A Coupled Ensemble Filtering and Probabilistic Collocation Approach for Uncertainty Quantification of Hydrological Models

Y.R. Fan[a], W.W. Huang[b], Y.P. Li[c], G.H. Huang[c,d], K. Huang[a]

[a] Faculty of Engineering and Applied Science, University of Regina, Regina, Saskatchewan, Canada S4S 0A2b

[b] Department of Civil Engineering, McMaster University, Hamilton, ON L8S 4L8, Canada

[c] MOE Key Laboratory of Regional Energy and Environmental Systems Optimization, North China Electric Power University, Beijing 102206, China

[d] Institute for Energy, Environment and Sustainability Research, UR-NCEPU, University of Regina, Regina, Saskatchewan, Canada S4S 0A2; Tel: +13065854095; Fax: +13065854855; E-mail: huang@iseis.org

*Correspondence: Dr. G. H. Huang

Faculty of Engineering and Applied Sciences, University of Regina

Regina, Saskatchewan, S4S 0A2, Canada

Tel: (306) 585-4095

Fax: (306) 585-4855

E-mail: huangg@uregina.ca

Abstract:

In this study, a coupled ensemble filtering and probabilistic collocation (EFPC) approach is proposed for uncertainty quantification of hydrologic models. This approach combines the capabilities of the ensemble Kalman filter (EnKF) and the probabilistic collocation method (PCM) to provide a better treatment of uncertainties in hydrologic models. The EnKF method would be employed to approximate the posterior probabilities of model parameters and improve the forecasting accuracy based on the observed measurements; the PCM approach is proposed to construct a model response surface in terms of the posterior probabilities of model parameters to reveal uncertainty propagation from model parameters to model outputs. The proposed method is applied to the Xiangxi River, located in the Three Gorges Reservoir area of China. The results indicate that the proposed EFPC approach can effectively quantify the uncertainty of hydrologic models. Even for a simple conceptual hydrological model, the efficiency of EFPC approach is about 10 times faster than traditional Monte Carlo method without obvious decrease in prediction accuracy. Finally, the results can explicitly reveal the contributions of model parameters to the total variance of model predictions during the simulation period.

Keywords: Uncertainty; Ensemble Kalman filter; Probabilistic collocation method; Gaussian anamorphosis; Hydrologic model; Monte Carlo

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). Such conceptualization and aggregation lead to extensive uncertainties involved in both model parameters and structures, and consequently produce significant uncertainties in hydrologic predictions. Uncertainty in hydrologic predictions can originate from several major sources, including model structures, parameters, and measurement errors in model inputs (Ajami et al., 2007; Liu et al., 2012). Therefore, effective uncertainty quantification and reduction methods are required to produce reliable hydrologic forecasts for many real-world water resources applications, such as flooding control, drought management and reservoir operation (Fan et al., 2012; Kong et al., 2015; Fan et al., 2015a).

Previously, a number of probabilistic estimation methods have been proposed for quantifying uncertainty in hydrologic predictions. The probabilistic estimation methods approximate the posterior probability distributions of the hydrological parameters through the Bayesian theorem, conditioned on the streamflow observations. The generalized likelihood uncertainty estimation (GLUE) (Beven and Binley, 1992), Markov Chain Monte Carlo (Vrugt et al., 2009; Han et al., 2014), Bayesian model averaging (BMA) (Diks and Vrugt., 2010), and approximate Bayesian computation (Vrugt and Sadegh, 2013) methods are those extensively used probabilistic estimation methods. For instance, Madadgar and Moradkhani (2014) improved Bayesian Multi-modeling predictions through integration of copulas and Bayesian model averaging methods. DeChant and Moradkhani (2014b) proposed a full review of uncertainty quantification methods.

In a separate line of research, sequential data assimilation methods have been developed to explicitly handle various uncertainties and optimally merging observations into uncertain model predictions (Xie and Zhang, 2013; Zhang et al., 2012a,b; Zhang and Yang, 2013, 2014; Chang and Sayemuzzaman, 2014; Assumaning and Chang, 2014). In contrast to classical model calibration strategies, sequential data assimilation approaches continuously update the states and parameters to improve model forecasts when new measurements become available (Vrugt et al., 2005). The prototype of sequential data assimilation techniques, the Kalman filter (KF) (Kalman, 1960) and the ensemble Kalman filter (EnKF) (Evensen, 1994), provide optimal frameworks for linear dynamic models with Gaussian uncertainties. The EnKF approach is one of the most frequently used data assimilation methods in hydrology due to its attractive features of real-time adjustment and easy implementation (Reichle et al., 2002). The EnKF method can provide a general framework for dynamic state, parameter, and joint state-parameter estimation in hydrologic models. For example, Moradkhani et al. (2005a) proposed a dual-state estimation approach based on EnKF for sequential estimation for both parameters and state variables of a hydrologic model. Weerts and El Serafy (2006) compared the capability of EnKF and particle filter (PF) methods in reducing uncertainty in the rainfall-runoff update and internal model state estimation for flooding forecasting purposes. Parrish et al. (2012) integrated Bayesian model averaging and data assimilation to reduce model uncertainty. DeChant and Moradkhani (2014a) combined ensemble data assimilation and sequential Bayesian methods to provide a reliable prediction of seasonal forecast uncertainty. Shi et al. (2014) conducted multiple parameter estimation using multivariate observations via the ensemble Kalman filter (EnKF) for a physically based land surface hydrologic model. However, due to the local complex characteristics of the watershed, some parameters in the hydrologic model may not be clearly identifiable and show slow convergence (Moradkhani et al., 2005b, 2012). Moreover, the same hydrologic model parameter may even show contrary convergence characteristics when different data assimilation methods are used. As shown by Moradkhani et al. (2005a, b), the C_{max} parameter for the Hymod was identifiable by using particle filter method but unidentifiable by using EnKF. Such unidentifiable parameters would lead to extensive uncertainties in hydrologic forecasts. Moreover, stochastic perturbations are usually added to the model inputs (e.g. precipitation, potential evapotranspiration etc.) and observations (e.g. streamflow) to account for uncertainties in actual measurements. Such random noise would results in uncertainties in model parameters. Consequently, efficient forward uncertainty quantification methods (i.e. from model parameters to model predictions) are still desired for further analyzing the uncertainty in hydrologic predictions. Such methods can reveal the uncertainty evolution and propagation in hydrologic simulation.

Previously, Monte Carlo simulations are usually employed to quantify the uncertainty of hydrologic predictions resulting from uncertain model parameters (Knighton et al., 2014; Houska et al., 2014). 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 distributed hydrologic models, this sampling approach is computationally intensive (Herman et al., 2013). The polynomial chaos expansions (PCEs) are effective for uncertainty propagation in stochastic processes, which represent the random variables through polynomial chaos basis and obtain the unknown expansion coefficients by the Galerkin technique or probabilistic collocation method (PCM) (Li and Zhang, 2007; Shi et al., 2009). The PCE-based methods have been widely used for uncertainty quantification of subsurface flow simulation in porous media (Li and Zhang, 2007; Shi et al., 2009), water quality modelling (Zheng et al., 2011), vehicle dynamics (Kewlani et al., 2012), mechanical systems (Blanchard, 2010), and so on. Fan et al. (2015c) integrated PCM into a hydrologic model for exploring the uncertainty propagation in hydrologic simulation, but it is only suitable for quantifying uncertainty of hydrologic models with specific distributions for model parameters (e.g. uniform, normal). However, in real-world hydrologic simulation, the posterior distributions of model parameters, after calibration through probabilistic estimation approaches, may be arbitrary.

In this study, a coupled ensemble filtering and probabilistic collocation (EFPC) method is proposed for uncertainty quantification of hydrologic models. In EFPC, the posterior distributions of model parameters will be approximated through EnKF; the obtained posterior distributions will be used as inputs for the probabilistic collocation method, in which PCEs will be constructed to connect the model parameters with the model responses. Such PCEs will reflect the uncertainty propagation between model parameters and its outputs. Therefore, the proposed EFPC will enable improved quantification of uncertainties existing in hydrologic predictions, model parameters, inputs and their interrelationships, and further reveal the uncertainty evolution through the obtained PCEs. Furthermore, a Gaussian anamorphosis (GA) approach will be presented to convert the obtained posterior distributions into standard normal random variables, which can be directly used as the inputs for PCM. The proposed approach will be applied to the Xiangxi River basin based on a conceptual rainfall-runoff model. The Xiangxi River basin, located in the Three Gorges Reservoir area of China, is one of the main tributaries in Hubei Province, with a draining area of about 3,200 km^2. The Hymod, which has been used in many catchments, will be employed in this study (van Delft, 2007; Wang et al., 2009; Dechant and Moradkhani, 2012; Moradkhani et al., 2012). This application will help demonstrate the strength and applicability of the proposed methodology.

2. Methodology

2.1. Ensemble Kalman Filter

The data assimilation methods have attracted increasing attention from hydrologists for exploring more accurate hydrological forecasts based on real-time observations (Moradkhani et al., 2005a; Weerts and EI Serafy, 2005; Wang et al., 2009; DeChant and Moradkhani, 2011a,b; Plaza Guingla et al., 2013). Sequential data assimilation is a general framework where system states and parameters are recursively estimated/corrected when new observations are available. In a sequential data assimilation process, the evolution of the simulated system states can be represented as follows:

$$x_t^- = f(x_{t-1}^+, u_t, \theta) + \omega_t \quad (1)$$

where f is a nonlinear function expressing the system transition from time t-1 to t, in response to model input vectors x_{t-1}^+ u_t and θ; x_{t-1}^+ is the analyzed (i.e. posteriori) estimation (after correction) of the state variable x at time step $t-1$; x_t^- is the forecasted (i.e. priori) estimation of the state variable x at time step t; θ represents time-invariant vectors, and ω_t is considered as process noise.

When new observations are available, the forecasted states can be corrected through assimilating the observations into the model, based on the output model responding to the state variables and parameters. The observation output model can be written as:

$$y_t^- = h(x_t^-, \theta) + v_t \tag{2}$$

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

The essential methods for states updating are based on Bayesian analysis, in which the probability density function of the current state given the observations is approximated by the recursive Bayesian law:

$$p(x_t, \theta_t \mid y_{1:t}) = \frac{p(y_t \mid x_t, \theta_t)\, p(x_t, \theta_t \mid y_{1:t-1})}{p(y_t \mid y_{1:t-1})} \tag{3}$$

where $p(x_t, \theta_t \mid y_{1:t-1})$ represents the prior information; $p(y_t \mid x_t, \theta_t)$ is the likelihood; $p(y_t \mid y_{1:t-1})$ represents the normalizing constant. If the model is assumed to be Markovian, the prior distribution can be estimated via the Chapman-Kolmogorov equation:

$$p(x_t, \theta_t \mid y_{1:t-1}) = \int p(x_t, \theta_t \mid x_{t-1}, \theta_{t-1})\, p(x_{t-1}, \theta_{t-1} \mid y_{1:t-1})\, dx_{t-1} d\theta_{t-1} \tag{4}$$

Similarly, the normalizing constant $p(y_t \mid y_{1:t-1})$ can be obtained as follows:

$$p(y_t \mid y_{1:t-1}) = \int p(y_t \mid x_t, \theta_t)\, p(x_t, \theta_t \mid y_{1:t-1})\, dx_t d\theta_t \tag{5}$$

The optimal Bayesian solutions (i.e. equations (3) and (4)) are difficult to determine since the evaluation of the integrals might be intractable (Plaza Guingla et al., 2013). Consequently, approximate methods are applied to treat above issues.

Ensemble Kalman Filter (EnKF) and particle filter (PF) are the two widely used methods, in which EnKF can recursively result in optimal estimation for linear dynamic models with Gaussian uncertainties, and PF is suitable for non-Gaussian nonlinear dynamical models (Xie and Zhang, 2013). Particularly, the PF can provide a more accurate update for model states and parameters by adjusting the hyperparameters (e.g., observation perturbation characteristics) based on the observations and ensemble predictions, which avoid excessive adjustment of the ensemble spread while still allowing for a relatively quick response when observations fall outside the prediction bound (Moradkhani, 2008; Leisenring and Moradkhani, 2012). The central idea of EnKF and PF is to quantify the probability density functions (PDF) of model states by a set of random samples. The difference between these two methods lies in the way of recursively generating an approximation for a state PDF (Weerts and EI Serafy, 2005). In EnKF, the distributions are considered to be Gaussian. The Monte Carlo approach is applied to approximate the error statistics and compute the Kalman gain matrix for updating model parameters and state variables.

Consider a general stochastic dynamic model with the transition equations of the system state expressed as:

$$x^{-}_{t+1,i} = f(x^{+}_{t,i}, u_{t,i}, \theta^{-}_{t+1,i}) + \omega_{t,i}\text{ , } i = 1, 2, \ldots, ne \tag{6}$$

where x_t is the states vector at time t; θ is the system parameters vector assumed to be known and time invariant; the superscript "-" indicates the "forecasted" sates; the superscript "+" indicates the "analyzed" states; *ne* represents the number of ensembles; u_t is the input vector (deterministic forcing data); *f* represents the model structure; ω_t is the model error term, which follows a Gaussian distribution with zero mean and covariance matrix $\sum_t^m$. For the evolution of the parameters, it is assumed that the parameters follow a random walk presented as:

$$\theta_{t+1,i}^{-} = \theta_{t,i}^{+} + \tau_{t,i} \,, \quad \tau_{t,i} \sim N(0, \textstyle\sum_t^\theta) \tag{7}$$

Prior to update of the model states and parameters, an observation equation is applied to transfer the states into the observation space, which can be characterized as:

$$y_{t+1,i}^{-} = h(x_{t+1,i}^{-}, \theta_{t+1,i}^{-}) + v_{t+1,i} \,, \quad v_{t+1,i} \sim N(0, \textstyle\sum_{t+1}^{y}) \tag{8}$$

where y_{t+1} is the observation vector at time t +1; h is the measurement function relating the state variables and parameters to the measured variables; $v_{k+1,i}$ reflects the measurement error, which is also assumed to be Gaussian with zero mean and covariance matix $\sum_{t+1}^{y}$. The model and observation errors are assumed to be uncorrelated, i.e. $E[\omega_t v_{t+1}^T] = 0$. After the prediction is obtained, the posterior states and parameters are estimated with the Kalman update equations as follows (DeChant and Moradkhani, 2012):

$$x_{t+1,i}^{+} = x_{t+1,i}^{-} + K_{xy}[y_{t+1} + \varepsilon_{t+1,i} - y_{t+1,i}^{-}] \tag{9}$$

$$\theta_{t+1,i}^{+} = \theta_{t+1,i}^{-} + K_{\theta y}[y_{t+1} + \varepsilon_{t+1,i} - y_{t+1,i}^{-}] \tag{10}$$

where y_t is the observed values; $\varepsilon_{t,i}$ represents the observation errors; K_{xy} and $K_{\theta y}$ are the Kalman gains for states and parameters, respectively (DeChant and Moradkhani, 2012):

$$K_{xy} = C_{xy}(C_{yy} + R_t)^{-1} \tag{11}$$

$$K_{\theta y} = C_{\theta y}(C_{yy} + R_t)^{-1} \tag{12}$$

Here C_{xy} is the cross covariance of the forecasted states $x_{t+1,i}^{-}$ and the forecasted output $y^{-}_{t+1,i}$; $C_{\theta y}$ is the cross covariance of the parameter ensembles $\theta^{-}_{t+1,i}$ with the predicted observation $y^{-}_{t+1,i}$; C_{yy} is the variance of the predicted observation; R_t is the observation error variance at time t.

