



Research papers

Towards robust quantification and reduction of uncertainty in hydrologic predictions: Integration of particle Markov chain Monte Carlo and factorial polynomial chaos expansion

S. Wang^{a,*}, G.H. Huang^b, B.W. Baetz^c, B.C. Ancell^a^a Department of Geosciences, Texas Tech University, Lubbock, TX, USA^b Institute for Energy, Environment and Sustainable Communities, University of Regina, Regina, Saskatchewan, Canada^c Department of Civil Engineering, McMaster University, Hamilton, Ontario, Canada

ARTICLE INFO

Article history:

Received 13 December 2016

Received in revised form 28 February 2017

Accepted 13 March 2017

Available online 15 March 2017

This manuscript was handled by Corrado Corradini, Editor-in-Chief, with the assistance of Renato Morbidelli, Associate Editor

Keywords:

Hydrologic prediction

Uncertainty

Data assimilation

Particle filter

Markov chain Monte Carlo

Polynomial chaos expansion

ABSTRACT

The particle filtering techniques have been receiving increasing attention from the hydrologic community due to its ability to properly estimate model parameters and states of nonlinear and non-Gaussian systems. To facilitate a robust quantification of uncertainty in hydrologic predictions, it is necessary to explicitly examine the forward propagation and evolution of parameter uncertainties and their interactions that affect the predictive performance. This paper presents a unified probabilistic framework that merges the strengths of particle Markov chain Monte Carlo (PMCMC) and factorial polynomial chaos expansion (FPCE) algorithms to robustly quantify and reduce uncertainties in hydrologic predictions. A Gaussian anamorphosis technique is used to establish a seamless bridge between the data assimilation using the PMCMC and the uncertainty propagation using the FPCE through a straightforward transformation of posterior distributions of model parameters. The unified probabilistic framework is applied to the Xiangxi River watershed of the Three Gorges Reservoir (TGR) region in China to demonstrate its validity and applicability. Results reveal that the degree of spatial variability of soil moisture capacity is the most identifiable model parameter with the fastest convergence through the streamflow assimilation process. The potential interaction between the spatial variability in soil moisture conditions and the maximum soil moisture capacity has the most significant effect on the performance of streamflow predictions. In addition, parameter sensitivities and interactions vary in magnitude and direction over time due to temporal and spatial dynamics of hydrologic processes.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Hydrologic systems involve dynamic interactions between water, climate, vegetation, and soil processes. Hydrologic models typically conceptualize the complex behavior of hydrologic systems by using mathematical equations, in which model parameters represent temporal and spatial variability in watershed characteristics that cannot be observed or measured explicitly. As a result, the predictive performance of hydrologic models is inevitably affected by various sources of uncertainty, including the descriptions of boundary and initial conditions, the errors in model forcing data and output observations, inaccurate parameter estimates, and model structural deficiencies (Ajami et al., 2007; DeChant and Moradkhani, 2014). Therefore, robust quantification

and reduction of uncertainties are necessary to improve the reliability of hydrologic predictions.

Over the past few decades, tremendous efforts have been made in the development of data assimilation techniques for addressing various sources of uncertainty in hydrologic modeling (Liu and Gupta, 2007; Ryu et al., 2009; Gharamti et al., 2013; Abaza et al., 2014; Panzeri et al., 2014; Randrianasolo et al., 2014; Sun et al., 2015; Xu and Gómez-Hernández, 2015). Data assimilation techniques are recognized as a powerful tool for probabilistic hydrologic predictions through recursively updating model states and parameters when new observations become available. Previously, the ensemble Kalman filter (EnKF) introduced by Evensen (1994) was the most commonly used technique for uncertainty assessment of hydrologic model parameters and state variables due to its attractive features of real-time adjustment and efficient implementation (Reichle et al., 2002; Moradkhani et al., 2005b; Xie and Zhang, 2010; Cammalleri and Ciraolo, 2012; Rafieeiniasab et al.,

* Corresponding author.

E-mail address: shuo.wang@ttu.edu (S. Wang).

2014; Samuel et al., 2014; Gharamti et al., 2015; Liu et al., 2016; Pathiraja et al., 2016). However, the EnKF relies on a Gaussian assumption that cannot be met in practical applications, resulting in unrealistic model simulations and unreliable hydrologic predictions.

In recent years, the particle filter (PF) technique has been introduced as an attractive alternative to remove the unrealistic Gaussian assumption of the EnKF, improving the reliability of hydrologic predictions. The PF is able to fully represent the posterior distributions of model parameters and state variables through a number of independent random samples called particles, and the particles are weighted and propagated sequentially by assimilating available observations. Over the past few years, the PF and its variants have been receiving increasing attention from the hydrologic community due to its ability to properly estimate the state of non-linear and non-Gaussian systems (Weerts and El Serafy, 2006; Smith et al., 2008; Salamon and Feyen, 2009; DeChant and Moradkhani, 2012; Dumedah and Coulibaly, 2013; Bi et al., 2015). Moradkhani et al. (2005a) introduced a PF with sampling importance resampling (PF-SIR) algorithm for uncertainty assessment of hydrologic model states and parameters. To further improve the PF-SIR algorithm, Moradkhani et al. (2012) proposed a PF with a Markov chain Monte Carlo (PF-MCMC) algorithm to increase parameter diversity within the posterior distribution, reducing the risk of sample impoverishment and leading to a more accurate streamflow forecast. Plaza Guingla et al. (2013) proposed two alternatives that included a resample-move step in the standard PF and an optimal importance density function in the Gaussian PF to improve the effectiveness of the PF. Vrugt et al. (2013) combined the strengths of sequential Monte Carlo sampling and the MCMC simulation with the Differential Evolution Adaptive Metropolis (DREAM) algorithm for the joint estimation of hydrologic model parameters and state variables. Yan et al. (2015) used the PF-SIR and PF-MCMC algorithms to estimate soil moisture states and soil hydraulic parameters through assimilating remotely sensed near-surface soil moisture measurements.

The PF has been introduced as a powerful tool to reduce uncertainty in the joint estimation of hydrologic model parameters and state variables through sequential assimilation of observations. When the posterior distributions of model parameters are derived from the PF, uncertainty is propagated through the hydrologic models, leading to probabilistic hydrologic predictions. In addition, model parameters are correlated with each other in the uncertainty propagation process, and their interactions have significant impacts on the predictive performance. Therefore, it is necessary to reveal the forward propagation and evolution of parameter uncertainties and their interactions after the data assimilation operation, leading to a robust quantification and reduction of uncertainty in hydrologic predictions.

Polynomial chaos expansion (PCE) techniques have been extensively used to represent stochastic processes through the propagation of random uncertainties in dynamic systems (Xiu and Karniadakis, 2002; Marzouk and Najm, 2009; Najm, 2009; Konda et al., 2010; Oladyshkin et al., 2011). In recent years, the PCE has attracted increased attention in hydrologic studies (Lin and Tartakovsky, 2009; Müller et al., 2011; Ciriello et al., 2013; Sochala and Le Maître, 2013; Dai et al., 2016). Fajraoui et al. (2011) used global sensitivity analysis in conjunction with the PCE methodology to provide valuable information for the interpretation of transport experiments in laboratory-scale heterogeneous porous media. Rajabi et al. (2015) used the non-intrusive PCE for efficient uncertainty propagation and moment independent sensitivity analysis of seawater intrusion simulations. Wang et al. (2015b) developed a polynomial chaos ensemble hydrologic prediction system for efficiently quantifying uncertainties in hydrologic predictions. Consequently, the PCE is recognized as a

promising technique for explicitly revealing the propagation of uncertainty through hydrologic models and for efficiently quantifying uncertainty in hydrologic predictions.

In this work, we propose a unified probabilistic framework to robustly quantify and reduce uncertainties in hydrologic predictions. In the unified probabilistic framework, a particle Markov chain Monte Carlo (PMCMC) algorithm will be first used to infer the posterior distributions of hydrologic model parameters through sequential assimilation of available observations. After the data assimilation operation, a factorial polynomial chaos expansion (FPCE) technique will then be introduced to examine the forward propagation and evolution of uncertainty by revealing complex parameter interactions and their impacts on the predictive performance of hydrologic models. In addition, a Gaussian anamorphosis technique will be used to establish a seamless bridge between the data assimilation using the PMCMC and the uncertainty propagation using the FPCE through a straightforward transformation of posterior distributions of model parameters. The proposed unified probabilistic framework will be applied to predict daily streamflow in the Xiangxi River watershed which is located in the Three Gorges Reservoir (TGR) region, China.

This paper is organized as follows. Section 2 introduces the proposed unified probabilistic framework for robustly quantifying uncertainties in hydrologic predictions. Section 3 provides details on the study area and the experimental setup. Section 4 presents a systematic analysis of sequential streamflow assimilation and uncertainty quantification along with a thorough discussion on the recursive inference of model parameters and the explicit characterization of parameter interactions. Finally, conclusions are drawn in Section 5.

2. Development of a unified probabilistic framework

The unified probabilistic framework merges the strengths of the PMCMC and the FPCE algorithms for robustly quantifying and reducing uncertainties in hydrologic predictions. A general overview of the steps involved within the unified probabilistic framework is provided as follows: (1) inference of model parameters and state variables through sequential data assimilation using the PMCMC; (2) transformation of irregular posterior distributions of model parameters into standard normal distributions using the Gaussian anamorphosis technique; (3) characterization of uncertainty propagation and evolution using the FPCE technique; (4) quantification of uncertainties in hydrologic predictions through the propagation of parameter uncertainties; and (5) examination of sensitivities of hydrologic model parameters and their interactions that affect the predictive performance using factorial analysis.

2.1. Particle Markov chain Monte Carlo

The PF technique is recognized as an effective means to remove the unrealistic assumption of Gaussian errors involved in hydrologic modeling, improving the robustness of hydrologic predictions. The PF is a sequential Monte Carlo approach that can be used to implement a recursive Bayesian filter through Monte Carlo simulations, by which the posterior distributions of state variables are represented by a set of random particles with corresponding importance weights (Bi et al., 2015). Thus, the most important concept in the particle filtering is the sequential importance sampling (SIS) used for estimating the particle weights in a recursive form.

