

## Research papers

## Towards reliable uncertainty quantification for hydrologic predictions, Part I: Development of a particle copula Metropolis Hastings method

Y.R. Fan<sup>a,\*</sup>, X. Shi<sup>b</sup>, Q.Y. Duan<sup>c</sup>, L. Yu<sup>d,\*</sup><sup>a</sup> Department of Civil and Environmental Engineering, Brunel University London, Uxbridge UB8 3PH, United Kingdom<sup>b</sup> School of Interdisciplinary Studies, University of Glasgow, Dumfries DG1 4ZL, United Kingdom<sup>c</sup> College of Hydrology and Water Resources, Hohai University, Nanjing 210024, China<sup>d</sup> School of Water Conservancy & Environment, Zhengzhou University, Zhengzhou 450001, China

## ARTICLE INFO

## Keywords:

Hydrological predictions  
 Sequential data assimilation  
 Particle filter  
 Uncertainty  
 Copula

## ABSTRACT

In this study, a particle copula Metropolis-Hastings (PCMH) approach was developed for reliable uncertainty quantification of hydrological predictions. The proposed PCMH approach employs a mixed particle evolution scheme, which integrates the Gaussian perturbation and copula-based dependent sampling methods. The Metropolis ratio is then employed to determine the acceptance of the candidate samples. The applicability of PCMH is elaborated for a long-term data assimilation case at the River Ouse in UK. Multiple hydrological models and different uncertainty settings in inputs, outputs and sample sizes are tested by the PCMH, particle filter (PF) and particle Markov chain Monte Carlo (PMCMC) approaches. The results indicate that the developed PCMH approach is able to generate more reliable results with less accuracy fluctuation than PF and PMCMC for both deterministic and probabilistic predictions from all the hydrological models. The mean values and the associated variation intervals of NSE over the total 270 runs for PCMH, PF and PMCMC are 0.752 (variation interval of [0.534, 0.866]), 0.661 (variation interval of [0.080, 0.879]), and 0.655 (variation interval of [0.247, 0.824]), respectively. For the probabilistic predictions evaluated by CRPS, the mean values and fluctuation ranges from PCMH, PF and PMCMC are respectively 15.215 ([8.624, 31.549]), 18.758 ([8.595, 43.536]), 19.308 ([10.848, 37.799]). These results suggested that the proposed PCMH method would be more robust than PF and PMCMC in generating reliable hydrologic predictions and be less influenced by the hydrologic model structures, uncertainty scenarios, and its inherent randomness. Moreover, the PCMH method can also show better robustness than the copula-based particle filter method since the particle evolution scheme of PCMH would balance extreme samples from copula sampling procedure by mixing samples from Gaussian perturbation and remove unacceptable candidates through the Metropolis acceptance criterion.

## 1. Introduction

Hydrologic models are proposed to conceptualize rainfall-runoff processes in a water cycle through a series of mathematical equations. Due to the rapid development of computer sciences, hydrological models have been widely used for a great number of water issues such as flood or drought prediction, water resources management and so on (e.g. Zhou et al., 2018a; Sun et al., 2019; Yu et al., 2020a; Yu et al., 2020b; Hu et al., 2021). However, extensive uncertainties may exist in hydrological models since they are generally developed based on amounts of simplifications and assumptions on water processes. These uncertainties may be embedded in different components of hydrologic models such as

model structures, parameters and forcing data (Liu and Gupta, 2007; Moradkhani and Sorooshian, 2008; Di et al., 2021). Uncertainty quantification is recognized as one of the most critical issues to be addressed in the applications of hydrologic models (Renard et al., 2010; Wu et al., 2019; Fan, 2019).

As one of the core issues in hydrologic prediction, uncertainty quantification has been addressed by a great number of studies, in which a variety of approaches have been developed such as Bayesian estimation and polynomial chaos expansion (e.g., Ghaith and Li, 2020; Ghaith et al., 2021; Zhou et al., 2022). Sequential data assimilation approaches have been applied for many hydrologic issues such as predictions of streamflow and soil moisture (e.g. Yan et al., 2015; Fan et al., 2015; Fan

\* Corresponding authors.