2.2. Probabilistic Collocation Method (PCM)

2.2.1. Polynomial chaos expression (PCE)

For a system dynamic model, its outputs are correlated to its input fields. In terms of random characteristics in model inputs, the outputs can be characterized by a nonlinear function with respect to the set of random variables. Polynomial chaos (PC) methods are usually applied to express the evolution of uncertainty in a dynamic system with random inputs. The PC method was first introduced by Wiener (1938), where the model stochastic process is decomposed by Hermite polynomials in terms of Gaussian random variables. The polynomial chaos expansion (PCE) can be seen as a mathematically optimal way to construct and obtain a model response surface in the form of a high-dimensional polynomial to uncertain model parameters (Oladyshkin and Nowak, 2012). This technique includes representing the system outputs through a polynomial chaos basis of random variables which are used to represent input stochasticity, and deriving the unknown expansion coefficients using intrusive (e.g. stochastic Galerkin technique) and non-intrusive (e.g. probabilistic collocation method) approaches. The original PCE is based on Hermite polynomials, which are optimal for normally distributed random variables (Oladyshkin and Nowak, 2012). However, for non-Gaussian random input variables (e.g. Gamma and uniform), the convergence of Herminte 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:

$$y = a_0 + \sum_{i_1=1}^{n} a_{i_1}\Gamma_1(\zeta_{i_1}) + \sum_{i_1=1}^{n}\sum_{i_2=1}^{i_1} a_{i_1 i_2}\Gamma_2(\zeta_{i_1},\zeta_{i_2}) + \sum_{i_1=1}^{n}\sum_{i_2=1}^{i_1}\sum_{i_3=1}^{i_2} a_{i_1 i_2 i_3}\Gamma_3(\zeta_{i_1},\zeta_{i_2},\zeta_{i_2}) + \ldots \tag{13}$$

where y is the output and $\Gamma_p(\zeta_{i_1},\zeta_{i_2},\ldots,\zeta_{i_p})$ are the polynomial chaos of order p 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:

$$\Gamma_p(\zeta_{i_1},\zeta_{i_2},\ldots,\zeta_{i_p}) = (-1)^p e^{1/2\zeta^T\zeta} \frac{\partial^M}{\partial\zeta_{i_1}\partial\zeta_{i_2}\ldots\partial\zeta_{i_p}} e^{-1/2\zeta^T\zeta} \tag{14}$$

where $(\zeta_{i_1},\zeta_{i_2},\ldots,\zeta_{i_p})$ (ζ is the vector form) are the standard normal random variables (SRV). The polynomial with an order greater than one has zero mean; polynomials of different orders are orthogonal to each other, and so are polynomials of the same order but with different arguments (Huang et al., 2007).

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). The computational requirement increases as the order of PCE increases. The total number of the truncated terms N for PCE is related to the dimension of the random variables M and the highest order of the polynomial p:

$$N = \frac{(M+p)!}{M!\,p!} \tag{15}$$

Table 1 contains some explicit values of N for given dimension of the random variables M and the order of the polynomial p. Thus Equation (10) can be written simply as:

$$y = \sum_{j=0}^{M-1} c_j \Psi_j(\zeta_i) \tag{16}$$

in which there is a one-to-one mapping between $\Gamma_p(\zeta_{i_1}, \zeta_{i_2}, ..., \zeta_{i_p})$ and $\Psi_j(\zeta_i)$, and also between c_j and $a_{i_1 i_2 ... i_p}$. For instance, the 2-order 2-dimensional PCE can be expressed as: $y = c_0 + c_1\zeta_1 + c_2\zeta_2 + c_3(\zeta_1^2 - 1) + c_4(\zeta_2^2 - 1) + c_5\zeta_1\zeta_2$; the 2-order 3-dimensional PCE can be written as: $y = c_0 + c_1\zeta_1 + c_2\zeta_2 + c_3\zeta_3 + c_4(\zeta_1^2 - 1) + c_5(\zeta_2^2 - 1) + c_6(\zeta_3^2 - 1) + c_7\zeta_1\zeta_2 + c_8\zeta_1\zeta_3 + c_9\zeta_2\zeta_3$.

Place Table 1 Here

2.2.2. Selection of collocation points for PCM

The basic idea of the probabilistic collocation method (PCM) is to let the polynomial chaos expansion (PCE) in terms of random inputs to be the same as the model simulation results at selected collocation points. The collocation points can be specified by various algorithms. In this study, the collocation points are derived from combinations of the roots of a Hermite polynomial with one order higher than the order of PCE. For a 2-order PCE, the collocation points are combinations of the roots of the 3-order Hermite polynomial $H_3(\zeta) = \zeta^3 - 3\zeta$, which are ($-\sqrt{3}$, 0, $\sqrt{3}$). For example, for a 2-order 2-dimensional PCE expressed as: y = $y = c_0 + c_1\zeta_1 + c_2\zeta_2 + c_3(\zeta_1^2 - 1) + c_4(\zeta_2^2 - 1) + c_5\zeta_1\zeta_2$, the collocation points ($\zeta_{1,i}$, $\zeta_{2,i}$) are chosen from the combinations of the three roots of the 3-order Hermite polynomial, which consists of a total of 9 collocation points which are expressed as: ($-\sqrt{3}$,$-\sqrt{3}$) ($-\sqrt{3}$, 0), ($-\sqrt{3}$, $\sqrt{3}$), (0, $-\sqrt{3}$), (0, 0), (0, $\sqrt{3}$), ($\sqrt{3}$,$-\sqrt{3}$), ($\sqrt{3}$, 0), ($\sqrt{3}$, $\sqrt{3}$). For a 3-order PCE, the collocation points are chosen based on the values of $\pm\sqrt{3 \pm \sqrt{6}}$, which are the roots of the 4-order Hermite polynomial $H_4(\zeta) = \zeta^4 - 6\zeta^2 + 3$. Furthermore, the selection is also expected to capture regions of high probability (Huang et al., 2007; Li and Zhang, 2007). The value of zero has the highest probability for a standard normal random variable, and thus the collocation points for 3-order PCE are the combinations of (0, $\pm\sqrt{3 \pm \sqrt{6}}$). The potential collocation points for the 2- and 3-order PCEs with two standard random variables are presented in Table 2.

Place Table 2 Here

2.2.3. Unknown Parameter Estimation

Probabilistic collocation method (PCM) is 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). Generally, there are two methods to obtain the unknown coefficients in PCE. The first one is to solve a linear equations system expressed as: $N \times a = f$, where N is a space-independent matrix of dimension P × P, consisting of Hermite polynomials evaluated at the selected collocation points; a is the unknown coefficient vector of the PCE; f is the realization of the simulation model at the selected collocation points. However, such a method may be unstable and the approximation results are highly dependent on the selection of the collocation points (Huang et al., 2007). Consequently, Huang et al. (2007) modified the collocation method to employ more collocation points than the number of unknown coefficients through a regression based method. In this study, we will employ the regression-based method to obtain the unknown coefficients in PCE. The detailed process for PCM method is illustrated in Figure 1.

Place Figure 1 here

2.3. Uncertainty Quantification for the Hydrological Model based on Coupled Ensemble Filtering and Probabilistic Collocation (EFPC) Method

Hydrologic models contain parameters that cannot be measured directly, and must therefore be estimated using measurements of the system inputs and outputs (Vrugt et al., 2005). Sequential data assimilation (SDA) is a class of methods that provide a general framework for explicitly dealing with input, output and model structural uncertainties. Of these SDA techniques, the ensemble Kalman filter (EnKF) is one of the most widely used methods in hydrologic community (Moradkhani et al., 2005a; DeChant and Moradkhani, 2012; Leisenring and Moradkhani, 2011; Li et al., 2013; Liu et al., 2012). The EnKF method is much more effective for reducing uncertainty and characterizing posterior distributions for model parameters as it can merge the observations and model outputs to improve the model predictions, and further characterize the initial condition of uncertainty of the catchment. However, uncertainty propagation and evolution from model parameters to model outputs can hardly be revealed just merely through EnKF. Consequently, in this study, we will integrate the ensemble Kalman filter (EnKF) and the probabilistic collocation methods (PCM) into a general framework to quantify the uncertainty of hydrological predictions. The posterior probability distributions of model parameters are estimated by EnKF, and the uncertainty propagation and evolution from uncertainty parameters to model outputs are further characterized by PCM.

2.3.1. Gaussian Anamorphosis Transformation for Non-Gaussian Distributions

When the polynomial chaos expansion (PCE) is applied to express the evolution of uncertainty in a dynamic system with random inputs, those random inputs should be transformed to random variables with specific distributions. For example, as proposed in Equation (14), for the stochastic process decomposed by Hermite polynomials, the random inputs should be first expressed through the standard Gaussian random variables. The EnKF method can continuously update the states and parameters in the model when new measurements become available. After the EnKF update process, the distributions of model parameters can hardly be normally distributed, even though their prior distributions are assumed to be normal. Moreover, the distributions of the updated parameters can hardly be expressed through some specific distributions (e.g. gamma, uniform, etc.) in many cases.

Consequently, in order to further quantify the inherent uncertainty of the hydrologic model after the data assimilation process, transformation techniques should applied to convert the posterior distributions of the updated parameters into standard Gaussian distributions. In this study, a nonlinear, monotonic transform technique known as Gaussian anamorphosis (GA), will be applied to transform the posterior distributions of model parameters to standard normal distributions. For the original random variable x and the transformed random variable $y = f(x)$, the idea of GA is to find a function f to define a change of the variable (anamorphosis) such that the random variable y obeys a standard Gaussian distribution. Such a transformation technique was applied in biogeochemical ocean model (Simon and Bertino, 2009), physical-biogeochemical ocean model (Béal et al., 2010) and subsurface hydraulic tomography model (Schöniger et al., 2012). In this study, the GA method will be applied to combine the EnKF and PCM method together to quantify the uncertainty of hydrologic models.

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

$$z = G^{-1}(F(y)) \tag{17}$$

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 (17) is called 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:

$$F_j = \frac{j-0.5}{N} \tag{18}$$

where j are the rank of the sample value of y; N is the sample size of y (rendered as the ensemble size of EnKF in this study). From Equations (17) and (18), the sample values of the Gaussian transform variable z can be obtained, which correspond to the sample values of y. Also, the sample range of z can be determined as follows:

$$z_{\min} = G^{-1}(\frac{1-0.5}{N}) \tag{19}$$

$$z_{\max} = G^{-1}(\frac{N-0.5}{N}) \tag{20}$$

2.3.2. The Detailed Procedures of the EFPC mehtod

The process of the proposed EFPC method mainly involves two components: the EnKF update procedures for uncertainty reduction and the PCM procedures for uncertainty quantification. The detailed process of EFPC includes the following steps:

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

Step (2). Model state forecast step: propagate the *ne* state variables and model parameters forward in time using model operator *f*:

$$x^-_{t+1,i} = f(x^-_{t,i}, u_{t,i}, \theta^-_{t+1,i}) + \omega_{t+1,i}\ , \omega_{t+1} \sim N(0, \Sigma^m_t)\ , i = 1, 2, \ldots, ne$$

Step (3). Observation simulation: use the observation operator *h* to propagate the model state forecast:

$$y^-_{t+1,i} = h(x^-_{t+1,i}, \theta^-_{t+1,i}) + v_{t+1,i}\ ,\ \ v_{t+1,i} \sim N(0, \Sigma^y_{t+1})\ , i = 1, 2, \ldots, ne$$

Step (4). Parameters and states updating: update the parameters and states via the EnKF updating equations:

$$x^+_{t+1,i} = x^-_{t+1,i} + K_{xy}[y_{t+1} + \varepsilon_{t+1,i} - y^-_{t+1,i}]$$

$$\theta^+_{t+1,i} = \theta^-_{t+1,i} + K_{\theta y}[y_{t+1} + \varepsilon_{t+1,i} - y^-_{t+1,i}]$$

Step (5) Parameter perturbation: take parameter evolution to the next stage through adding small stochastic error around the sample:

$$\theta^-_{t+2,i} = \theta^+_{t+1,i} + \tau_{t+1,i}\ ,\ \ \tau_{t+1,i} \sim N(0, \Sigma^\theta_{t+1})$$

Step (6). 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 (7). Convert the parameter θ into standard Gaussian variables through GA.

Step (8). Approximate the outputs of interest using the polynomial chaos expansion in terms of the standard Gaussian variables.

Step (9). Select the collocation points according to the dimensions of the stochastic vector and the order of the applied polynomial chaos expansion.

Step (10). Determine the unknown coefficients in the polynomial expansion through statistical regression techniques.

Step (11). Evaluate the inherent statistical properties of the outputs stemming from the uncertainty of the parameters.

3. Experimental Setup

3.1. The Conceptual Hydrologic Model

The Hymod, which is a well-known conceptual hydrologic model, will be used in this study. Hymod is a non-linear rainfall-runoff conceptual model which can be run in a minute/hour/daily time step (Moore, 1985). The general concept of the model is based on the 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 (Moore, 1985, 2007):

$$F(c) = 1 - \left[1 - \frac{c}{C_{\max}}\right]^{b_{\exp}}, \ 0 \le c \le C_{max} \tag{21}$$

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 Figure 2, the 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, (iii) the factor partitioning the flow between the two series of linear reservoir tanks α, (iv) the residence time of the linear quick-flow tank R_q, and (v) the residence time of the slow-tank R_s. The model uses two input variables: mean areal precipitation, P (mm/day), and potential evapotranspiration, ET (mm/day).

Place Figure 2 Here

3.2. Site Description

The Xiangxi River basin, located in the Three Gorges Reservoir area (Figure 3), China, is selected to demonstrate the effectiveness of the proposed forecasting algorithm. The Xiangxi River is located between 30.96 ~ 31.67 ^{0}N and 110.47 ~ 111.13^{0}E in the Hubei part of the China Three Gorges Reservoir (TGR) region, with a draining area of approximately 3,200 km^2. The Xiangxi River originates in the Shennongjia Nature Reserve with a main stream length of 94 km and a catchment area of 3,099 km^2 and is one of the main tributaries of the Yangtze River (Han et al., 2014; Yang and Yang, 2014; Miao et al., 2014). The watershed experiences a northern subtropical climate. The annual precipitation is about 1,100 mm and ranges from 670 to 1,700 mm with considerable spatial and temporal variability (Xu et al., 2010; Zhang et al., 2014). The main rainfall season is from May through September, with a flooding season from July to August. The annual average temperature in this region is 15.6 ^{0}C and ranges from 12 ^{0}C to 20 ^{0}C.

