



# Parameter uncertainty and temporal dynamics of sensitivity for hydrologic models: A hybrid sequential data assimilation and probabilistic collocation method

Y.R. Fan<sup>a</sup>, G.H. Huang<sup>a, b, \*</sup>, B.W. Baetz<sup>c</sup>, Y.P. Li<sup>b</sup>, K. Huang<sup>d</sup>, Z. Li<sup>c</sup>, X. Chen<sup>e</sup>, L.H. Xiong<sup>f</sup>

<sup>a</sup> Institute for Energy, Environment and Sustainable Communities, University of Regina, Regina, Saskatchewan S4S 0A2, Canada

<sup>b</sup> School of Environment, Beijing Normal University, Beijing, 100875, China

<sup>c</sup> Department of Civil Engineering, McMaster University, Hamilton, ON L8S 4L8, Canada

<sup>d</sup> Faculty of Engineering and Applied Science, University of Regina, Regina, Saskatchewan S4S 0A2, Canada

<sup>e</sup> State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing, 210098, China

<sup>f</sup> State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan, 430072, China

## ARTICLE INFO

### Article history:

Received 9 November 2015

Received in revised form 8 September 2016

Accepted 12 September 2016

Available online xxx

### Keywords:

Uncertainty

Particle filter

Probabilistic collocation method

Sensitivity analysis

Maximal information coefficient

Hydrologic model

Monte Carlo simulation

## ABSTRACT

In this study, a hybrid sequential data assimilation and probabilistic collocation (HSDAPC) approach is proposed for analyzing uncertainty propagation and parameter sensitivity of hydrologic models. In HSDAPC, the posterior probability distributions of model parameters are first estimated through a particle filter method based on streamflow discharge data. A probabilistic collocation method (PCM) is further employed to show uncertainty propagation from model parameters to model outputs. The temporal dynamics of parameter sensitivities are then generated based on the polynomial chaos expansion (PCE) generated by PCM, which can reveal the dominant model components for different catchment conditions. The maximal information coefficient (MIC) is finally employed to characterize the correlation/association between model parameter sensitivity and catchment precipitation, potential evapotranspiration and observed discharge. The proposed method is applied to the Xiangxi River located in the Three Gorges Reservoir area. The results show that: (i) the proposed HSDAPC approach can generate effective 2nd and 3rd PCE models which provide accuracy predictions; (ii) 2<sup>nd</sup>-order PCE, which can run nearly ten time faster than the hydrologic model, can capably represent the original hydrological model to show the uncertainty propagation in a hydrologic simulation; (iii) the slow ( $R_s$ ) and quick flows ( $R_q$ ) in Hymod show significant sensitivities during the simulation periods but the distribution factor ( $\alpha$ ) shows a least sensitivity to model performance; (iv) the model parameter sensitivities show significant correlation with the catchment hydro-meteorological conditions, especially during the rainy period with MIC values larger than 0.5. Overall, the results in this paper indicate that uncertainty propagation and temporal sensitivities of parameters can be effectively characterized through the proposed HSDAPC approach.

© 2016 Published by Elsevier Ltd.

## 1. Introduction

Hydrologic models are simplified, conceptual representations of a part of the hydrologic cycle, which use relatively simple mathematical equations to conceptualize and aggregate the complex, spatially distributed, and highly interrelated water, energy, and vegetation processes in a watershed (Vrugt et al., 2005). Hydrologic models are increasingly used in real world applications due to the growing availability of both computing power and hydrological data observed at fine spatial and temporal scales (Montanari and Brath, 2004). However, significant uncertainties are associated with rainfall-runoff simulation resulting from uncertainties in model parameters, structures and inputs. Due to the inherent complexities (e.g. Space-time variability of climatic inputs, Heterogeneity of the land surface condition etc.) in the rainfall-runoff process, uncertainties in the hydrological model are inevitable. In addition, the uncertainty can be divided into epistemic and aleatory uncertainty, in which the aleatory uncertainty

cannot be reduced (Parrish et al., 2012; Gong et al., 2013). Consequently, probabilistic approaches are desired to characterize the uncertainty and provide reliable hydrologic forecasts for estimating designated variables in engineering practice, mitigating hydrological risks and improving water resource management policies (Fan et al., 2012; Sikorska et al., 2014; DeChant and Moradkhani, 2014; Yan et al., 2015; Fan et al., 2015b,c; Kong et al., 2015; Rakovec et al., 2015).

Previously, numerous approaches have been proposed for quantifying the uncertainty in hydrologic predictions (Parrish et al., 2012; DeChant and Moradkhani, 2014; Madadgar and Moradkhani, 2014). Among these uncertainty quantification approaches, data assimilation methods, especially sequential data assimilation techniques, have been developed for explicitly dealing with various uncertainties and for optimally merging observations into uncertain model predictions (Reichle et al., 2002; Moradkhani et al., 2005a; Clark et al., 2008; Xie and Zhang, 2013; Pathiraja et al., 2016a,b). Sequential data assimilation methods continuously update the states and parameters in a model when new measurements become available to improve the model forecast and evaluate the forecast accuracy (Vrugt et al., 2005). Particle filter (PF) method, as the most common example of sequential Monte Carlo (SMC) methods, has been widely used for

\* Corresponding author. Institute for Energy, Environment and Sustainable Communities, University of Regina, Regina, Saskatchewan S4S 0A2, Canada.

Email addresses: huangg@uregina.ca (G.H. Huang); xichen@hhu.edu.cn (X. Chen)

quantifying uncertainties in hydrologic simulation (Moradkhani et al., 2005b; Weerts and EI Serafy, 2006; Zhang et al., 2012a,b; Zhang and Yang, 2013, 2014; Noh et al., 2014). PF evolves a sample of the state space forward using the SMC method to approximate the predictive distribution (Liu et al., 2012). The most significant advantage of PF is that it relaxes the Gaussian distribution in state-space model errors, which is required for the ensemble Kalman filter method. Furthermore, the PF method performs updating on the particle weights instead of the state variables, which can reduce numerical instability especially in physically-based or process-based models (Liu et al., 2012). The initial implementation of PF was based on sequential importance sampling, which led to severe deterioration for particles (i.e. only several or even one particle would be available). Consequently, sampling importance resampling (SIR) was proposed to mitigate the above problem (Moradkhani et al., 2005b). Previous studies in other fields concluded that the PF method usually requires more samples than other filtering methods and the sample size would increase exponentially with the number of state variables (Liu and Chen, 1998; Fearnhead and Clifford, 2003; Snyder et al., 2008). Specifically, hundreds or thousands of ensemble members may be needed for reliable characterization of the posterior PDFs even for small problems with only a few unknown states and parameters (Liu et al., 2012). Some studies found that, for conceptual hydrologic models, PF would perform better than ensemble Kalman filter (EnKF) method when the sample size is more than a hundred, and EnKF performance was found to be suboptimal due to high non-linear non-Gaussian hydrologic system (Weerts and EI Serafy, 2006; DeChant and Moradkhani, 2012). However, the number requirement of particles for physically-based distributed hydrologic models may limit operational applications of PF (Liu et al., 2012). A recent improvement for PF is to combine the strengths of sequential Monte Carlo sampling and Markov chain Monte Carlo simulation (Moradkhani et al., 2012; Vrugt et al., 2013), which can allow a more complete representation of the posterior distribution, reduce the chance of sample impoverishment (i.e. decrease in the diversity of the particles or even a single particle available after resampling steps) and lead to a more accurate streamflow forecast with small, manageable ensemble sizes (Moradkhani et al., 2012).

The PF approach can sequentially merge observations into uncertain model predictions and quantify the posterior probabilities of model parameters through a set of random samples. Further characterization for uncertainty propagation from model parameters to model outputs are desired (Beven, 2006; Samaniego et al., 2013; Westra et al., 2014; Chaney et al., 2015). For instance, Samaniego et al. (2013) evaluated the implications of parameter uncertainty on soil moisture drought analysis in Germany. Results of this study emphasize the importance of accounting for the parametric uncertainty for identifying benchmark drought events in Germany. Chaney et al. (2015) analyzed the role of model parameter uncertainty in flood and drought monitoring, which parameter uncertainty remains an important concern for predicting extreme events even after applying monthly and annual constraints to the ensemble. Consequently, previous studies demonstrate that parameter uncertainties always pose significant impacts on hydrological prediction and need to be addressed clearly. However, most of previous methods are based on Monte Carlo simulation, which may not explicitly present the uncertainty evolution in model simulation. In such a MC simulation process, model parameters would be sampled from known distributions, and each sample of model parameters would be entered into the hydrologic model to obtain statistics or density estimates of the model predictions. However, with complex hydrologic models such as distrib-

uted hydrologic models, this sampling approach is computationally intensive (Herman et al., 2013; Fan et al., 2015d). In addition, the uncertainty propagation from model parameters to model outputs can hardly be explicit since no expressions exist in MC process to show how parameter uncertainties influence the variations of model outputs. Furthermore, due to the temporal-spatial variations in rainfall-runoff processes, the parameter sensitivity may be time-variant. Consequently, characterization of temporal dynamics of parameter sensitivity is also desired to indicate the dominant model component under different hydro-meteorological conditions. Therefore, this work proposes a hybrid sequential data assimilation and probabilistic collocation (HSDAPC) method to explicitly show uncertainty propagation in hydrologic simulation and further identify the temporal dynamics of parameter sensitivity. The HSDAPC approach will approximate the posterior probabilities of hydrologic model parameters through the use of PF and then facilitate uncertainty propagation from model parameters to model predictions through a probabilistic collocation method (PCM). The PCM-based temporal dynamics of parameter sensitivity will be derived to identify the dominant model components that impact model predictions under different hydro-meteorological conditions. The association between parameter sensitivity and catchment conditions will be finally revealed through the maximal information coefficient. The proposed approach is applied to the Xiangxi River basin based on a conceptual rainfall-runoff model. This application can help demonstrate the strength and applicability of the proposed methodology.

## 2. Methodology

### 2.1. Sequential data assimilation through the use of a particle filter

#### 2.1.1. Bayesian filtering

In a sequential data assimilation process, the evolution of the simulated system states can be represented as follows:

$$x_t = f(x_{t-1}, u_{t-1}, \theta) + \omega_{t-1} \quad (1)$$

where the subscript  $t$  denotes the time step;  $f$  is a nonlinear function expressing the system transition from time  $t - 1$  to  $t$ ;  $x_t$  denote the state variables, and  $\theta$  are the model parameters;  $\omega_{t-1}$  is considered as process noise (i.e. model error).

When new observations are available, the forecasted states can be corrected through assimilating the observations into the model, resulting in the updating process described by:

$$y_t = h(x_t, \theta) + v_t \quad (2)$$

where  $h$  is the nonlinear function producing forecasted observations;  $v_t$  is the observation noise.

The essence of the state estimation problem in the Bayesian filtering framework is to construct the posterior probability density function (PDF)  $p(x_t|y_{1:t})$  of a state based on all of the available information (Gordon et al., 1993). The posterior PDF can be calculated in two steps theoretically: prediction and update, in which the state PDF from the previous state would be integrated through the system model, and the update operation modifies the prediction PDF making use of the latest observations (Han and Li, 2008). The prediction step aims to obtain the prior  $p(x_t|y_{1:t-1})$  based on the fact that the transition  $p(x_t|x_{t-1})$  and the posterior  $p(x_{t-1}|y_{1:t-1})$  at time step  $t-1$  are known, which can be expressed as:

$$p(x_t|y_{1:t-1}) = \int p(x_t|x_{t-1}) p(x_{t-1}|y_{1:t-1}) dx_{t-1} \quad (3)$$

where the transition is the probabilistic model of the system described by Equation (1). When new observations at time  $t$  are available, the prior can be corrected according to Bayes' rule, formulated as follows:

$$p(x_t|y_{1:t}) = \frac{p(y_t|x_t) p(x_t|y_{1:t-1})}{\int p(y_t|x_t) p(x_t|y_{1:t-1}) dx_t} \quad (4)$$

where  $p(x_t|y_{1:t-1})$  represents the prior information;  $p(y_t|x_t)$  is the likelihood.

The optimal Bayesian solution (i.e. Equations (3) and (4)) is difficult to determine since the evaluation of the integrals may be intractable (Plaza Guingla et al., 2013). Consequently, approximation methods are applied to address the above issues. Ensemble Kalman filter (EnKF) and PF approaches are the two most widely used methods. The central idea of EnKF and PF is to represent the state probability density function (pdf) as a set of random samples and the difference between these two methods lies in the way of recursively generating an approximation to the state PDF (Weerts and EI Serafy, 2006).

### 2.1.2. Particle filter method

The PF approach is a sequential Monte Carlo method that calculates the posterior distributions of states and parameters by a set of random samples. The advantage of PF, in comparison with EnKF, is that it relaxes the assumption of a Gaussian error structure, which allows the PF to more accurately predict the posterior distribution in the presence of skewed distributions (Moradkhani et al., 2005b; DeChant and Moradkhani, 2012). Moreover, in hydrologic simulation, PF can preserve the water balance while EnKF may not achieve it (Moradkhani et al., 2012; Yan et al., 2015). In detail, consider  $ne$  independent and identically distributed random variables  $x_{t,i} \sim p(x_t|y_{1:t})$  for  $i = 1, 2, \dots, ne$ , the posterior density based on the sequential importance sampling (SIS) method can be approximated as a discrete function:

$$p(x_t|y_{1:t}) = \sum_{i=1}^{ne} w_{t,i} \delta(x_t - x_{t,i}) \quad (5)$$

where  $w_{t,i}$  are the normalized weight of the  $i$ th particle drawn from the proposal distribution;  $\delta$  is the Dirac delta function. Assuming the system state to be a Markov process, and applying the Bayesian recursive expression to the filtering problem, the updating expression for the importance weights (non-normalized) is expressed as:

$$w_{t,i} = w_{t-1,i} \cdot \frac{p(y_t|x_{t,i}) p(x_{t,i}|x_{t-1,i})}{q(x_{t,i}|x_{t-1,i}, y_t)} \quad (6)$$

Equation (6) provides the mechanism to sequentially update the importance weights, given an appropriate choice of the proposal distribution  $q(x_{t,i}|x_{t-1,i}, y_t)$ . Consequently, the expression of the pro-

positional distribution will significantly affect the efficiency and complexity of the PF method. An appropriate choice for the proposal density function is expressed as follows (Doucet et al., 2000; Doucet and Johansen, 2011):

$$q(x_{t,i}|x_{t-1,i}, y_t) = p(x_{t,i}|x_{t-1,i}) \quad (7)$$

When the transition prior is chosen as the proposal distribution, the importance weights depend on their past values and the likelihood  $p(y_t|x_{t,i})$ , which is expressed as:

$$w_{t,i} = w_{t-1,i} p(y_t|x_{t,i}) \quad (8)$$

For the likelihood  $p(y_t|x_{t,i})$ , a common choice of the likelihood density function is the Gaussian distribution that describes the differences between the observation predictions and the observations, scaled by the (usually defined *a priori*) observation errors (Plaza Guingla et al., 2013).

For the PF through SIS, a serious limitation is the depletion of the particle set, which means that, after a few iterations (time steps), all the particles except one are discarded because their importance weights are insignificant (Doucet et al., 2000). To address the above issue, a resampling step is usually adopted in PF to eliminate the particles with small importance weights and replace them by particles with large importance weights. Various resampling methods have been developed, and the most commonly used ones include multinomial resampling, systematic resampling, stratified resampling, and residual resampling methods (Bi et al., ). In this study, the multinomial resampling scheme is employed.