E-mail addresses: [yurui.fan@brunel.ac.uk](mailto:yurui.fan@brunel.ac.uk) (Y.R. Fan), [yulei2018@zzu.edu.cn](mailto:yulei2018@zzu.edu.cn) (L. Yu).

et al., 2017b; Lyu and Fan, 2021). These methods can continuously assimilate available measurements into the modelling process to update probabilistic features of states and parameters and generate probabilistic predictions (Vrugt et al., 2005). Ensemble Kalman filter (EnKF) and particle filter (PF) methods are two typical data assimilation techniques widely used in the field of hydrology. The EnKF and its variants use ensembles generated through the Monte Carlo technique to approximate state covariance matrices to achieve suboptimal estimations for parameters and states (Evensen, 2003; Moradkhani et al., 2005a; Shen and Tang, 2015; Wang et al., 2022). Thus, the model derivatives, which is challenging for complex hydrological models, are not required in EnKF. Moreover, the EnKF explicitly characterizes the uncertainty with the ensemble, in which each ensemble member is equally weighted (Moradkhani et al., 2019). However, the Gaussian error assumption may limit the applicability of EnKF-based approaches, and may not fully represent the statistical properties of the model simulations (Khaki et al., 2017).

In a parallel line, the particle filter (PF) method has also been proposed for uncertainty quantification of hydrologic predictions. In the PF framework, a weighted function would be employed to represent the probabilistic distributions for state and parameter variables. These weights, other than the original samples, would be updated through sequential importance sampling (SIS) algorithm (Liu et al., 2001; Moradkhani et al., 2012; Liu et al., 2012). Moreover, resampling algorithms are employed after SIS in order to mitigate weight degeneration in PF (Fan et al., 2016; Fan et al., 2017a). However, the resampling procedure in PF may also lead to limited sample diversity, which means that only a few samples are available with significant weights and thus cannot fully describe the probabilistic features in model parameters and states. The particle Markov chain Monte Carlo (PMCMC) approaches have been proposed to address the above sample impoverishment issue through integrating MCMC sampling strategy into PF (Moradkhani et al., 2012; Vrugt et al., 2013; Yan et al., 2015). Abbaszadeh et al. (2018) proposed a new data assimilation technique based on PF, MCMC and genetic algorithm (named as EPFM) to enhance hydrological data assimilation, which has been demonstrated to be more effective and reliable than PMCMC in state and parameter estimations. Abbaszadeh et al., (2019) further coupled a deterministic four-dimensional variational (4DVAR) assimilation method with the particle filter (PF) ensemble data assimilation system to produce a robust approach for dual-state-parameter estimation. This method is able to account for all sources of uncertainties involved in hydrologic predictions and thus lead to more accurate and reliable posterior distributions for both state variables and parameters in data assimilation applications (Abbaszadeh et al., 2019). Hernández and Liang (2018) introduced a hybrid algorithm called OPTIMISTS (Optimized PareTo Inverse Modeling through Integrated Stochastic Search) to bring together ideas from particle filters (PFs), four-dimensional variational methods (4D-Var), evolutionary Pareto optimization, and kernel density estimation in a unique way, which has shown to efficiently produce probabilistic forecasts with comparable accuracy to that obtained from using a particle filter. However, the interdependencies in model parameters can hardly be well considered in these approaches, which may limit the efficiency and robustness in particle evolution. Recently, Fan et al. (2017a) introduced a copula-based particle filter (CopPF) method, in which the copula function is employed to reflect the interdependence of model parameters and generate new samples for the next step. The CopPF method has shown better performances than traditional PF methods, especially for small sample size scenarios. Nevertheless, one potential issue in CopPF is that extreme samples from the outlier of the copula model may exist in parameter evolution, which may influence the robustness of CopPF.

Consequently, this study aims to develop a particle copula Metropolis-Hastings (PCMH) approach to provide reliable quantifications for model states and parameters and further provide robust predictions. The proposed PCMH approach will employ a mixed particle evolution scheme, consisting of a Gaussian perturbation and a copula-based dependent sampling methods. The Metropolis ratio is then

employed to determine the acceptance of the candidate samples. The applicability and reliability of the proposed PCMH approach will be extensively demonstrated through multiple hydrological models and different uncertainty settings in inputs, outputs and sample sizes at the River Ouse located in North Youksire in UK.