Place Figure 3 here

3.3. Synthetic Data Experiment

In this study, a synthetic case will be initially applied to demonstrate the applicability of the EFPC method in quantifying prediction uncertainty. For the synthetic experiment, "truth" is defined when the model is run for a set of meteorological and initial conditions (Moradkhani, 2008). In detail, the model parameter values are predefined as the "true" values presented in Table 3. The model inputs, including the potential evapotranspiration, ET (mm/day), and mean areal precipitation, P (mm/day), are the observed data collected at Xingshan Hydrologic Station (110^{0}45'0'' E, 31^{0}13'0'' N) on the main stream of the Xiangxi River. These data are provided by the Water Conservancy Bureau of Xiangshan County. Using these model inputs and parameter values, the "true states" and "true streamflow observations" can be generated by running Hymod. Such generated streamflow values are considered as the observations in the EnKF updating process. Moreover, as with any data assimilation framework, it is necessary to assume error values for any quantity that contains uncertainties (DeChant and Moradkhani, 2012). In the synthetic experiment, the model structure is assumed to be perfect. Thus, random perturbations would be added to precipitation and potential evapotranspiration (ET) observations to account for their uncertainties. In this study, these random perturbations are assumed to be normally distributed with the mean values being 0 and the standard errors being proportional to the magnitude of true values. The proportional coefficients for precipitation, potential evapotranspiration, and streamflow observations are all set to be 0.1. This means that precipitation, ET, and streamflow observations are assumed to have normal distributions with relative errors of 10%. However the study proposed by DeChant and Moradkhani (2011a; 2011b) showed that the log-norm perturbation for precipitation is more appropriate. The comparison among norm and log-norm perturbation for precipitation will be conducted in the subsequent real-case study.

Place Table 3 Here

3.4. Evaluation Criteria

To evaluate the performance of the proposed EFPC approach, some indices are introduced. In detail, root-mean-square error (RMSE), the Nash-Sutcliffe efficiency (NSE) coefficient and the percent bias (%BIAS) will be employed to evaluate the performance of the proposed method, which are expressed as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Q_i - P_i)^2} \tag{22}$$

$$NSE = 1 - \frac{\sum_{i=1}^{N}(Q_i - P_i)^2}{\sum_{i=1}^{N}(Q_i - \overline{Q})^2} \tag{23}$$

$$PBIAS = \frac{\sum_{i=1}^{N}(Q_i - P_i)*100}{\sum_{i=1}^{N}Q_i} \tag{24}$$

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

4. Results Analysis of Synthetic Experiment

4.1. Uncertainty Characterization of Hymod through EnKF

To demonstrate the capability of EnKF in model parameter estimation and uncertainty reduction, the five parameters of Hymod (i.e. C_{max}, b_{exp}, α, R_q R_s) are initialized to be varied within predefined intervals, as presented in Table 3. The ensemble size in this study was set to be 50. This ensemble size is set based on the conclusion from Yin et al. (2015). They tested the optimal ensemble size of EnKF in sequential soil moisture data and found that the standard deviation decreases sharply with ensemble size increasing when the ensemble size was less than 10, and this tendency was to slow down when the ensemble size was greater than 10 (Yin et al., 2015). Particularly, for larger ensemble sizes, the error variance did not decrease much further, suggesting that the EnKF estimates at the final times might not converge to the optimal smoothing solution when the ensemble size became too large (Yin et al., 2015). The random perturbation for parameter evolution in Equation (7) is set to have a normal distribution with a relative error of 10%. The initial samples of the five parameters are uniformly sampled from those predefined intervals and the total data assimilation steps would be one year (i.e. 365 days).

Figure 4 shows the comparison between the ensembles of the forecasted streamflow and the synthetic-generated true discharge. The results indicate that the ensemble means of streamflow predictions can track the observed discharge data. The ranges formulated by 5 and 95% percentiles (i.e. 90% confidence intervals) of streamflow predictions can adequately bracket the observations. Figure 5 depicts the evolution of the sampled marginal posterior distributions for the five parameters of Hymod during the EnKF assimilation period. From Figure 5, it is observed that b_{exp}, α, R_q and R_s are identifiable, while in comparison, the C_{max} parameter is less identifiable than the other four parameters. This means that the marginal distribution of C_{max} exhibits considerable uncertainty and move intermittently throughout the feasible parameter space. For b_{exp}, α, R_q and R_s, one year discharge observations are deemed sufficient to estimate their values. Table 3 presents the final fluctuating intervals for these five parameters after one year data assimilation period. It is indicated that the EnKF method estimated C_{max} b_{exp}, R_s accurately, while there are small differences between the true values and the final estimated intervals for α and R_q. The extensive uncertainty of C_{max} indicates that, in this synthetic experiment, the C_{max} is low sensitivity to the model prediction performance.

Place Figures 4 and 5 Here

In this study, we set the ensemble size to be 50. To confirm the effectiveness of this ensemble size, we compare the performance of EnKF under different ensemble sizes. In detail, six ensemble size scenarios are assumed, and under each scenario, the synthetic experiment is run 10 times. The results of the mean values of NSE, RMSE, and PBIAS are presented in Table 4. The results show that as the increase in ensemble size, the performance of EnKF would not be improved significantly; conversely, EnKF performed slightly worse as ensemble size larger than 150. This may because that the EnKF estimates at the final times might not converge to the optimal smoothing solution when the ensemble size became too large (Yin et al., 2015). Therefore, in this study, the ensemble size being 50 seems to be appropriate in this study.

Place Table 4 Here

EnKF can merge the observations and model outputs to improve the model predictions, and further characterize the initial condition uncertainty. The posterior probability distributions for model parameters can be estimated through EnKF, and the uncertainty in model parameters can be significantly reduced. However, as presented in Table 3, the parameters of Hymod still contain some uncertainties. These uncertainties may result from random errors in the precipitation, potential evaporation, streamflow observation and model prediction. Consequently, further exploration would be required to characterize uncertainty propagation in hydrologic simulation and analyze the inherent statistic characteristics of the hydrologic predictions after data assimilation.

4.2. Uncertainty Quantification of Hymod through the Probabilistic Collocation Method.

In this study, the Hermite polynomial chaos expansion is employed to quantify the evolution of uncertainty in Hymod stemming from the uncertain parameters.

Consequently, the posterior distributions of model parameter estimated by EnKF would be firstly converted into standard Gaussian distribution. As presented in Table 3, after the data assimilation process through EnKF, there is still some extent of uncertainty existing in the five parameters of Hymod. Since the value of R_q changes within a very small interval (i.e. [0.75, 0.76]), it will be considered to be deterministic in further uncertainty quantification through PCM. The other four parameters (i.e. C_{max}, b_{exp}, α, R_s) are transformed to standard Gaussian distributions according to GA method proposed by Equations (17) - (19). Figure 6 shows the histogram of original data, empirical anamorphosis function, histogram of transformed data, and normal probability plot of transformed data for C_{max}. Obviously, after transformation through GA, the sample values of C_{max} are well fitted to a standard Gaussian distribution. Similarly, the posterior distributions of b_{exp}, α, R_s can also be converted to standard Gaussian distributions through the GA method. These transformed data can be introduced into the PCM method to further quantify the uncertainty of Hymod.

Place Figure 6 Here

The 2-order polynomial chaos expansion (PCE) is employed to quantify the uncertainty in the Hymod predictions. Since there are four parameters in Hymod (i.e. C_{max}, b_{exp}, α, R_s), the PCE used to represent the output of interest (i.e. streamflow) would be four-dimensional and two order. The detailed polynomials of the 4-dimensional 2-order PCE are expressed by Equation (16). There are total of 15 unknown coefficients in this 4-dimensional 2-order PCE. The potential collocation points are obtained through combining the roots (i.e. (-$\sqrt{3}$, 0, $\sqrt{3}$)) of the 3-order Hermite polynomial $H_3(\zeta) = \zeta^3 - 3\zeta$. For a 4-dimensional 2-order PCE, there are 81 (i.e. 3^4) potential collocation points. For each collocation point, the probability can be obtained through the standard CDF *G* in Equation (17), and consequently, the corresponding rank *j* can be calculated through Equation (18). Since *j* may not be an integer, the original value of C_{max}, b_{exp}, α, or R_s corresponding to the collocation point of ζ would be obtained through linear interpolation method based on the two adjacent original data. In this paper, all the collocation points would be used to establish the linear regression equations and generate the values of unknown coefficients of PCE. Afterward, 2,000 values are independently sampled from the standard Gaussian distribution for ζ_1, ζ_2, ζ_3, and ζ_4, respectively, and 2,000 realizations would be generated through both the obtained PCE and Hymod. The latter 2,000 realizations obtained through Hymod are considered as Monte Carlo simulation results.

Figure 7 shows the comparison for the mean values of the streamflow obtained through 2-order PCE and Monte Carlo (MC) simulation methods. It indicates that the mean values obtained through 2-oder PCE are highly identical to the MC simulation results. This means that the 2-order PCE can generally replace the hydrologic model (i.e. Hymod) to reflect the temporal variations for the streamflow. Figure 8 compares the standard deviations of the streamflow, at each time step, obtained through 2-order PCE and MC simulation methods, respectively. It suggests that the standard deviation of 2-order PCE and MC simulation is identical at low uncertain conditions (i.e. low standard deviation values). During the high streamflow periods, the standard deviation obtained by the 2-order PCE would be slightly less than the actual values (i.e. MC results). However, the PCE results would generally fit well with the MC simulation results in both means and standard deviations. As shown in Figure 9, the relative errors between the standard deviations from MC simulation and 2-order PCE prediction results are relatively small, and most of them are located within [-0.10, 0.10]. Moreover, Figure 10 shows the comparison between the 90% confidence intervals from the MC simulation and 2-order PCE prediction results. It indicates that the predicted intervals of streamflow from MC simulation and 2-order PCE are highly consistent under the 90% confidence level.

Place Figures 7 to 10 Here

To further compare the accuracy between 2-order PCE and MC simulation results, the detailed statistical characteristics would be analyzed at specific time periods. The specific time periods are selected artificially through screening the mean streamflow values, as shown in Figure 7, over the simulation period so that the low, medium, and high streamflow levels are all considered. Consequently, the streamflow predictions from MC simulation and PCE at the day 23, 145, 181, 182, 218, 350 are chosen, and their inherent statistical properties are further analyzed. These statistical properties, including mean, standard deviation, kurtosis and skewness, are presented in Table 5. The results show that the probability density distributions obtained through 2-order PCE would be similar with those obtained by MC simulation. However, the shape of those probability density distributions generated by 2-order PCE would be slightly steeper (i.e. lower standard deviation and higher kurtosis) than those from MC simulation method. For example, at the 181th day, the mean, standard deviation, kurtosis, skewness values obtained by 2-order PCE would be 613.59, 76.32, 3.01, 0.23, respectively, while those values generated by MC simulation method would be 615.01, 84.43, 2.07, 0.12, respectively. Figure 11 shows the histograms of 2-order PCE and MC simulation results at the selected time periods. In Figure 11, the left column in each subfigure represents the histogram obtained through MC method, while the right one express the histogram obtained by PCE results. It can be seen from Figure 11 that the shapes of the probability distributions obtained by 2-order PCE have similar shapes with those obtained from the MC simulation results. This suggests that the PCE model obtained by the proposed EFPC can be effective to replace the original hydrologic model to characterize the uncertainty in hydrologic predictions.

Place Table 5 and Figure 11 Here

Generally, after the data assimilation process by EnKF, the uncertainty of Hymod would be significantly reduced, and the posterior probability of model parameters would be estimated. The probabilistic collocation method (PCM) can further characterize the uncertainty propagation through establishing a PCE model between the model parameters and model outputs. Such a model can well reveal uncertainty evolution in hydrologic simulations. Even based on the 2-order PCE, the mean and standard deviation values of this PCE model would be consistent with those obtained by MC simulation method. Moreover, the detailed probability densities generated by 2-order PCE at each time step would have similar shapes than those obtained through MC simulation method.

5. Real Case Study

5.1. Model Setup

A real-case study will be performed to further demonstrate the applicability of the proposed EFPC method in quantifying uncertainty for hydrologic models. This real-case study is set up based on on-site measurements for daily precipitation, potential evapotranspiration, and streamflow discharge from 1991 to 1993 at the Xingshan Hydrologic Station on the Xiangxi River.

The EnKF method can quantify model errors, which may be caused by uncertainties in model inputs, structures, and parameter values, by using the variance of streamflow predictions from an ensemble of model realizations (McMillan, 2013). Random perturbations are added to model inputs, outputs, and parameters to reflect their inherent uncertainties. In the synthetic experiment, random perturbations were added to precipitation and potential evapotranspiration (ET) observations, which were normally distributed with standard errors being 10% of the true values. In order to investigate the impact of relative errors on the performance of EnKF, five relative error scenarios would be assumed. In detail, precipitation is assumed to be normally distributed with relative error being 10, 15, 20, 25, and 30% of the true values, respectively, and ET is also normally distributed having the same relative errors. For the streamflow measurements, several studies set the standard deviation of the observed error to be proportional to the true discharge (Dechant and Moradkhai, 2012; Moradkhani et al., 2012; Abaze, et al., 2014), while some research works assumed the error to be proportional to the log discharge (Clark et al, 2008; McMillan et al., 2013). In our study, five relative errors would be selected (i.e. 10, 15, 20, 25 and

30%) in order to characterize their impacts on the performance of EnKF. Also, these five error scenarios are assumed to account for the uncertainty in the model predictions.

5.2. Impact of Stochastic Perturbation on the Performance of EnKF

Table 6 shows the performance of EnKF under different relative error scenarios. The results indicate that the stochastic perturbation can influence the performance of EnKF. In detail, large relative errors may better reflect the uncertainties in the catchment, and thus leading to better model performance. In this study, the performance of EnKF would be improved as the relative error increases from 10% to 20%. However, such a trend would not keep going as the relative larger than 20%. Consequently, for the Xiangxi River, the relative error of 20% may be the appropriate stochastic perturbation to account for the uncertainties in the precipitation, potential evapotranspiration and streamflow observation.

Place Table 6 Here

5.3. Uncertainty Quantification

Based on the EnKF approach, the posterior probabilities of model parameters would be identified. However, uncertainties in hydrologic predictions, stemming from the uncertainties in hydrologic parameters, are still required to be characterized. Previous research works mainly address this issue through the Monte Carlo method, in which random samples are drawn from the posterior distributions of hydrologic parameters to run the original hydrologic model (Lu and Zhang, 2003; Khu and Werner, 2003; Demaria et al., 2007). This approach may be insufficient, especially for complex hydrologic models, which requires a large number of runs to establish a reliable estimate of model uncertainties (Khu and Werner, 2003). Moreover, traditional Monte Carlo method can hardly reveal how these model parameters would affect the uncertainties in model predictions. Therefore, the developed ensemble filtering and probabilistic collocation (EFPC) method can better address the above issues, in which the posterior probabilities of model parameters would be estimated through EnKF and the probabilistic collocation method (PCM) would be further proposed to establish a proxy for the hydrologic model, with respect to the posterior distributions of model parameters, to reveal the uncertainty evolution in the hydrologic simulation.