To better understand the SIS algorithm, the hydrologic model can be formulated as follows:

$$x_{i,t+1}^- = f(x_{i,t}^+, u_{i,t+1}^-, \theta_{i,t+1}^-) + \omega_{i,t+1}, \omega_{i,t+1} \sim N(0, \text{Var}_{t+1}) \quad (1)$$

where i and t denote the ensemble number and the time step, respectively; $x_{i,t}^+$ and $x_{i,t-1}^-$ represent the posterior model states at the previous time step and the predicted model states at the current time step, respectively; f represents the forward model that propagates the system states from time t to $t + 1$ in response to model inputs $u_{i,t+1}$ and parameters $\theta_{i,t+1}$; $\omega_{i,t+1}$ represent normally distributed random noise with zero mean and variance Var_{t+1} . As the model outputs are characterized as a function of state variables and parameters, model predictions can be made by:

$$y_{i,t+1}^* = h(x_{i,t+1}^-, \theta_{i,t+1}^-) + v_{i,t+1}, \quad v_{i,t+1} \sim N(0, Q_{t+1}) \quad (2)$$

where $y_{i,t+1}^*$ is the prediction for ensemble member i ; h is the operator that relates state variables and parameters to observations (streamflow) and yields the expected value of the prediction given model states and parameters; $v_{i,t+1}$ represents the prediction error with zero mean and variance Q_{t+1} .

The SIS algorithm begins with Monte Carlo sampling to generate random particles from a proposal density (Moradkhani et al., 2005a). When an observation becomes available, the filtering posterior density can then be approximated by:

$$p(x_{t+1}, \theta_{t+1} | y_{t+1}) \approx \sum_{i=1}^N w_{i,t+1}^+ \delta(x_{t+1} - x_{i,t+1}^-, \theta_{t+1} - \theta_{i,t+1}^-) \quad (3)$$

where N represents the ensemble size; $w_{i,t+1}^+$ is the posterior weight for ensemble member i at time $t + 1$; $\delta(\cdot)$ denotes the Dirac delta function. The posterior weight is estimated based on the normalized likelihood function that can be calculated by:

$$p(y_{i,t+1} | x_{i,t+1}^-, \theta_{i,t+1}^-) = \frac{L(y_{i,t+1} | x_{i,t+1}^-, \theta_{i,t+1}^-)}{\sum_{i=1}^N L(y_{i,t+1} | x_{i,t+1}^-, \theta_{i,t+1}^-)} = p(y_{i,t+1} - y_{i,t+1}^* | R_k) \quad (4)$$

where $y_{i,t+1}$ is the perturbed observation with an assumed error at time $t + 1$ (Moradkhani et al., 2012). When the normalized likelihood is calculated, the posterior weight for each ensemble member can be obtained by:

$$w_{i,t+1}^+ = \frac{w_{i,t+1}^- \cdot p(y_{i,t+1} | x_{i,t+1}^-, \theta_{i,t+1}^-)}{\sum_{i=1}^N w_{i,t+1}^- \cdot p(y_{i,t+1} | x_{i,t+1}^-, \theta_{i,t+1}^-)} \quad (5)$$

where $w_{i,t+1}^-$ is the prior weight at time $t + 1$ which is equal to the posterior weight at time t . A weighted sample of model realizations can thus be generated, leading to an estimation of the posterior distribution (DeChant and Moradkhani, 2012).

The weights of particles are updated according to the normalized likelihood function in a recursive manner. However, most particles will carry negligible weights after a few iterations, resulting in severe particle degeneracy. The occurrence of particle degeneracy can be examined by calculating the effective sample size:

$$N_{\text{eff}} = \left(\sum_{i=1}^N (w_{i,t+1}^+)^2 \right)^{-1} \quad (6)$$

To address the problem of particle degeneracy, the sampling importance resampling (SIR) technique can be performed when the effective sample size N_{eff} drops below a pre-specified threshold which has been set to $N_{\text{eff}}/2$ in this study (Vrugt et al., 2013). After resampling, the MCMC algorithm can then be used to maintain particle diversity and quality. The MCMC is able to maximize the search of the parameter space by creating a proposal parameter distribution, and then the metropolis acceptance ratio is used to avoid parameter samples moving outside the filtering posterior distribution so as to ensure that parameters do not diverge

(Moradkhani et al., 2012). The metropolis acceptance ratio is calculated by comparing the probabilities of proposed joint state parameters and resampled joint state parameters:

$$\alpha = \min \left(1, \frac{p(x_{i,t+1}^p, \theta_{i,t+1}^p | y_{1:t+1})}{p(x_{i,t+1}^+, \theta_{i,t+1}^+ | y_{1:t+1})} \right) \quad (7)$$

where the probability of the proposed joint state parameters, $p(x_{i,t+1}^p, \theta_{i,t+1}^p | y_{1:t+1})$, can be calculated as follows:

$$p(x_{i,t+1}^p, \theta_{i,t+1}^p | y_{1:t+1}) = p(y_{1:t+1} | x_{i,t+1}^p, \theta_{i,t+1}^p) p(x_{i,t+1}^p | \theta_{i,t+1}^p, y_{1:t}) p(\theta_{i,t+1}^p | y_{1:t}) \quad (8)$$

The probability of the resampled joint state parameters is then calculated similarly. The particle Markov chain Monte Carlo (PMCMC) with the SIR algorithm is able not only to ensure that the particles remain in the filtering posterior density through the acceptance/rejection step but also to promote sample diversity, reducing both particle degeneracy and impoverishment problems.

The PMCMC is capable of deriving posterior distributions of model parameters with reduced uncertainty through sequential assimilation of available observations. To further quantify uncertainties in hydrologic predictions, it is necessary to explore the forward propagation and evolution of uncertainty in model parameters after the data assimilation operation.

2.2. Factorial polynomial chaos expansion

To explicitly characterize the propagation of uncertainty through complex dynamic systems, the response variable can be represented by a nonlinear function with a set of random variables. The PCE is recognized as a powerful tool to express the evolution of uncertainty in random dynamic systems. The PCE was first introduced by Wiener (1938), which can be used as a functional approximation of the hydrologic model through a series expansion of orthogonal polynomials. When model parameters are normally distributed random variables, the probabilistic output can be represented by multivariate Hermite polynomials (Xiu and Karniadakis, 2002), and written as:

$$y = a_0 + \sum_{i_1=1}^{\infty} a_{i_1} \Gamma_1(\xi_{i_1}) + \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{i_1} a_{i_1 i_2} \Gamma_2(\xi_{i_1}, \xi_{i_2}) + \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{i_1} \sum_{i_3=1}^{i_2} a_{i_1 i_2 i_3} \Gamma_3(\xi_{i_1}, \xi_{i_2}, \xi_{i_3}) + \dots \quad (9)$$

where a_0 and $a_{i_1 i_2 \dots i_d}$ are the PCE coefficients to be estimated through model simulations, and $\Gamma_d(\xi_{i_1}, \dots, \xi_{i_d})$ are Hermite polynomials of order d in terms of standard normal random variables ξ . The general expression of Hermite polynomials is given by:

$$\Gamma_d(\xi_{i_1}, \dots, \xi_{i_d}) = (-1)^d e^{\frac{1}{2}\xi^T \xi} \frac{\partial^d}{\partial \xi_{i_1} \dots \partial \xi_{i_d}} e^{-\frac{1}{2}\xi^T \xi} \quad (10)$$

Eq. (9) can be rewritten in a straightforward way as:

$$y = \sum_{j=1}^P \hat{a}_j \Psi_j(\xi) \quad (11)$$

where there is an one-to-one mapping between functions $\Gamma_d(\xi_{i_1}, \dots, \xi_{i_d})$ and $\Psi_j(\xi)$ (Li and Zhang, 2007), and the total number of PCE terms P can be determined by the dimensionality N and the degree d of the PCE as:

$$P = \frac{(N + d)!}{N!d!} \quad (12)$$

The unknown coefficients in the PCE can be determined by the probabilistic collocation method that equates the model outputs and the estimates of the corresponding PCE at a set of selected collocation points. The selection of collocation points follows a standard procedure (Tatang et al., 1997), in which collocation points are selected from the roots of the Hermite polynomial of one degree higher than the order of the given PCE. For example, to solve for a two-dimensional second-order PCE given by $y = a_0 + a_1\xi_1 + a_2\xi_2 + a_3(\xi_1^2 - 1) + a_4(\xi_2^2 - 1) + a_5\xi_1\xi_2$, the roots of the third-order Hermite polynomial expressed as $H_3(\xi) = \xi^3 - 3\xi$ are $-\sqrt{3}, 0$, and $\sqrt{3}$. Thus, the collocation points are chosen from all combinations of the roots of the third-order Hermite polynomial, including $(-\sqrt{3}, -\sqrt{3})$, $(-\sqrt{3}, 0)$, $(-\sqrt{3}, \sqrt{3})$, $(0, -\sqrt{3})$, $(0, 0)$, $(0, \sqrt{3})$, $(\sqrt{3}, -\sqrt{3})$, $(\sqrt{3}, 0)$, and $(\sqrt{3}, \sqrt{3})$. The number of collocation points for an N -dimensional PCE of order d can be determined as $(d + 1)^N$.

To obtain a robust estimate of the unknown coefficients in the PCE, a multi-level factorial experimental design of collocation points was proposed by Wang et al. (2015c), leading to a factorial polynomial chaos expansion (FPCE). The FPCE can not only quantify uncertainties in model outputs through the propagation of parameter uncertainties in dynamic systems, but also reveal the statistical significance of model parameters and their interactions that affect the predictive performance.

Factorial design of experiments is a powerful tool for planning numerical experiments, so that appropriate data can be collected and then analyzed by statistical methods for drawing valid and objective conclusions (Montgomery and Runger, 2013). Factorial analysis can thus be used to examine the effects of multiple factor variables and their interactions on response variables by conducting hypothesis tests with the F -statistic (Wang and Huang, 2015; Wang et al., 2016). In a factorial experiment, all combinations of levels (e.g., collocation points) of factors are investigated through a factorial design. For example, if there are a levels of factor A , b levels of factor B , and c levels of factor C , there will be a total of abc observations in a complete factorial experiment. The effects model for such a factorial experiment can be expressed as:

$$y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijk} + \varepsilon_{ijkl} \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, b \\ k = 1, 2, \dots, c \\ l = 1, 2, \dots, n \end{cases} \quad (13)$$

where μ is the overall mean effect; τ_i is the effect of the i th level of factor A , β_j is the effect of the j th level of factor B , γ_k is the effect of the k th level of factor C , $(\tau\beta)_{ij}$ is the effect of the interaction between factors A and B , $(\tau\gamma)_{ik}$ is the effect of the interaction between factors A and C , $(\beta\gamma)_{jk}$ is the effect of the interaction between factors B and C , $(\tau\beta\gamma)_{ijk}$ is the effect of the interaction between factors A , B , and C , and ε_{ijkl} is the random error component. The effects model contains three main effects, three two-factor interactions, a three-factor interaction, and an error term. The effects are defined as deviations from the overall mean, so $\sum_{i=1}^a \tau_i = 0$, $\sum_{j=1}^b \beta_j = 0$, $\sum_{k=1}^c \gamma_k = 0$, $\sum_{i=1}^a (\tau\beta)_{ij} = \sum_{j=1}^b (\tau\beta)_{ij} = 0$, $\sum_{i=1}^a (\tau\gamma)_{ik} = \sum_{k=1}^c (\tau\gamma)_{ik} = 0$, $\sum_{j=1}^b (\beta\gamma)_{jk} = \sum_{k=1}^c (\beta\gamma)_{jk} = 0$, and $\sum_{i=1}^a (\tau\beta\gamma)_{ijk} = \sum_{j=1}^b (\tau\beta\gamma)_{ijk} = \sum_{k=1}^c (\tau\beta\gamma)_{ijk} = 0$ (Montgomery, 2000).