### 2.2. Probabilistic collocation method (PCM)

The polynomial chaos expansion (PCE), first introduced by Wiener (1938), is typically applied to express the evolution of uncertainty in a dynamical system with random inputs, in which the model stochastic process is decomposed by Hermite polynomials in terms of Gaussian random variables. For non-Gaussian random input variables (e.g. Gamma and uniform), the convergence of Hermite polynomial expansion is not optimal (Xiu and Karniadakis, 2003). Xiu and Karniadakis (2002) proposed generalized polynomial chaos expansions for non-Gaussian distributions. The general polynomial chaos expansion can be written in the form:

$$Y = a_0 + \sum_{i_1=1}^n a_{i_1} \Gamma_1(\zeta_{i_1}) + \sum_{i_1=1}^n \sum_{i_2=1}^{i_1} a_{i_1 i_2} \Gamma_2(\zeta_{i_1}, \zeta_{i_2}) + \sum_{i_1=1}^n \sum_{i_2=1}^{i_1} \sum_{i_3=1}^{i_2} a_{i_1 i_2 i_3} \Gamma_3(\zeta_{i_1}, \zeta_{i_2}, \zeta_{i_3}) + \dots \quad (9)$$

where  $Y$  is the output and  $\Gamma_p(\zeta_{i_1}, \zeta_{i_2}, \dots, \zeta_{i_p})$  are the  $p$ th order polynomials in terms of the multi-dimensional random variables  $\{\zeta_{i_k}\}_{k=1}^M$ . For standard normal variables, the Hermite polynomial will be used, which is expressed as:

$$\Gamma_p(\zeta_{i_1}, \zeta_{i_2}, \dots, \zeta_{i_M}) = (-1)^p e^{1/2\zeta^T \zeta} \frac{\partial^M}{\partial \zeta_{i_1} \partial \zeta_{i_2} \dots \partial \zeta_{i_M}} e^{-1/2\zeta^T \zeta} \quad (10)$$

where  $(\zeta_{i_1}, \zeta_{i_2}, \dots, \zeta_{i_p})$  ( $\zeta$  is the vector form) are the standard normal random variables (SNVs). Consequently, Equation (9) is often written in a simple formulation as:

$$Y = a_0 + \sum_{i=1}^{\infty} a_i \Gamma_i(\zeta) \quad (11)$$

where  $a_i$  are the unknown expansion coefficients.

Previous studies have demonstrated that accurate approximations can be obtained through a truncated PCE with only low order terms (Lucas and Prinn, 2005; Li and Zhang, 2007; Shi et al., 2009; Zheng et al., 2011; Fan et al., 2015a,d; Xiao et al., 2015; El Mocayd et al., 2016). Therefore, if the degree of a truncated PCE is predefined as the highest order (denoted as  $p$ ) of the involved Hermite polynomials, the truncated PCE for  $M$  dimensional random variables can be expressed as:

$$Y \approx a_0 + \sum_{i=1}^{n-1} a_i \Gamma_i(\zeta) \quad (12)$$

where  $n = (M + p)!/M!p!$ .

The basic premise of the probabilistic collocation method is to let the polynomial chaos expansion(PCE) in terms of random inputs be the same as the model simulation results at selected collocation points. Collocation points can be specified by various algorithms. In this study, the algorithm proposed by Webster et al. (1996) is adopted, in which the collocation points are selected so that each SNV takes zero or one of the roots for the higher-order Hermite polynomial (Huang et al., 2007; Li and Zhang, 2007). For example, for the 2<sup>nd</sup>-order polynomial chaos expansion, the collocation points are combinations the values of  $-\sqrt{3}$ , 0 and  $\sqrt{3}$ , which are the roots of the 3<sup>rd</sup>-order Hermite polynomial  $H_3(\zeta) = \zeta^3 - 3\zeta$ . For the 3<sup>rd</sup>-order polynomial chaos expansion, the collocation points are chosen from zero and the roots of the 4-order Hermite polynomial  $H_4(\zeta) = \zeta^4 - 6\zeta^2 + 3$  (i.e.  $\pm\sqrt{3 \pm \sqrt{6}}$ ).

A Probabilistic collocation method (PCM) can be implemented through approximating a model output with a polynomial chaos expansion (PCE) in terms of random inputs (Zheng et al., 2011). The unknown coefficients contained in the expansion can be determined based on model simulations at selected collocation points (each collocation point is a realization of the random inputs). The number of collocation points is much larger than the number of unknown coefficients, leading to two main methods for estimating these unknown coefficients: linear equations and regression-based methods (Huang et al., 2007; Fan et al., 2015a,d). In this study, the regression-based method is employed in which all collocation points are used to form an over-determined equation system, and further generate the unknown coefficients through linear regression method.

### 2.3. Uncertainty quantification for the hydrological model based on a hybrid sequential data assimilation and probabilistic collocation method

Based on data assimilation through PF, the posterior probability distributions of model parameters can be quantified through a set of

random samples. Consequently, these obtained posterior distributions can hardly be quantified through some specific probability distributions (e.g. Gaussian, Gamma etc.). In comparison, the PCE model can merely be established to reflect uncertainty propagation in a dynamic system in terms of some specific distributions. For example, for the stochastic process decomposed by Hermite polynomials, the random inputs should be first expressed as standard normal random variables (SNVs). Consequently, if the PCE model is to be employed to explicitly express uncertainty evolution from the posterior distributions obtained by PF to the hydrologic outputs, those obtained posterior distributions would be firstly transformed to SNVs. Several methods have been developed to transform the non-Gaussian random variables to Gaussian random variables such as Box-Cox transformation (Box and Cox, 1964), Gaussian anamorphosis (GA) (Simon and Bertino, 2009), and log-sinh transformation (Wang et al., 2012). In this study, the GA method is adopted since it can directly transform any random variable into a SNV (see Appendix).

The process of the HSDAPC method involves two main components: parameter estimation through PF and uncertainty quantification through PCM. The detailed process includes the following steps (Fig. 1):

*Step (1).* Model state initialization: Initialize  $N_x$ -dimensional model state variables and parameters for  $ne$  samples:  $x_{t,i}^-$ ,  $i = 1, 2, \dots, ne$ ,  $x \in \mathbb{R}^{N_x}$ ,  $\theta_{t,i}$ ,  $\theta \in \mathbb{R}^{N_\theta}$ .

*Step (2).* Sample weight assignment: Assign the particle weights uniformly:  $w_{t,1} = 1/ne$ .

*Step (3).* Model state forecast step: Propagate the  $ne$  state variables and model parameters forward in time using model operator  $f$ :

$$\begin{aligned} x_{t+1,i} &= f(x_{t,i}, u_{t,i}, \theta_i) + \omega_t, \omega_t \\ &\sim N(0, Q_t), i \\ &= 1, 2, \&, ne \end{aligned}$$

where  $x_{t+1,i}$  is the forecasted value for particle  $i$  at time  $t + 1$ ,  $x_{t,i}$  and  $\theta_{t,i}$  are the values of state variables and parameters at time  $t$ .

*Step (4).* Observation simulation: Use the observation operator  $h$  to propagate the model state forecast:

$$\begin{aligned} y_{t+1,i} &= h(x_{t+1,i}, \theta_i) + v_{t+1}, v_{t+1} \\ &\sim N(0, R_{t+1}), i \\ &= 1, 2, \&, ne \end{aligned}$$

*Step (5).* Estimate the likelihood for the selected particles:

$$L(y_{t+1}|x_{t+1,i}, \theta_{t,i}) = \frac{1}{\sqrt{2\pi R_{t+1}}} \exp\left(-\frac{1}{2R_{t+1}} [y_{t+1} - h(x_{t+1,i}, \theta_{t,i})]^2\right)$$

$$p(y_{t+1}|x_{t+1,i}, \theta_{t,i}) = \frac{L(y_{t+1}|x_{t+1,i}, \theta_{t,i})}{\sum_{i=1}^{ne} L(y_{t+1}|x_{t+1,i}, \theta_{t,i})} = p(y_{t+1} - h(x_{t+1,i}, \theta_{t,i}))$$

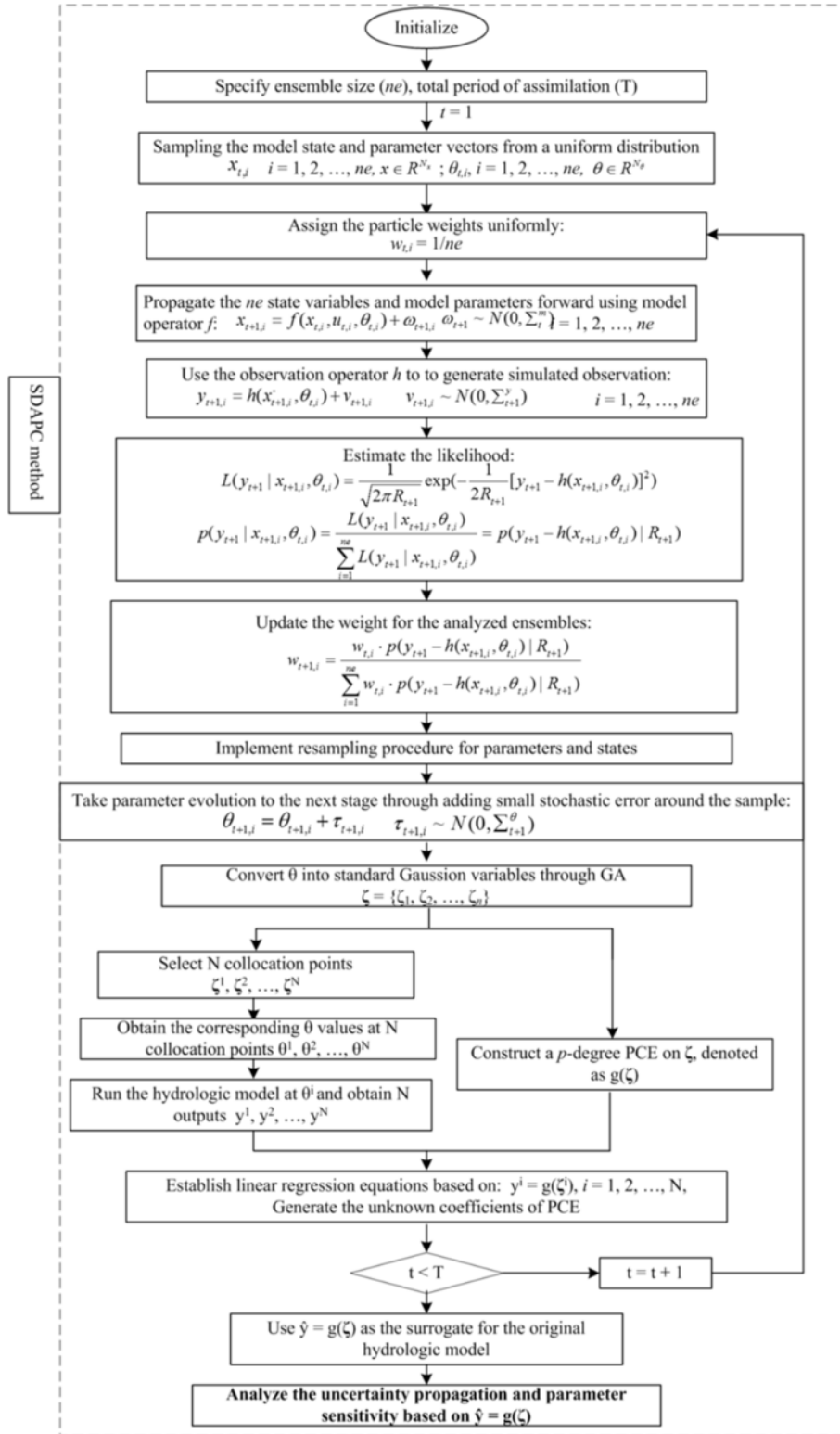


Fig. 1. The flowchart of the proposed SDAPC method.

Step (6). Obtain the updated weight  $w_{t+1,i}$  for the analyzed ensemble values:

$$w_{t+1,i} = \frac{w_{t,i} \cdot p(y_{t+1} - h(x_{t+1,i}, \theta_{t,i}) | R_{t+1})}{\sum_{i=1}^{ne} w_{t,i} \cdot p(y_{t+1} - h(x_{t+1,i}, \theta_{t,i}) | R_{t+1})}$$

Step (7). Resampling: Apply the resampling procedure proposed by Moradkhani et al. (2005b) for all states and parameters, and store the resulting particles as:  $\theta_{t+1-resamp,i} \times x_{t+1-resamp,i}$ .

Step (8). Parameter perturbation: Take the parameter evolution to the next stage by adding small stochastic errors around the sample:

$$\theta_{t+1,i} = \theta_{t+1-resamp,i} + \tau_{t+1,i}, \quad \tau_{t+1,i} \sim N(0, \eta S(\theta_{t,i}))$$

where  $\eta$  is a hyper-parameter which determines the radius around each sample being explored, which is set to be 0.15 in this study. Also, the hyper-parameter can be adjusted through variance multiplier approach (Leisenring and Moradkhani, 2012).  $S(\theta_{t+1,i})$  is the standard deviation of the analyzed particle values.

Step (9). Set  $w_{t+1,i} = 1/ne$ .

Step (10). Convert the parameter  $\theta$  into standard Gaussian variables through GA.

Step (11). Approximate the outputs of interest using the polynomial chaos expansion in terms of the standard Gaussian variables.

Step (12). Select the collocation points according to the dimensions of the stochastic vector and the order of the applied polynomial chaos expansion.

Step (13). Determine the unknown coefficients in the polynomial expansion through a statistical regression technique.

Step (14). Check the stopping criterion: if measurement data is still available in the next stage,  $t = t + 1$  and return to step 2; otherwise, continue to the next step.

Step (15). Evaluate the inherent statistical properties of the outputs stemming from the uncertainty of the parameters through the obtained PCE.

#### 2.4. PCE-based temporal dynamics of parameter sensitivity

Sensitivity analysis (SA) evaluates the impact of model parameters on the model outputs, and is therefore a convenient tool to assess model behavior and particularly the importance of certain parameterizations within the model (Reusser et al., 2011). Generally, SA is widely adopted in the model calibration process and attempts to identify the most important parameters for hydrologic model calibration and the unimportant parameters which can be prefixed as a constant value. Some objective functions are adopted for sensitivity analysis in hydrology, such as RMSE and NSE. In contrast to classical sensitivity analysis, the temporal dynamics of parameter sensitivity (TEDPAS) analyze the model output variables (such as discharge, groundwater level or snow water equivalent) to quantify which model components dominate the catchment response and can be considered as dominant indicators for functioning of the model (Reusser et al., 2011). Many SA methods as Sobol's method can be adopted for both classic SA and TEDPAS processes (Guse et al., 2014; Song et al., 2015). The main difference between these two processes is that SA is performed for each time step individually in TEDPAS, while classic SA is conducted only once over the simulation period.