The results in Table 6 show that a relative error of 0.2 may be appropriate to account for the inherent uncertainty in the Xiangxi River. The potential evapotranspiration, streamflow observations, and model predictions are normally distributed with the standard errors being 20% of the true values. For the precipitation, it is first assumed to be normally distributed with a relative error of 20%. Based on the proposed EFPC approach, a polynomial chaos expansion (PCE) can be obtained at each time period, which expresses the relationship between the discharge prediction and the uncertain model parameters.

Figure 12 shows the comparison between predicting means of hydrologic model and observations as well as PCE results and observations. This figure is obtained under the assumption of normal error distribution for precipitation. Figure 12(a) indicates the mean predictions of hydrologic model and observations. The mean predictions in Figure 12(a) are obtained through Monte Carlos method in which the parameters values of the hydrologic model are sampled based on their posterior probabilities estimated through EnKF. Figure 12(b) shows the mean predictions of PCE and observations. This figure suggests that the predictions from hydrologic model and PCE show similar trend. The mean predictions from both hydrologic model and PCE can well track the observed streamflow data, except some underestimates during some extreme flow periods. To evaluate the performance of hydrologic model and PCE obtained by the proposed EFPC method, the values of RMSE, PBIAS, and NSE are calculated based on the prediction means and observations. Table 7 compares the results of RMSE, PBIAS, and NSE values obtained through the original hydrologic model and PCE. The comparison process is as follows: (i) choosing N samples from the standard Gaussian distribution, (ii) generating the associated parameter values of the hydrologic model based on the relationships between posterior distributions and standard Gaussian distribution established by the GA approach, (iii) running PCE and hydrologic model respectively, (iv) obtaining the evaluation criteria results. The results in Table 7 indicate good performance of hydrologic model and PCE in tracking the streamflow dynamics in the Xiangxi River, with high NSE values and low PBIAS and RMSE values. Particularly, the hydrologic model performs slightly better than the PCE approach. This is because the PCEs generated by the proposed EFPC method is a proxy of the hydrologic model. However, the results in Table 7 suggest that the PCE can adequately represent the hydrologic model. Figure 13 compares the 90% confidence intervals of hydrologic model vs. observations and 90% confidence intervals of PCE predictions vs. observations. This figure shows that 90% prediction intervals from hydrologic model and PCE can encompass most observations.

Place Table 7 and Figures 12 and 13 Here

As recommended by DeChant and Moradkhani (2011a; 2011b), the log-norm perturbation for precipitation is more appropriate. Thus the proposed EFPC approach is further tested through adding 20% log-normal perturbation to the precipitation and 20% normal perturbations for the model prediction, streamflow observation, and potential evapotranspiration. Table 8 shows related RMSE, PBIAS, and NSE values. Compared with results in Table 7, adding log-normal perturbation in the precipitation can improve the performance of the proposed method, with the NSE value larger than 0.7. Figure 14 presents the comparison between predictions from the hydrologic model and observations as well as PCE results and observations. Figure 15 compares prediction intervals from the hydrologic model and PCE with observations. Both of them show good agreement between model predictions and real observations.

Place Table 8 and Figures 14 and 15 Here

5.4. Computational Efficiency of the EFPC Method

The essential ideal of the EFPC approach is to use the ensemble Kalman filter method to estimate the posterior distributions of model parameters and then apply probabilistic collocation method (PCM) to reveal the uncertainty evolution of hydrologic models. Such a method has two advantages in quantifying the uncertainty in hydrologic simulation: (i) the original samples can be drawn from the standard

Gaussian distribution, which is easily conducted; (ii) the computational efficiency can be highly improved.

The first advantage is straightforward. The second advantage of EFPC will be illustrated through comparing it with traditional Monte Carlo (MC) method. Tables 6 and 7 shows the computation efficiency of Monte Carlo method and PCE which are obtained through the proposed EFPC method. In this study, five sample sizes (n = 500, 1,000, 1,500, 2,000, 2,500) are selected to compare the computation efficiency of MC and the obtained PCE through EFPC. As the sample size increases, the performance of the hydrologic model and PCE would not vary significantly. Both the hydrologic model and PCE produce satisfactory streamflow forecasting in the Xiangxi River. However, the computational efficiency of PCE would be more than ten times faster than the MC method. For example, when n = 500, the computational time of MC method would be 54.7 (s), as shown in Table 7, while the computational time of PCE is just 5.3 (s). The ratio of computational efficiency between PCE and MC (time (MC)/time (PCE)) is 10.3. Such a ratio would increase for larger sample sizes (e.g. the ratio is 11.9 for n = 2,500). Consequently, the proposed EFPC approach would greatly improve the computational efficiency for uncertainty quantification of hydrologic models

In this study, the Hymod was applied to demonstrate the efficiency of the proposed approach. This model is a simple conceptual hydrologic model with five parameters to calibrate. Consequently, the computational requirement for this model is relatively low when compared with other sophisticated models such as semi-distributed and distributed hydrologic models. However, the proposed EFPC approach is more than 10 times faster in computational efficiency for such a simple hydrologic model. The computational efficiency would be improved even more significantly for other complex hydrologic models.

5.4. Uncertainty Assessment of Model Parameters

One of the most attraction features for the proposed method is that the polynomial chaos expansion (PCE), with respect to the posterior probabilities of model parameters, can be obtained through the proposed EFPC approach. Such a PCE model can explicitly reveal the contributions of model parameters and their interactions to the total variation in model predictions.

In this study, the 5-dimensional 2-order PCE is advanced to reflect the uncertainty propagation of model uncertainty resulting from uncertainty in model parameters. The detailed expression for a 5-dimensional 2-order PCE can be expressed as: $y = a_0 + a_1\zeta_1 + a_2\zeta_2 + a_3\zeta_3 + a_4\zeta_4 + a_5\zeta_5 + a_6(\zeta_1^2 - 1) + a_7(\zeta_2^2 - 1) + a_8(\zeta_3^2 - 1) + a_9(\zeta_4^2 - 1) + a_{10}(\zeta_5^2 - 1) + a_{11}\zeta_1\zeta_2 + a_{12}\zeta_1\zeta_3 + a_{13}\zeta_1\zeta_4 + a_{14}\zeta_1\zeta_5 + a_{15}\zeta_2\zeta_3 + a_{16}\zeta_2\zeta_4 + a_{17}\zeta_2\zeta_5 + a_{18}\zeta_3\zeta_4 + a_{19}\zeta_3\zeta_5 + a_{20}\zeta_4\zeta_5$, where ζ_1, ζ_2, ζ_3, ζ_4, ζ_5 are independent standard normal variable representing C_{max}, b_{exp}, α, R_q and R_s, respectively. Since the variables ζ_1, ζ_2, ζ_3, ζ_4, ζ_5 are standard normal variables, the variance of y can be easily derived, which can be obtained as: $\mathrm{Var}(y) = \mathrm{Var}(a_0 + a_1\zeta_1 + a_2\zeta_2 + a_3\zeta_3 + a_4\zeta_4 + a_5\zeta_5 + a_6(\zeta_1^2 - 1) + a_7(\zeta_2^2 - 1) + a_8(\zeta_3^2 - 1) + a_9(\zeta_4^2 - 1) + a_{10}(\zeta_5^2 - 1) + a_{11}\zeta_1\zeta_2 + a_{12}\zeta_1\zeta_3 + a_{13}\zeta_1\zeta_4 + a_{14}\zeta_1\zeta_5 + a_{15}\zeta_2\zeta_3 + a_{16}\zeta_2\zeta_4 + a_{17}\zeta_2\zeta_5 + a_{18}\zeta_3\zeta_4 + a_{19}\zeta_3\zeta_5 + a_{20}\zeta_4\zeta_5) = a_1^2 + a_2^2 + a_3^2 + a_4^2 + a_5^2 + 2a_6^2 + 2a_7^2 + 2a_8^2 + 2a_9^2 + 2a_{10}^2 + a_{11}^2 + a_{12}^2 + a_{13}^2 + a_{14}^2 + a_{15}^2 + a_{16}^2 + a_{17}^2 + a_{18}^2 + a_{19}^2 + a_{20}^2$. Such an expression can explicitly reflect the contribution of the variation in model parameters to the uncertainty of model predictions.

Figure 16 shows the comparison of the contributions for different parameters to the total uncertainty in model predictions. The variance ratio is calculated through the coefficients of the obtained PCE and the total variance. For instance the variance ratio of the main effect for C_{max} is generated by a_1^2/ Var(y). As shown in Figure 16, for the main effect of each parameter, namely ζ_1, ζ_2, ζ_3, ζ_4, ζ_5, the variable of ζ_5, indicating the parameter R_q, contributes most to the total variance in model predictions, and also ζ_3 and ζ_4, which respectively represent α and R_s, present apparent contributions to the uncertainty in model outputs. For the quadratic terms, ζ_5^2 would be most sensitive to the uncertainty in model predictions, but other quadratic terms do not show apparent contributions, with all the values less than 0.1 in most simulation periods. Moreover, as shown in Figure 16(c), the interactions among those five parameters only contribute slightly to the variance in model predictions, with the highest variance ratio less than 0.06. Among these interactive effects, the interaction between ζ_3 and ζ_5 contributes most to the total variance, followed by the interaction between ζ_3 and ζ_4.

Place Figure 16 Here

The proposed EFPC approach can effectively quantify the uncertainty propagation in model simulation resulting from uncertainty model parameters. Particularly, the obtained PCEs are able to express how the uncertainty in model parameters can affect the uncertainty in model predictions, and further identify the main, quadratic and interactive effects of model parameters on the variation in model outputs. Moreover, based on the obtained PCEs, the global sensitivity analysis can be easily conducted without running the original hydrologic model through Monte Carlo method. Such PCE-based global sensitivity analysis has been conducted in our forthcoming paper (Fan et al., 2015b).

6. Conclusions

Hydrologic models are designed to simulate the rainfall-runoff processes through conceptualizing and aggregating the complex, spatially distributed and highly interrelated water, energy, and vegetation processes in a watershed into relatively simple mathematical equations. A significant consequence of process conceptualization is that the model parameters exhibit extensive uncertainties, leading to significant uncertainty in hydrologic forecasts. This study proposed an integrated framework for uncertainty quantification of hydrologic models through a coupled ensemble filtering and probabilistic collocation (EFPC) approach. This developed EFPC method combined the backward and forward uncertainty quantification methods together, in which the backward uncertainty quantification method (i.e. EnKF) was employed to reduce model uncertainty and improve the forecast accuracy based on the observed measurements, and the forward method (i.e. PCM) was further used to quantify the inherent uncertainty of the hydrologic model after a data assimilation process.

The conceptual hydrologic model, Hymod, was used to demonstrate the applicability of the proposed method in quantifying uncertainties of the hydrologic forecasts. A synthetic experiment was firstly conducted based on a short simulation period (i.e. 365 days). A set of predefined values for model parameters of Hymod were provided to generate streamflows which were considered as the observations in the EnKF adjusting process. After one-year data assimilation process by EnKF, the uncertainty of model parameters (i.e. b_{exp}, α, R_s, R_q) was significantly reduced except the parameter C_{max}. Meanwhile, the uncertainty of the Hymod predictions was also reduced. Afterward, a probabilistic collocation method (PCM) was used to quantify the uncertainty in the Hymod predictions. In PCM, a 4-dimensional 2-order polynomial chaos expansion (PCE) (R_q is considered to be deterministic) was used to approximate the forecasted streamflow, and all potential collocation points were applied to formulate linear regression equations to estimate the unknown coefficients in PCE. The results indicated that the PCE reflected the uncertainty of the streamflow results. The mean and standard deviation values of PCE were consistent with those obtained by Monte Carlo (MC) simulation method, except slight errors existing in the standard deviation values. For the detailed probability density functions, the histograms formulated by the PCE predictions hold similar but slightly steeper shapes to the MC simulation results.

The proposed EFPC method was then applied to a real-world watershed in the Three Gorges Reservoir area in China. The impact of relative errors was evaluated for the performance of EnKF for estimating the posterior distributions of hydrologic model parameters. The results showed that 20% of relative error may be appropriate to account for the uncertainties in precipitation, potential evapotranspiration, and streamflow observations in Xiangxi River. The results showed that the polynomial chaos expansion (PCE) is a good representation of the hydrologic model for streamflow forecasting and uncertainty quantification. Specifically, the efficiency of the PCE would be more than 10 times faster than the hydrologic model.

This study proposed a coupled ensemble filtering and probabilistic collocation (EFPC) method for quantifying the uncertainty of hydrologic models. The innovation of this study is to integrate EnKF and PCM into a framework, in which the posterior distributions of model parameters are estimated through EnKF, and the uncertainty propagation and evolution from model parameters to hydrologic predictions are characterized by the probabilistic collocation method. Compared with a classic Monte Carlo simulation method, the proposed method can be easily implemented, avoiding drawing samples from arbitrary probability distributions. The computation efficiency can be highly improved by the proposed method.

Acknowledgments: This research was supported by the Natural Science Foundation of China (No. 51190095 and 51225904) and the Program for Innovative Research Team in University (IRT1127).

References

Abaza M., Anctil F., Fortin V., Turcotte R., (2014). Sequential streamflow assimilation for short-term hydrological ensemble forecasting, Journal of Hydrology, doi 10.1016/j.jhydrol.2014.08.038

Ajami N.K., Duan Q.Y., Sorooshian S., (2007). An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction. *Water Resources Research*, 43, W01403.

Assumaning G.A., Chang S.Y., (2014). State and Parameter Estimation in Three-Dimensional Subsurface Contaminant Transport Modeling using Kalman Filter Coupled with Monte Carlo Sampling, Journal of Environmental Informatics, 24(2), 80-89.

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 Science*, 6, 247-263.

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

Blanchard E.D., (2010). Polynomial Chaos Approaches to Parameter Estimation and Control Design for Mechanical Systems with Uncertain Parameters. Ph.D Thesis, Department of Mechanical Engineering, Virginia Tech University.

Chang S.Y., Sayemuzzaman M., (2014). Using Unscented Kalman Filter in Subsurface Contaminant Transport Models. *Journal of Environmental Informatics* 23(1), 14-22

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. *Advances in Water Resources*, 31, 1309-1324.

Dechant, C., & Moradkhani, H. (2011a). Radiance data assimilation for operational snow and streamflow forecasting. *Advances in Water Resources*, 34(3), 351-364.

DeChant, C. M., & Moradkhani, H. (2011b). Improving the characterization of initial condition for ensemble streamflow prediction using data assimilation. *Hydrology and Earth System Sciences*, 15(11), 3399-3410.

