has higher skills in capturing higher flows.

[55] The percentage of observations which are bracketed by the estimated 95% uncertainty bounds is given in Figure 18. Ninety-five percent uncertainty bounds estimated through IBUNE cover 74% of the observation over the whole study period, which is significantly higher than any single model. This 74% excludes the points which are in the discontinuity sections with zero probability. However, only considering whatever point falls within the upper and lower uncertainty bounds at each time step will give us a percent convergence equal to 83% (Figure 18).

[56] To demonstrate usefulness of IBUNE as a streamflow prediction framework, we evaluated its performance using data from an independent 3-year validation period (1958–1960). Table 5 presents summary statistics of the validation results comparing all three scenarios, including SCEM, Extended SCEM, and IBUNE. The results in Table 5 indicate that IBUNE consistently provides better values of

Table 5. Summary of Statistics of the Streamflow Prediction During the Validation Period (Water Year 1958–1960) for All Three Scenarios Presented

Statistic	SCEM (par)			Extended SCEM (par + input)			
	SAC	HYM	SWB	SAC	HYM	SWB	IBUNE
DRMS	16.68	15.48	24.05	16.09	13.23	22.27	12.59
Percent bias	−12.38	−7.83	18.89	−0.45	−3.67	10.64	−2.6
Correlation	0.89	0.89	0.86	0.90	0.92	0.87	0.94
Percent of Observations[a]	11.5	17.5	7.6	22	20.5	14.5	76.3

[a]Percentage of observations being in 95% uncertainty bounds.

the DRMS error, bias (percent bias), correlation, and percent of observations fall within 95% uncertainty bounds (percent of observations) statistics than both SCEM and Extended SCEM for all three models. It is also interesting to point out that as we account for input uncertainty along with parameter uncertainty (extended SCEM) in the hydrological models, the statistics tend to improve compared with the conventional SCEM approach, which just accounts for parameter uncertainty. Therefore explicitly accounting for input, parameter, and model structural uncertainty during calibration period can lead to improved assessment of predictive uncertainty as well as model forecasts.

6.   Summary and Conclusions

[57] The prevailing approach in hydrological modeling and the assessment of related uncertainty has been the use of sophisticated calibration techniques to estimate an optimal set of parameters for a single model. Through these processes, all other sources of uncertainty, including input and model structural uncertainty, are generally ignored, and the uncertainty in the model estimation of the system is primarily assigned to the uncertainty in model parameters. Nevertheless, we know that a single-model structure is incapable of representing all of the hydrological processes within a watershed and all of the system observations including input forcing contain measurement error. Consequently, these assumptions lead to an incorrect estimation of total uncertainty in the model predictions.

[58] The objectives of this paper were threefold: (1) to demonstrate that the classic uncertainty assessment approach in hydrology which relays all of the uncertainty within the system on the parameter estimation is not reliable and accurate, (2) to introduce a new approach to simultaneously address model parameter estimation and input forcing uncertainty, and (3) to propose a new framework that tackles three major sources of uncertainty, including uncertainty inherited in input forcings, parameter estimation, and model structure. The conclusion of this work can be summarized as follows:

[59] 1. The underlying approach for uncertainty assessment in hydrological modeling has been to treat the model and observation data unbiased and precise and treat the uncertainty in the modeling processes as being explicitly attributed to the uncertainty in the parameter estimates. In this study we verified that such an assumption will lead to biased and corrupted parameter estimates. Hence the result is unrealistic model simulations and their associated uncertainty bounds which does not consistently capture and represent the real-world behavior of the watershed. This was demonstrated through two separate case studies using Shuffled Complex Evolution Metropolis (SCEM) [*Vrugt et al.*, 2003], the newly developed probabilistic parameter estimation algorithm, to calibrate three selected hydrologic models for the Leaf River basin in Mississippi. The understudy models included the Sacramento soil moisture accounting (SAC-SMA) model, the soil water balance (SWB) model, and the hydrologic model (HYMOD).