In this study, Sobol's method is employed for the temporal dynamics of parameter sensitivity. Sobol's method is a global SA method derived from variance decomposition, attempting to quantify the contribution to the total variance of the model output by both an individual parameter and its interactions with other parameters (Dai et al., 2014). In Sobol's method, a variance decomposition from the random variable  $y$  can be formulated as (Zheng et al., 2011):

$$V(y) = \sum_i V_i + \sum_i \sum_{j>i} V_{ij} + \dots + V_{12\dots k} \quad (13)$$

where  $V_i$  is the variance attributed to the single effect of input  $x_i$  and  $V_i = V(E(y | x_i))$ ;  $V_{ij}$  is the variance attributed to the interaction effect of  $x_i$  and  $x_j$ , and  $V_{ij} = V(E(y | x_i, x_j)) - V(E(y | x_i)) - V(E(y | x_j))$ ; higher-order variances have similar expressions (Zheng et al., 2011; Wu et al., 2014). The Sobol's sensitivity indices are defined as the ratios of partial variances to the total variance, indicating the contribution of each individual parameter and its interactions to the total uncertainty (Dai et al., 2014):

$$S_i = \frac{V_i}{V(y)} \quad (14)$$

$$S_{ij} = \frac{V_{ij}}{V(y)} \quad (15)$$

$$S_{12\dots k} = \frac{V_{12\dots k}}{V(y)} \quad (16)$$

The total sensitivity index is defined as the sum of all partial sensitivity indices for a parameter and provides the total effect of the parameter, including the interactions (Dai et al., 2014):

$$S_i^T = S_i + \sum_{j \neq i} S_{ij} + \dots + S_{12\dots k} \quad (17)$$

The Sobol's indices are mathematically rigorous but time-consuming (Zheng et al., 2011). Consequently, Zheng et al. (2011) integrated PCE into Sobol's variance-decomposition and derived the total variance  $V(y)$  as:

$$V(y) = \sum_{i=1}^{\infty} a_i^2 \prod_{j=1}^{M_i} p_{ij}! \quad (18)$$

where  $p_{ij}!$  is the order of  $j$ th univariate Hermite polynomial (UHP). An example for the PCE-based sensitivity analysis indices are provided in Appendix.

#### 2.5. Temporal association between model sensitivity and catchment conditions

TEDPAS calculates the parameter sensitivity of the model output (discharge) for each time step, with the goal to determine the dominance of different model components for different periods (Reusser et al., 2011; Guse et al., 2014). Moreover, values of TEDPAS will

change over time since the dominant hydrologic process may vary temporally, e.g. between wet and dry periods (Wagener et al., 2003; Reusser et al., 2011; Guse et al., 2014; Rahmani and Zarghami, 2015). Consequently, the values of TEDPAS may be highly correlated with the catchment conditions (e.g. precipitation, discharge). Such correlation may be nonlinear due to the nonlinear behavior of hydrologic systems.

To identify the correlation between model sensitivity and catchment conditions, the maximal information coefficient (MIC) proposed by Reshef et al. (2011) will be adopted. MIC captures a wide range of associations both functional and non-functional, and for linear relationships it provides a score that approximately equals the coefficient of determination ( $R^2$ ) (Reshef et al., 2011). The expression of MIC can be formulated as (Zhang et al., 2014):

$$MIC = \max \left\{ I(x, y) / \log_2 \min(n_x, n_y) \right\} \quad (19)$$

where

$$I(x, y) = H(x) + H(y) - H(x, y)$$

$$= \sum_{i=1}^{n_x} p(x_i) \log_2 \frac{1}{p(x_i)} + \sum_{j=1}^{n_y} p(y_j) \log_2 \frac{1}{p(y_j)}$$

$$- \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} p(x_i, y_j) \log_2 \frac{1}{p(x_i, y_j)}$$

$$n_x n_y < B(n) \text{ and } B(n) = n^{0.6}$$

Based on MIC, the nonlinearity between two variables  $x$  and  $y$  can be defined by  $MIC - R_{xy}^2$  (Reshef et al., 2011), where  $R_{xy}$  is the Pearson's  $r$  value between  $x$  and  $y$ . Through MIC values, the associations between a parameter's sensitivity and catchment conditions is identified, which can answer (i) which factor (e.g. precipitation, potential evapotranspiration) poses the most significant impact on parameter sensitivity, and (ii) does the association between parameter sensitivity and the catchment factor change under different hydro-meteorological conditions (e.g. wet and dry periods for precipitation)?

### 3. Case study

#### 3.1. Study catchment

The Xiangxi River basin, located in the Three Gorges Reservoir area of China (as shown in Fig. 2), was selected to demonstrate the applicability of the proposed method. The Xiangxi River is located between  $30.96$  and  $31.67^\circ\text{N}$  and  $110.47$ – $111.13^\circ\text{E}$  in the Hubei area of China, Three Gorges Reservoir (TGR) region, which drains an area of about  $3200 \text{ km}^2$ . The Xiangxi River originates in the Shennongjia Nature Reserve, with a main stream length of  $94 \text{ km}$  and a catchment area of  $3099 \text{ km}^2$ , and is one of the main tributaries of the Yangtze River (Han et al., 2014; Fan et al., 2015b,c). The watershed experiences a northern subtropical climate. The annual average temperature in this region is  $15.6^\circ\text{C}$  and ranges from  $12^\circ\text{C}$  to  $20^\circ\text{C}$ . The historical measurements record a maximum temperature of  $43.1^\circ\text{C}$  and a minimum temperature of  $-9.3^\circ\text{C}$  (Fan et al., 2016). The average frost-free days at low, middle and high elevations are 272, 215 and 163 days, respectively (Li, 2012). The average annual quantity of solar radiation value (heat units) is  $2.90 \times 10^8 \text{ kW/m}^2$ , with values during April to September reaching as high as  $1.88 \times 10^8 \text{ kW/m}^2$  (Li, 2012; Li et al., 2015). Annual precipitation is  $1100 \text{ mm}$ , ranging from  $670$  to  $1700 \text{ mm}$  with considerable spatial and temporal variability. The major rainfall season is May–September, with a flooding season from July to August. The precipitation in the north is higher than that in the south, with an average of sixty-nine percent of the annual precipitation occurring as rain, resulting in an average streamflow of  $40.18 \text{ m}^3/\text{s}$  (Li, 2012) in the Xiangxi River.

The study area consists of a mixed coniferous-deciduous forest which demonstrates an explicit vertical gradient with elevation. Vegetation changes from broadleaf forest (below  $800 \text{ m}$ ) to coniferous forest ( $800$ – $1000 \text{ m}$ ) and shrub-grassland (above  $1800 \text{ m}$ ) as elevation increases (Li, 2012). The land use is characterized by mixed grain and cash-crop farming on terraced farmland. Crops include rape, wheat, maize and rice, nuts, and garden fruits respectively (Li, 2012).

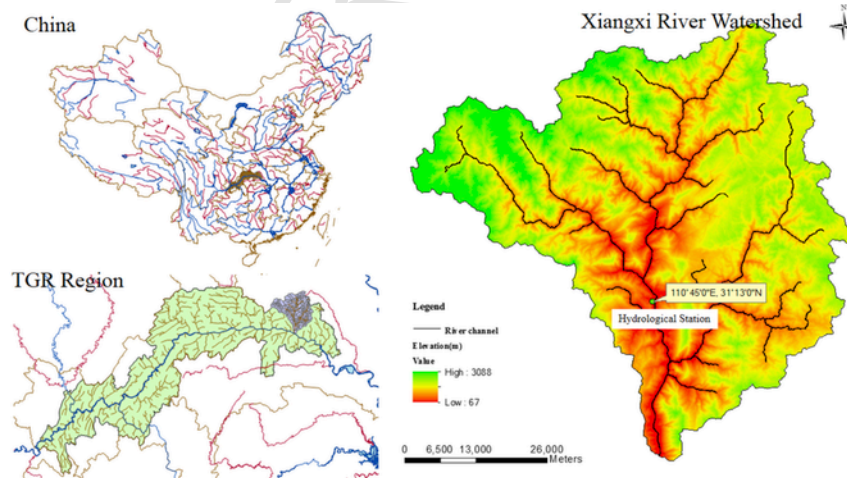


Fig. 2. The location of the studied watershed.

### 3.2. Hydrologic model

Hymod, which is a well-known conceptual hydrologic model, will be used in this study. Hymod is a non-linear rainfall-runoff conceptual model run in any time scale (Moor, 1985). The general concept of the model is based on a probability distribution of soil moisture modeling proposed by Moore (1985, 2007). In Hymod the catchment is considered as an infinite amount of points, each of which has a certain soil moisture capacity denoted as  $c$  [L] (Wang et al., 2009). Soil moisture capacities vary within the catchment due to spatial variability such as soil type and depth and a cumulative distribution function (CDF) is proposed to describe such variability, expressed as (Moor, 1985, 2007):

$$F(c) = 1 - \left[ 1 - \frac{c}{C_{\max}} \right]^{b_{\exp}}, \quad 0 \leq c \leq C_{\max} \quad (20)$$

where  $C_{\max}$  [L] is the maximum soil moisture capacity within the catchment and  $b_{\exp}$  [-] is the degree of spatial variability of soil moisture capacities and affects the shape of the CDF.

As shown in Fig. 3, Hymod conceptualizes the rainfall-runoff process through a nonlinear rainfall excess model connected with two series of reservoirs (three identical quick-flow tanks representing the surface flow, in parallel with a slow-flow tank representing the groundwater flow). The Hymod has five parameters to be calibrated: (i) the maximum storage capacity in the catchment  $C_{\max}$ ; (ii) the degree of spatial variability of the soil moisture capacity within the catchment  $b_{\exp}$ ; (iii) the factor partitioning the flow between the two series of linear reservoir tanks  $\alpha$ ; (iv) the residence time of the linear quick-flow tank  $R_q$ ; and (v) the residence time of the slow-flow tank  $R_s$ . The model uses two input variables: mean areal precipitation,  $P$  (mm/day), and potential evapotranspiration,  $ET$  (mm/day).

### 3.3. Synthetic experiment setup

A Synthetic experiment is to be proposed to demonstrate the effectiveness of the proposed HDAPC approach in uncertainty quantification. In this synthetic experiment, the synthetic streamflow data are generated based on the potential evapotranspiration,  $ET$  (mm/day), and mean areal precipitation,  $P$  (mm/day) from the Xiangxi River basin. In the generation process of the synthetic streamflow, the values of the five parameters of Hymod would be predefined, as given in Table 1. These generated streamflow values are considered as the "true" observations in the updating process of particle filter (PF). In any data assimilation framework, one must assume error values for any quantity that contains uncertainties (DeChant and Moradkhani,

2012). In this study, random perturbations would be added to the precipitation, potential evapotranspiration (ET) and model predictions to account for their uncertainties. For the potential evapotranspiration, the Gaussian noise is recommended by a number of literature (e.g. DeChant and Moradkhani, 2012; Moradkhani et al., 2012; Chen et al., 2013; Rasmussen et al., 2015; Yang et al., 2015; Nourani et al., 2015). For precipitation, some studies applied Gaussian noise (e.g. Rasmussen et al., 2015), while other studies concluded that log-normal noise may perform better (e.g. DeChant and Moradkhani, 2012; Moradkhani et al., 2012). In this study, the Gaussian noise is adopted in the synthetic experiment, while log-normal noise is employed in the real-case study. These random perturbations are assumed to have their standard errors being proportional to the magnitude of the true values. The proportional coefficients for precipitation, potential evapotranspiration, streamflow observation and model predictions, are set to be 0.3 in the synthetic experiment. This means that precipitation, PET, streamflow observation, and model predictions are assumed to have normal distributions with relative errors of 30%.

### 3.4. Real-case study

To demonstrate the applicability of the proposed method in quantifying uncertainty for hydrologic models, the daily precipitation, potential evapotranspiration, and streamflow measurements from 1994 to 1996 were applied to evaluate the performance of the proposed algorithm. The uncertainty of model parameters were first characterized through the PF method.

The sequential data assimilation approach can quantify model prediction errors stemming from various sources such as inputs, model structures and parameters. To account for actual uncertainties in real world climatic variables such as precipitation and potential evapotranspiration, random perturbations are usually added into real measurements in climatic variables and discharges. Several studies have assumed the standard deviation of the observed errors to be proportional to the true discharge (Dechant and Moradkhani, 2012; Moradkhani et al., 2012; Abaza et al., 2014), while some studies set the errors to be proportional to the log discharge (Clark et al., 2008; McMillan et al., 2013). In our study, we primarily focus on the capability of the developed HSDAPC approach to quantify the uncertainty of hydrologic models. Consequently, the observed errors in measurements are assumed to be independent and set to be proportional to the true values. In detail, according to the study by Fan et al. (2015d), the precipitation is assumed to be log-normally distributed with a relative error of 20% of the true values. The ET is normally distributed with a relative error of 20%, respectively. The error in streamflow is assumed to be normally distributed, with a proportional rate of 20% of the true discharges.

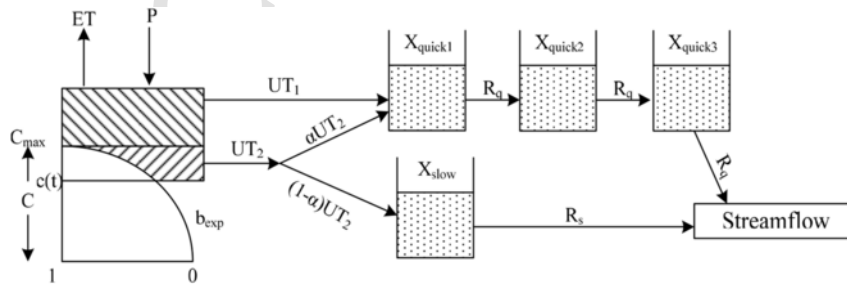


Fig. 3. Description of Hymod.



**Table 1**

The predefined true values and fluctuating ranges for the parameters of Hymod through PF.

	Parameters				
	$C_{max}$	$b_{exp}$	$\alpha$	$R_s$	$R_q$
True	175.40	1.5	0.46	0.11	0.82
Primary range	[120, 250]	[0, 2]	[0.20, 0.70]	[0.05, 0.20]	[0.60, 0.99]

## 4. Result analysis and discussion

### 4.1. Result analysis of synthetic experiment

#### 4.1.1. Performance of the HSDAPC approach

In the proposed HSDAPC approach, the particle filter method recursively updates the posterior probabilities of model parameters and state variables through representative samples when new observations become available; the probabilistic collocation method (PCM) is proposed after each update to explicitly describe uncertainty propagation from model parameters to model outputs. However, one major challenge for connecting PF and PCM is that the posterior probabilities of model parameters obtained by PF are arbitrary, while the PCM requires inputs with specific probability distributions (e.g. standard normal variable for PCE with Hermite polynomials). Consequently, the GA approach is employed to convert the posterior distributions into standard normal variables.

In the synthetic experiment, four sample size scenarios, involving {50, 100, 200, 500} particles are adopted to evaluate the performance of HSDAPC. Under each scenario the synthetic experiment is tested for 30 trials to ensure the reliability of the proposed approach. For each sample size, the prior probabilities of model parameters are assumed to follow uniform distribution within predefined intervals, as presented in Table 1. The posterior probabilities of those five model parameters are estimated through particle filter, and the associated 2nd and 3<sup>rd</sup>-order PCE models are derived to represent the hydrologic model. The average values of NSE and RMSE for these 30 trials are listed in Table 2. The results suggest that, as the sample size increases, the performance of particle filter method is improved, leading to improvement for the following PCE models. Moreover, the variations of the NSE, RMSE and PBIAS values for Hymod and the associated 2nd and 3<sup>rd</sup>-order PCE models for the 30 trials are plotted in Fig. 4. This indicates that the performance of HSDAPC approach is mainly determined by the posteriors estimation through PF. Once the parameter posterior probabilities are obtained, the differences between the PCE models and the original hydrologic model can even be neglected, indicating the accuracy for the 2nd and 3<sup>rd</sup>-order PCE models in representing the original hydrological model. Consequently, the generated PCE models can represent the original hydro-

**Table 2**

Comparison between the performance of Hymod and two and three-order PCEs under different sample size scenarios of data assimilation process.