DeChant C.M., Moradkhani H., (2012). Examining the effectiveness and robustness of sequential data assimilation methods for quantification of uncertainty in hydrologic forecasting. *Water Resources Research*, 48, W04518, doi:10.1029/2011WR011011

DeChant C.M., and H. Moradkhani (2014a), Toward a Reliable Prediction of Seasonal Forecast Uncertainty: Addressing Model and Initial Condition Uncertainty with Ensemble Data Assimilation and Sequential Bayesian Combination, Journal of Hydrology , 519, 2967-2977, doi: 10.1016/j.jhydrol.2014.05.045.

DeChant, C.M. and H. Moradkhani (2014b), *Hydrologic Prediction and Uncertainty Quantification*. Handbook of Engineering Hydrology, Modeling, Climate Change and Variability, CRC press, Taylor & Francis Group, PP 387-414.

Demaria E.M., Nijssen B., Wagener, T., (2007). Monte Carlo sensitivity analysis of land surface parameters using the Variable Infiltration Capacity model. *Journal of Geophysical Research*, 112, D11113.

Diks C.G.H., Vrugt J.A., (2010). Comparison of point forecast accuracy of model averaging methods in hydrologic applications. *Stochastic Environmental Research and Risk Assessment*, 24, 809-820.

Evensen, G. (1994), Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *Journal Geophysical Research*, 99(C5), 10,143–10,162.

Fan Y.R., Huang G.H., Huang K., Baetz B.W., (2015a). Planning Water Resources Allocation under Multiple Uncertainties through A Generalized Fuzzy Two-Stage Stochastic Programming Method. IEEE Transactions on Fuzzy Systems, doi 10.1109/TFUZZ.2014.2362550

Fan Y.R., Huang G.H., Li Y.P., (2012). Robust interval linear programming for environmental decision making under uncertainty. *Engineering Optimization* 44 (11), 1321-1336

Fan Y.R., Huang G.H., Huang K., Li Y.P., (2015b). Screening Uncertainty and Parameter Sensitivities of Hydrologic Models through a Hybrid Sequential Data Assimilation and Probabilistic Collocation Method. Submitted to *Journal of Hydrometeorology*

Fan Y.R., Huang W.W., Huang G.H., Huang K., Zhou X., (2015c). A PCM-based stochastic hydrological model for uncertainty quantification in watershed systems. *Stochastic Environmental Research and Risk Assessment*, 29, 915-927

Han J.C., Huang G.H., Zhang H., Li Z., Li Y.P., (2014). Bayesian uncertainty analysis in hydrological modeling associated with watershed subdivision level: a case study of SLURP model applied to the Xiangxi River watershed, China. *Stochastic Environmental Research and Risk Assessment*. 28(4), 973-989

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. *Hydrology and Earth System Sciences*, 17, 2893-2903.

Houska T., Multsch S., Kraft P., Frede H.G., Breuer L., (2014). Monte Carlo-based calibration and uncertainty analysis of a coupled plant growth and hydrologic model. Biogeosciences, 11, 2069-2082.

Huang S., Mahadevan S., Rebba R., (2007). Collocation-based stochastic finite element analysis from random field problems. *Probabilistic Engineering Mechanics* 22, 194–205

Johnson R.A., Wichern D.W., (1988). *Applied multivariate statistical analysis*, Prentice Hall International, NJ, USA

Kalman R.E., (1960). A new approach to linear filtering and prediction problems. *Transactions of*

the ASME–Journal of Basic Engineering, 82(1) 35-46.
Kewlani G., Crawford J., Iagnemma K., (2012). A polynomial chaos approach to the analysis of vehicle dynamics under uncertainty. Vehicle System Dynamics, 50, 749-774.
Khu S.T., Werner M.G.F., (2003). Reduction of Monte-Carlo simulation runs for uncertainty estimation in hydrological modelling. *Hydrology and Earth System Sciences*, 7(5), 680-692
Knighton J., White E., Lennon E., Rajan R., (2014). Development of probability distributions for urban hydrologic model parameters and a Monte Carlo analysis of model sensitivity, Hydrologic Processes, 28(19), 5131-5139.
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. *Stochastic Environmental Research and Risk Assessment* 29, 833-846.
Leisenring M., Moradkhani H., (2011). Snow water equivalent prediction using Bayesian data assimilation methods. *Stochastic Environmental Research and Risk Assessment*, 25, 253-270
Leisenring, M., & Moradkhani, H. (2012). Analyzing the uncertainty of suspended sediment load prediction using sequential data assimilation. Journal of Hydrology, 468, 268-282.
Li H., Zhang D., (2007). Probabilistic collocation method for flow in porous media: comparisons with other stochastic method. *Water Resources Research*, 43.
Li Y., Ryu D., Western A. W., Wang Q. J., (2013). Assimilation of stream discharge for flood forecasting: The benefits of accounting for routing time lags, *Water Resources Research*, 49, 1887–1900, doi:10.1002/wrcr.20169.
Liu Y., Weerts A.H., Clark M., Hendricks Franssen H.-J., Kumar S., Moradkhani H., Seo D.-J., Schwanenberg 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. *Hydrology and Earth System Sciences*, 16, 3863-3887.
Lu Z., Zhang D., (2003). On importance sampling Monte Carlo approach to uncertainty analysis for flow and transport in porous media. *Advances in Water Resources* 26, 1177–1188
Lucas D.D., Prinn R.G., (2005). Parametric sensitivity and uncertainty analysis of dimethylsulfide oxidation in the clear-sky remote marine boundary layer. Atmospheric Chemistry Physics, 5, 1505-1525.
Madadgar, S. and H. Moradkhani (2014). Improved Bayesian Multi-modeling: Integration of Copulas and Bayesian Model Averaging. *Water Resources Research*, 50, 9586-9603, doi: 10.1002/2014WR015965.
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, Hydrology and Earth System Science, 17, 21-38.
Miao D.Y., Huang W.W., Li Y.P., Yang Z.F., (2014). Planning Water Resources Systems under Uncertainty Using an Interval-Fuzzy De Novo Programming Method, Journal of Environmental

Informatics, 24(1), 11-23.

Moore R.J., (1985). The probability-distributed principle and runoff production at point and basin scales. *Hydrological Sciences Journal*, 30, 273-297

Moore R.J., (2007). The PDM rainfall-runoff model. *Hydrology & Earth System Sciences*, 11(1), 483-499.

Moradkhani, H. (2008). Hydrologic remote sensing and land surface data assimilation. Sensors, 8(5), 2986-3004.

Moradkhani, H., S. Sorooshian, H. V. Gupta, and P. Houser (2005a), Dual state – parameter estimation of hydrologic models using ensemble Kalman filter. *Advances in Water Resources*, 28, 135 – 147.

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 Resources Research, 41(5)

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 Resources Research, 48, W12520, doi:10.1029/2012WR012144.

Oladyshkin S., Nowak W., (2012). Data-driven uncertainty quantification using the arbitrary polynomial chaos expansion, *Reliability Engineering and System Safety*, 106, 179-190.

Parrish, M., H. Moradkhani, and C.M. DeChant (2012). Towards Reduction of Model Uncertainty: Integration of Bayesian Model Averaging and Data Assimilation, Water Resources Research, 48, W03519, doi:10.1029/2011WR011116.

Plaza Guingla D. A., De Keyser R., De Lannoy G. J. M., Giustarini L., Matgen P., and 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 Resource Research*, 49, doi:10.1002/wrcr.20291.

Reichle R., Mclaughlin D., Entekhabi D., (2002). Hydrologic data assimilation with the ensemble Kalman filter. *Monthly Weather Review*, 130(1), 103-114.

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 Resources Research*, 48, W04502.

Shi L., Yang J., Zhang D., Li H., (2009). Probabilistic collocation method for unconfined flow in heterogeneous media. *Journal of Hydrology* 365 4–10

Shi, Y., K. J. Davis, F. Zhang, C. J. Duffy, and X. Yu (2014), Parameter estimation of a physically based land surface hydrologic model using the ensemble Kalman filter: A synthetic experiment. Water Resources Research, 50, 1-19, doi:10.1002/2013WR014070

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 Science*, 5, 495-510 van Delft G., (2007). The Ensemble Particle Filter in Rainfall-Runoff Models. M.S. Thesis, Delft University of Technology, Delft, Netherlands.

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 Resources Research*, 41, W01017.

Vrugt J.A., Sadegh M., (2013). Toward diagnostic model calibration and evaluation: Approximate Bayesian computation. *Water Resources Research* 49, 4335-4345.

Vrugt J.A., ter Braak C.J.F., Gupta H.V., Robinson B.A., (2009). Equifinality of formal (DREAM) and informal (GLUE) Bayesian approchches in hydrologic modeling? *Stochastic Environmental Research and Risk Assessment*, 23, 1011-1026.

Wang Dingbao, Chen Yuguo, Cai Ximing, (2009). State and parameter estimation of hydrologic models using the constrained ensemble Kalman filter. *Water Resources Research*, 45, W11416

Weerts A.H., EI Serafy G.Y.H., (2006). Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models. *Water Resources Research*, 42, W09403. doi:10.1029/2005WR004093

Wiener N., (1938). The homogeneous chaos. American Journal of Mathematics, 60, 897-936.

Xie X., Zhang D., (2013). A partitioned update scheme for state-parameter estimation of distributed hydrologic models based on the ensemble Kalman filter. *Water Resources Research*, 49, 7530-7365

Xiu, D., Karniadakis, G.E., (2002). The Wiener–Askey polynomial chaos for stochastic differential equations. *SIAM Journal on Scientific Computing* 24, 619–644.

Xiu, D., Karniadakis, G.E., (2003). Modeling uncertainty in flow simulations via generalized polynomial chaos. *Journal of Computational Physics* 187, 137–167.

Xu H.M., Taylor R.G., Kingston D.G., Jiang T., Thompson J.R., Todd M.C., (2010). Hydrological modeling of River Xiangxi using SWAT2005: A comparison of model parameterizations using station and gridded meteorological observations. Q*uaternary International*, 226, 54-59

Yang W., Yang Z.F., (2014). Evaluation of Sustainable Environmental Flows Based on the Valuation of Ecosystem Services: a Case Study for the Baiyangdian Wetland, China. *Journal of Environmental Informatics*, 24(2), 90-100

Yin J., Zhan X., Zheng Y., Hain C.R., Liu J., Fang L., (2015). Optimal ensemble size of ensemble Kalman filter in sequential soil moisture data assimilation. *Geophysical Research Letters*, DOI: 10.1002/2015GL063366

Zhang N., Li Y.P., Huang W.W., Liu J., An Inexact Two-Stage Water Quality Management Model for Supporting Sustainable Development in a Rural System. Journal of Environmental Informatics, 24(1), 52-64.

Zhang Y., Li H., Yang D., (2012a). Simultaneous estimation of relative permeability and capillary pressure using ensemble-based history matching techniques. *Transport in 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., Yang D., (2013). Simultaneous estimation of relative permeability and capillary pressure for tight formations using ensemble-based history matching method. *Computers & 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 Problems in Science and Engineering* 22 (7), 1150-1175

Zheng Y., Wang W., Han F., Ping J., (2011). Uncertainty assessment for watershed water quality modeling: A Probabilistic Collocation Method based approach. *Advances in Water Resources* 34 887–898

List of Figure Captions

Figure 1. The process of the PCM approach

Figure 2. Description of Hymod

Figure 3. The location of the studied watershed

Figure 4. Comparison between the ensembles of the forecasted and synthetic-generated true discharge

Figure 5. Convergence of the parameter through the EnKF for the synthetic experiment over data assimilation period

Figure 6. Histogram of untransformed variable, empirical CDF, histogram of transformed variable, and normal probability plot for C_{max} (unit (mm)).

Figure 7. The comparison between the mean values of MC simulation and 2-order PCE results

Figure 8. The comparison between the standard deviation values of MC simulation and 2-order PCE results

Figure 9. The distribution of the relative errors between the standard deviations from MC simulation and 2-order PCE prediction results

Figure 10. The comparison between the percentile values of MC simulation and 2-order PCE results

Figure 11. The comparison of histograms between MC simulation and 2-order PCE results (note: the left is for MC and the right is for PCE)

Figure 12 Comparison between the predication means and observations under normal error assumption for precipitation: (a) hydrologic model predictions vs. observations, (b) PCE results vs. observation

Figure 13. Comparison between the predication intervals and observations under normal error assumption for precipitation: (a) hydrologic model prediction intervals vs. observation, (b) PCE predicting intervals vs. observation

Figure 14. Comparison between the predication means and observations under lognormal error assumption for precipitation: (a) hydrologic model predictions vs. observations, (b) PCE results vs. observation

Figure 15. Comparison between the predication intervals and observations under lognormal error assumption for precipitation: (a) hydrologic model prediction intervals vs. observation, (b) PCE predicting intervals vs. observation

Figure 16. Contributions of model parameters to the uncertainty in model predictions over the simulation period