To test the statistical significance for design factors and their interactions, the F -statistic can be used as follows:

$$F_A = \frac{SS_A/a - 1}{SS_E/abc(n - 1)}; F_B = \frac{SS_B/b - 1}{SS_E/abc(n - 1)}; F_C = \frac{SS_C/c - 1}{SS_E/abc(n - 1)} \quad (14)$$

$$F_{AB} = \frac{SS_{AB}/(a - 1)(b - 1)}{SS_E/abc(n - 1)}; F_{AC} = \frac{SS_{AC}/(a - 1)(c - 1)}{SS_E/abc(n - 1)}; F_{BC} = \frac{SS_{BC}/(b - 1)(c - 1)}{SS_E/abc(n - 1)} \quad (15)$$

$$F_{ABC} = \frac{SS_{ABC}/(a - 1)(b - 1)(c - 1)}{SS_E/abc(n - 1)} \quad (16)$$

where SS_A , SS_B , SS_C , SS_{AB} , SS_{AC} , SS_{BC} , SS_{ABC} , and SS_E are the sum of squares for factors A , B , C , and the $A \times B$, $A \times C$, $B \times C$, $A \times B \times C$ interactions, as well as the error component, respectively; SS_T represents the total sum of squares. They can be calculated by:

$$SS_A = \frac{1}{bcn} \sum_{i=1}^a y_{i..}^2 - \frac{y_{...}^2}{abcn}; SS_B = \frac{1}{acn} \sum_{j=1}^b y_{.j.}^2 - \frac{y_{...}^2}{abcn}; SS_C = \frac{1}{abn} \sum_{k=1}^c y_{..k.}^2 - \frac{y_{...}^2}{abcn} \quad (17)$$

$$SS_{AB} = \frac{1}{cn} \sum_{i=1}^a \sum_{j=1}^b y_{ij.}^2 - \frac{y_{...}^2}{abcn} - SS_A - SS_B \quad (18)$$

$$SS_{AC} = \frac{1}{bn} \sum_{i=1}^a \sum_{k=1}^c y_{i.k.}^2 - \frac{y_{...}^2}{abcn} - SS_A - SS_C \quad (19)$$

$$SS_{BC} = \frac{1}{an} \sum_{j=1}^b \sum_{k=1}^c y_{.jk.}^2 - \frac{y_{...}^2}{abcn} - SS_B - SS_C \quad (20)$$

$$SS_{ABC} = \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c y_{ijk.}^2 - \frac{y_{...}^2}{abcn} - SS_A - SS_B - SS_C - SS_{AB} - SS_{AC} - SS_{BC} \quad (21)$$

$$SS_T = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^n y_{ijkl}^2 - \frac{y_{...}^2}{abcn} \quad (22)$$

$$SS_E = SS_T - SS_A - SS_B - SS_C - SS_{AB} - SS_{AC} - SS_{BC} - SS_{ABC} \quad (23)$$

where $y_{i..}$, $y_{.j.}$, $y_{..k.}$, $y_{ij.}$, $y_{i.k.}$, $y_{.jk.}$, and $y_{ijk.}$ denote the total of all observations under the i th level of factor A , the j th level of factor B , the k th level of factor C , the ij th interaction between factors A and B , the ik th interaction between factors A and C , the jk th interaction between factors B and C , and the ijk th interaction between factors A , B , and C , respectively; $y_{...}$ denotes the grand total of all observations. These statistics are useful for decomposing the total variance into its contributing components, which can explicitly reveal the statistical significance of design factors and their interactions that affect the model performance (Wu and Hamada, 2009; Shen et al., 2012; Wei et al., 2013; Wang et al., 2015a; Zeng et al., 2016).

2.3. Particle Markov chain Monte Carlo simulation coupled with factorial polynomial chaos expansion

The PMCMC algorithm recursively updates model states and parameters through sequential assimilation of observations, greatly reducing the uncertainty and improving the accuracy in hydrologic predictions. When the posterior distributions of model parameters are derived from the sequential data assimilation using the PMCMC, the predictive uncertainty can be further characterized through the forward propagation and evolution of parameter uncertainty. The FPCE can be used to reveal the statistical significance of model parameters and their interactions that influence the predictive performance in the uncertainty propagation process,

advancing our understanding of complex dynamics and chaos in hydrologic prediction systems. However, the posterior parameter densities are often derived with irregular distributions through the PMCMC, resulting in a potential challenge in implementing the further quantification of uncertainties in hydrologic predictions. As a result, data transformation techniques are needed to convert the irregular posterior distributions of model parameters into specific distributions (e.g., Gaussian, uniform, and gamma), establishing a seamless bridge between hydrologic data assimilation and forward uncertainty quantification.

The Gaussian anamorphosis technique is applied to transform the posterior distributions of model parameters to standard normal distributions after the data assimilation operation. The Gaussian anamorphosis is a nonlinear and monotonic transformation technique, which is able to link an arbitrarily distributed variable y and its Gaussian transform variable z through their cumulative distribution functions (CDFs):

$$z = G^{-1}(F(y)) \tag{24}$$

where $F(y)$ is the empirical CDF of y , and G is the theoretical standard normal CDF of z . As G is monotonously increasing, the inverse G^{-1} exists (Schöniger et al., 2012). Eq. (24) is called the Gaussian anamorphosis function.

Following Johnson and Wichern (1988), the empirical CDF of y can be obtained as follows:

$$F_j = \frac{j - 0.5}{N} \tag{25}$$

where j is the rank of the sample data, and N is the sample size. The sample data of the Gaussian transform variable z can then be obtained according to the empirical CDF F_j , and their data ranges can be determined as:

$$z_{\min} = G^{-1}((1 - 0.5)/N) \tag{26}$$

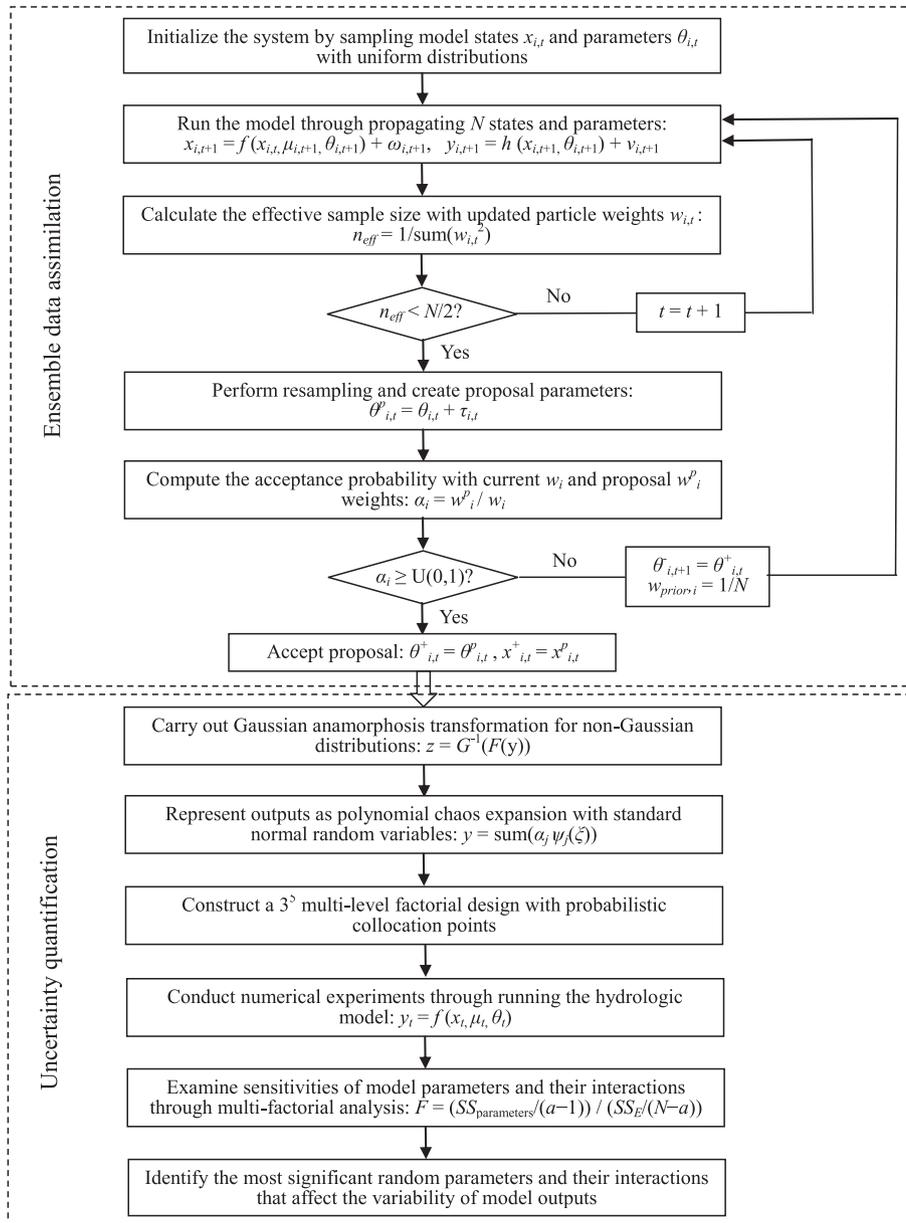


Fig. 1. The unified probabilistic framework for hydrologic data assimilation and uncertainty quantification.