Sample size	NSE		RMSE			
	Hymod	2 <sup>th</sup> -order PCE	3 <sup>th</sup> -order PCE	Hymod	2 <sup>th</sup> -order PCE	3 <sup>th</sup> -order PCE
50	0.8564	0.8566	0.8561	12.6264	12.6055	12.6416
100	0.8831	0.8840	0.8829	10.7984	10.7595	10.8041
200	0.8878	0.8875	0.8829	11.0099	11.0180	11.0083
500	0.8903	0.8902	0.8900	10.5155	10.5147	10.5372

logic model for further uncertainty propagation and sensitivity analysis.

#### 4.1.2. Comparison with Monte Carlo simulation

To demonstrate the accuracy of the obtained 2nd and 3<sup>rd</sup>-order PCE models from SDAPC, the inherent probabilistic characteristics of the predictions from PCEs and Monte Carlo approach are compared. In detail, the streamflow predictions in six days (i.e. 23, 145, 181, 182, 218, 350) are chosen, which cover low, medium and high flow conditions. Their inherent statistical properties and the corresponding histograms are obtained through MC and the 2nd and 3<sup>rd</sup>-order PCE models. Table 3 lists the mean, standard deviation, kurtosis and skewness values of uncertainty predictions from MC simulation and the 2nd and 3<sup>rd</sup>-order PCEs. The results show that the probability functions generated by the 2nd and 3<sup>rd</sup>-order PCEs would be similar to those probability functions generated by MC. Fig. 5 shows the histograms of the 2nd and 3<sup>rd</sup>-order PCEs and MC simulation results at the selected time periods. It also indicates that the probability distributions obtained by the 2nd and 3<sup>rd</sup>-order PCEs have similar shapes with the "true" distributions from the MC simulation method.

### 4.2. Results analysis of real case study

#### 4.2.1. Uncertainty characterization of hydrologic model through HSDAPC

In the uncertainty quantification process through HSDAPC for the Xiangxi River, the prior probability of each model parameters are assumed to be uniformly distributed within an interval. The five predefined intervals for Hymod are obtained based on the calibrated parameter values through SCE-UA algorithm. The sample size is set to be 500 in the real case study, based on the results of synthetic experiment. The posterior probabilities of model parameters in the hydrologic model are first estimated through the PF approach based on available observations. The uncertainty of model predictions, stemming from uncertain model parameters, is then characterized through the probabilistic collocation method. In detail, as presented in Fig. 1, once the observed discharge is available, the posterior probabilities of the model parameters are updated, and then the coefficients of the obtained PCE model are changed. In this study, the 2nd and 3<sup>rd</sup>-order PCE models are employed, and so there are 21 and 56 coefficients for these two models, as presented in Table 4. For instance, Fig. 6 shows the variability of posterior probabilities in different temporal periods. Correspondingly, Table 4 presents the detailed 2nd and 3<sup>rd</sup>-order PCE models and provides the corresponding coefficients on 100<sup>th</sup> and 200<sup>th</sup> day. The results show that, even though 3<sup>rd</sup>-order PCE model has more polynomials, the coefficients of its low order polynomials (1st and 2<sup>nd</sup>-order) have similar values with the coefficients of the 2<sup>nd</sup>-order PCE, while in comparison most coefficients of the 3<sup>rd</sup>-order polynomials are zero. This suggests that, for the Hymod, 2<sup>nd</sup>-order PCE may be adequate to serve as the proxy. Moreover, the proposed HSDAPC approach improves upon previous methods where random samples are drawn directly from the posterior distributions through the inefficient Monte Carlo method. In comparison, a proxy model of the original hydrologic model is established through PCM with respect to the uncertain model parameters. Such a proxy model can explicitly reflect the uncertainty propagation from uncertain parameters to uncertain outputs and thus can be employed to reveal the prediction uncertainty of the hydrologic model.

Fig. 7 shows the comparison between predictions of the hydrologic model, PCEs and observations. The polynomial chaos expansions (PCEs) are the proxy models for the original hydrologic model obtained by PCM. Fig. 7(a) indicates the comparison between the

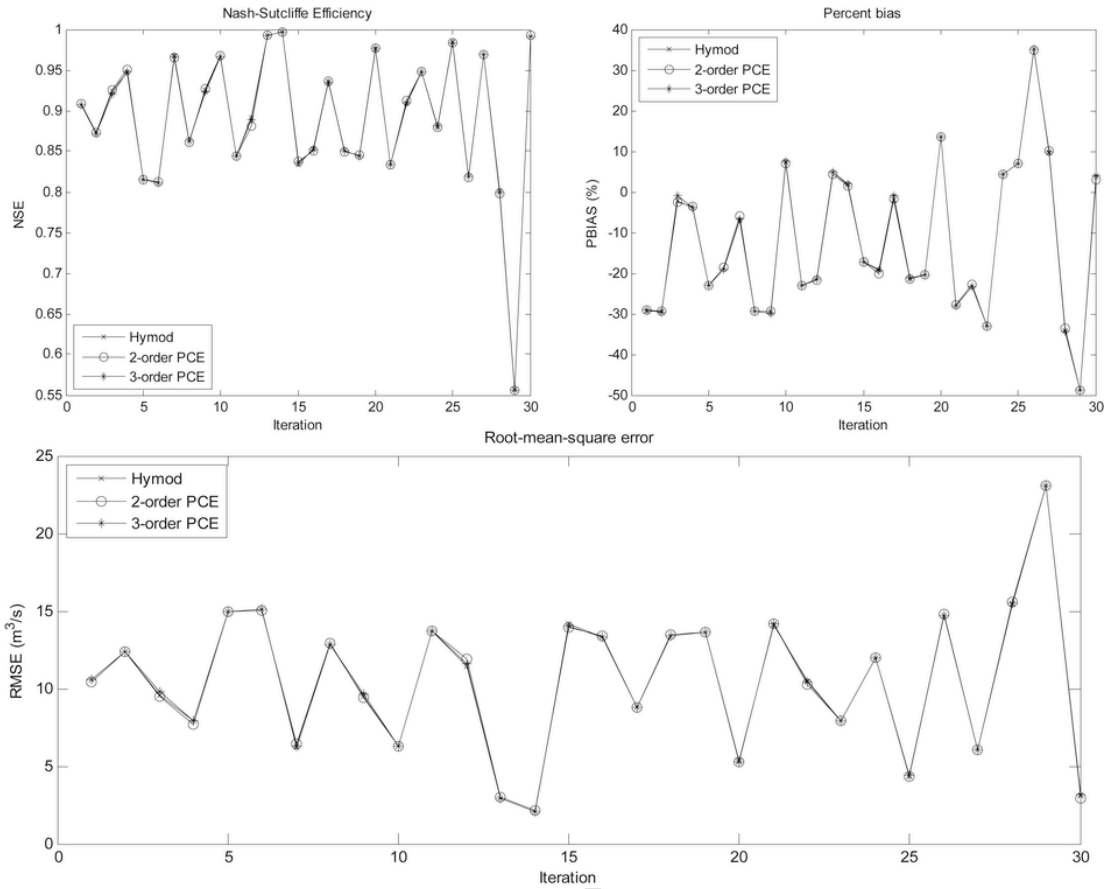


Fig. 4. Performance comparison between hydrologic model and the associated PCE models.

Table 3

Comparison of statistic characteristics among two and three-order PCEs and MC simulation results at specific time periods.

Time (d)	Mean			Standard deviation			Kurtosis			Skewness		
	2-Order PCE	3-Order PCE	MC	2-Order PCE	3-Order PCE	MC	2-Order PCE	3-Order PCE	MC	2-Order PCE	3-Order PCE	MC
23	1.67	1.66	1.67	0.18	0.20	0.18	3.55	3.87	3.26	0.76	0.74	0.64
145	111.15	110.25	110.95	12.42	13.26	12.46	3.47	3.77	3.31	0.70	0.68	0.62
181	336.92	334.41	336.37	34.27	36.37	34.43	3.43	3.71	3.27	0.67	0.64	0.59
182	289.41	287.68	288.33	19.85	25.10	21.13	3.70	3.98	2.91	0.80	0.72	0.56
218	159.18	158.12	158.72	11.37	13.54	11.84	3.42	3.80	2.91	0.65	0.63	0.49
350	0.01	0.01	0.01	0.0012	0.0014	0.0012	3.73	3.88	3.45	0.71	0.68	0.63

mean predictions of the hydrologic model and observations. Comparisons between the mean predictions of 2<sup>nd</sup>-order PCE and observations and the mean predictions of 3<sup>rd</sup>-order PCE and observations are presented in Fig. 7(b) and (c). From Fig. 7, the predictions from the original hydrologic model and the two and three-order PCEs exhibit no significant differences. All three approaches can trace the variations in observed streamflow data while the underestimates or overestimates in the original hydrologic model will also lead to similar deviations in the two and three-order PCE models.

To further compare the performance of the hydrologic model and PCEs in discharge predictions, the indices of RMSE, PBIAS, NSE were calculated based on the prediction means and observations. The comparison process between the Monte Carlo method and PCEs was implemented through: (i) choosing *N* samples from the standard Gaussian distribution, (ii) deriving the associated parameter values in

the hydrologic model through the GA approach, (iii) running the PCEs and hydrologic model respectively, (iv) obtaining the evaluation criteria results. Table 5 shows the results of RMSE, PBIAS, and NSE values obtained through the Monte Carlo method and 2<sup>nd</sup> and 3<sup>rd</sup>-order PCEs. These values indicate a satisfactory performance for the hydrologic model and its corresponding 2<sup>nd</sup> and 3<sup>rd</sup>-order PCE models in tracking the streamflow dynamics in the Xiangxi River. Specifically, there are no obvious differences in the performance of the Hymod and the 2<sup>nd</sup> and 3<sup>rd</sup>-order PCEs in predicting the streamflow. Although the two and three-order PCEs are established as proxy models for the original Hymod, the results in Table 5 suggest the proxy models based on the two and three-order PCEs can well represent the original Hymod. The 90% predictive intervals obtained by the original hydrologic model and the 2<sup>nd</sup> and 3<sup>rd</sup>-order PCEs are presented in Fig. 8. These predictive intervals are obtained based on

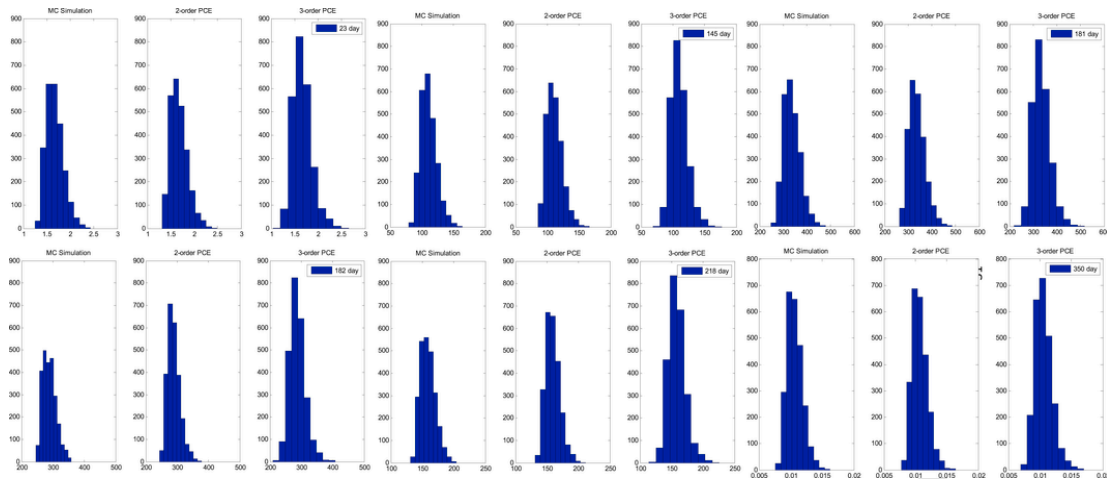


Fig. 5. The comparison of histograms among MC simulation, 2-order and 3-order PCE results.

the posterior probabilities of model parameters at each time step. This meaning that the parameters are time-variant. The results indicate the 90% predictive intervals from all three models can bracket most observations, especially for high flow periods. Moreover, the predictive intervals obtained by the 2nd and 3<sup>rd</sup>-order PCE models show consistent varying trends with the predictive interval obtained by the original hydrological model, indicating the accuracy of the 2nd and 3<sup>rd</sup>-order PCE models in quantifying uncertainty in the hydrologic model.

#### 4.2.2. Computational efficiency of the HSDAPC method

The basis of the hybrid sequential data assimilation and probabilistic collocation (HSDAPC) approach for quantifying the uncertainty of hydrologic models is to estimate the posterior distributions of the hydrologic model parameters through a PF method and then reveal the uncertainty of hydrologic models through the probabilistic collocation method (PCM). Previously, once the posterior distributions were obtained through PF, the Monte Carlo method was employed to draw samples and run the hydrologic model again to generate prediction uncertainty ranges. In the HSDAPC approach, such uncertainty quantification can be conducted through the obtained PCE models of the hydrologic model. This method has three advantages: (i) the original samples can be drawn from the standard Gaussian distribution, which is easily conducted; (ii) the computational efficiency can be highly improved; (iii) the inherent relationship between model parameters and model outputs can be explicitly expressed by the PCE models.

The first advantage of the HSDAPC approach is straightforward since the inputs of the PCE models with Hermite polynomials obey the standard Gaussian distribution. The second advantage is illustrated through comparison between the traditional Monte Carlo method and the PCE models. Table 1 shows the computational efficiency and the associated performance of the Monte Carlo method and PCEs. In this study, five sample sizes ( $n = 500, 1,000, 1,500, 2,000, 2500$ ) are selected to compare the computational efficiency of the hydrologic model and PCEs. As shown in Table 5, as the sample size increases, the performance of the hydrologic model and PCEs does not vary significantly. However, the computational efficiency of PCEs would be much faster than the Monte Carlo (MC) method. Specifically, the computational efficiency of 2<sup>nd</sup>-order PCE model would be ten times faster than the original hydrologic model. For example, when  $n = 500$ , the computational time of MC method would be 60.01 (s), while the computational time of two-order PCE is just 2.85 (s). For the 3<sup>rd</sup>-order PCE, its computational efficiency is also

more efficient than the original hydrologic model. Consequently, the proposed HSDAPC approach can significantly improve the computational efficiency for uncertainty quantification of hydrologic models.

In this study, Hymod was applied to demonstrate the efficiency of the proposed approach. This model is a simple conceptual hydrologic model, with five parameters to be calibrated. Consequently, the computational requirement for this model is quite low, when compared with other sophisticated models such as semi-distributed and distributed hydrologic models. However, as presented in Table 5, the obtained 2nd and 3<sup>rd</sup>-order PCE models are much faster in computational efficiency when compared with Hymod. Consequently, the computational efficiency would be improved significantly for more complex hydrologic models.