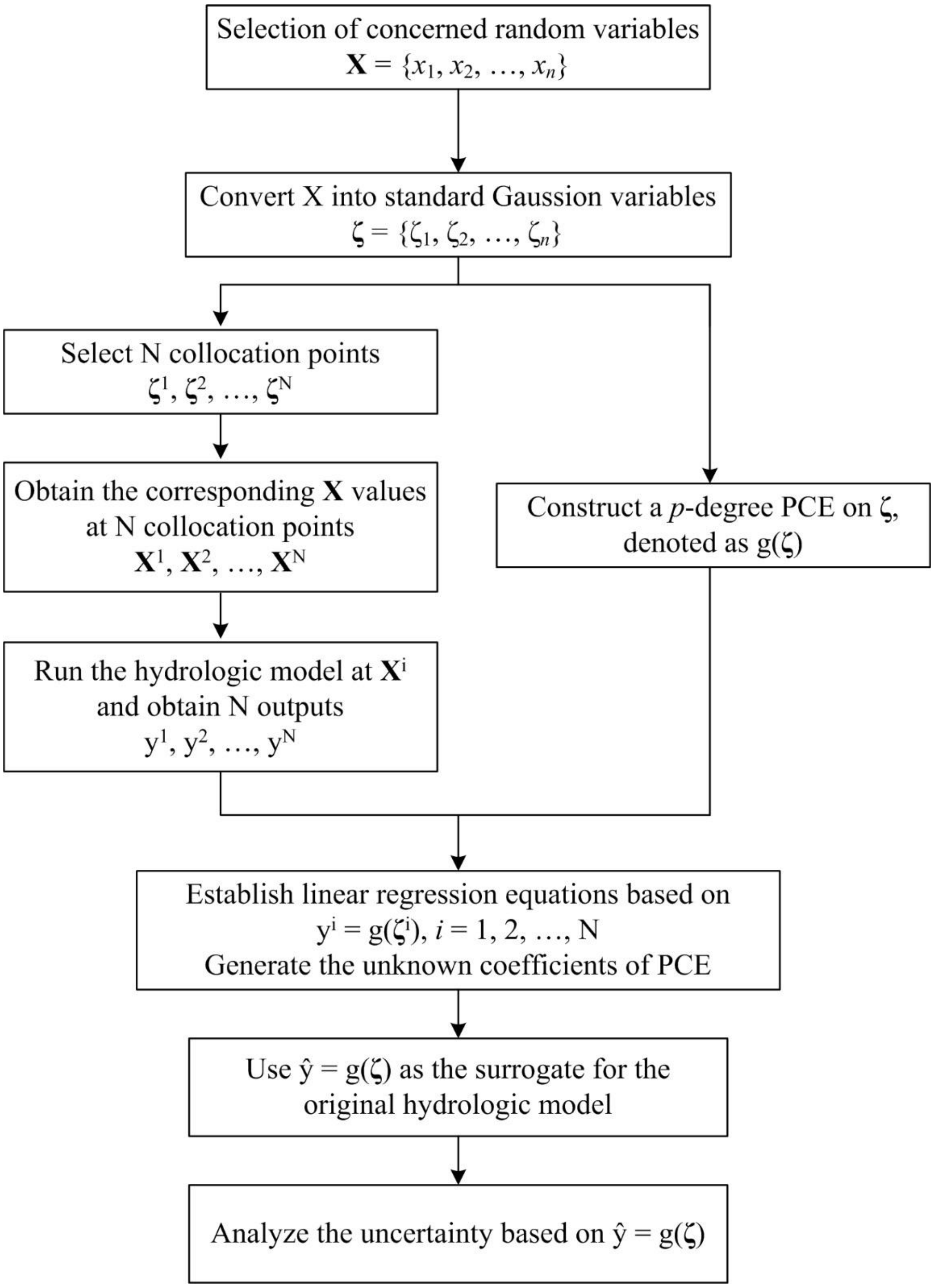
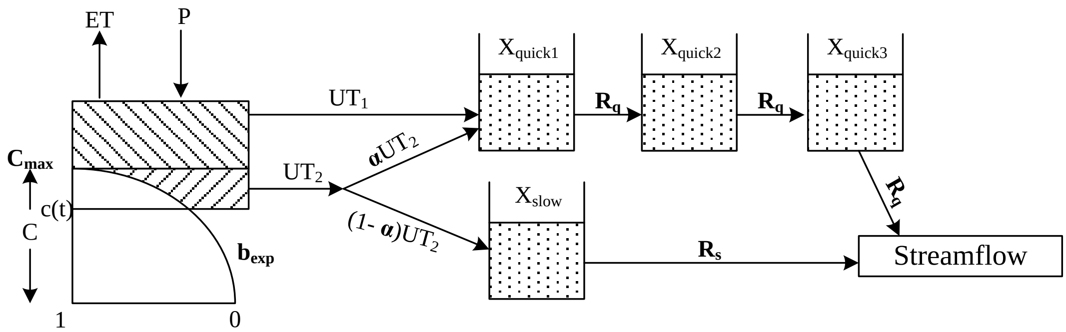


Figure 1. The process of the PCM approach

1191



1192 Figure 2 Description of Hymod (modified from Vrugt et al., 2003)

1193

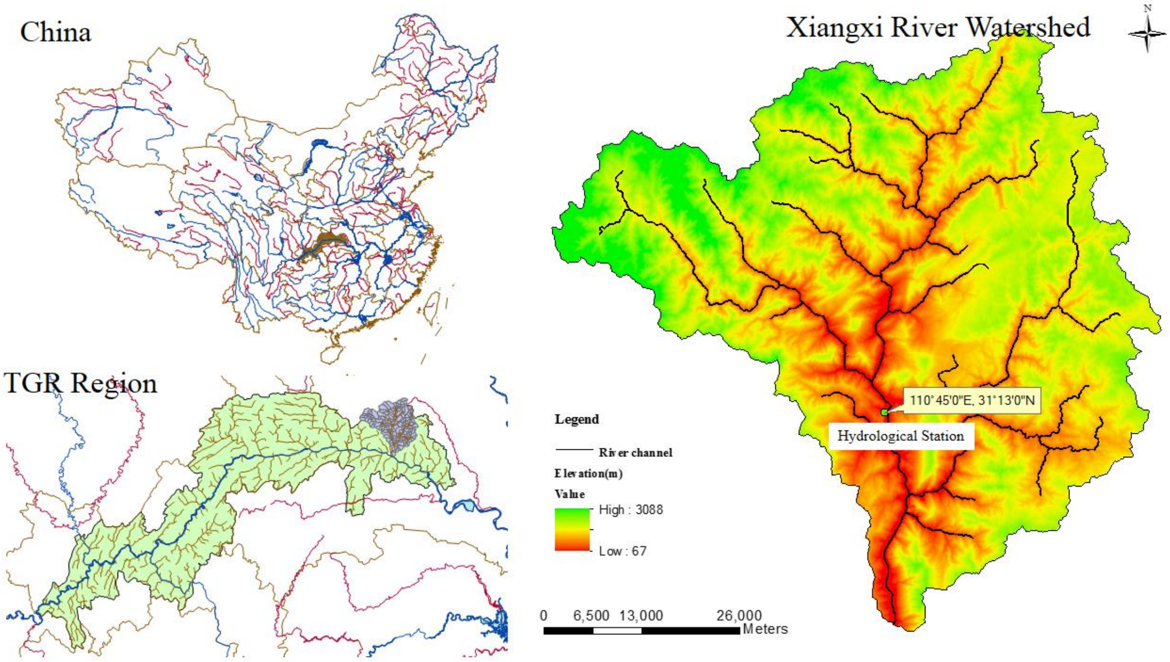


Figure 3: The location of the studied watershed

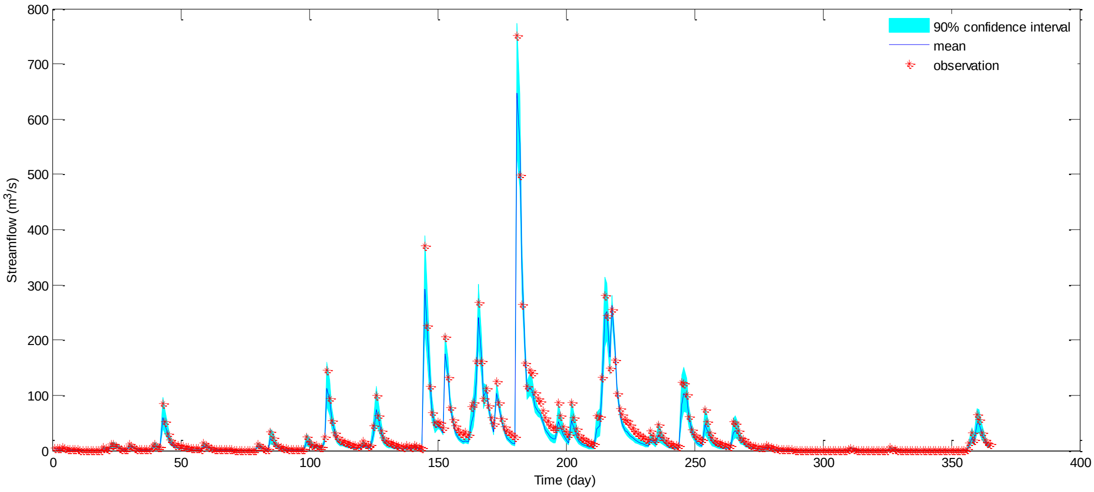


Figure 4. Comparison between the ensembles of the forecasted and synthetic-generated true discharge

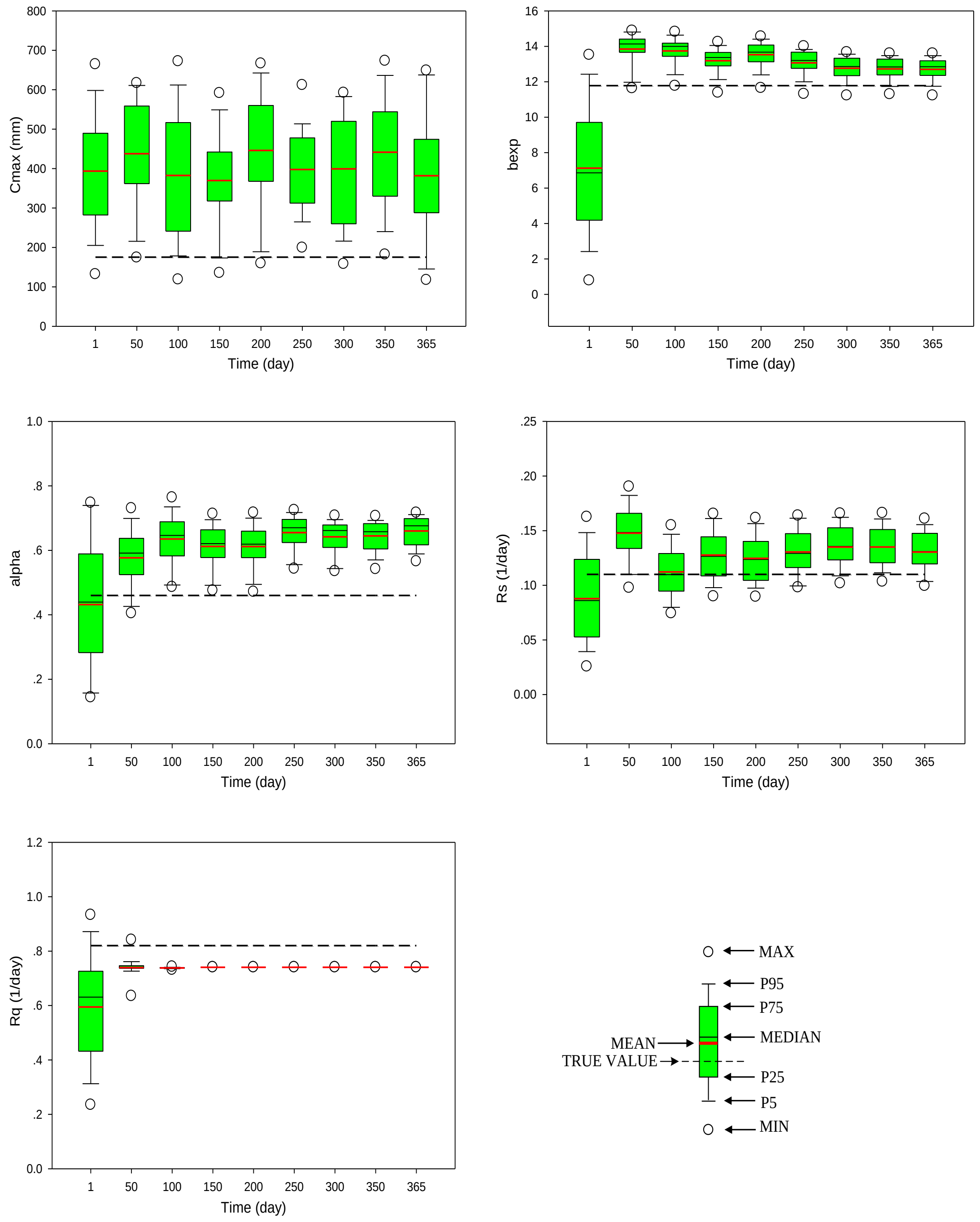


Figure 5. Convergence of the parameter through the EnKF for the synthetic experiment over data assimilation period

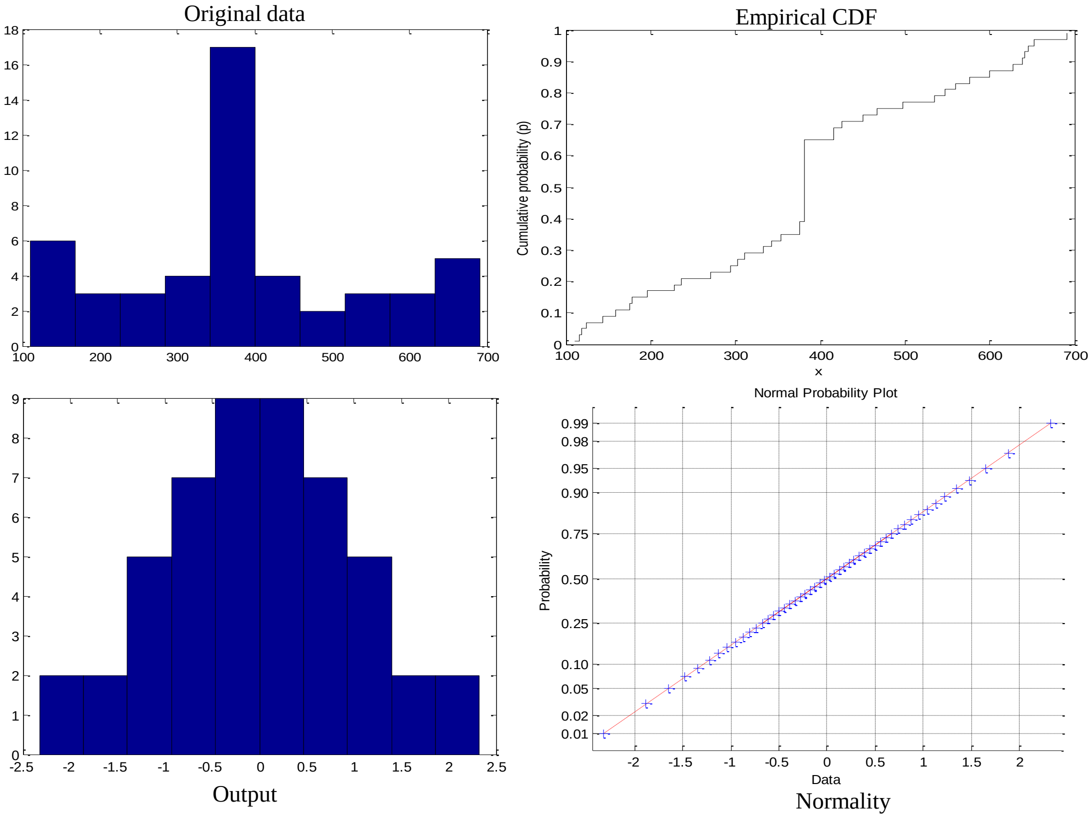


Figure 6. Histogram of untransformed variable, empirical CDF, histogram of transformed variable, and normal probability plot for C_{max} (unit (mm)).