## 2. Methodology

### 2.1. Sequential data assimilation

In a sequential data assimilation process, the state-space model for a general hydrological system can be described by generic non-linear functions as follows (Moradkhani, 2008; Yan et al., 2016):

$$\mathbf{x}_t = g(\mathbf{x}_{t-1}, \mathbf{u}_t, \boldsymbol{\theta}_t) + \boldsymbol{\omega}_t, \quad (1)$$

$$\mathbf{y}_t = h(\mathbf{x}_t, \boldsymbol{\theta}_t) + \mathbf{v}_t, \quad (2)$$

where  $g$  is a nonlinear function expressing states transition in the hydrologic model;  $\mathbf{x}_t$  is the state vector at time step  $t$ ;  $\mathbf{u}_t$  represents the forcing data;  $\boldsymbol{\theta}_t$  is parameter vector for the hydrological model;  $h(\cdot)$  is the non-linear function relates the state vector  $\mathbf{x}_t$  to the observation vector  $\mathbf{y}_t$ ;  $\boldsymbol{\omega}_t$  and  $\mathbf{v}_t$  respectively describe the random errors in hydrologic model and observations.

In the Bayesian-based filtering methods such as PF, the posterior probabilities of model parameters and states can be established conditionally on both previous observations ( $\mathbf{y}_{1:t-1}$ ) and current observations ( $\mathbf{y}_t$ ) (Gordon et al., 1993; Fan et al., 2016; Fan et al., 2017a; Wang et al., 2018), which can be formulated as follows:

$$p(\mathbf{x}_t, \boldsymbol{\theta}_t | \mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_t | \mathbf{x}_t, \boldsymbol{\theta}_t) p(\mathbf{x}_t, \boldsymbol{\theta}_t | \mathbf{y}_{1:t-1})}{\int p(\mathbf{y}_t | \mathbf{x}_t, \boldsymbol{\theta}_t) p(\mathbf{x}_t, \boldsymbol{\theta}_t | \mathbf{y}_{1:t-1}) d\mathbf{x}_t d\boldsymbol{\theta}_t} \quad (3)$$

$$p(\mathbf{x}_t, \boldsymbol{\theta}_t | \mathbf{y}_{1:t-1}) = \int p(\mathbf{x}_t, \boldsymbol{\theta}_t | \mathbf{x}_{t-1}, \boldsymbol{\theta}_{t-1}) p(\mathbf{x}_{t-1}, \boldsymbol{\theta}_{t-1} | \mathbf{y}_{1:t-1}) d\mathbf{x}_{t-1} d\boldsymbol{\theta}_{t-1} \quad (4)$$

where  $p(\mathbf{x}_t, \boldsymbol{\theta}_t | \mathbf{y}_{1:t-1})$  represents the prior information;  $p(\mathbf{y}_t | \mathbf{x}_t, \boldsymbol{\theta}_t)$  denotes the likelihood function;  $p(\mathbf{x}_t, \boldsymbol{\theta}_t | \mathbf{y}_{1:t})$  and  $p(\mathbf{x}_{t-1}, \boldsymbol{\theta}_{t-1} | \mathbf{y}_{1:t-1})$  respectively describes the posterior probabilities of model parameters and states at time step  $t$  and  $t-1$ ;  $p(\mathbf{x}_t, \boldsymbol{\theta}_t | \mathbf{x}_{t-1}, \boldsymbol{\theta}_{t-1})$  represents model states and parameters evolution which can be obtained by Eq. (1). In practice, the Eq. (3) does not have an analytic solution except for few special cases (e.g., the linear system with Gaussian assumption) since the evaluation of the integrals may be intractable (Yan et al., 2016). Consequently, the estimation of posterior distributions for model states and parameters is approximated by some Monte Carlo-based methods.

### 2.2. Particle filter method

The PF method is a sequential Monte Carlo method that estimates the posterior distributions of states and parameters through random samples. In PF, a weighted function will be employed to denote the posterior distributions for state and parameter variables, given observations  $\mathbf{y}$  at previous  $t$  steps, which can be formulated as follows (Arulampalam et al., 2002; Moradkhani et al., 2005b; Moradkhani et al., 2012):

$$p(\mathbf{x}_t, \boldsymbol{\theta}_t | \mathbf{y}_{1:t}) = \sum_{i=1}^N w_t^{i+} \delta(\mathbf{x}_t - \mathbf{x}_t^{i-}, \boldsymbol{\theta}_t - \boldsymbol{\theta}_t^{i-}) \quad (5)$$

where  $\delta$  is the Dirac delta function, and  $N$  is the sample size.  $w_t^{i+}$  denotes the posterior weight for the  $i$ th particle, which can be updated as follows:

$$w_t^{i+} = w_t^{i-} \frac{p(\mathbf{y}_t | \mathbf{x}_t^{i-}, \boldsymbol{\theta}_t^{i-})}{\sum_{i=1}^N w_t^{i-} p(\mathbf{y}_t | \mathbf{x}_t^{i-}, \boldsymbol{\theta}_t^{i-})} \quad (6)$$