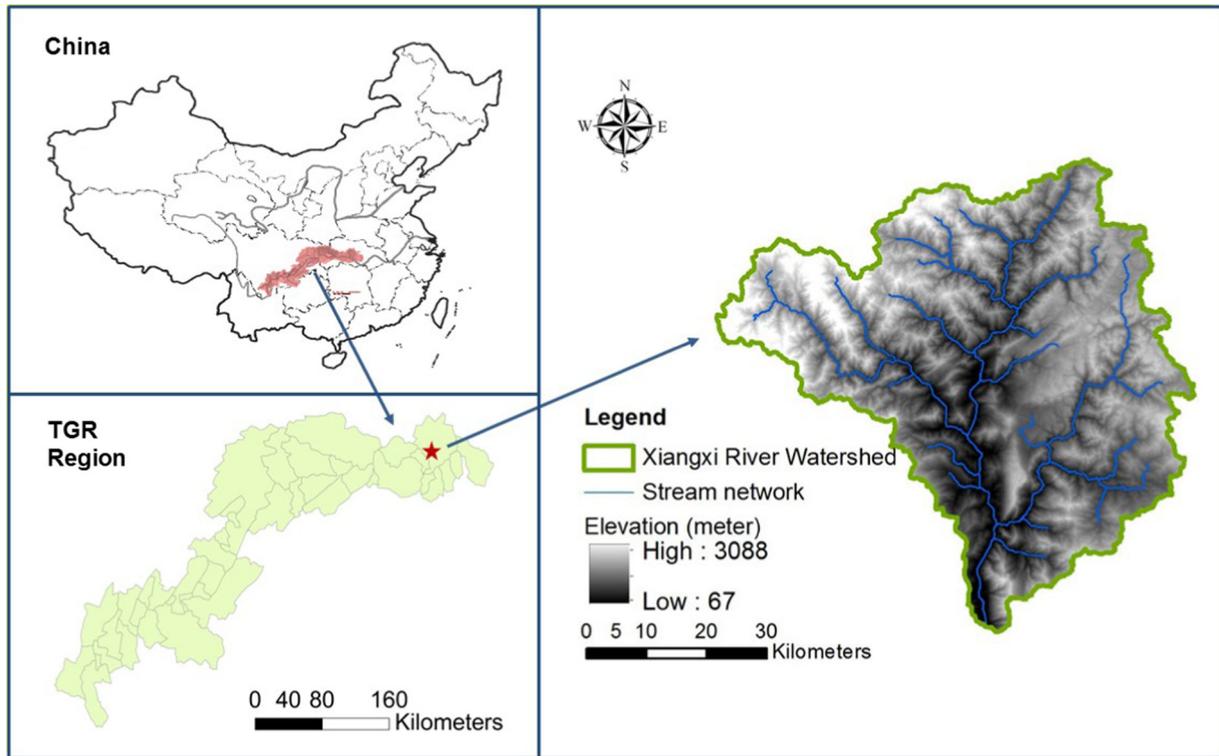


Fig. 2. Geographical location and topographic characteristics of the Xiangxi River watershed.

$$z_{\max} = G^{-1}((N - 0.5)/N) \quad (27)$$

The Gaussian anamorphosis technique is able to combine the PMCMC and the FPCE algorithms within a unified probabilistic framework. As shown in Fig. 1, the unified probabilistic framework consists of two parts, including the sequential data assimilation using the PMCMC and the forward uncertainty quantification using the FPCE. Such a unified framework is able to systematically quantify and reduce uncertainties in hydrologic predictions and explicitly reveal potential interactions among system components in the uncertainty propagation process. This developed approach greatly enhances the robustness of hydrologic modeling and facilitates our understanding of the dynamic behavior of hydrologic systems.

3. Experimental setup

The unified probabilistic framework is applied to predict daily streamflow in the Xiangxi River watershed. The Xiangxi River is the largest tributary of the TGR in Central China's Hubei Province (as shown in Fig. 2). The Xiangxi River watershed with a total area of 3099 km² lies in the subtropical region, and experiences a typical continental monsoon climate with substantial temperature variations in Spring and concentrated rainfalls in Summer. The weather is rainy in Autumn and snowy in Winter. Temperature and precipitation are also influenced by the mountainous topography of the watershed, and vary significantly with altitude. The Xiangxi River watershed is characterized by a large difference in elevation. The river originates at an elevation of 3088 m near the source of the watershed in the Shennongjia Mountains, while it reaches an elevation as low as 67 m at the outlet of the watershed where the Xiangxi River discharges into the Yangtze River. It is one of the most representative watersheds in the TGR region in terms of topographic properties, runoff volumes, and economic conditions (Han et al., 2014).

In this study, data assimilation experiments using the PMCMC algorithm were undertaken through HYMOD which is a well-known rainfall-runoff model with a daily time step (Moore, 1985). The runoff production in HYMOD is represented as a rainfall excess process, and the runoff is determined according to a probability-distributed storage capacity model (Moore, 2007; Bulygina and Gupta, 2011; Young, 2013). The catchment is considered as an infinite amount of points, and each of them has a certain soil moisture capacity denoted as c [L]. Due to spatial variability such as soil type and depth within the catchment, the variability of soil moisture capacities can be characterized by a CDF defined as:

$$F(c) = 1 - \left(1 - \frac{c}{C_{\max}}\right)^{b_{\exp}} \quad 0 \leq c \leq C_{\max} \quad (28)$$

where C_{\max} [L] is the maximum soil moisture capacity within the watershed, and b_{\exp} [-] is the degree of spatial variability in soil moisture capacities and affects the shape of the CDF. The CDF indicates the probability of occurrence of a specific soil moisture capacity across the catchment. The HYMOD model partitions excess rainfall into surface and subsurface storage through a partitioning factor, β [-]. The surface storage is characterized by three quick-flow tanks, and the subsurface storage is represented by a single slow-flow tank. The residence time of slow- and quick-flow tanks are represented as R_s and R_q [1/T], respectively. The generated streamflow is thus the addition of discharges from the slow- and quick-flow tanks. The input data of daily precipitation P [mm/d] and potential evapotranspiration ET [mm/d] are used to drive the HYMOD model.

The HYMOD is characterized by the five aforementioned model parameters, including C_{\max} , b_{\exp} , β , R_s , and R_q . The initial ranges of model parameters are given in Table 1. In addition, the HYMOD consists of five state variables, including x_{loss} , x_{slow} , x_{quick1} , x_{quick2} , and x_{quick3} . x_{loss} represents the soil moisture storage within the

Table 1
Initial uncertainty ranges of model parameters.

Parameter	Description	Unit	Minimum	Maximum
C_{\max}	Maximum storage capacity of watershed	mm	100	700
b_{\exp}	Degree of spatial variability of soil moisture capacity	–	0.10	15.00
β	Factor distributing flow to the quick-flow tank	–	0.10	0.99
R_s	Residence time of the slow-flow tank	day	0.01	0.20
R_q	Residence time of the quick-flow tank	day	0.10	0.90

watershed, x_{slow} denotes the slow-flow tank storage (subsurface storage), x_{quick1} , x_{quick2} , and x_{quick3} are the quick-flow tank storages that represent temporary detentions (depression storages).

To properly assess the performance of the PMCMC algorithm, a synthetic experiment was conducted to generate the “true” model states and streamflow observations based on predefined model parameters and observed forcing data. The data assimilation using the PMCMC was then performed on the synthetic data set, and the convergence of model parameters to the predefined values was evaluated accordingly. To account for various sources of uncertainty, stochastic perturbations were imposed by adding noise to the forcing data (precipitation and potential evapotranspiration) and observations, leading to an ensemble of model variables. When the posterior parameter distributions were derived from the ensemble data assimilation experiment, the FPCE technique was then used to quantify uncertainties in hydrologic predictions through the propagation of parameter uncertainties. The PMCMC coupled with the FPCE is able to take into account hydrologic data assimilation and uncertainty quantification within a unified probabilistic framework.

4. Results and discussion

4.1. Sequential streamflow assimilation through particle Markov chain Monte Carlo

The PMCMC algorithm was used to assimilate daily streamflow into the HYMOD model over a time period of two years from January 1994 to December 1995. To account for uncertainties in forcing data and observations, the precipitation, potential evapotranspiration and streamflow observation were perturbed by adding random noise in the data assimilation experiment. Precipitation was assumed to have a log-normal error distribution with a relative error of 20% while the potential evapotranspiration and the streamflow observation errors were assumed to be normally distributed with a relative error of 20%. The parameter particles were sampled with an ensemble size of 150 at each time step, leading to posterior parameter distributions. The experimental factors were chosen based on sensitivity analysis.

Fig. 3 shows the temporal evolution of model parameters with 90% confidence intervals for streamflow assimilation. As the

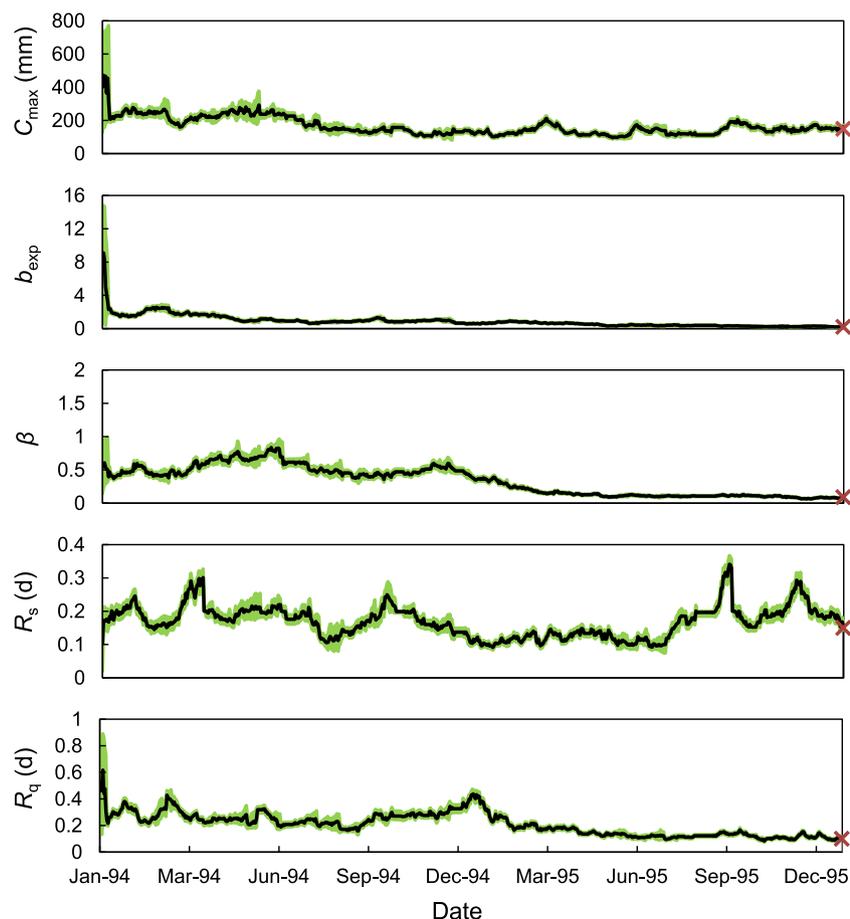


Fig. 3. Temporal evolution of model parameters with 90% confidence intervals for streamflow assimilation. Shaded areas and solid lines represent uncertainty bounds and mean values, respectively; crosses denote true parameter values.

assimilation proceeds, the estimated mean values of posterior distributions for all parameters converge toward the true values given as: $C_{max} = 149$ mm, $b_{exp} = 0.22$, $\beta = 0.09$, $R_s = 0.15$ day, and $R_q = 0.10$ day. Since the PMCMC is a recursive data assimilation method and its performance depends on the availability of new data and information in updating (correcting) parameter samples, the uncertainty degree (i.e. difference between maximum and minimum values) of model parameters is relatively high at the early stage of filtering, and the parameter uncertainty diminishes when new observations become available.