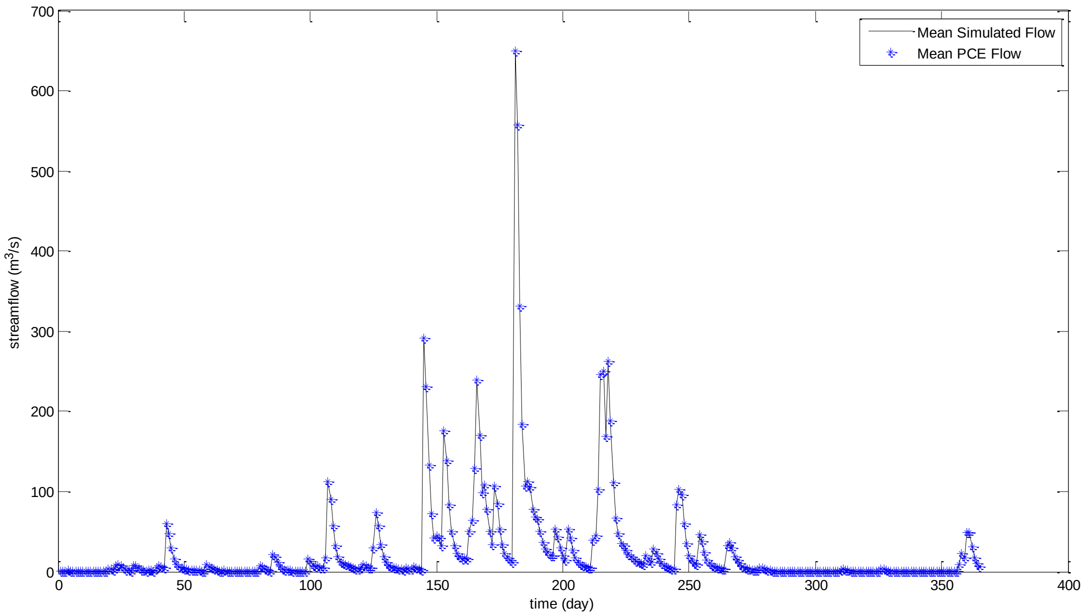


Figure 7. The comparison between the mean values of the MC simulation and 2-order PCE results

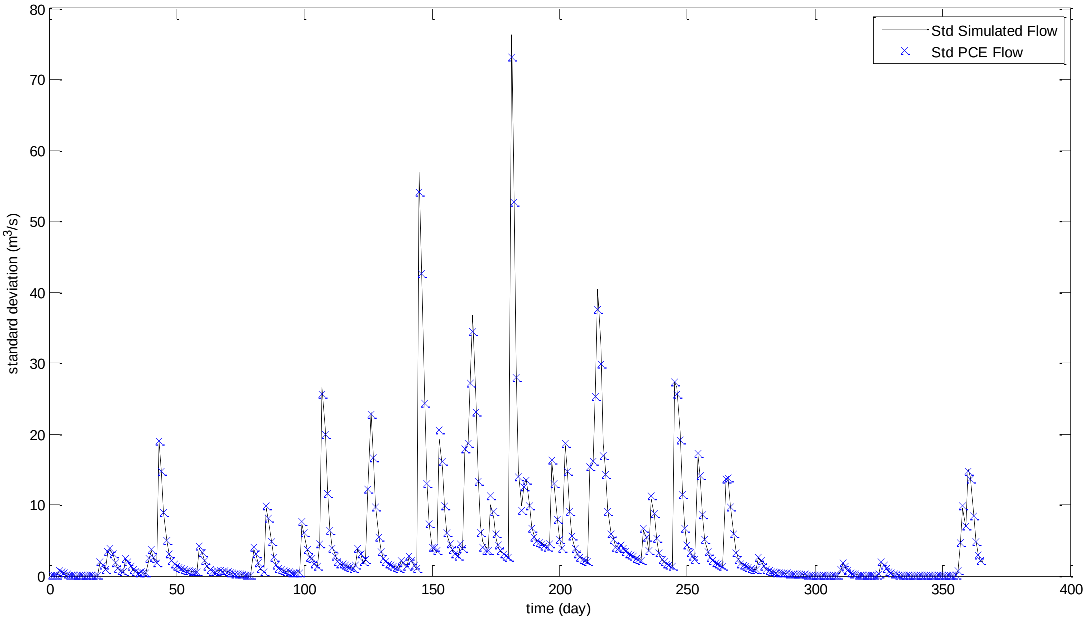


Figure 8. The comparison between the standard deviation values of MC simulation and 2-order PCE results

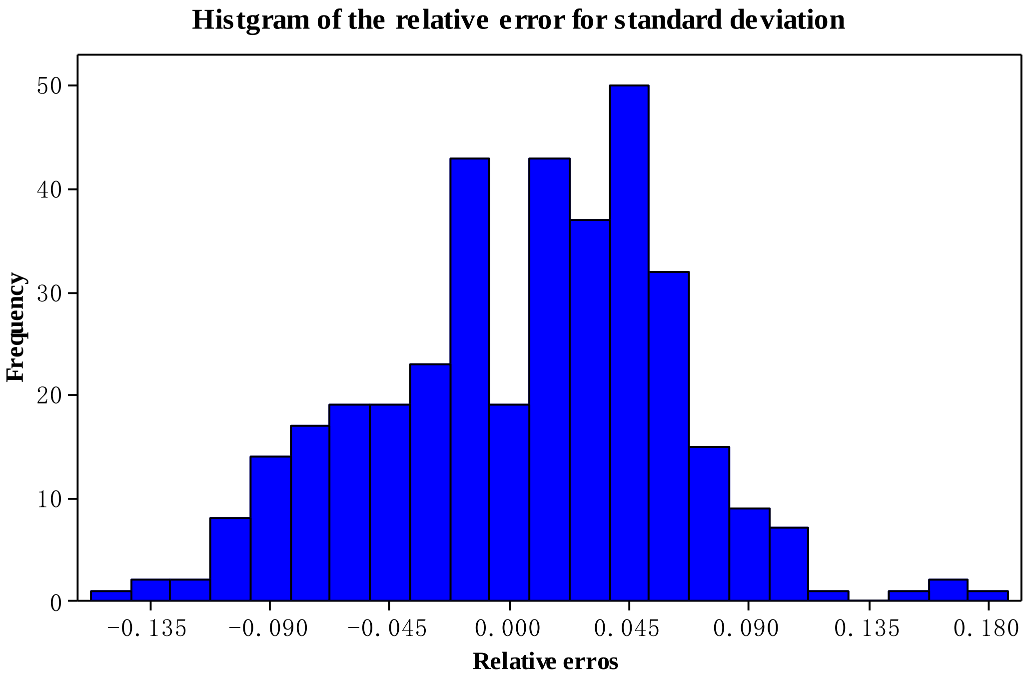


Figure 9. The distribution of the relative errors between the standard deviations from MC simulation and 2-order PCE prediction results

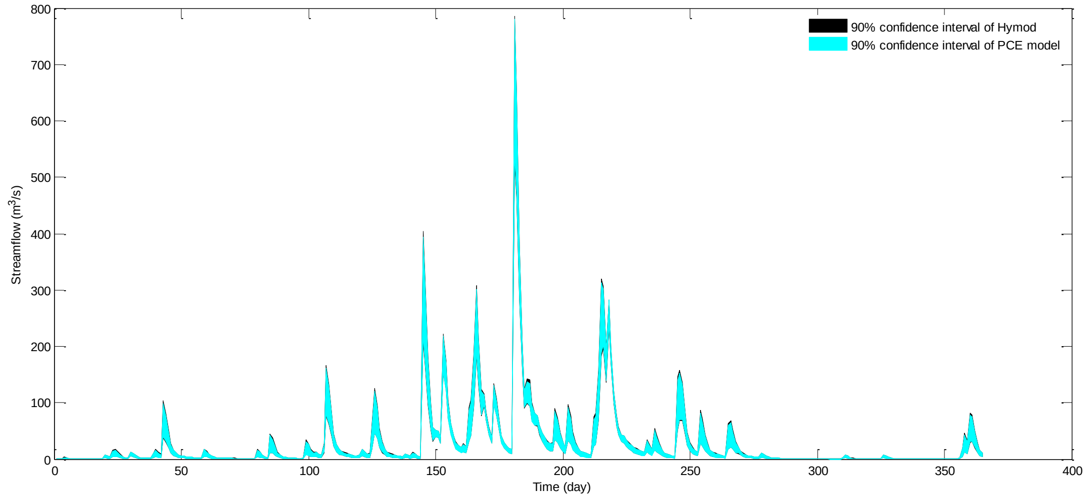


Figure 10. The comparison between the prediction intervals of MC simulation and 2-order PCE results

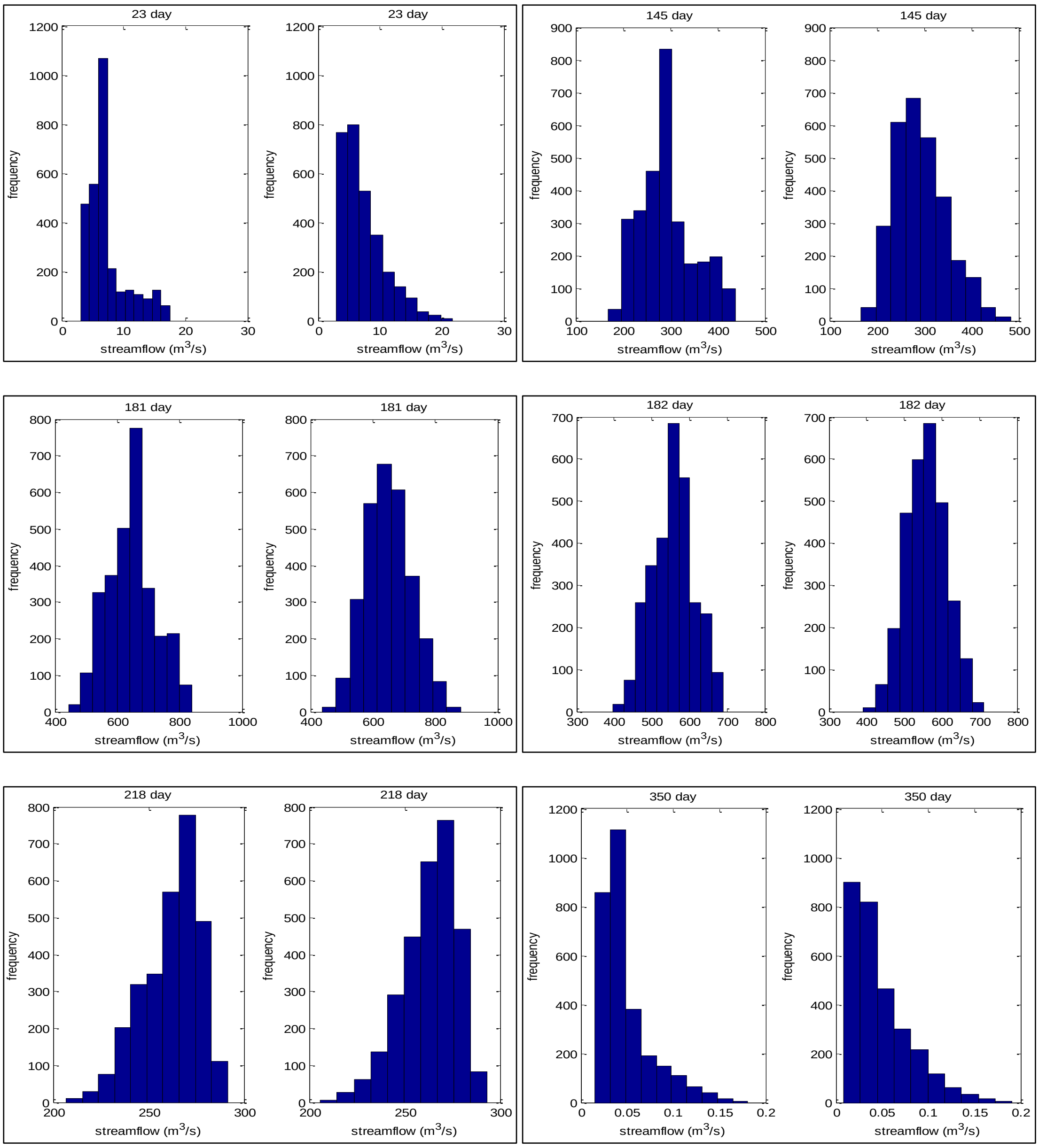


Figure 11. The comparison of histograms between MC simulation and 2-order PCE results (note: in each subfigure, the left column represent MC results and the right one represents the PCE results)

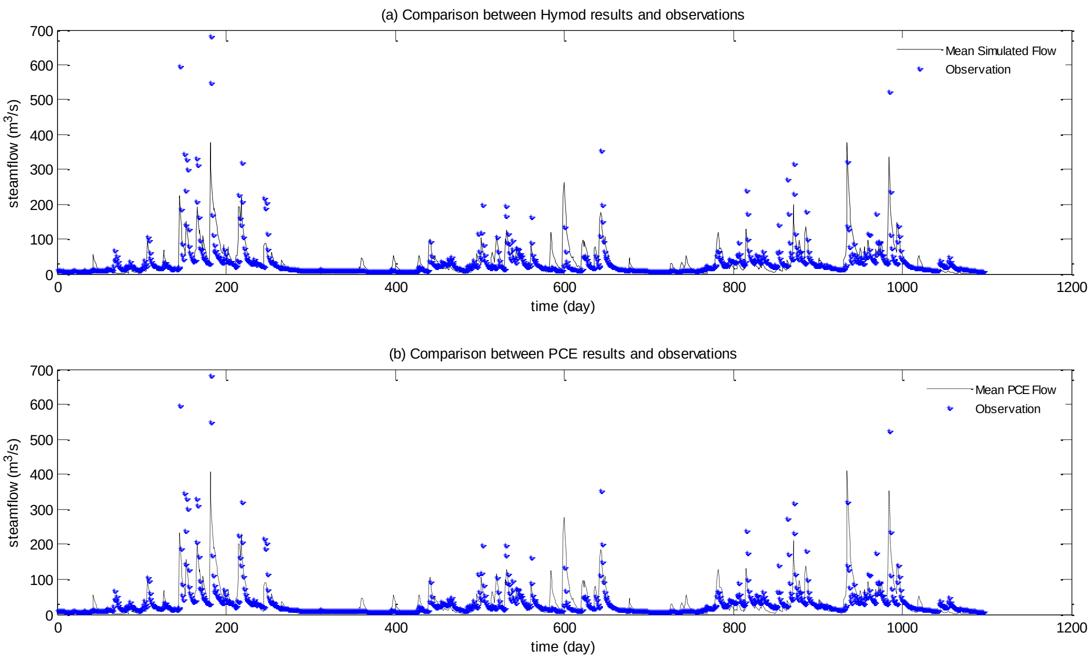


Figure 12. Comparison between the predication means and observations under normal error assumption for precipitation: (a) hydrologic model predictions vs. observations, (b) PCE results vs. observation