[60] 2. In the second attempt to estimate more accurate and less corrupt uncertainty bounds for the hydrologic model simulation, we proposed a new approach to account for associated uncertainty in the input forcings. We simply introduced an input error model which assumed a random Gaussian error as a multiplier for every input observation. The common ground for these multipliers is that they are all from an identical distribution with unknown first two moments (mean and variance). Therefore we extended SCEM to estimate these two new unknown parameters with the hydrologic model parameters and their associated uncertainty. We demonstrated that undertaking such a simple approach to address input uncertainty improved the accuracy and reliability of the hydrologic simulations and their associated uncertainty bounds significantly.

[61] 3. Although accounting for the input uncertainty generated more reliable results, these results were still suffering from a very common limitation in hydrologic modeling attitudes that the model understudy is the best model in hand. However, the most sophisticated models are still simple representations of the real world and cannot capture all of the processes with the catchments. In order to take into account this source of uncertainty, we exploit the newly developed technique, called Bayesian model averaging (BMA) [*Hoeting et al.*, 1999]. BMA disregards the traditional belief in hydrological modeling and explores multiple model structures to represent the processes within the system. We merged this method (BMA) with the extended SCEM presented in this paper which accounts for both input and parameter uncertainty and proposed a new hybrid framework entitled, Integrated Bayesian Uncertainty Estimator (IBUNE). IBUNE combines and exploits the strengths of the SCEM as an efficient and effective probabilistic model parameter estimator algorithm and the introduced input error model, as well as Bayesian model combination techniques, to provide an integrated assessment of uncertainty propagating through the system from parameter estimation, input forcing, and model structure.

[62] 4. The usefulness and applicability of IBUNE has also been demonstrated via a validation study over a 3-year-period from 1958 to 1960. The results confirmed that both extended SCME and IBUNE framework are convincing tools to improve the accuracy and reliability of the predictions and their associated uncertainty bounds even during validation period. The results presented here were obtained through simulation experiment; however, it would be very interesting to test the performance of this framework through a set of forecast experiments which uses forecasted inputs such as precipitation to force the hydrologic models.

[63] To demonstrate the usefulness and applicability of IBUNE, we used the same hydrologic models considered earlier. The strength of these three models was combined through IBUNE. We showed that IBUNE is a very useful and applicable technique which accounts for all of the different sources of uncertainty within the hydrologic system and results in improved model prediction uncertainty bounds that bracket higher percentage of system observations.

[64] IBUNE is a flexible framework which can be expanded by including many more hydrologic models. All three major components of the framework, SCEM, input error model, and BMA, investigate different limitations in hydrologic modeling processes and provide more precise estimation of uncertainty bounds by confronting all of these different sources of uncertainty.

[65] Although accounting for all sources of uncertainty is very important in forecasting future devastating events, all

of the at-hand techniques including the work presented here are still too expensive to be used for real-time operational application. However, the ever increasing pace of computational power will soon provide the opportunity for operational communities to take advantage of these state-of-the-art methods to address uncertainty associated with their forecasts in a more reliable and accurate manner.

References

Abrahart, R. J., and L. See (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.

Ajami, N. K., Q. Duan, H. Moradkhani, and S. Sorooshian (2005), Recursive Bayesian model combination for streamflow forecasting, paper presented at the American Meteorological Society meeting, San Diego, Calif., Jan.

Ajami, N. K., Q. Duan, X. Gao, and S. Sorooshian (2006), Multi-model combination techniques for hydrological forecasting: Application to Distributed Model Intercomparison Project results, *J. Hydrometeorol.*, *7*(4), 755–768.

Beck, M. B. (1987), Water quality modeling: Review of the analysis uncertainty, *Water Resour. Res.*, *23*, 1393–1442.

Beven, K., and A. Binley (1992), The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Processes*, *6*, 279–298.

Box, G. E. P., and G. C. Tiao (1973), *Bayesian Inference in Statistical Analysis*, Addison-Wesley, Boston, Mass.

Boyle, D. (2001), Multicriteria calibration of hydrological models, Ph.D. dissertation, Univ. of Ariz., Tucson.

Brazil, L. E. (1988), Multilevel calibration strategy for complex hydrologic simulation models, Ph.D. dissertation, 217 pp., Dep. of Civ. Eng., Colo. State Univ., Fort Collins.