In terms of identifiability of model parameters, the slow-flow tank parameter R_s is less identifiable than the others because the slow-flow tank has the minimum contribution to the volume of generated streamflow. Contrarily, the degree of spatial variability of soil moisture capacity, denoted by b_{exp} , shows the fastest convergence with the smallest uncertainty bound. This is because the spatial variability of soil moisture capacity is strongly correlated with the streamflow generation. The identifiability of model parameters and the recursive pattern of parameter evolution would vary across case studies due to the difference in hydrologic time series and interactions between state variables and parameters.

4.2. Quantification of predictive uncertainty through factorial polynomial chaos expansion

Since the posterior distributions of all parameters are derived from the sequential data assimilation experiment with the PMCMC, it is necessary to quantify uncertainties in hydrologic

predictions through a forward propagation of parameter uncertainties. In addition, model parameters are not independent in the uncertainty propagation process; instead, they are correlated with each other, and their interactions have a significant impact on the predictive performance. As a result, it is indispensable to explore potential interactions among hydrologic parameters for advancing our understanding of the dynamic behavior of hydrologic systems.

The FPCE technique was used in this study to efficiently represent probabilistic streamflow predictions derived from the HYMOD model after the sequential data assimilation operation. As the rate of convergence of the Hermite orthogonal polynomials used in the FPCE was optimal for normally distributed stochastic processes, the posterior distributions of model parameters were transformed into normal distributions by using Gaussian anamorphosis. For illustrative purposes, Fig. 4 shows the histogram of original data, the empirical anamorphosis function, the histogram of transformed data for the slow-flow tank parameter R_s . The posterior distribution of R_s was converted to the standard normal distribution through the Gaussian anamorphosis method in a straightforward way. When all parameters are normally distributed, they can be used to quantify uncertainties in hydrologic predictions through the FPCE.

Since there were five random parameters in the HYMOD model, a five-dimensional second-order FPCE was used to characterize uncertainties in streamflow predictions. To solve the second-order FPCE, the roots of the third-order Hermite polynomial $H_3(\zeta) = \zeta^3 - 3\zeta$, namely $0, -\sqrt{3}$, and $\sqrt{3}$ were selected as the collocation points for estimating the coefficients of the FPCE.

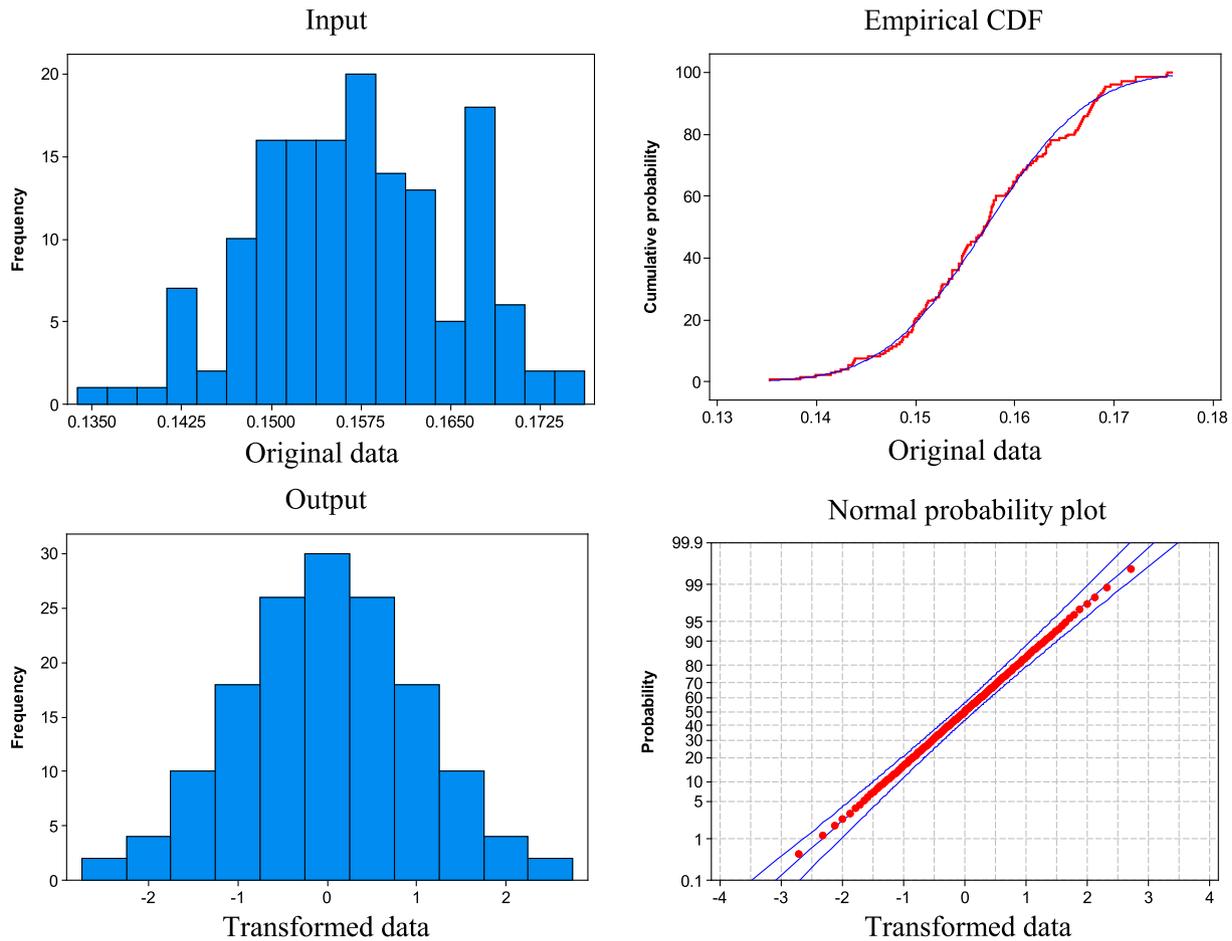


Fig. 4. Data transformation by using Gaussian anamorphosis.

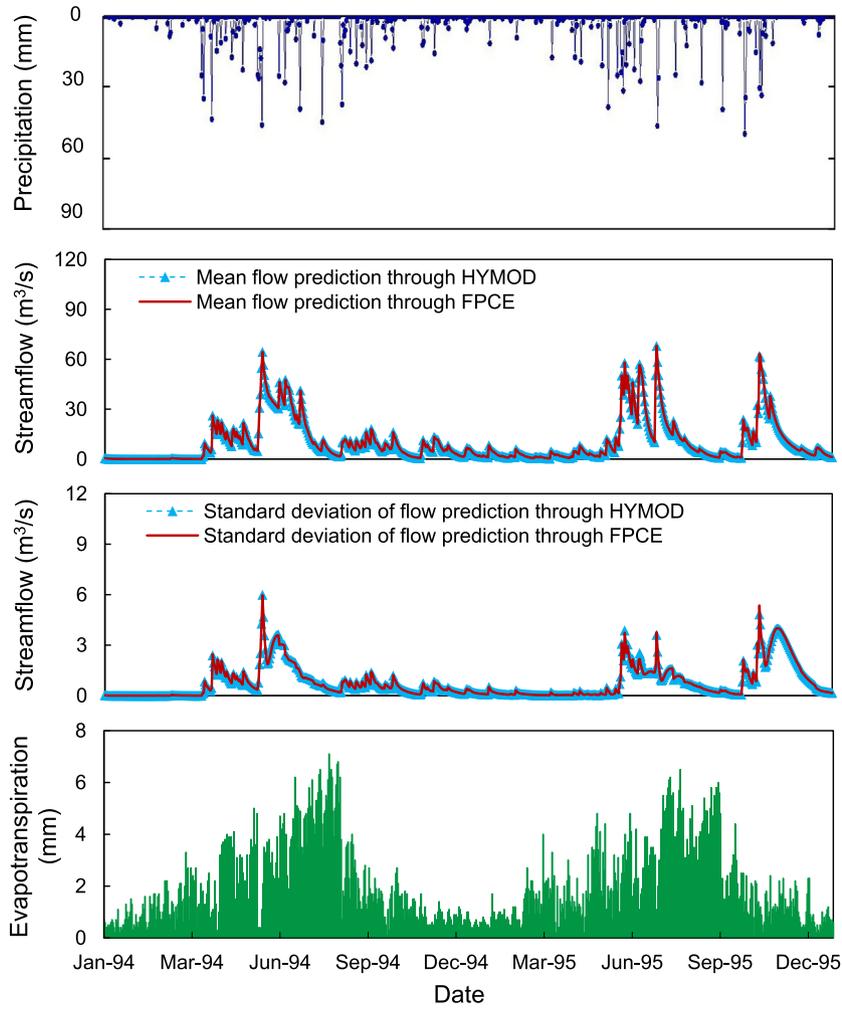


Fig. 5. Comparison of probabilistic streamflow time series generated through HYMOD and FPCE.

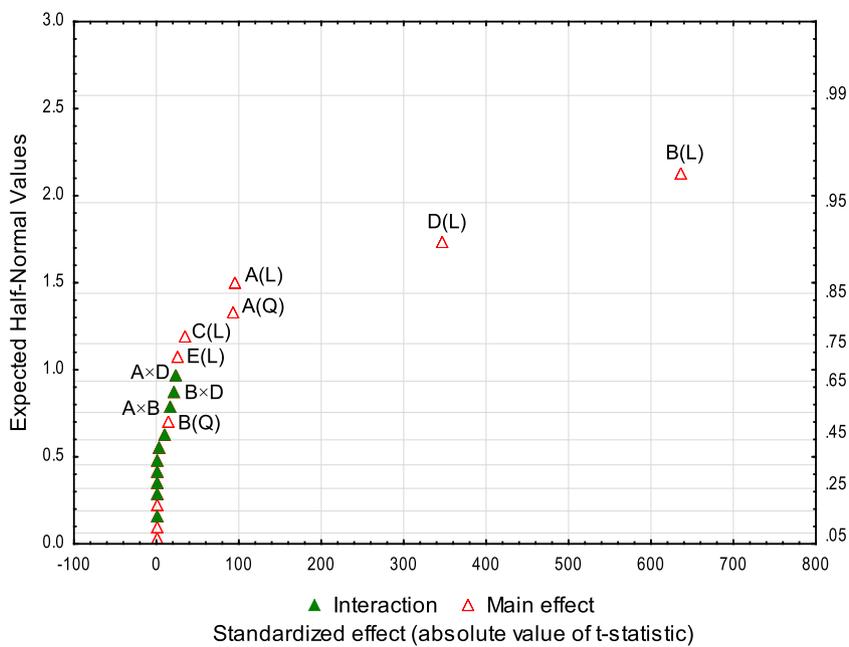


Fig. 6. Probability plots of the 10 most significant effects. A, B, C, D, and E represent model parameters of C_{max} , b_{exp} , β , R_s , and R_q , respectively. (L) and (Q) denote first- and second-order effects, respectively.