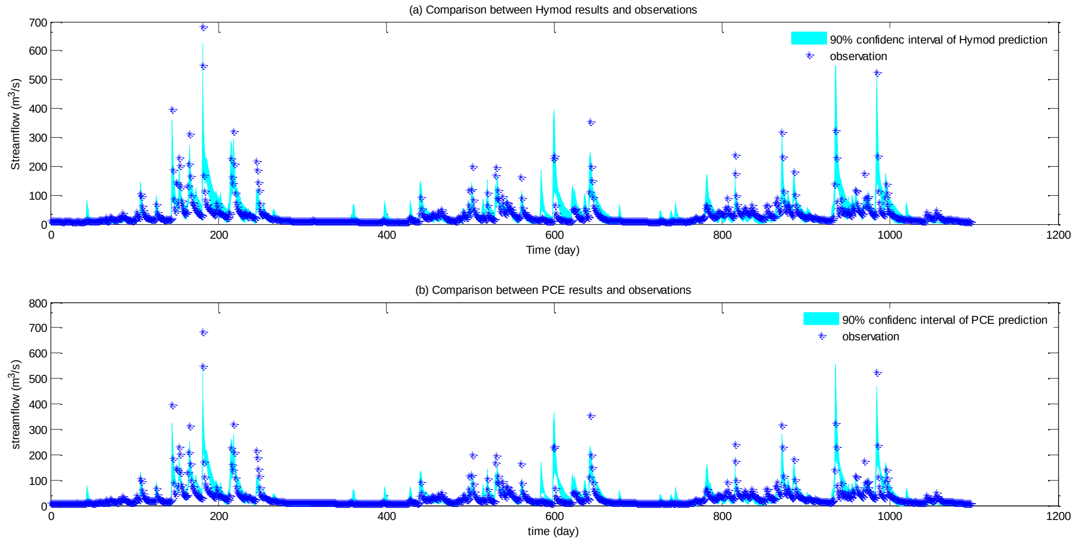


Figure 13. Comparison between the predication intervals and observations under normal error assumption for precipitation: (a) hydrologic model prediction intervals vs. observation, (b) PCE predicting intervals vs. observation

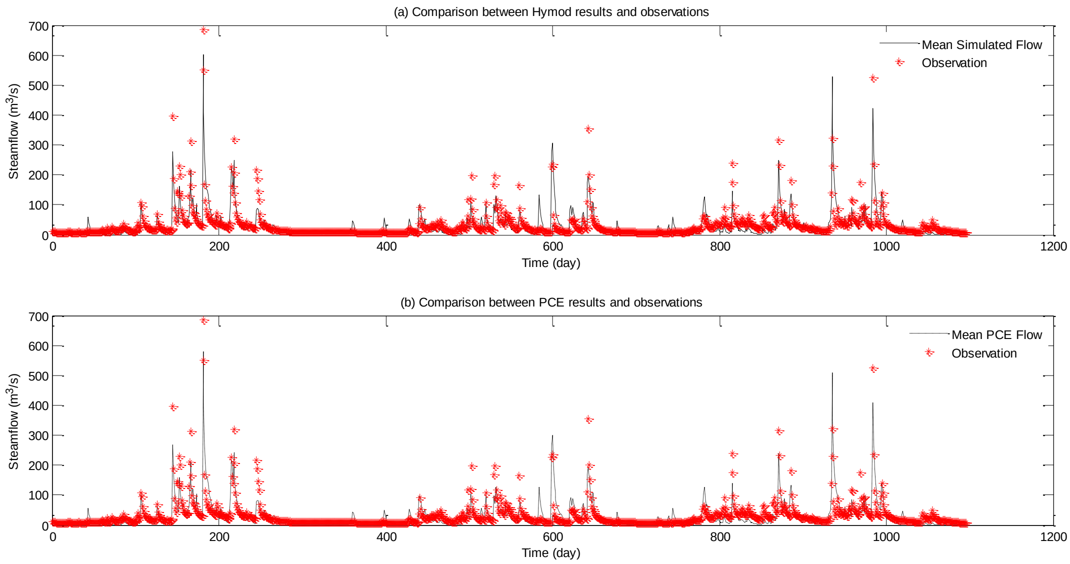


Figure 14. Comparison between the predication means and observations under lognormal error assumption for precipitation: (a) hydrologic model predictions vs. observations, (b) PCE results vs. observation

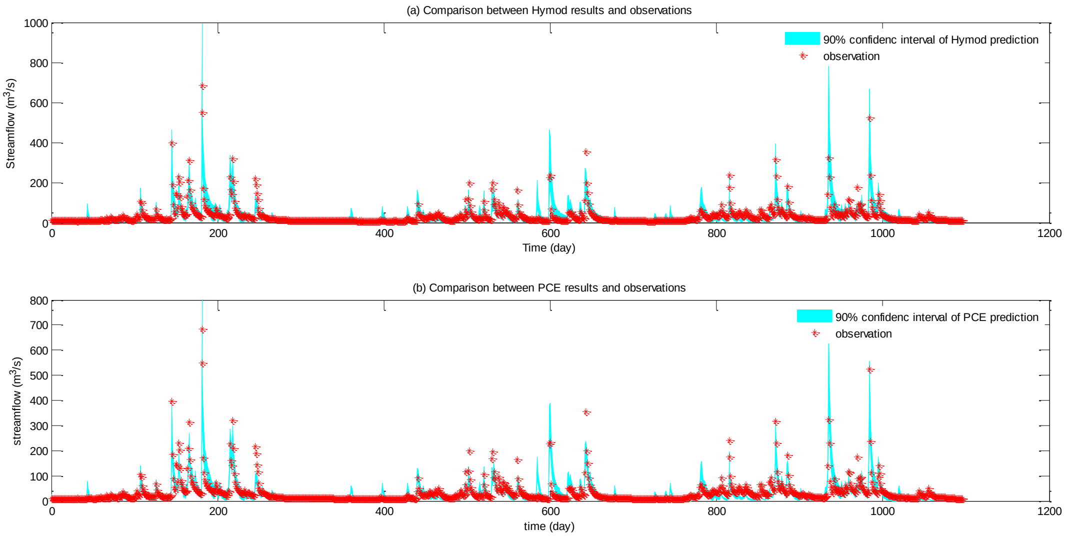


Figure 15. Comparison between the predication intervals and observations under lognormal error assumption for precipitation: (a) hydrologic model prediction intervals vs. observation, (b) PCE predicting intervals vs. observation

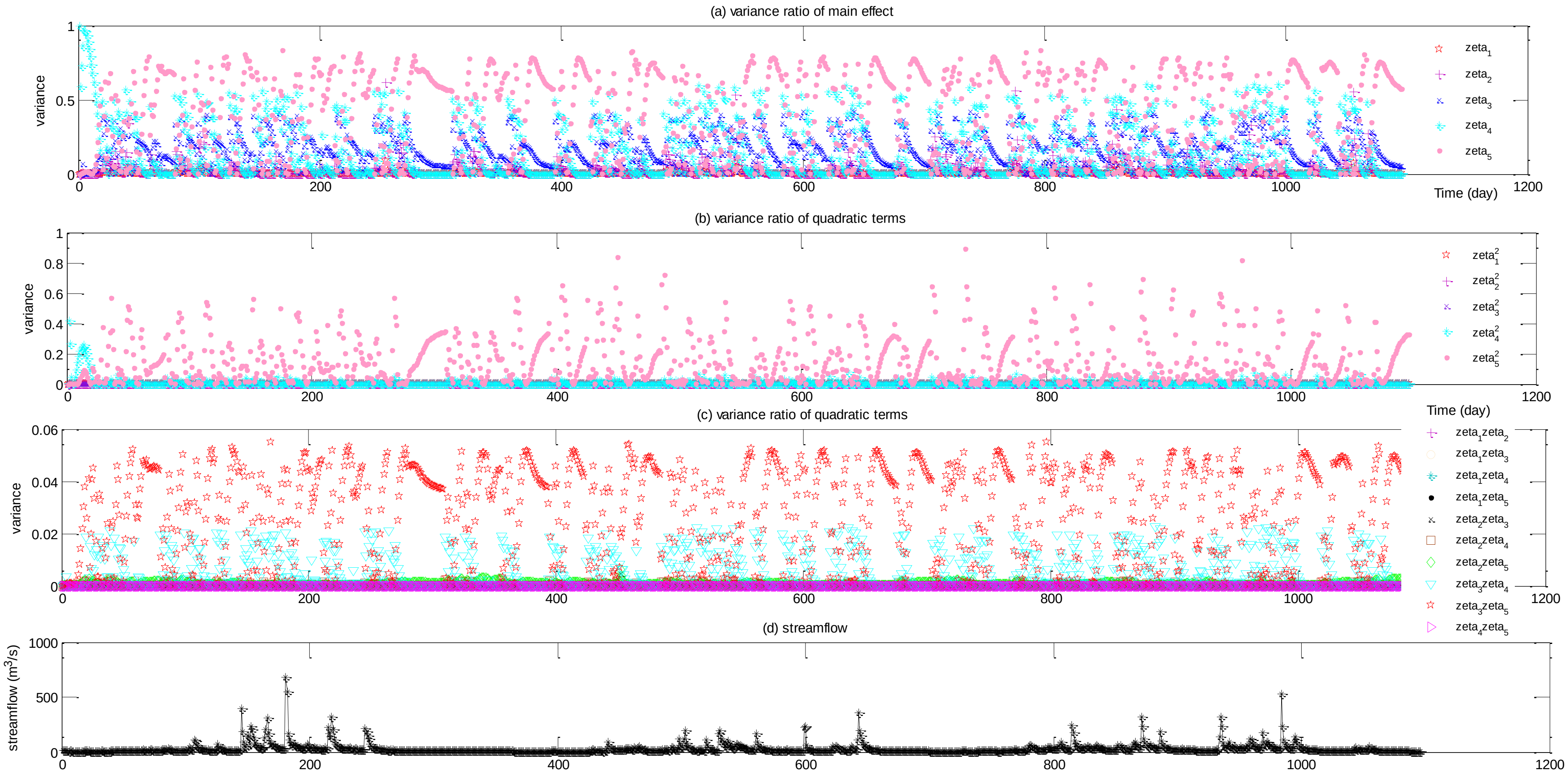


Figure 16. Contributions of model parameters to the uncertainty in model predictions over the simulation period

List of Table Captions

Table 1. Number of the truncated terms for *M*-dimensional *p*th order PCE

Table 2. All collocation points for 2-dimensional 2- and 3-ord PCEs

Table 3. The predefined true values and fluctuating ranges for the parameters of Hymod

Table 4. Comparison of the performance of EnKF under different ensemble sizes

Table 5. Comparison of statistic characteristics of the 2-order PCE and MC simulation results at specific time periods

Table 6 Performance of EnKF under different relative error scenarios

Table 7. Comparison between Monte Carlo method and PCE with normal error perturbation for precipitation

Table 8. Comparison between Monte Carlo method and PCE with lognormal error perturbation for precipitation

Table 1. Number of the truncated terms for M-dimensional pth order PCE

	$M = 1$	$M = 2$	$M = 3$	$M = 4$	$M = 5$
$p = 1$	2	3	4	5	6
$p = 2$	3	6	10	15	21
$p = 3$	4	10	20	35	56

Table 2. All collocation points for the 2-dimensional 2- and 3-ord PCEs

Collocation points	Second order		Third order	
	ζ_1	ζ_2	ζ_1	ζ_2
1	-1.73	-1.73	0.00	0.00
2	-1.73	0.00	0.00	-2.33
3	-1.73	1.73	0.00	-0.74
4	0.00	-1.73	0.00	0.74
5	0.00	0.00	0.00	2.33
6	0.00	1.73	-2.33	0.00
7	1.73	-1.73	-2.33	-2.33
8	1.73	0.00	-2.33	-0.74
9	1.73	1.73	-2.33	0.74
10			-2.33	2.33
11			-0.74	0.00
12			-0.74	-2.33
13			-0.74	-0.74
14			-0.74	0.74
15			-0.74	2.33
16			0.74	0.00
17			0.74	-2.33
18			0.74	-0.74
19			0.74	0.74
20			0.74	2.33
21			2.33	0.00
22			2.33	-2.33
23			2.33	-0.74
24			2.33	0.74
25			2.33	2.33

Table 3. The predefined true values and fluctuating ranges for the parameters of Hymod

	Parameters				
	C_{max} (mm)	b_{exp}	α	R_s (1/day)	R_q (1/day)
True	175.40	11.68	0.46	0.11	0.82
Primary range	[100, 700]	[0.10, 15]	[0.10, 0.80]	[0.001, 0.20]	[0.10, 0.99]
EnKF results	[110.9, 690.6]	[10.2, 13.8]	[0.56, 0.73]	[0.10, 0.16]	[0.75, 0.76]

Table 4. Comparison of the performance of EnKF under different ensemble sizes

Ensemble Size	30	50	100	150	200	300
NSE	0.771	0.731	0.727	0.672	0.652	0.738
PBIAS	8.917	10.172	10.424	13.023	10.429	12.029
RMSE	32.186	34.880	35.236	38.415	39.480	49.831

Table 5. Comparison of statistic characteristics of the 2-order PCE and MC simulation results at specific time periods

Time (d)	Mean		Standard Deviation		Kurtosis		Skewness	
	PCE	MC	PCE	MC	PCE	MC	PCE	MC
23	7.38	7.35	3.35	3.22	4.30	4.07	1.27	1.38
145	292.05	292.17	54.04	56.88	2.93	2.63	0.56	0.48
181	649.71	647.20	73.11	76.28	2.70	2.56	0.20	0.13
182	558.05	555.92	52.64	55.47	2.67	2.53	0.02	-0.04
218	263.00	261.77	14.19	15.00	3.27	2.98	-0.68	-0.70
350	0.05	0.05	0.03	0.03	4.64	5.59	1.35	1.67

Table 6 Performance of EnKF under different relative error scenarios

Relative error	10%	15%	20%	25%	30%
RMSE	42.4	43.8	37.1	37.4	39.2
PBIAS(%)	27.4	22.5	6.0	13.8	13.6
NSE	0.63	0.64	0.65	0.64	0.64

Table 7 Comparison between hydrologic model and PCE with normal error perturbation for precipitation

Sample size		500	1000	1500	2000	2500
Hydrologic Model	RMSE	37.118	37.134	37.107	37.099	37.101
	PBIAS(%)	6.043	6.124	6.053	5.755	5.857
	NSE	0.6475	0.6473	0.6476	0.6478	0.6468
	Time (s)	54.697	111.478	166.210	232.847	334.471
PCE	RMSE	37.394	37.349	37.360	37.310	37.339
	PBIAS(%)	7.062	7.444	7.257	7.222	7.238
	NSE	0.6441	0.6417	0.6433	0.6429	0.6423
	Time (s)	5.278	8.750	14.044	19.050	28.232

Table 8. Comparison between hydrologic model and PCE with lognormal error perturbation for precipitation

Sample size		500	1000	1500	2000	2500
Hydrologic Model	RMSE	27.1144	27.1248	27.1438	27.1004	27.1379
	PBIAS(%)	18.7209	18.5018	18.4552	18.5887	18.4069
	NSE	0.7185	0.7182	0.7178	0.7187	0.7179
	Time (s)	56.8370	107.4660	173.3930	240.7350	305.1020
PCE	RMSE	27.3754	27.4964	27.4632	27.3709	27.4515
	PBIAS(%)	18.6222	18.5811	18.5557	18.6772	18.6420
	NSE	0.7130	0.7105	0.7111	0.7131	0.7114
	Time (s)	5.7430	9.1590	16.3160	19.0140	22.1320