where  $w_t^{i-}$  is the prior weight for the  $i$ th particle, and  $\mathbf{x}_t^{i-}$  and  $\boldsymbol{\theta}_t^{i-}$  denotes

the prior samples for the model parameter and state vector, respectively.  $p(y_t|x_t^-, \theta_t^-)$  can be obtained through a Gaussian likelihood function  $L(y_t|x_t^-, \theta_t^-)$ , which has been widely used in a number of studies (Moradkhani et al., 2005b; Salamon and Feyen, 2010; Fan et al., 2016; Yan and Moradkhani, 2016), and is formulated as:

$$L(y_t|x_t^-, \theta_t^-) = \frac{1}{\sqrt{(2\pi)^m |R_t|}} \exp\left(-\frac{1}{2R_t} [y_{o,t}^i - y_t^i]^2\right) \quad (7)$$

where  $y_{o,t}^i$  is the randomized observations and  $R_t$  is the standard deviation of the residuals between observations and model simulations.

Sampling importance resampling (SIR) algorithms (Moradkhani et al., 2005b) are usually employed to replace the particles with small importance weights by those particles having large importance weights. Moreover, a perturbation of the resampled parameters is recommended to avoid the sample impoverishment (Yan and Moradkhani, 2016). One of the well-known perturbation methods is based on a random walk process as follows (Moradkhani et al., 2005b; Moradkhani et al., 2012):

$$\theta_{t+1}^- = \theta_{t-resamp}^- + \varepsilon_t^i, \quad \varepsilon_t^i \sim N(0, sVar(\theta_t^-)) \quad (8)$$

where  $Var(\theta_t^-)$  is the variance of the prior parameters at the current time step and  $s$  is a small tuning parameter ranging from 0.005 to 0.025 (Moradkhani et al., 2012).

### 2.3. Particle Markov chain Monte Carlo method

The particle Markov chain Monte Carlo methods (PMCMC) (e.g. Moradkhani et al., 2012; Vrugi et al., 2012) improve upon traditional PF method through integrating MCMC steps into the parameter evolution process in Eq. (8). The benefit of using MCMC moves is that larger noise values can be used since the Metropolis acceptance criterion is employed to avoid moving outside the filtering posterior (Moradkhani et al., 2012). This implies that, even though larger noise values may produce larger particle spread and thus have more outlier samples, the Metropolis acceptance criterion is able to detect and eliminate those samples from the filtering posteriors. Therefore, introducing MCMC into PF procedures can lead to better descriptions for the parameter posteriors and also mitigating sample impoverishment.

In the PMCMC method, a candidate sample/particle would be generated through perturbation of the resampled parameters as follows (Moradkhani et al. 2012; Yan and Moradkhani, 2016):

$$\theta_t^{i,p} = \theta_{t-resamp}^i + \varepsilon_t^i, \quad \varepsilon_t^i \sim N(0, sVar(\theta_t^-)) \quad (9)$$

The Metropolis acceptance ratio is then employed to determine the acceptance of the proposed parameter candidate as follows:

$$\alpha = \min\left[1, \frac{p(x_{t-1}^{i+}, \theta_t^{i,p} | y_{1:t})}{p(x_{t-1}^{i+}, \theta_t^i | y_{1:t})}\right] \quad (10)$$

If  $\alpha \geq U(0, 1)$ , the posterior state ( $x_t^{i+}$ ) and parameter ( $\theta_t^{i+}$ ) samples can be obtained as:

$$x_t^{i+} = x_t^{i,p}, \text{ and } x_t^{i,p} = g(x_{t-1}^{i+}, u_t^i, \theta_t^{i,p}) + \omega_t \quad (11)$$

$$\theta_t^{i+} = \theta_t^{i,p} \quad (12)$$

Otherwise:

$$x_t^{i+} = x_{t-resamp}^i \text{ and } \theta_t^{i+} = \theta_{t-resamp}^i \quad (13)$$

Then prior weights and prior parameter samples for the next time step can be obtained as:

$$w_{t+1}^{j-} = \frac{1}{N} \text{ and } \theta_{t+1}^{j-} = \theta_t^{j+} \quad (14)$$

The main difference between PF and PMCMC is that the Metropolis

acceptance ratio is included in PMCMC. This means that, after parameters are perturbed through random walk, the candidate samples will be judged through the Metropolis acceptance ratio to determine whether these samples will be accepted in the parameter posteriors. Such a process can eliminate the outlier/abnormal samples generated in the random walk process and enhance quantification of parameter posteriors in data assimilation.