Fig. 5 presents a comparison of probabilistic streamflow time series generated through the HYMOD model and the FPCE. The FPCE results agree well with the HYMOD simulation results in terms of mean values and standard deviations. This verifies that the FPCE is able to represent probabilistic outputs of streamflow predictions generated from the HYMOD model, and can then be used as an efficient alternative to quantify uncertainties and their interactions in hydrologic predictions.

4.3. Multivariate sensitivity analysis and interaction detection

In addition to the efficient quantification of predictive uncertainty, the FPCE is also able to estimate the statistical significance of model parameters and their interactions affecting the predictive

performance, based on a 3^5 factorial design with selected collocation points. Fig. 6 shows the probability plot of the 10 most significant effects. The spatial variability in soil moisture capacity, denoted as b_{exp} , has the largest linear effect on the performance of hydrologic prediction. This is because the runoff production is represented as a rainfall excess process in HYMOD, and the volume of surface runoff is estimated through a probabilistic soil moisture capacity model which is driven by rainfall and potential evapotranspiration data. As a result, the spatial variability in soil moisture conditions greatly influences the performance of streamflow predictions. In addition, the variability of soil moisture is closely correlated with the maximum soil moisture capacity within the catchment, and their pairwise interaction has a significant contribution to the variability in model performance. To conduct a robust

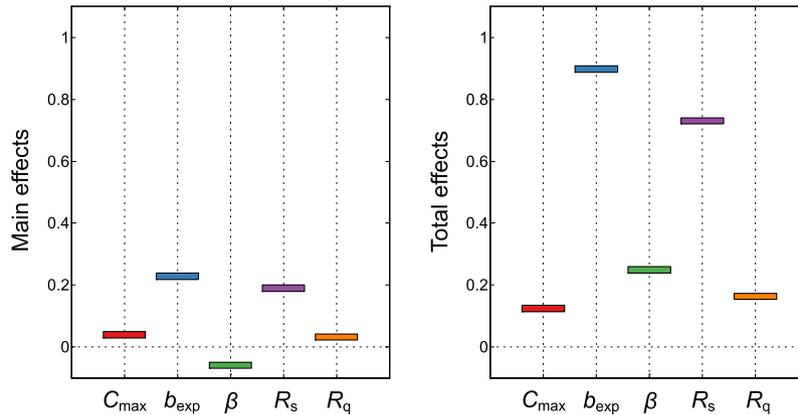


Fig. 7. Main and total effects of model parameters by using the Sobol's method.

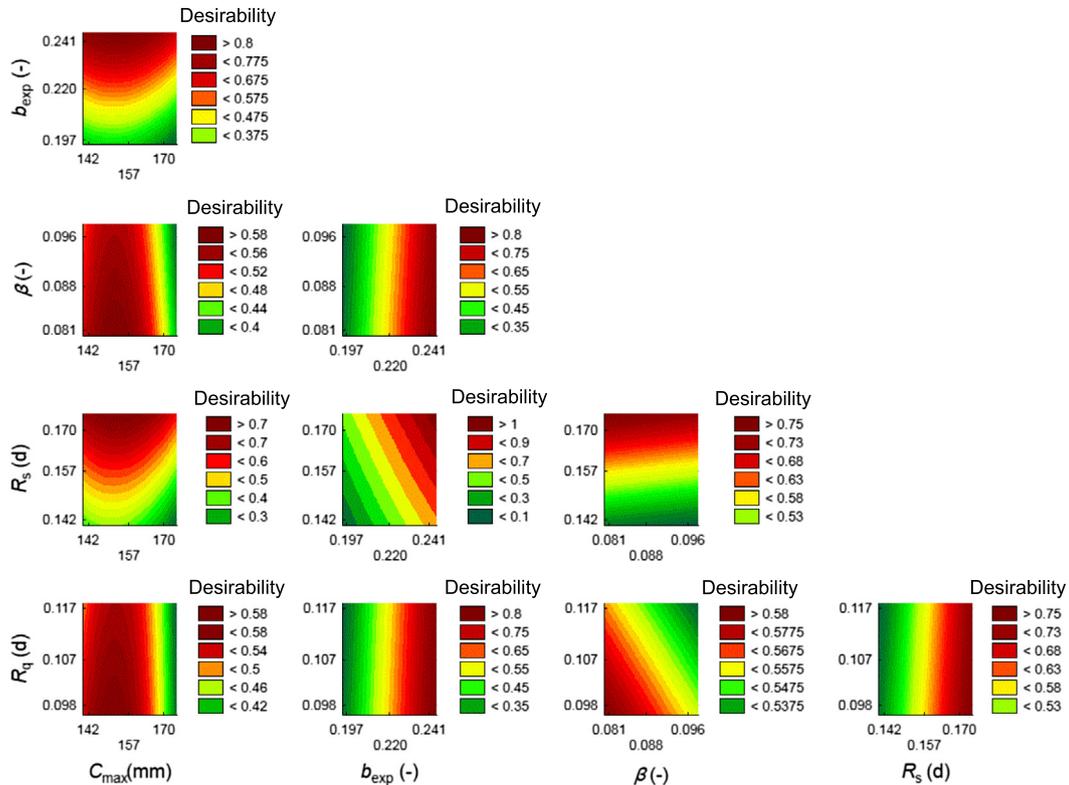


Fig. 8. Desirability of model response for pairwise parameter interactions. Desirability values range from 0.0 for an undesirable response to 1.0 for a highly desirable response.

assessment of parameter sensitivities and interactions, the variance-based sensitivity analysis based on the Sobol’s method was also performed by using the SAFE toolbox developed by Pianosi et al. (2015). As shown in Fig. 7, the results obtained from the Sobol’s method agree with those from the FPCE (Pianosi and Wagener, 2015; Sarrazin et al., 2016). The spatial variability in soil moisture capacity, denoted as b_{exp} , has the largest main and total (interaction) effects on the predictive performance.

Fig. 8 depicts the desirability of model response in terms of the root mean square error (RMSE) derived from various interactions between model parameters with each having three collocation points. The response desirability shows which parameter values produce the most desirable predicted response on RMSE, and the desirability values range from 0.0 for an undesirable response to 1.0 for a highly desirable response. The pairwise interactions between the degree of spatial variability in soil moisture capacity and the other parameters tend to produce relatively high desirability values of model response. When the maximum soil moisture capacity is 157 mm and the degree of spatial variability in soil moisture capacity is 0.24, the corresponding pairwise interaction would generate the most desirable response on the RMSE. Fig. 9 shows the variations of marginal means of the RMSE for the most significant three-way interaction. These plots indicate that the three-way interaction would reach the best predictive performance with the minimum RMSE value when the settings of the maximum storage capacity, the degree of spatial variability in soil moisture capacity, and the residence time of the slow-flow tank are 157 mm, 0.24, and 0.17 day, respectively. Our findings reveal dynamic interactions among model parameters and their contributions to the variation in the overall desirability of model response. These findings play a curial role in advancing our understanding of nonlinear dynamics of hydrologic systems and maximizing the performance of hydrologic predictions through identifying the best parameter settings. Fig. 10 depicts the temporal dynamics in the sensitivity of the most significant model parameters and pairwise interactions. Results reveal that the parameter sensitivities vary

in magnitude and direction over time due to the dynamic characteristics of hydrologic systems. In general, the performance of hydrologic predictions is more sensitive to the variation in model parameters and potential interactions when high streamflow values occur. Therefore, minimization of the errors in simulating high-flow events would significantly improve the overall accuracy of hydrologic predictions.

5. Conclusions

In this study, we developed a unified probabilistic framework that merges the strengths of the PMCMC and the FPCE algorithms for robustly quantifying and reducing uncertainties in hydrologic predictions. As an extension of the particle filter algorithm, the PMCMC was introduced to increase the diversity in posterior parameters through enhancing the posterior sampling with the MCMC, reducing the potential risk of sample impoverishment in sequential importance resampling and providing an accurate representation of uncertainty in hydrologic predictions.

When the posterior parameters were derived from the hydrologic data assimilation experiment using the PMCMC, the FPCE was introduced to quantify uncertainties in hydrologic predictions through the forward propagation of parameter uncertainties. In addition, the Gaussian anamorphosis technique was used to establish a statistical link between the PMCMC and the FPCE through a straightforward transformation of the posterior distributions of model parameters. Such a unified computational framework enables a systematic integration of data assimilation and uncertainty propagation techniques for robust hydrologic ensemble predictions.