Burnash, R. J., R. L. Ferral, and R. A. McGuire (1973), A generalized streamflow simulation system conceptual modeling for digital computers, report, U. S. Dep. of Commer. Natl. Weather Serv. and State of Calif. Dep. of Water Resour.

Clyde, M. A. (1999), Bayesian model averaging and model search strategies, in *Bayesian Statistics*, vol. 6, edited by J. M. Benardo et al., pp. 157–185, Oxford Univ. Press, New York.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977), Maximum likelihood from incomplete data via the EM algorithm, *J. R. Stat. Soc., Ser. B*, *39*, 1–39.

Duan, Q., S. Sorooshian, and V. K. Gupta (1992), Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resour. Res.*, *28*(4), 265–284.

Ellison, A. M. (2004), Bayesian inference in ecology, *Ecol. Lett.*, *7*, 509–520.

Evenson, G. (1992), Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model, *J. Geophys. Res.*, *97*, 17,905–17,924.

Fernandez, C., E. Ley, and M. Steel (2001), Benchmark priors for Bayesian model averaging, *J. Econometrics*, *100*, 381–427.

Gelman, A., and D. B. Rubin (1992), Inference from iterative simulation using multiple sequences, *Stat. Sci.*, *7*, 457–472.

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

George, E. I., and R. E. McCulloch (1993), Variable selection via Gibbs sampling, *J. Am. Stat. Assoc.*, *88*, 881–889.

Gupta, V. H., S. Sorooshian, and P. O. Yapo (1998), Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, *Water Resour. Res.*, *34*(4), 751–763.

Hastings, W. K. (1970), Monte-Carlo sampling methods using Markov Chains and their applications, *Biometrika*, *57*, 97–109.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999), Bayesian model averaging: A tutorial, *Stat. Sci.*, *14*(4), 382–417.

Hogue, T. S., H. V. Gupta, S. Sorooshian, and C. D. Tomkins (2003), A multi-step automatic calibration scheme for watershed models, in *Calibration of Watershed Models*, *Water Sci. Appl. Ser.*, vol. 6, edited by Q. Duan et al., pp. 165–174, AGU, Washington, D. C.

Kavetski, D., S. W. Franks, and G. Kuczera (2003), Confronting input uncertainty in environmental modeling, in *Calibration of Watershed Models*, *Water Sci. Appl. Ser.*, vol. 6, edited by Q. Duan et al., pp. 49–68, AGU, Washington, D. C.

Kitanidis, P. K., and R. L. Bras (1980a), Adaptive filtering through detection of isolated transient errors in rainfall-runoff models, *Water Resour. Res.*, *16*, 740–748.

Kitanidis, P. K., and R. L. Bras (1980b), Real-time forecasting with a conceptual hydrological model: 1. Analysis of uncertainty, *Water Resour. Res.*, *16*, 1025–1033.

Kuczera, G., and E. Parent (1998), Monte Carlo assessment of parameter uncertainty in conceptual catchment models: The metropolis algorithm, *J. Hydrol.*, *211*, 69–85.

Madigan, D., A. E. Raftery, C. Volinsky, and J. Hoeting (1996), Bayesian model averaging, in *Proceedings of the AAAI Workshop on Integrating Multiple Learned Models*, pp. 77–83, AAAI Press, Portland, Oreg.

Marshall, L., A. Sharma, and D. J. Nott (2003), Efficient Bayesian model selection in hydrological modeling, in *Proceedings of the International Congress on Modelling and Simulation (MODSIM2003)*, edited by D. A. Post, pp. 831–836, Modell. and Simul. Soc. of Australia and N. Z., Townsville, Australia. Australia.

McLachlan, G. J., and T. Krishnan (1997), *The EM Algorithm and Extensions*, John Wiley, Hoboken, N. J.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953), Equation of state calculations by fast computing machines, *J. Chem. Phys.*, *21*, 1087–1091.

Miller, R. N., M. Ghil, and F. Ghautiez (1994), Advanced data assimilation in a strongly nonlinear dynamical system, *J. Atmos. Sci.*, *51*, 1037–1055.

Montanari, A., and A. Brath (2004), A stochastic approach for assessing the uncertainty of rainfall-runoff simulations, *Water Resour. Res.*, *40*, W01106, doi:10.1029/2003WR002540.