### 2.4. Particle copula Metropolis-Hastings method

However, one potential challenge for PF and PMCMC approaches is that the parameter evolution scheme in Eqs. (8) and (9) can hardly consider interdependence among model parameters and thus may influence the reliability of those two approaches, especially for a long-term data assimilation process. To address this challenge, this study is going to develop a particle copula Metropolis-Hastings (PCMH) method. The major difference between PCMH and previous approaches (PF and PMCMC) is that a mixed sample evolution scheme, consisting of Gaussian evolution (expressed by Eq. (8)) and copula-based evolution schemes, is employed in PCMH to enhance the robustness of the data assimilation process.

Copula functions have been widely used for modelling dependence structures among correlated random variables. Since introduced into hydrology by Favre et al. (2004), the copula functions have been applied to addressing a number of hydrologic issues such as multivariate risk analysis for natural hazards (e.g. Sun et al., 2019; Fan et al., 2020; Huang and Fan, 2021; Qing et al., 2022), hydrologic simulation (e.g. Bárdossy and Hörning, 2016; Fan et al., 2017a), and hydroclimatic projections (e.g. Zhou et al., 2018b; Zhang et al., 2021). The copula functions are extensively applied in the past decade since, in a copula model, the marginal distributions and the dependence model can be established in separated processes, giving flexibility in choosing both marginal and dependent models (Brechman and Schepsmeier, 2013). A number of copula functions are available for low-dimensional cases (e.g. two or three), such as Gaussian, Student-t and Archimedean copulas (Nelsen 2006). For high-dimensional dependent variables, their joint distribution can be established through a number of bivariate copulas (Aas et al., 2009). The vine copula methods (e.g. C-vine, D-vine) have been developed to address the above issue (Brechman and Schepsmeier, 2013).

In PCMH, the parameter interdependence at each time step will be quantified through a copula function. A set of samples based on the established copula model will be generated. This sample set will be mixed with the sample set generated by the Gaussian perturbation process (denoted as Eq. (8)) to formulate the candidate samples for the next step. The Metropolis acceptance ratio will finally be employed to determine whether the candidate sample would be accepted. The detailed parameter evolution process is described as follows:

For the resampled parameter particles  $\theta_{t-resamp}^i$  ( $i = 1, \dots, ns$ , and  $ns$  is the sample size), the marginal distribution for any parameter  $\theta_d \in \theta$  is firstly constructed as:  $\theta_d \sim G_d(\theta|\gamma_d)$ . Here, the nonparametric kernel estimator is recommended since the distribution for the resampled particles can hardly be quantified through some parametric models (e.g. normal distribution).

The joint probability distribution for the parameter vector  $\theta = (\theta_1, \theta_2, \dots, \theta_D)$  can be established through a vine copula model as follows:

$$c_{1:D}(\mathbf{u}|\boldsymbol{\eta}) = \prod_{j=1}^{D-1} \prod_{i=1}^{D-j} c_{i,j|i+1,\dots,i+j-1}(F(u_i|u_{i+1}, \dots, u_{i+j-1}), F(u_{i+j}|u_{i+1}, \dots, u_{i+j-1})|\boldsymbol{\eta}) \quad (15)$$

$$u_d = G_d(\theta_{d,t-resamp}|\gamma_d) \quad (16)$$

The new sample set based on the copula model can then be generated through two steps: Firstly, the sample set  $\mathbf{u}^{i,c} = (u_1^{i,c}, \dots, u_D^{i,c})$  ( $i = 1, 2, \dots, ns$ ) can be produced through simulation of the copula model denoted in