Our results uncover the temporal evolution of model parameters for streamflow assimilation in the Xiangxi River watershed. As the PMCMC is a recursive data assimilation algorithm for estimating model states and parameters, the degree of uncertainty is relatively high at the early stage of filtering, but diminishes

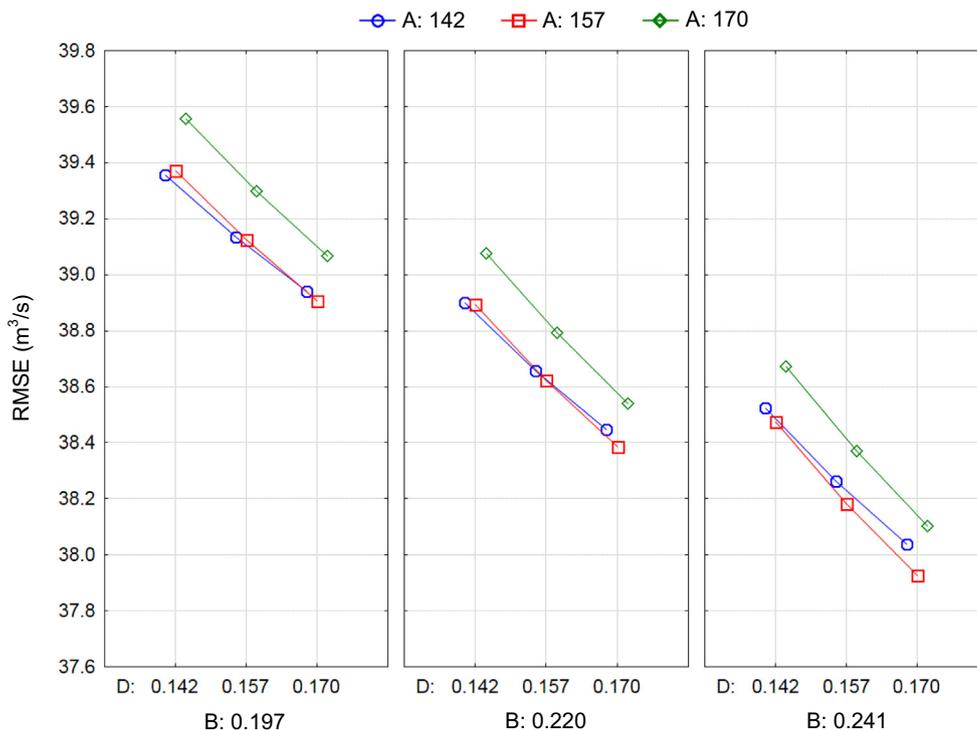


Fig. 9. Variations of marginal means of RMSE for the most significant three-way interaction. Factors A, B, and D represent C_{max} , b_{exp} , and R_s , respectively.

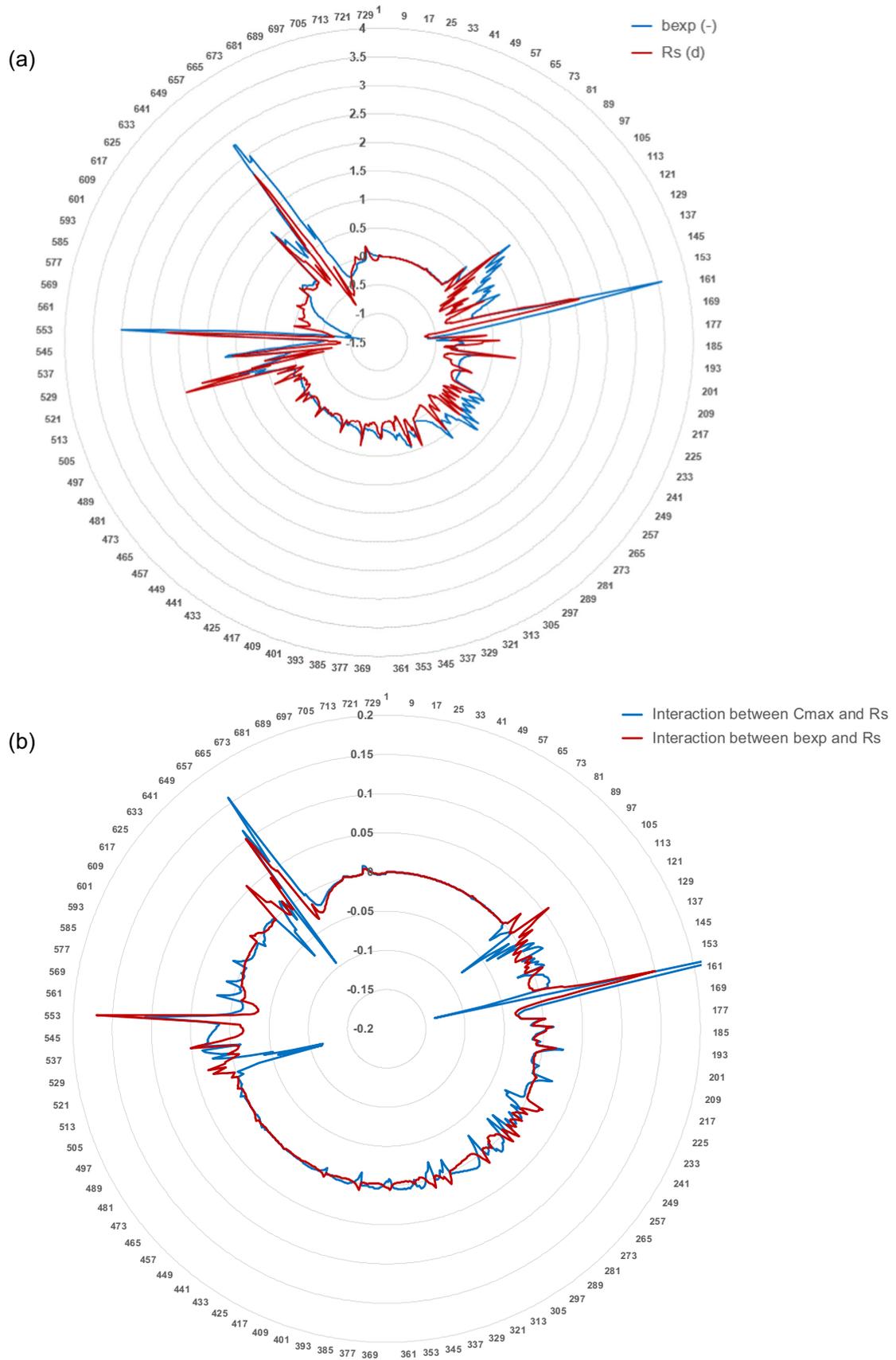


Fig. 10. Temporal variation in the sensitivity of model parameters and their interactions affecting the predictive performance.

gradually over time when new observations become available. Since soil moisture capacity plays an important role in predicting daily streamflow, the degree of spatial variability of soil moisture capacity is the most identifiable model parameter with the fastest convergence through the streamflow assimilation process. Our findings are useful for providing meaningful insights into the sequential data assimilation process, leading to robust streamflow predictions.

In addition to the recursive estimation of model parameters and state variables through a streamflow assimilation experiment, the FPCE was used to efficiently quantify uncertainties and their interactions in hydrologic predictions. Our results reveal that the spatial variability in soil moisture conditions has the most significant effect on the performance of streamflow predictions, and its interaction with the maximum soil moisture capacity in the Xiangxi River watershed has a considerable contribution to the variability in predictive accuracy. Furthermore, parameter sensitivities and interactions vary in magnitude and direction over time due to the temporal and spatial dynamics of hydrologic processes. In general, the performance of streamflow predictions is more sensitive for high-flow events, and thus the accuracy of high-flow predictions plays a crucial role in maximizing the overall predictive performance. In addition, we realize that evaluation of predictive performance and identification of model parameters are conditional on different hydrologic metrics (Wang et al., 2015b). It is necessary to take into account various evaluation metrics for addressing different hydrologic characteristics, leading to a robust assessment of predictive performance. Our findings are useful for advancing our understanding of complex dynamics and chaos in hydrologic systems, and for improving the reliability and accuracy of streamflow predictions.

The proposed unified computational framework has strong potential to strengthen our capability in providing robust hydroclimatic forecasting through integrating data assimilation with uncertainty quantification techniques in a systematic fashion. Future studies will be undertaken to examine the effectiveness and efficiency of the proposed methodology for more parametrized and complex hydroclimatic models.

Acknowledgments

This research was supported by the Major Project Program of the Natural Sciences Foundation (51190095), the National Natural Science Foundation (51225904), and the Natural Sciences and Engineering Research Council of Canada. The authors appreciate Dr. Francesca Pianosi for providing the SAFE toolbox to perform variance-based sensitivity analysis. The authors would also like to express thanks to the editor and the anonymous reviewers for their constructive comments and suggestions.