Moradkhani, H., K. L. Hsu, H. V. Gupta, and S. Sorooshian (2005), Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter, *Water Resour. Res.*, *41*, W05012, doi:10.1029/2004WR003604.

Neuman, S. P. (2003), Maximum likelihood Bayesian averaging of uncertain model predictions, *Stochast. Environ. Res. Risk Assess.*, *17*, 291–305.

Neuman, S. P., and P. J. Wierenga (2003), A comprehensive strategy of hydrologic modeling and uncertainty analysis for nuclear facilities and sites, *NUREG/CR-6805*, U. S. Nucl. Regul. Comm., Washington, D. C.

Raftery, A. E., and Y. Zheng (2003), Discussion: Performance of Bayesian model averaging, *J. Am. Stat. Assoc.*, *98*(464), 931–938.

Raftery, A. E., D. Madigan, and C. T. Volinsky (1997), Model selection and accounting for model uncertainty in linear regression models, *J. Am. Stat. Assoc.*, *92*, 179–191.

Raftery, A. E., F. Balabdaoui, T. Gneiting, and M. Polakowski (2003), Using Bayesian model averaging to calibrate forecast ensembles, *Tech. Rep. 440*, Dep. of Stat., Univ. of Wash., Seattle.

Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski (2005), Using Bayesian model averaging to calibrate forecast ensembles, *Mon. Weather Rev.*, *133*, 1155–1174.

Schaake, J. C., V. I. Koren, Q. Y. Duan, K. Mitchell, and F. Chen (1996), Simple water balance model for estimating runoff at different spatial and temporal scales, *J. Geophys. Res.*, *101*(D3), 7461–7475.

Shamseldin, A. Y., K. M. O'Connor, and G. C. Liang (1997), Methods for combining the outputs of different rainfall-runoff models, *J. Hydrol.*, *197*, 203–229.

Sorooshian, S., and J. A. Dracup (1980), Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: Correlated and heteroscedastic error cases, *Water Resour. Res.*, *16*(2), 430–442.

Thiemann, M., M. Trosset, H. Gupta, and S. Sorooshian (2001), Bayesian recursive parameter estimation for hydrologic models, *Water Resour. Res.*, *37*(10), 2521–2535.

Viallefont, V., A. E. Raftery, and S. Richardson (2001), Variable selection and Bayesian model averaging in epidemiological case-control studies, *Stat. Med.*, *20*, 3215–3230.

Vrugt, J. A., H. V. Gupta, W. Bouten, and S. Sorooshian (2003), A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, *Water Resour. Res.*, *39*(8), 1201, doi:10.1029/2002WR001642.

Vrugt, J. A., C. G. H. Diks, H. V. Gupta, W. Bouten, and J. M. Verstraten (2005), Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation, *Water Resour. Res.*, *41*, W01017, doi:10.1029/2004WR003059.

Wagener, T., D. P. Boyle, M. J. Lees, H. S. Wheater, H. V. Gupta, and S. Sorooshian (2001), A framework for development and application of hydrological models, *Hydrol. Earth Syst. Sci.*, *5*(1), 13–26.

Wintle, B. A., M. A. McCarthy, C. T. Volinsky, and R. P. Kavanagh (2003), The use of Bayesian model averaging to better represent uncertainty in ecological models, *Conserv. Biol.*, *17*(6), 1579–1590.

Yapo, P. O., H. V. Gupta, and S. Sorooshian (1998), Multi-objective global optimization for hydrologic models, *J. Hydrol.*, *204*, 83–97.

————————————

N. K. Ajami, University of California at Berkeley, Department of Civil and Environmental Eng., University of California, Berkeley, 626 Davis Hall, Berkeley, CA 94720, USA. (newshaajami@berkeley.edu)

Q. Duan, Lawrence Livermore National Laboratory, P.O. Box 808, L0631, 7000 East Avenue, Livermore, CA 94551, USA.

S. Sorooshian, Center for Hydrometeorology and Remote Sensing, Department of Civil and Environmental Engineering, University of California, Irvine, E-4130 Engineering Gateway, Irvine, CA 92679, USA.