#### 4.2.3. Temporal dynamics of parameter sensitivity (TEDPAS)

Based on the HSDAPC approach, the posterior probability distributions of model parameters were obtained based on three years of measurements for the Xiangxi River. The two and three-order PCE models were further derived to characterize the uncertainty propagation from model parameters to model forecasts. Even though the test model is a simple conceptual model having only five parameters, a temporal diagnostic analysis for Hymod is useful to characterize which model components control the performance under different hydrologic conditions, and to further explore the dominant runoff mechanisms under different hydro-meteorological conditions.

As expressed by Equations (A9)–(A11), Sobol's indices can be theoretically approximated through the PCE model. Based on the HSDAPC approach, the PCE model was generated to serve as a proxy for the original hydrologic model. The performance of the PCE was consistent with the original hydrologic model. Consequently, Sobol's indices can be used to analyze the temporal dynamics of parameter sensitivity. Figs. 9 and 10 present the temporal dynamics of parameter sensitivity obtained from the 2<sup>nd</sup>-order and 3<sup>rd</sup>-order PCE respectively, across the simulation period. The results in Figs. 5 and 6 show a consistent trend between each other due to the accurate approximation of the two- and three-order PCEs for the original hydrologic model.

From Figs. 9 and 10, the dominant model components can be characterized under different hydro-meteorological conditions. As shown in Fig. 9, the parameter alpha, in general, posed less influence on model performance over the simulation period, with its values of less than 0.2. This means that the distribution factor of water flowing

**Table 4**  
The coefficients of 2<sup>nd</sup> and 3<sup>rd</sup>-order PCE model under different periods.

Polynomials	2 <sup>nd</sup> -order PCE		3 <sup>rd</sup> -order PCE	
	100 (day)	200 (day)	100 (day)	200 (day)
1	41.64	28.96	41.61	28.96
$\xi_1$	-0.36	-0.09	-0.32	-0.08
$\xi_2$	0.34	0.08	0.32	0.07
$\xi_3$	0.33	0.11	0.29	0.10
$\xi_4$	0.06	0.18	0.06	0.16
$\xi_5$	0.30	-0.82	0.28	-0.77
$\xi_1^2 - 1$	0.00	0.00	0.00	0.00
$\xi_2^2 - 1$	0.01	0.00	0.01	0.00
$\xi_3^2 - 1$	-0.01	0.00	0.00	0.00
$\xi_4^2 - 1$	0.00	-0.01	0.00	-0.01
$\xi_5^2 - 1$	0.00	-0.03	-0.03	0.03
$\xi_1 \xi_2$	0.00	0.00	0.00	0.00
$\xi_1 \xi_3$	0.00	0.00	0.00	0.00
$\xi_1 \xi_4$	0.00	0.00	0.00	0.00
$\xi_1 \xi_5$	0.00	0.01	0.00	0.00
$\xi_2 \xi_3$	0.00	0.00	0.00	0.00
$\xi_2 \xi_4$	0.00	0.00	0.00	0.00
$\xi_2 \xi_5$	0.00	-0.01	0.00	0.00
$\xi_3 \xi_4$	0.00	0.00	0.00	0.00
$\xi_3 \xi_5$	0.00	-0.01	0.00	0.00
$\xi_4 \xi_5$	0.00	0.00	0.00	0.00
$\xi_1^3 - 3\xi_1$			0.01	0.00
$\xi_2^3 - 3\xi_2$			-0.01	0.00
$\xi_3^3 - 3\xi_3$			0.00	0.00
$\xi_4^3 - 3\xi_4$			0.00	0.00
$\xi_5^3 - 3\xi_5$			0.00	0.01
$\xi_1(\xi_2^2 - 1)$			0.00	0.00
$\xi_1(\xi_3^2 - 1)$			0.00	0.00
$\xi_1(\xi_4^2 - 1)$			0.00	0.00
$\xi_1(\xi_5^2 - 1)$			0.00	0.01
$\xi_2(\xi_1^2 - 1)$			0.00	0.00
$\xi_2(\xi_3^2 - 1)$			0.00	0.00
$\xi_2(\xi_4^2 - 1)$			0.00	-0.01
$\xi_2(\xi_5^2 - 1)$			0.00	0.00
$\xi_3(\xi_1^2 - 1)$			0.00	-0.01
$\xi_3(\xi_2^2 - 1)$			0.00	0.00
$\xi_3(\xi_4^2 - 1)$			0.00	0.00
$\xi_3(\xi_5^2 - 1)$			0.00	0.00
$\xi_4(\xi_1^2 - 1)$			0.00	0.00
$\xi_4(\xi_2^2 - 1)$			0.00	0.00
$\xi_4(\xi_3^2 - 1)$			0.00	0.00
$\xi_4(\xi_5^2 - 1)$			0.00	0.00
$\xi_5(\xi_1^2 - 1)$			0.00	0.00
$\xi_5(\xi_2^2 - 1)$			0.00	0.00
$\xi_5(\xi_3^2 - 1)$			0.00	0.00
$\xi_5(\xi_4^2 - 1)$			0.00	0.00
$\xi_1 \xi_2 \xi_3$			0.00	0.00
$\xi_1 \xi_2 \xi_4$			0.00	0.00
$\xi_1 \xi_2 \xi_5$			0.00	0.00
$\xi_1 \xi_3 \xi_4$			0.00	0.00
$\xi_1 \xi_3 \xi_5$			0.00	0.00
$\xi_1 \xi_4 \xi_5$			0.00	0.00
$\xi_2 \xi_3 \xi_4$			0.00	0.00
$\xi_2 \xi_3 \xi_5$			0.00	0.00
$\xi_2 \xi_4 \xi_5$			0.00	0.00
$\xi_3 \xi_4 \xi_5$			0.00	0.00

to the quick-flow reservoir would not influence the model performance significantly, regardless of the variation in actual hydro-meteorological conditions. Conversely, the parameters  $R_s$  and  $R_q$  exhibit significant impacts on the model streamflow predictions. Specifically, the values of  $R_s$  and  $R_q$  show significant fluctuation, with the minimum value being zero and maximum value approaching one. Furthermore, the temporal sensitivity of  $R_q$  is generally consistent with the variation of precipitation, while the value of  $R_s$  shows an op-

posite trend with the precipitation. This is because  $R_q$  indicates the residence time of quick-flow reservoirs, while the  $R_s$  represents the residence time of slow-flow reservoirs. For  $C_{max}$  and  $b_{exp}$ , they exhibit similar variation trends between each other since they are indicators for soil moisture but in general,  $C_{max}$  plays a more importance role in model performance.

Table 6 presents the sensitivities of the five parameters in low and high flow periods. In this study, the high flow is characterized as a flow amount larger than the 95% quantile value of the historical flow records while the low flow is less than the 5% quantile value of the historical records. The results in Table 6 mean that the parameter  $R_q$  pose most significant impact on the model predictions in high flow period, while the parameter  $R_s$  is the dominant impact factor for model predictions in low flow period. This suggest that the quick flow process contributes most in high flow periods but the slow flow process dominates in low flow periods. Particularly, the results in Table 6 show that there is only slight differences for parameter sensitivities obtained by 2<sup>nd</sup>-order and 3<sup>rd</sup>-order PCE models, indicating the accuracy for these two models in reflecting uncertainty propagation in hydrologic simulation.

Furthermore, the differences of parameter sensitivities under precipitation and non-precipitation conditions are characterized. Table 7 shows the parameter sensitivities under dry and rainy conditions. The results indicate that the quick flow process (i.e.  $R_q$ ) contribute most to the streamflow when precipitation occurs, followed by the maximum soil storage (i.e.  $C_{max}$ ) and the slow flow (i.e.  $R_s$ ). Conversely, the parameter  $R_s$  dominates during the dry periods with a parameter sensitivity value more than 0.4. The parameters of  $R_q$  and  $C_{max}$  also pose apparent impacts on the model predictions with the parameter sensitivities more than 0.3 and 0.1, respectively.

Fig. 11 shows the comparison of the cumulative probability for the TEDPAS obtained through 2<sup>nd</sup>-order and 3<sup>rd</sup>-order PCEs. The CDFs for the five parameters obtained by two PCEs show similar trends, indicating the accuracy of the proposed method for quantifying the temporal sensitivities of the hydrologic model. Similar to the results presented in Figs. 9 and 10, the parameter  $\alpha$  in Hymod shows the least sensitivity to the model performance, with the maximum value less than 0.2, followed by  $C_{max}$  and  $b_{exp}$ , with their values fluctuating within [0, 0.6] and [0, 0.5].  $R_q$  had a sensitivity higher than 0.4 for more than 40% of the days, illustrating that it has a high sensitivity for a considerable period of time. Moreover,  $R_s$  also shows high sensitivity for a considerable period since it exhibits a high sensitivity higher than 0.4 for nearly 40% of days in the simulation period.

As shown in Figs. 9 and 10, the model parameter sensitivities will vary over time. One issue to characterize the variation in model parameter sensitivity is to determine how catchment conditions will influence the model parameter sensitivity. Fig. 12 shows the correlation between model inputs and outputs (i.e. precipitation, potential evapotranspiration, and streamflow) and model parameter sensitivities. The results show that the association between model inputs and outputs and model parameter sensitivity is not fully clear. Consequently, to further reveal such associations, the MIC will be adopted which is expressed as Equation (19). Table 8 presents correlation between precipitation, potential evapotranspiration, observed discharge and model parameter sensitivities. It can be concluded that, over the simulation period, the model parameter sensitivities are correlated with the precipitation, potential evapotranspiration and streamflow discharge, with most MIC values being larger than 0.1. Moreover, during the rainy periods, the correlation between model parameter sensitivities and catchment meteorological conditions is quite significant, with most MIC values larger than 0.5. Particularly, such correlation/

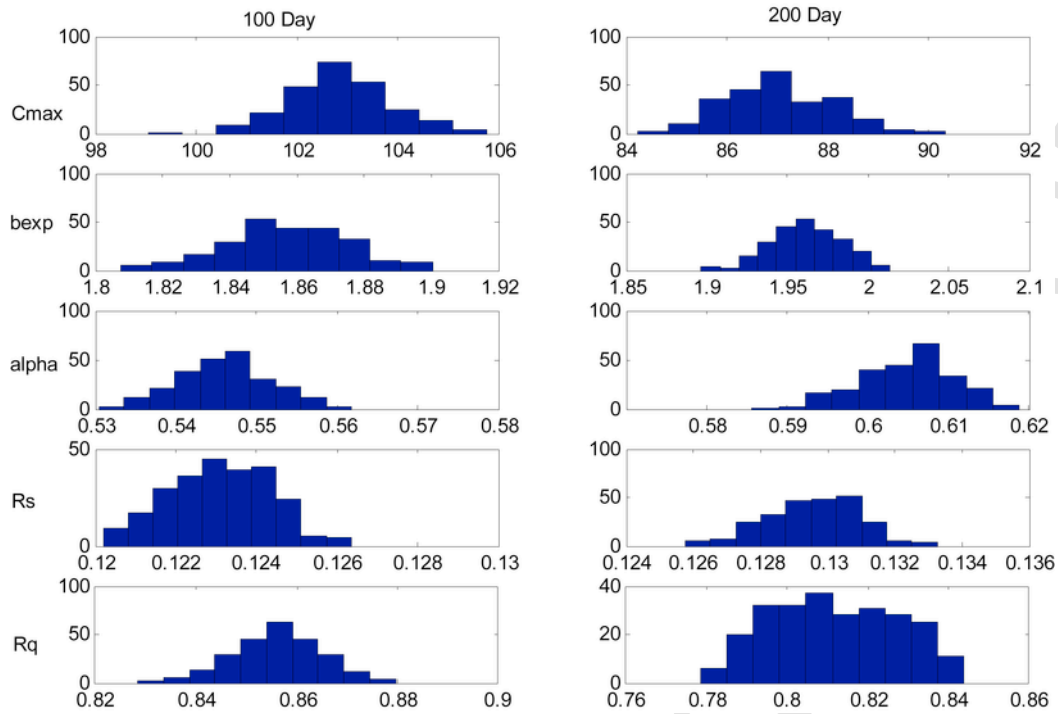


Fig. 6. Variability of posterior probabilities under different temporal periods.

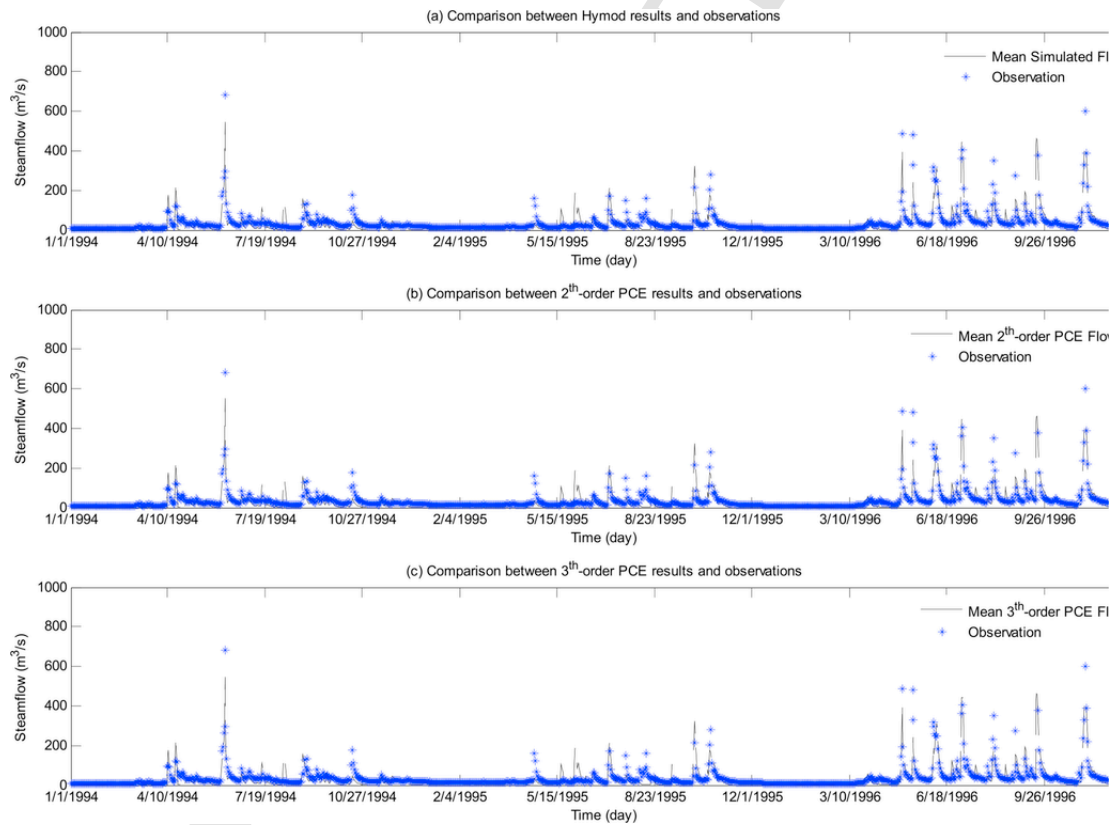


Fig. 7. Comparison between the prediction means and observations: (a) hydrologic model predictions vs. observations, (b) 2<sup>nd</sup>-order PCE results vs. observations, (c) 3<sup>rd</sup>-order PCE results vs. observations.