References

- Abaza, M., Anctil, F., Fortin, V., Turcotte, R., 2014. Sequential streamflow assimilation for short-term hydrological ensemble forecasting. *J. Hydrol.* 519, 2692–2706.
- 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), W01403.
- Bi, H., Ma, J., Wang, F., 2015. An improved particle filter algorithm based on ensemble Kalman filter and Markov chain Monte Carlo method. *IEEE J. Selected Top. Appl. Earth Observations Remote Sensing* 8 (2), 447–459.
- Bulygina, N., Gupta, H., 2011. Correcting the mathematical structure of a hydrological model via Bayesian data assimilation. *Water Resour. Res.* 47 (5), W05514.
- Cammalleri, C., Ciraolo, G., 2012. State and parameter update in a coupled energy/hydrologic balance model using ensemble Kalman filtering. *J. Hydrol.* 416–417, 171–181.
- Ciriello, V., Federico, V.D., Riva, M., Cadini, F., Sanctis, J.D., Zio, E., Guadagnini, A., 2013. Polynomial chaos expansion for global sensitivity analysis applied to a model of radionuclide migration in a randomly heterogeneous aquifer. *Stoch. Environ. Res. Risk Assess.* 27 (4), 945–954.
- Dai, C., Xue, L., Zhang, D., Guadagnini, A., 2016. Data-worth analysis through probabilistic collocation-based Ensemble Kalman Filter. *J. Hydrol.* 540, 488–503.
- DeChant, C.M., Moradkhani, H., 2012. Examining the effectiveness and robustness of sequential data assimilation methods for quantification of uncertainty in hydrologic forecasting. *Water Resour. Res.* 48 (4), W04518.
- DeChant, C.M., Moradkhani, H., 2014. Toward a reliable prediction of seasonal forecast uncertainty: addressing model and initial condition uncertainty with ensemble data assimilation and Sequential Bayesian Combination. *J. Hydrol.* 519, 2967–2977.
- Dumedah, G., Coulibaly, P., 2013. Evaluating forecasting performance for data assimilation methods: the ensemble Kalman filter, the particle filter, and the evolutionary-based assimilation. *Adv. Water Resour.* 60, 47–63.
- Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res. Oceans* 99 (C5), 10143–10162.
- Fajraoui, N., Ramasomanana, F., Younes, A., Mara, T.A., Ackerer, P., Guadagnini, A., 2011. Use of global sensitivity analysis and polynomial chaos expansion for interpretation of nonreactive transport experiments in laboratory-scale porous media. *Water Resour. Res.* 47 (2), W02521.
- Gharamti, M.E., Hoteit, I., Valstar, J., 2013. Dual states estimation of a subsurface flow-transport coupled model using ensemble Kalman filtering. *Adv. Water Resour.* 60, 75–88.
- Gharamti, M.E., Ait-El-Fquih, B., Hoteit, I., 2015. An iterative ensemble Kalman filter with one-step-ahead smoothing for state-parameters estimation of contaminant transport models. *J. Hydrol.* 527, 442–457.
- Han, J., Huang, G., Zhang, H., Li, Z., Li, Y., 2014. Heterogeneous precipitation and streamflow trends in the Xiangxi River watershed, 1961–2010. *J. Hydrol. Eng.* 19 (6), 1247–1258.
- Johnson, R.A., Wichern, D.W., 1988. *Applied Multivariate Statistical Analysis*. Prentice-Hall, Englewood Cliffs.
- Konda, U., Singh, T., Singla, P., Scott, P., 2010. Uncertainty propagation in puff-based dispersion models using polynomial chaos. *Environ. Modell. Softw.* 25 (12), 1608–1618.
- Li, H., Zhang, D., 2007. Probabilistic collocation method for flow in porous media: comparisons with other stochastic methods. *Water Resour. Res.* 43 (9), W09409.
- Lin, G., Tartakovsky, A.M., 2009. An efficient, high-order probabilistic collocation method on sparse grids for three-dimensional flow and solute transport in randomly heterogeneous porous media. *Adv. Water Resour.* 32 (5), 712–722.
- Liu, D., Mishra, A.K., Yu, Z., 2016. Evaluating uncertainties in multi-layer soil moisture estimation with support vector machines and ensemble Kalman filtering. *J. Hydrol.* 538, 243–255.
- Liu, Y., Gupta, H.V., 2007. Uncertainty in hydrologic modeling: toward an integrated data assimilation framework. *Water Resour. Res.* 43 (7), W07401.
- Marzouk, Y.M., Najm, H.N., 2009. Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems. *J. Comput. Phys.* 228 (6), 1862–1902.
- Montgomery, D., 2000. *Design and Analysis of Experiments*. John Wiley & Sons, New York.
- Montgomery, D.C., Runger, G.C., 2013. *Applied Statistics and Probability for Engineers*. John Wiley & Sons, New York.
- Moore, R.J., 2007. The PDM rainfall-runoff model. *Hydrol. Earth Syst. Sci.* 11 (1), 483–499.
- Moore, R.J., 1985. The probability-distributed principle and runoff production at point and basin scales. *Hydrolog. Sci. J.* 30 (2), 273–297.
- Moradkhani, H., DeChant, C.M., Sorooshian, S., 2012. Evolution of ensemble data assimilation for uncertainty quantification using the particle filter-Markov chain Monte Carlo method. *Water Resour. Res.* 48 (12), W12520.
- Moradkhani, H., Hsu, K.L., Gupta, H., Sorooshian, S., 2005a. Uncertainty assessment of hydrologic model states and parameters: sequential data assimilation using the particle filter. *Water Resour. Res.* 41 (5), W05012.
- Moradkhani, H., Sorooshian, S., Gupta, H.V., Houser, P.R., 2005b. Dual state-parameter estimation of hydrological models using ensemble Kalman filter. *Adv. Water Resour.* 28 (2), 135–147.
- Müller, F., Jenny, P., Meyer, D.W., 2011. Probabilistic collocation and lagrangian sampling for advective tracer transport in randomly heterogeneous porous media. *Adv. Water Resour.* 34 (12), 1527–1538.
- Najm, H.N., 2009. Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics. *Annu. Rev. Fluid. Mech.* 41 (1), 35–52.
- Oladyshkin, S., Class, H., Helmig, R., Nowak, W., 2011. A concept for data-driven uncertainty quantification and its application to carbon dioxide storage in geological formations. *Adv. Water Resour.* 34 (11), 1508–1518.
- Pathiraja, S., Marshall, L., Sharma, A., Moradkhani, H., 2016. Detecting non-stationary hydrologic model parameters in a paired catchment system using data assimilation. *Adv. Water Resour.* 94, 103–119.
- Panzeri, M., Riva, M., Guadagnini, A., Neuman, S.P., 2014. Comparison of Ensemble Kalman Filter groundwater-data assimilation methods based on stochastic moment equations and Monte Carlo simulation. *Adv. Water Resour.* 66, 8–18.
- Rafieeiniasab, A., Seo, D.-J., Lee, H., Kim, S., 2014. Comparative evaluation of maximum likelihood ensemble filter and ensemble Kalman filter for real-time

- assimilation of streamflow data into operational hydrologic models. *J. Hydrol.* 519, 2663–2675.
- Randrianasolo, A., Thirel, G., Ramos, M.H., Martin, E., 2014. Impact of streamflow data assimilation and length of the verification period on the quality of short-term ensemble hydrologic forecasts. *J. Hydrol.* 519, 2676–2691.
- Pianosi, F., Sarrazin, F., Wagener, T., 2015. A Matlab toolbox for Global Sensitivity Analysis. *Environ. Modell. Softw.* 70, 80–85.
- Pianosi, F., Wagener, T., 2015. A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. *Environ. Modell. Softw.* 67, 1–11.
- Plaza Guingla, D.A., De Keyser, R., De Lannoy, G.J.M., Giustarini, L., Matgen, P., Pauwels, V.R.N., 2013. Improving particle filters in rainfall-runoff models: application of the resample-move step and the ensemble Gaussian particle filter. *Water Resour. Res.* 49 (7), 4005–4021.
- Rajabi, M.M., Ataie-Ashtiani, B., Simmons, C.T., 2015. Polynomial chaos expansions for uncertainty propagation and moment independent sensitivity analysis of seawater intrusion simulations. *J. Hydrol.* 520, 101–122.
- Reichle, R.H., McLaughlin, D.B., Entekhabi, D., 2002. Hydrologic data assimilation with the ensemble Kalman filter. *Mon. Weather. Rev.* 130, 103–114.
- Ryu, D., Crow, W.T., Zhan, X., Jackson, T.J., 2009. Correcting unintended perturbation biases in hydrologic data assimilation. *J. Hydrometeorol.* 10, 734–750.
- Salamon, P., Feyen, L., 2009. Assessing parameter, precipitation, and predictive uncertainty in a distributed hydrological model using sequential data assimilation with the particle filter. *J. Hydrol.* 376 (3–4), 428–442.
- Samuel, J., Coulibaly, P., Dumedah, G., Moradkhani, H., 2014. Assessing model state and forecasts variation in hydrologic data assimilation. *J. Hydrol.* 513, 127–141.
- Sarrazin, F., Pianosi, F., Wagener, T., 2016. Global sensitivity analysis of environmental models: convergence and validation. *Environ. Modell. Softw.* 79, 135–152.
- Schöniger, A., Nowak, W., Hendricks Franssen, H.J., 2012. Parameter estimation by ensemble Kalman filters with transformed data: approach and application to hydraulic tomography. *Water Resour. Res.* 48 (4), W04502.
- Shen, Y.S., Tan, Z.F., Shen, X.L., Bai, J.J., Li, Q.Z., Wang, S., 2012. Study of energy saving and emission reduction based on the OLAP multi-indicator relational model. *J. Environ. Inform.* 20 (2), 115–122.
- Smith, P.J., Beven, K.J., Tawn, J.A., 2008. Detection of structural inadequacy in process-based hydrological models: a particle-filtering approach. *Water Resour. Res.* 44 (1), W01410.
- Sochala, P., Le Maître, O.P., 2013. Polynomial Chaos expansion for subsurface flows with uncertain soil parameters. *Adv. Water Resour.* 62, 139–154.
- Sun, L., Nistor, L., Seidou, O., 2015. Streamflow data assimilation in SWAT model using Extended Kalman Filter. *J. Hydrol.* 531, 671–684.
- Tatang, M.A., Pan, W., Prinn, R.G., McRae, G.J., 1997. An efficient method for parametric uncertainty analysis of numerical geophysical models. *J. Geophys. Res.* 102 (D18), 21925–21932.
- Vrugt, J.A., ter Braak, C.J.F., Diks, C.G.H., Schoups, G., 2013. Hydrologic data assimilation using particle Markov chain Monte Carlo simulation: theory, concepts and applications. *Adv. Water Resour.* 51, 457–478.
- Wang, S., Huang, G.H., 2015. A multi-level Taguchi-factorial two-stage stochastic programming approach for characterization of parameter uncertainties and their interactions: an application to water resources management. *Eur. J. Oper. Res.* 240 (2), 572–581.
- Wang, S., Huang, G.H., Baetz, B.W., 2015a. An inexact probabilistic-possibilistic optimization framework for flood management in a hybrid uncertain environment. *IEEE Trans. Fuzzy Syst.* 23 (4), 897–908.
- Wang, S., Huang, G.H., Baetz, B.W., Huang, W., 2015b. A polynomial chaos ensemble hydrologic prediction system for efficient parameter inference and robust uncertainty assessment. *J. Hydrol.* 530, 716–733.
- Wang, S., Huang, G.H., Baetz, B.W., Huang, W., 2016. Probabilistic inference coupled with possibilistic reasoning for robust estimation of hydrologic parameters and piecewise characterization of interactive uncertainties. *J. Hydrometeorol.* 17, 1243–1260.
- Wang, S., Huang, G.H., Huang, W., Fan, Y.R., Li, Z., 2015c. A fractional factorial probabilistic collocation method for uncertainty propagation of hydrologic model parameters in a reduced dimensional space. *J. Hydrol.* 529 (3), 1129–1146.
- Weerts, A.H., El Serafy, G.Y.H., 2006. Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models. *Water Resour. Res.* 42 (9), W09403.
- Wei, J., Huang, G.H., Wang, S., Zhao, S., Yao, Y., 2013. Improved solubilities of PAHs by multi-component Gemini surfactant systems with different spacer lengths. *Colloid Surface A* 423, 50–57.
- Wiener, N., 1938. The homogeneous chaos. *Am. J. Math.* 60 (4), 897–936.
- Wu, C.F.J., Hamada, M.S., 2009. *Experiments: Planning, Analysis, and Optimization*. John Wiley & Sons, New Jersey.
- Xie, X., Zhang, D., 2010. Data assimilation for distributed hydrological catchment modeling via ensemble Kalman filter. *Adv. Water Resour.* 33 (6), 678–690.
- Xiu, D., Karniadakis, G., 2002. The Wiener-Askey polynomial chaos for stochastic differential equations. *Siam. J. Sci. Comput.* 24 (2), 619–644.
- Xu, T., Gómez-Hernández, J.J., 2015. Probability fields revisited in the context of ensemble Kalman filtering. *J. Hydrol.* 531, 40–52.
- Yan, H., DeChant, C.M., Moradkhani, H., 2015. Improving soil moisture profile prediction with the particle filter-Markov chain Monte Carlo method. *IEEE Trans. Geosci. Remote Sensing* 53 (11), 6134–6147.
- Young, P.C., 2013. Hypothetico-inductive data-based mechanistic modeling of hydrological systems. *Water Resour. Res.* 49 (2), 915–935.
- Zeng, X.T., Li, Y.P., Huang, G.H., Liu, J., 2016. Modeling water trading under uncertainty for supporting water resources management in an arid region. *J. Water Res. Planning ASCE* 142 (2), 04015058.