**Table 5**  
Comparison between Monte Carlo method and PCEs.

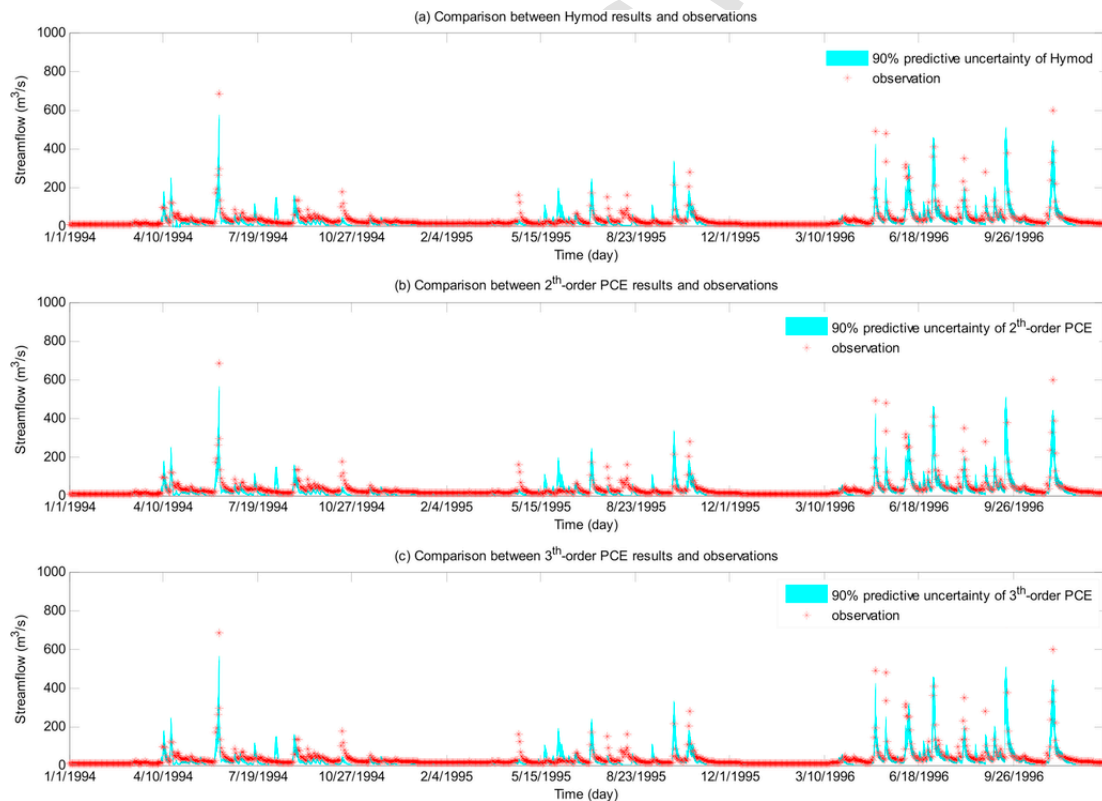
Sample size		500	1000	1500	2000	2500
Hydrologic Model	RMSE	33.076	33.119	33.111	33.108	33.111
	PBIAS(%)	25.510	25.526	25.521	25.511	25.541
	NSE	0.7293	0.7286	0.7287	0.7288	0.7287
	Time (s)	60.01	119.39	179.05	239.33	300.40
2 <sup>th</sup> -order PCE	RMSE	33.055	33.093	33.084	33.084	33.088
	PBIAS(%)	25.480	25.491	25.485	25.480	25.510
	NSE	0.7296	0.7290	0.7292	0.7291	0.7291
	Time (s)	2.85	5.15	7.58	10.17	12.74
3 <sup>th</sup> -order PCE	RMSE	33.053	33.086	33.078	33.078	33.081
	PBIAS(%)	25.497	25.504	25.499	25.495	25.520
	NSE	0.7297	0.7291	0.7293	0.7293	0.7292
	Time (s)	22.89	45.28	67.95	90.90	114.29

association will mainly increase as the precipitation increases from 0 to 1 mm/day. The increasing trends would not apparent as the precipitation increases from 1 to 5 mm/day. This may be due to the fact that the hydrologic model uses saturation excess overland flow mechanism in the rainfall-runoff process. Moreover, Table 9 characterizes the nonlinearity between model parameter sensitivities and observed precipitation, potential evapotranspiration and streamflow discharge. The results in Tables 8 and 9 show that the correlation/association between model parameter sensitivities and catchment conditions (i.e. precipitation) are mainly nonlinear regardless of precipitation levels. Moreover, the nonlinearity will increase as the precipitation increases. For instance, under light ( $P > 0$ ) and heavy ( $P > 5$ ) rain days, the nonlinearity between  $C_{max}$  and  $Q$  is about 0.52 and 0.62 respectively.

## 5. Conclusions

In this study, a hybrid sequential data assimilation and probabilistic collocation (HSDAPC) approach was developed to quantify the uncertainty of hydrologic models and screen the sensitivities of model parameters. The proposed HSDAPC method integrates the capability of sequential data assimilation and the probabilistic collocation method into its framework, in which the posterior probabilities of hydrologic model parameters are estimated through the PF method, and the uncertainty propagation is showed through the probabilistic collocation method. The temporal dynamics of parameter sensitivity (TEDPAS) based on Sobol's indices are obtained through the obtained PCE. The maximal information coefficient (MIC) is then adopted to identify the association between model parameter sensitivities and catchment conditions.

The developed HSDAPC approach was applied for a small catchment in the Three Gorges Reservoir area, China. The results showed that, the 2<sup>nd</sup> and 3<sup>rd</sup>-order polynomial chaos expansion (PCE) models can well represent the hydrologic model for streamflow forecasting and uncertainty quantification. The performance of the two and three-order PCE models did not deteriorate when compared with the original hydrologic model. In this study, the original hydrologic model produced satisfactory predictions with best NSE value being about 0.73 based on multiple trials. The associated 2<sup>nd</sup> and 3<sup>rd</sup>-order PCE models also generated satisfactory predictions with NSE values being about 0.73. Specifically, the efficiency of the PCEs was considerably more efficient than the hydrologic model, as the 2<sup>nd</sup>-order PCE model ran ten times faster than the simplified conceptual model (i.e. Hymod). Moreover, temporal dynamics of parameter sensitivity (TED-



**Fig. 8.** Comparison between the prediction intervals and observations: (a) hydrologic model prediction intervals vs. observations, (b) 2<sup>nd</sup>-order PCE predicting intervals vs. observations, (c) 3<sup>rd</sup>-order PCE predicting intervals vs. observations.

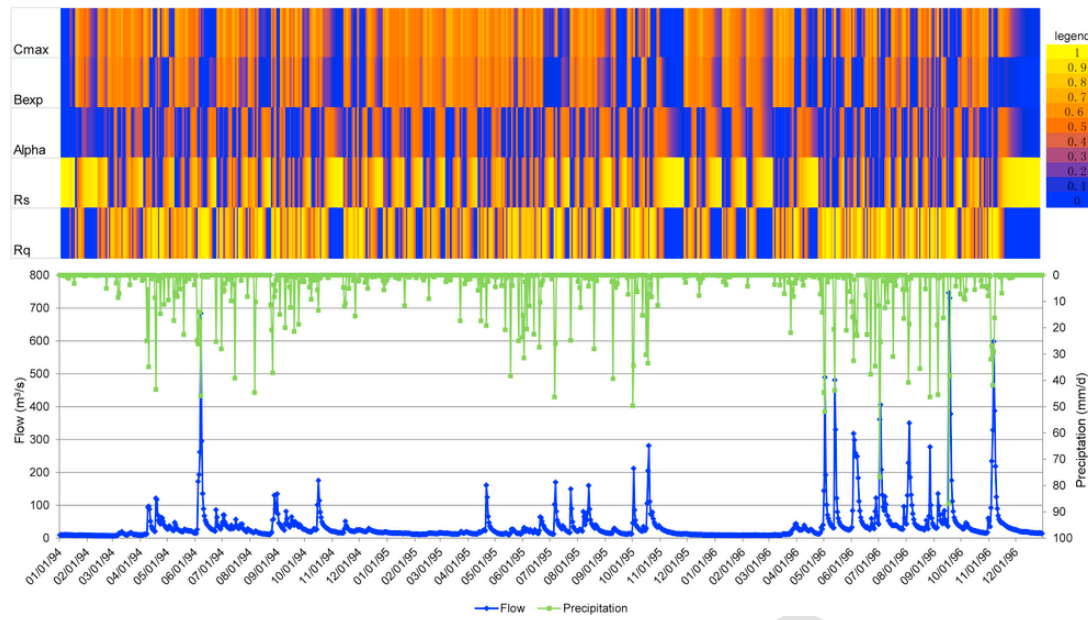


Fig. 9. Paramter sensitivity analysis through 2<sup>nd</sup>-order PCE.

Table 6  
Parameter sensitivity under different streamflow periods.

Period	Parameter	2-Order PCE		3-Order PCE	
		Mean	Std	Mean	Std
High flow	Cmax	0.132	0.160	0.128	0.160
	bexp	0.045	0.070	0.040	0.063
	alpha	0.031	0.050	0.029	0.047
	Rs	0.072	0.085	0.073	0.087
	Rq	0.721	0.267	0.730	0.263
Low Flow	Cmax	0.148	0.124	0.147	0.127
	bexp	0.117	0.097	0.105	0.090
	alpha	0.036	0.035	0.034	0.033
	Rs	0.567	0.348	0.577	0.350
	Rq	0.131	0.198	0.138	0.207

Table 7  
Parameter sensitivity under different precipitation conditions.

Period	Parameter	2-Order PCE		3-Order PCE	
		Mean	Std	Mean	Std
P > 0	Cmax	0.200	0.161	0.198	0.164
	bexp	0.129	0.109	0.116	0.101
	alpha	0.036	0.048	0.034	0.045
	Rs	0.159	0.244	0.163	0.249
	Rq	0.477	0.312	0.489	0.315
P = 0	Cmax	0.136	0.150	0.134	0.152
	bexp	0.094	0.106	0.085	0.098
	alpha	0.057	0.051	0.053	0.048
	Rs	0.412	0.361	0.419	0.364
	Rq	0.301	0.323	0.309	0.326

PAS) analysis were then performed to characterize the dominant model component under different hydro-meteorological conditions. Sobol's sensitivity indices were employed for the above sensitivity analysis since they could be well approximated through the coefficients in the obtained PCE models. The results showed that the slow-flow and quick-flow, in Hymod, would be the dominant model components, in which slow-flow was most important under dry meteorological conditions and quick-flow contributed most when precipita-

tion occurred. Soil moisture contributed more to runoff generation under wet conditions than that under dry conditions. The association/correlation between model parameter sensitivities and catchment conditions was revealed through the maximal information coefficient (MIC). These results showed that the correlations between model parameter sensitivity and precipitation, potential evapotranspiration and streamflow discharge were significant and such correlations mainly showed nonlinear features. Particularly, these correlations will be quite significant during rainy periods, with most MIC values being larger than 0.5. Also, the correlations were dominated by nonlinearity during the rainy periods.

The developed HSDAPC method integrates the capability of the PF method and the probabilistic collocation method for quantifying the uncertainty of hydrologic predictions. This method could also be applied to other hydrological models (conceptual, semi-distributed, and distributed hydrologic models). The innovations of this study are that: (i) after the data assimilation process by PF, the probabilistic collocation method was further used, other than the classic Monte Carlo simulation method, to show uncertainty propagation in the hydrologic model; (ii) the temporal dynamics of parameter sensitivity were derived through the obtained PCE to reveal the dominant components of the hydrologic model in different simulation conditions; and, (iii) the correlation between model parameter sensitivities and catchment conditions (i.e. precipitation, potential evapotranspiration, and streamflow discharge) was identified through the use of the maximal information coefficient.

In the HSDAPC approach, the resulting PCE model is a proxy for the original hydrologic model, which means that the accuracy of the PCE model can hardly perform better than the original hydrologic model. Consequently, once the parameters in a hydrological model are not well identified in the data assimilation procedure of HSDAPC, the resulting PCE model would also provide inaccuracy predictions. Moreover, the applicability of the proposed method is demonstrated through a lumped hydrologic model with only 5 parameters. For other high-dimensional cases, such as distributed hydrologic models, the number of polynomials used in PCE will increase remarkably. For instance, the 3<sup>rd</sup>-order PCE for a hydrological model with 10 parameters has 286 polynomials. This may lead to difficulties

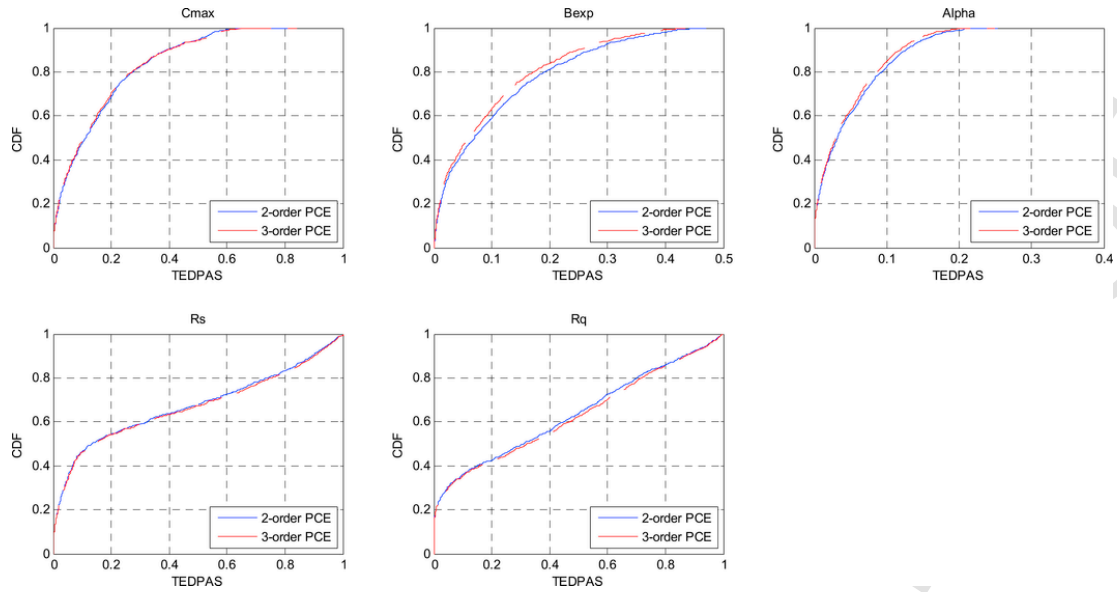


Fig 11. Cumulative probability of the temporal sensitivity for all five parameters obtained through 2-order and 3-order PCE.

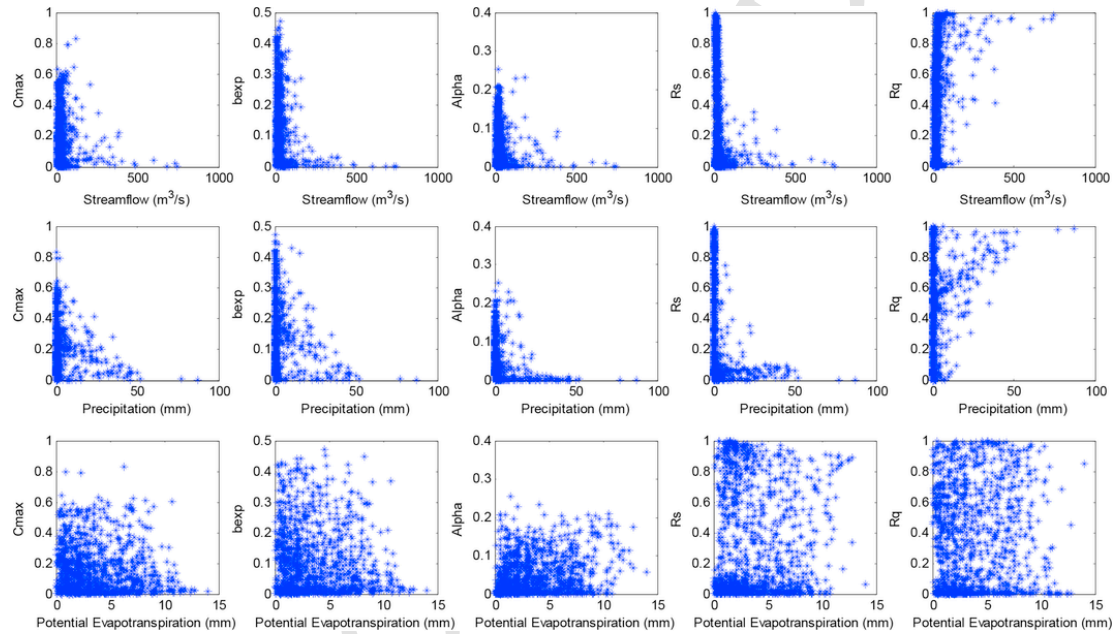


Fig. 12. Correlation between parameter sensitivity and models inputs.

in coefficients estimation for the PCE model. However, the proposed HSDAPC can still be applied for high-dimensional cases through some improvements: (i) as demonstrated in this study, a 2<sup>nd</sup>-order PCE model can adequately represent the original hydrologic model, which has less polynomials than the 3<sup>rd</sup>-order PCE (i.e. 66 items for 10 dimensional cases). Consequently, 2<sup>nd</sup>-order PCE model can be applicable for some high-dimensional cases. (ii) For high-dimensional cases, some pre-processing approaches can be employed to identify the low sensitive parameters, and the HSDAPC is then applied which only consider high sensitive parameters. For instance, Wang et al. (2015) addressed this issue through introducing fractional

factorial analysis into the probabilistic collocation method to analyze uncertainty propagation of hydrologic model parameters in a reduced dimensional space.

**Acknowledgments**

This research was supported by the Natural Science Foundation of China (Nos. 51190095 and 51225904) and the Program for Innovative Research Team in University (IRT1127).

**Appendix.**



**Table 8**

The MIC values between model inputs and outputs and model parameter sensitivity.

Precipitation		$C_{\max}$	$b_{\exp}$	$alpha$	$R_s$	$R_q$
All Pre	Q	0.1759	0.1924	0.1498	0.2551	0.3083
	P	0.1682	0.1552	0.1836	0.2293	0.1946
$P > 0$	ET	0.1384	0.1395	0.1574	0.1756	0.1284
	Q	0.5601	0.5889	0.5281	0.5080	0.5740
$P > 1$	P	0.5141	0.4970	0.4877	0.5540	0.5176
	ET	0.4599	0.5085	0.5333	0.4257	0.4557
$P > 5$	Q	0.7254	0.7496	0.6560	0.6322	0.7010
	P	0.6288	0.6543	0.6161	0.6600	0.7603
	ET	0.5943	0.6507	0.6608	0.6656	0.6920
	Q	0.7151	0.7713	0.6550	0.6472	0.6962
	P	0.6369	0.6578	0.6222	0.6606	0.7603
	ET	0.5953	0.6515	0.6633	0.6654	0.6920

**Table 9**

The nonlinearity between model inputs and outputs and model parameter sensitivity.

		$C_{\max}$	$b_{\exp}$	$alpha$	$R_s$	$R_q$
All Pre	Q	0.1721	0.1667	0.1384	0.1969	0.1844
	P	0.1653	0.1504	0.1384	0.1712	0.0826
$P > 0$	ET	0.1272	0.1258	0.1049	0.1755	0.1265
	Qt	0.5228	0.4934	0.5268	0.4874	0.4515
$P > 1$	Pt	0.4653	0.4455	0.4434	0.5037	0.3342
	ET	0.4582	0.4969	0.5319	0.3934	0.3976
$P > 5$	Qt	0.6354	0.6078	0.6501	0.6236	0.5582
	Pt	0.4731	0.5061	0.5846	0.6300	0.5010
	ET	0.5918	0.6341	0.6597	0.6566	0.6666
	Qt	0.6209	0.6259	0.6482	0.6371	0.5559
	Pt	0.4747	0.5070	0.5870	0.6232	0.5021
	ET	0.5922	0.6343	0.6628	0.6500	0.6668

### A1. Gaussian anamorphosis (GA) transformation

For an original random variable  $x$  and the transformed random variable  $x' = f(x)$ , the premise of GA is to find a function  $f$  to define a change of variables (anamorphosis) such that the random variable  $x'$  obeys a standard normal distribution. Such a transformation technique was applied previously for biogeochemical and physical-biogeochemical simulations in ocean (Simon and Bertino, 2009; Béal et al., 2010) and subsurface hydraulic tomography simulation (Schöniger et al., 2012).

Consider an arbitrarily distributed variable  $x$  and its Gaussian transform variable  $z$ , they can be linked through their cumulative distribution functions (CDFs) as follows:

$$z = G^{-1}(F(y)) \quad (A1)$$

where  $F(y)$  is the empirical CDF of  $y$ ,  $G$  is the theoretical standard normal CDF of  $z$ . Since  $G$  is monotonously increasing, the inverse  $G^{-1}$  exists. Equation (A1) is called the Gaussian anamorphosis function.

Following the method proposed by Johnson and Wichern (1988), the empirical CDF of  $y$  can be obtained, based on its sample values as follows:

$$F_j = \frac{j - 0.5}{N} \quad (A2)$$

where  $j$  are the ranks of the sample values of  $x$ ;  $N$  is the sample size of  $x$  (rendered as the number of particles in this study). From Equations (A1) and (A2), the sample values of the Gaussian trans-

form variable  $z$  can be obtained, which correspond to the sample values of  $x$ . Also, the sample range of  $z$  can be determined as follows:

$$z_{\min} = G^{-1}\left(\frac{1 - 0.5}{N}\right) \quad (A3)$$

$$z_{\max} = G^{-1}\left(\frac{N - 0.5}{N}\right) \quad (A4)$$

### A2. PCE-based sensitivity analysis

For example, consider a truncated two-dimensional 2<sup>th</sup>-order PCE expressed as:

$$\hat{y} = a_0 + a_1\zeta_1 + a_2\zeta_2 + a_3(\zeta_1^2 - 1) + a_4(\zeta_2^2 - 1) + a_5\zeta_1\zeta_2 \quad (A5).$$

the total variance can be obtained as:

$$V(\hat{y}) = a_1^2 + a_2^2 + 2a_3^2 + 2a_4^2 + a_5^2 \quad (A6)$$

If  $\zeta_1$  is fixed, the variance would be:

$$V(\hat{y}|\zeta_1) = a_1^2 + 2a_3^2 \quad (A7)$$

Similarly  $V(\hat{y}|\zeta_2) = a_2^2 + 2a_4^2$  and  $V(\hat{y}|\zeta_1, \zeta_2) = V(\hat{y})$  (A8).

Consequently, the first and second-order Sobol's sensitivity indices can be expressed as (Zheng et al., 2011):

$$S_1 \cong \frac{V(E(\hat{y}|\zeta_1))}{V(\hat{y})} = \frac{a_1^2 + 2a_3^2}{a_1^2 + a_2^2 + 2a_3^2 + 2a_4^2 + a_5^2} \quad (A9)$$

$$S_2 \cong \frac{V(E(\hat{y}|\zeta_2))}{V(\hat{y})} = \frac{a_2^2 + 2a_4^2}{a_1^2 + a_2^2 + 2a_3^2 + 2a_4^2 + a_5^2} \quad (A10)$$

$$S_{12} \cong \frac{V(\hat{y}|\zeta_1, \zeta_2) - V(E(\hat{y}|\zeta_1)) - V(E(\hat{y}|\zeta_2))}{V(\hat{y})} = \frac{a_5^2}{a_1^2 + a_2^2 + 2a_3^2 + 2a_4^2 + a_5^2}$$

Also, the total-effect indices can be obtained as:  $S_1^T = S_1 + S_{12}$ , and  $S_2^T = S_2 + S_{12}$ .

### A3. Evaluation methods

A number of evaluation criteria are proposed to identify the performance of hydrologic forecasting accuracy. In this study, the root-mean-square error (RMSE) and the Nash-Sutcliffe efficiency (NSE), are employed to evaluate the performance of HSDAPC approach. The formulation of the above four indices are presented as:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (Q_i - P)_i^2} \quad (A12)$$

$$NSE = 1 - \frac{\sum_{i=1}^N (Q_i - P_i)^2}{\sum_{i=1}^N (Q_i - \bar{Q})^2} \quad (A13)$$

where  $N$  is the total number of observations (or predictions),  $Q_i$  is the observed value,  $P_i$  is the estimated value, and  $\bar{Q}$  is the mean of all observed and estimated values.

$$PBIAS = \frac{\sum_{i=1}^N (Q_i - P_i) * 100}{\sum_{i=1}^N Q_i} \quad (A14)$$

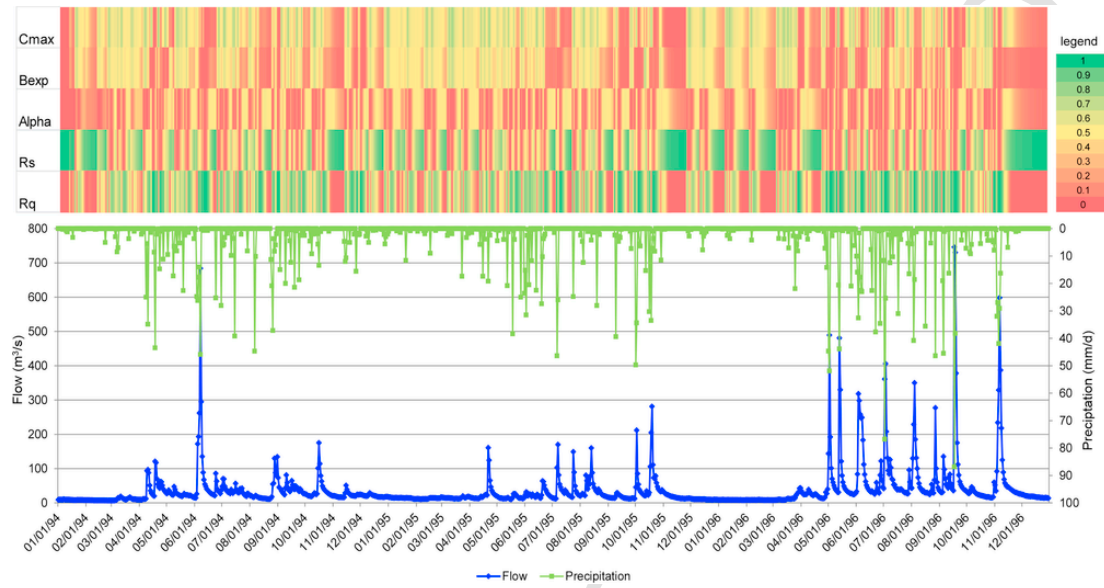


Fig. 10. Parameter sensitivity analysis through 3<sup>rd</sup>-order PCE.

## References

- Abaza, M., Anctil, F., Fortin, V., Turcotte, R., 519, 2014, 2692–2706. Sequential streamflow assimilation for short-term hydrological ensemble forecasting. *J. Hydrol.* <http://dx.doi.org/10.1016/j.jhydrol.2014.08.038>.
- Béal, D., Brasseur, P., Brankart, J.-M., Ourmières, Y., Verron, J., 2010. Characterization of mixing errors in a coupled physical biogeochemical model of the North Atlantic: implications for nonlinear estimation using Gaussian anamorphosis. *Ocean Sci.* 6, 247–263.
- Beven, K., 2006. A Manif. Equifinality Thesis, *J. Hydrol.* 320, 18–36.
- Bi, H., Ma, J., Wang, F., An improved particle filter algorithm based on ensemble Kalman filter and Markov Chain Monte Carlo method. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 8(2), 2015, 447–459 <http://dx.doi.org/10.1109/JSTARS.2014.2322096>.
- Box, G.E.P., Cox, D.R., 1964. An analysis of transformation. *J. R. Stat. Soc. Ser. B* 26 (2), 211–252.
- Chaney, N.W., Herman, J.D., Reed, P.M., Wood, E.F., 2015. Flood and drought hydrologic monitoring: the role of model parameter uncertainty. *Hydrol. Earth Syst. Sci.* 19 (7), 3239–3251. <http://dx.doi.org/10.5194/hess-19-3239-2015>.
- Chen, H., Yang, D., Hong, Y., Gourley, J.J., Zhang, Y., 2013. Hydrological data assimilation with the Ensemble Square-Root-Filter: use of streamflow observations to update model states for real-time flash flood forecasting. *Adv. Water Resour.* 59, 209–220.
- Clark, M.P., Rupp, D.E., Woods, R.A., Zheng, X., Ibbitt, R.P., Schmidt, A.G., Uddstrom, M.J., 2008. Hydrological data assimilation with the ensemble Kalman filter: use of streamflow observations to update states in a distributed model. *Adv. Water Resour.* 31, 1309–1324.
- Dai, C., Li, H., Zhang, D.X., 2014. Efficient and accurate global sensitivity analysis for reservoir simulations by use of the probabilistic collocation method. *SPE J.* 19 (4), 621–635.
- 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, W04518. <http://dx.doi.org/10.1029/2011WR011011>.
- 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. <http://dx.doi.org/10.1016/j.jhydrol.2014.05.045>.
- Wang, Dingbao, Yuguang, Chen, Ximing, Cai, 2009. State and parameter estimation of hydrologic models using the constrained ensemble Kalman filter. *Water Resour. Res.* 45, W11416.
- Doucet, A., Johansen, A., 2011. A tutorial on particle filtering and smoothing: fifteen years later. In: *Handbook of Nonlinear Filtering*. pp. 656–704.
- Doucet, A., Godsill, S., Andrieu, C., 2000. On sequential Monte Carlo sampling methods for Bayesian filtering. *Stat. Comput.* 10 (3), 197–208. <http://dx.doi.org/10.1023/A:1008935410038>.
- El Mocaed, N., Ricci, S., Goutal, N., Rochoux, M.C., Boyaval, S., Goeury, C., Lucor, D., Thual, O., 2016. Polynomial-chaos representation of 1D water level in open-channel flows with uncertain physical characteristics - detailed analysis of steady flows. In: *Proceedings of 12th International Conference on Hydroinformatics (IHC 2016)*, Incheon, Republic of Korea, Aug. pp. 21–25.
- Fan, Y.R., Huang, G.H., Li, Y.P., 2012. Robust interval linear programming for environmental decision making under uncertainty. *Eng. Optim.* 44 (11), 1321–1336.
- Fan, Y.R., Huang, W., Huang, G.H., Huang, K., Zhou, X., 2015a. A PCM-based stochastic hydrological model for uncertainty quantification in watershed systems. *Stoch. Environ. Res. Risk Assess.* 29 (3), 915–927.
- Fan, Y.R., Huang, W., Huang, G.H., Li, Z., Li, Y.P., Wang, X.Q., Cheng, G.H., Jin, L., 2015b. A stepwise-cluster forecasting approach for monthly streamflows based on climate teleconnections. *Stoch. Environ. Res. Risk Assess.* 29 (6), 1557–1569.
- Fan, Y.R., Huang, G.H., Huang, K., Baetz, B.W., 2015c. Planning water resources allocation under multiple uncertainties through a Generalized Fuzzy two-stage stochastic programming method. *IEEE Trans. Fuzzy Syst.* 23 (5), 1488–1504.
- Fan, Y.R., Huang, W.W., Li, Y.P., Huang, G.H., Huang, K., 2015d. A coupled ensemble filtering and probabilistic collocation approach for uncertainty quantification of hydrological models. *J. Hydrol.* 530, 255–272.
- Fan, Y.R., Huang, G.H., Li, Y.P., Kong, X.M., 2016. Bivariate hydrologic risk analysis based on coupled entropy-copula method for the Xiangxi river in three Gorges reservoir area, China. *Theor. Appl. Climatol.* 125, 381–397.
- Fearnhead, P., Clifford, P., 2003. On-line inference for hidden Markov models via particle filters. *J. R. Stat. Soc. Ser. B, Methodol.* 65, 887–899.
- Gong, W., Gupta, H.V., Yang, D., Sricharan, K., Hero, A.O., 2013. Estimating episodic and aleatory uncertainties during hydrologic modeling: an information theoretic approach. *Water Resour. Res.* 49 (4), 2253–2273. <http://dx.doi.org/10.1002/wrcr.20161>.
- Gordon, N.J., Salmond, D.J., Smith, A.F.M., 1993. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEEE Proc. Radar Signal Process.* 140 (2), 107–113.
- Guse, B., Reusser, D.E., Fohrer, N., 2014. How to improve the representation of hydrological processes in SWAT for a lowland catchment—temporal analysis of parameter sensitivity and model performance. *Hydrol. Process.* 28, 2651–2670.
- Han, X., Li, X., 2008. An evaluation of the nonlinear/non-Gaussian filters for the sequential data assimilation. *Remote Sens. Environ.* 112, 1434–1449.
- Han, J.C., Huang, G.H., Zhang, H., Li, Z., Li, Y.P., 28(4), 2014, 973–989. Bayesian uncertainty analysis in hydrological modeling associated with watershed subdivision level: a case study of SLURP model applied to the Xiangxi River watershed, China. *Stoch. Environ. Res. Risk Assess.* <http://dx.doi.org/10.1007/s00477-013-0792-0>.
- Herman, J.D., Kollat, J.B., Reed, P.M., Wagener, T., 2013. Technical note: method of Morris effectively reduces the computational demands of global sensitivity analysis for distributed watershed models. *Hydrol. Earth Syst. Sci.* 17, 2893–2903.
- Huang, S., Mahadevan, S., Rebba, R., 2007. Collocation-based stochastic finite element analysis from random field problems. *Probab. Eng. Mech.* 22, 194–205.
- Johnson, R.A., Wichern, D.W., 1988. *Applied Multivariate Statistical Analysis*. Prentice-Hall, Englewood Cliffs, N.J.
- Kong, X.M., Huang, G.H., Fan, Y.R., Li, Y.P., 2015. Maximum entropy-Gumbel-Hougaard copula method for simulation of monthly streamflow in Xiangxi river, China. *Stoch. Environ. Res. Risk Assess.* 29, 833–846.
- Leisenring, M., Moradkhani, H., 2012. Analyzing the uncertainty of suspended sediment load prediction using sequential Monte Carlo methods. *J. Hydrol.* 468–469, 268–282. <http://dx.doi.org/10.1016/j.jhydrol.2012.08.049>.
- Li, Z., 2012. *Inexact Optimization Modelling for Water Quality Management*. Master's Thesis University of Regina, Regina, SK.
- Li, H., Zhang, D., 2007. Probabilistic collocation method for flow in porous media: comparisons with other stochastic method. *Water Resour. Res.* 43.
- Li, Z., Huang, G.H., Fan, Y.R., Xu, J.L., 2015. Hydrologic risk analysis for Nonstationary streamflow records under uncertainty. *J. Environ. Inf.* 26 (1), 41–51.
- Liu, J.S., Chen, R., 1998. Sequential Monte Carlo methods for dynamic systems. *J. Am. Stat. Assoc.* 93, 1032–1044. 1998.
- Liu, Y., Weerts, A.H., Clark, M., Hendricks Franssen, H.-J., Kumar, S., Moradkhani, H., Seo, D.-J., Schwabenberg, D., Smith, P., van Dijk, A.I.J.M., van Velzen, N., He, M., Lee, H., Noh, S.J., Rakovec, O., Restrepo, P., 2012. Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities. *Hydrol. Earth Syst. Sci.* 16, 3863–3887.
- Lucas, D.D., Prinn, R.G., 2005. Parametric sensitivity and uncertainty analysis of dimethylsulfide oxidation in the clear-sky remote marine boundary layer. *Atmos. Chem. Phys.* 5, 1505–1525.
- Madagascar, S., Moradkhani, H., 2014. Improved Bayesian multimodeling: integration of copulas and Bayesian model averaging. *Water Resour. Res.* 50 (12), 9586–9603.
- McMillan, H.K., Hreinsson, E.O., Clark, M.P., Singh, S.K., Zammit, C., Uddstrom, M.J., 2013. Operational Hydrological data assimilation with the recursive ensemble Kalman filter. *Hydrol. Earth Syst. Sci.* 17, 21–38.
- Montanari, A., Brath, A., 2004. A stochastic approach for assessing the uncertainty of rainfall-runoff simulations. *Water Resour. Res.* 40, W01106.
- Moor, R.J., 1985. The probability-distributed principle and runoff production at point and basin scales. *Hydrol. Sci. J.* 30, 273–297.
- Moor, R.J., 2007. The PDM rainfall-runoff model. *Hydrol. Earth Syst. Sci.* 11 (1), 483–499.
- Moradkhani, H., Sorooshian, S., Gupta, H.V., Houser, P.R., 2005a. Dual state-parameter estimation of hydrological models using ensemble Kalman filter. *Adv. Water Resour.* 28, 135–147. <http://dx.doi.org/10.1016/j.advwatres.2004.09.002>.
- Moradkhani, H., Hsu, K.-L., Gupta, H., Sorooshian, S., 2005b. Uncertainty assessment of hydrologic model states and parameters: sequential data assimilation using the particle filter. *Water Resour. Res.* 41 (5), W05012. <http://dx.doi.org/10.1029/2004WR003604>.
- 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, W12520. <http://dx.doi.org/10.1029/2012WR012144>.
- Noh, S.J., Rakovec, O., Weerts, A.H., Tachikawa, Y., 2014. On noise specification in data assimilation schemes for improved flood forecasting using distributed hydrological models. *J. Hydrol.*, 519, Part D 2707–2721. <http://dx.doi.org/10.1016/j.jhydrol.2014.07.049>.
- Nourani, V., Khanghah, T.R., Baghanam, A.H., 2015. Application of entropy concept for input selection of wavelet-ANN based rainfall-runoff modeling. *J. Environ. Inf.* 26 (1), 52–70.
- Parrish, M.A., Moradkhani, H., Dechant, C.M., 2012. Toward reduction of model uncertainty: integration of Bayesian model averaging and data assimilation. *Water Resour. Res.* 48, <http://dx.doi.org/10.1029/2011WR011116>.
- Pathiraja, S., Marshall, L., Sharma, A., Moradkhani, H., 2016a. Detecting non-stationary hydrologic model parameters in a paired catchment system using Data Assimilation. *Adv. Water Resour.* 94, 103–119. <http://dx.doi.org/10.1016/j.advwatres.2016.04.021>.
- Pathiraja, S., Marshall, L., Sharma, A., Moradkhani, H., 52, 2016b, 3350–3372. Hydrologic modeling in dynamic catchments: a data assimilation approach. *Water Resour. Res.* <http://dx.doi.org/10.1002/2015WR017192>.

- 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, <http://dx.doi.org/10.1002/wrcr.20291>.
- Rahmani, M.A., Zarghami, M., 2015. The use of statistical weather generator, hybrid data driven and system dynamics models for water resources management under climate change. *J. Environ. Inf.* 25 (1), 23–35.
- Rakovec, O., Weerts, A.H., Sumihar, J., Uijlenhoet, R., 2015. Operational aspects of asynchronous filtering for flood forecasting. *Hydrol. Earth Syst. Sci.* 19 (6), 2911–2924. <http://dx.doi.org/10.5194/hess-19-2911-2015>.
- Rasmussen, J., Madsen, H., Jensen, K.H., Refsgaard, J.C., 2015. Data assimilation in integrated hydrological modeling using ensemble Kalman filtering: evaluating the effect of ensemble size and localization on filter performance. *Hydrol. Earth Syst. Sci.* 19, 2999–3013.
- Reichle, R., McLaughlin, D.B., Entekhabi, D., 2002. Hydrologic data assimilation with the ensemble Kalman filter. *Mon. Weather Rev.* 130 (1), 103–114.
- Reshef, D.N., Reshef, Y.A., Finucane, H.K., Grossman, S.R., McVean, G., Turnbaugh, P., Lander, E.S., Mitzenmacher, M., Sabeti, P.C., 2011. Detecting novel associations in large data sets. *Science* 334, 1518. <http://dx.doi.org/10.1126/science.1205438>.
- Reusser, D.E., Buytaert, W., Zehe, E., 2011. Temporal dynamics of model parameter sensitivity for computationally expensive models with the Fourier amplitude sensitivity test. *Water Resour. Res.* 47, W07551.
- Samaniego, L., Kumar, R., Zink, M., 2013. Implications of parameter uncertainty on soil moisture drought analysis in Germany. *J. Hydrometeorol.* 14 (1), 47–68. <http://dx.doi.org/10.1175/JHM-D-12-075.1>.
- 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, W04502.
- Shi, L., Yang, J., Zhang, D., Li, H., 2009. Probabilistic collocation method for unconfined flow in heterogeneous media. *J. Hydrol.* 365, 4–10.
- Sikorska, A.E., Montanari, A., Koutsoyiannis, D., 2014. A4014009. Estimating the uncertainty of hydrological predictions through data-driven resampling techniques. *J. Hydrol. Eng.* [http://dx.doi.org/10.1061/\(ASCE\)HE.1943-5584.0000926](http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0000926).
- Simon, E., Bertino, L., 2009. Application of the Gaussian anamorphosis to assimilation in a 3-D coupled physical-ecosystem model of the North Atlantic with the EnKF: a twin experiment. *Ocean Sci.* 5, 495–510.
- Snyder, C., Bengtsson, T., Bickel, P., Anderson, J., 2008. Obstacles to high-dimensional particle filtering. *Mon. Weather Rev.* 136, 4629–4640.
- Song, X., Zhang, J., Zhang, C., Xuan, Y., Ye, M., Xu, C., 2015. Global sensitivity analysis in hydrological modeling: review of concepts, methods, theoretical framework, and applications. *J. Hydrol.* 523, 739–757.
- Vrugt, J.A., Diks, C.G.H., Gupta, H.V., Bouten, W., Verstraten, J.M., 2005. Improved treatment of uncertainty in hydrologic modelling: combining the strengths of global optimization and data assimilation. *Water Resour. Res.* 41, W01017.
- 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.
- Wagener, T., McIntyre, N., Lees, M.J., Wheatley, H.S., Gupta, H.V., 2003. Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis. *Hydrol. Process.* 17, 455–476.
- Wang S., Huang G.H., Huang W., Fan Y.R., Li Z., (2015). A fractional factorial probabilistic collocation method for uncertainty propagation of hydrologic model parameters in a reduced dimensional space. *Journal of Hydrology*, 529, 1129–1146.
- Wang, W.J., Shrestha, D.L., Robertson, D.E., Pokhrel, P., 2012. A log-sinh transformation for data normalization and variance stabilization. *Water Resour. Res.* 48, W05514.
- Webster, M., Tatang, M.A., Mcrae, G.J., 1996. Application of the probabilistic collocation method for an uncertainty analysis of a simple ocean model. In: MIT Joint Program on the Science and Policy of Global Change, Report Series No. 4. Massachusetts Institute of Technology.
- 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, W09403. <http://dx.doi.org/10.1029/2005WR004093>.
- Westra, S., Thyer, M., Leonard, M., Kavetski, D., Lambert, M., 2014. A strategy for diagnosing and interpreting hydrological model nonstationarity. *Water Resour. Res.* 50 (6), 5090–5113.
- Wiener, N., 1938. The homogeneous chaos. *Am. J. Math.* 60, 897–936.
- Wu, B., Zheng, Y., Tian, Y., Wu, X., Yao, Y.Y., Han, F., Liu, J., Zheng, C., 2014. 5848–5865. Systematic assessment of the uncertainty in integrated surface water-groundwater modelling based on probabilistic collocation method. *Water Resour. Res.* <http://dx.doi.org/10.1002/2014WR015366>.
- Xiao, W.H., Wang, J.H., Huang, Y.F., Sun, S.C., Zhou, Y.Y., 2015. An approach for estimating the Nitrobenzene (NB) emission effect in Frozen rivers: a case study of Nitrobenzene pollution in the Songhua river, China. *J. Environ. Inf.* 26 (2), 140–147.
- Xie, X., Zhang, D., 2013. A partitioned update scheme for state-parameter estimation of distributed hydrologic models based on the ensemble Kalman filter. *Water Resour. Res.* 49, 7530–7565.
- Xiu, D., Karniadakis, G.E., 2002. The Wiener–Askey polynomial chaos for stochastic differential equations. *SIAM J. Sci. Comput.* 24, 619–644.
- Xiu, D., Karniadakis, G.E., 2003. Modeling uncertainty in flow simulations via generalized polynomial chaos. *J. Comput. Phys.* 187, 137–167.
- 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 Sens.* 53 (11), 6134–6147. <http://dx.doi.org/10.1109/TGRS.2015.2432067>.
- Yang, Y., Zhang, C.T., Zhang, R.X., Christakos, G., 2015. Improving environmental prediction by assimilating auxiliary information. *J. Environ. Inf.* 26 (2), 91–105.
- Zhang, Y., Yang, D., 2013. Simultaneous estimation of relative permeability and capillary pressure for tight formations using ensemble-based history matching method. *Comput. Fluids* 71, 446–460.
- Zhang, Y., Yang, D., 2014. Estimation of relative permeability and capillary pressure for tight formations by assimilating field production data. *Inverse Probl. Sci. Eng.* 22 (7), 1150–1175.
- Zhang, Y., Li, H., Yang, D., 2012a. Simultaneous estimation of relative permeability and capillary pressure using ensemble-based history matching techniques. *Transp. Porous Media* 94 (1), 259–276.
- Zhang, Y., Song, C., Zheng, S., Yang, D. (2012b). Simultaneous estimation of relative permeability and capillary pressure for tight formations from displacement experiments. SPE paper 162663, the SPE Canadian Unconventional Resources Conference; 2012 October 30–November 1. Calgary, AB.
- Zhang, Y., Jia, S., Huang, H., Qiu, J., Zhou, C., 2014. A novel algorithm for the precise calculation of the maximal information coefficient. *Sci. Rep.* 4, 6662. <http://dx.doi.org/10.1038/srep06662>.
- Zheng, Y., Wang, W., Han, F., Ping, J., 2011. Uncertainty assessment for watershed water quality modeling: a Probabilistic Collocation Method based approach. *Adv. Water Resour.* 34, 887–898.