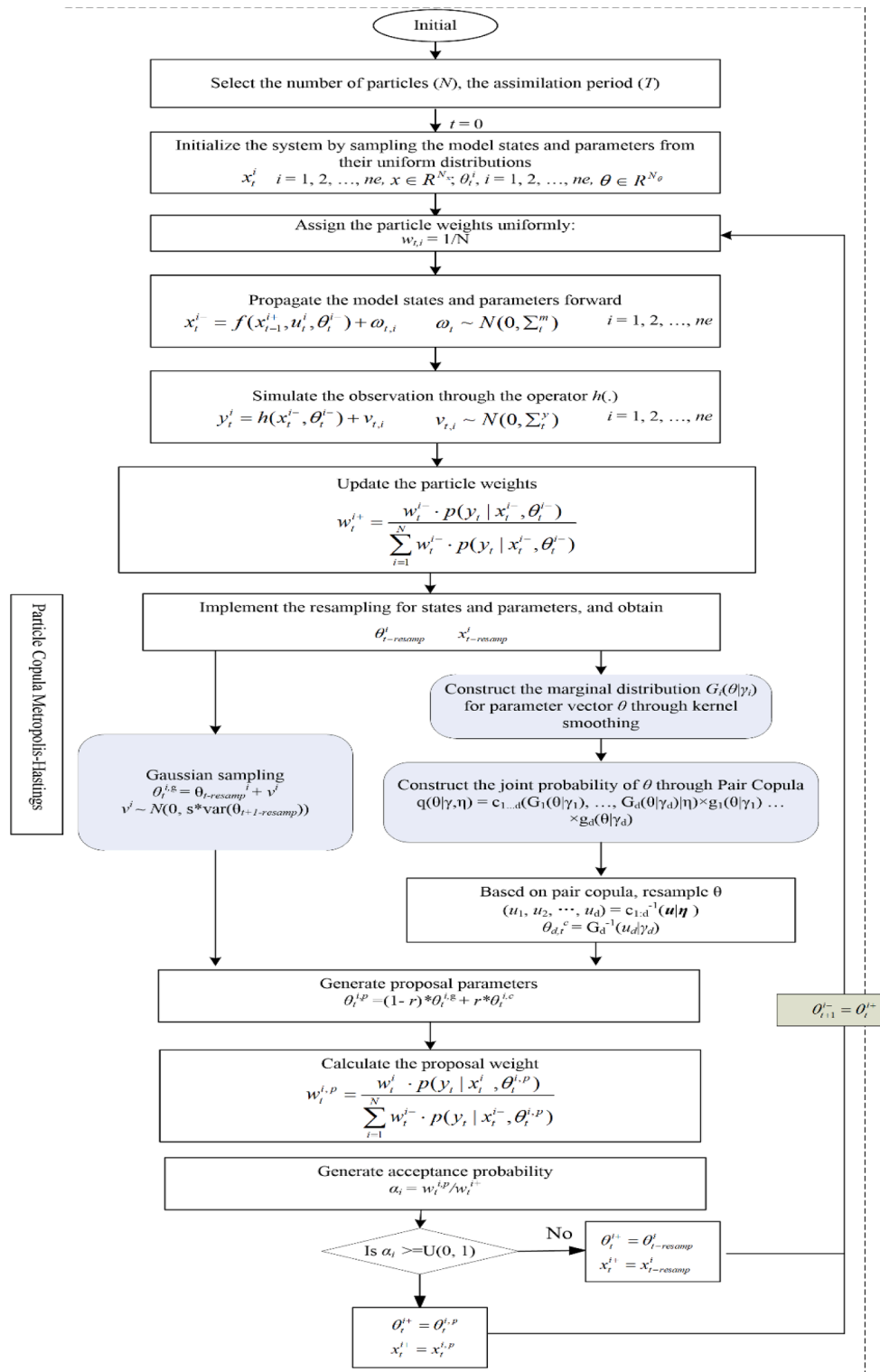


Fig. 1. The procedures for the PCMH method.

Eq. (15). There are several algorithms to generate samples from a multivariate copula model (Aas et al., 2009). Such a process can also be implemented based on the R package “CDVine” (Brechmann and Schepsmeier, 2013). Secondly, the new samples for the original model parameters  $\theta_{d,t}$  ( $d = 1, 2, \dots, D$ ) can be generated through their inverse probability functions:

$$\theta_{d,t}^{i,c} = G_d^{-1}(u_d^{i,c} | \gamma_d), \quad d = 1, 2, \dots, D \quad (17)$$

A second sample set  $\theta_t^{i,g}$  will be generated through the Gaussian perturbation scheme expressed as Eq. (8). The candidate samples can then be obtained by mixing the sample sets of  $\theta_t^{i,g}$  and  $\theta_t^{i,c}$  ( $\theta_t^{i,c} = (\theta_{1,t}^{i,c}, \dots, \theta_{D,t}^{i,c})$ ) as follows:

$$\theta_t^{i,p} = r\theta_t^{i,c} + (1-r)\theta_t^{i,g} \quad (18)$$

where  $r$  is a value between  $[0, 1]$ .  $r = 1$  means all the samples are

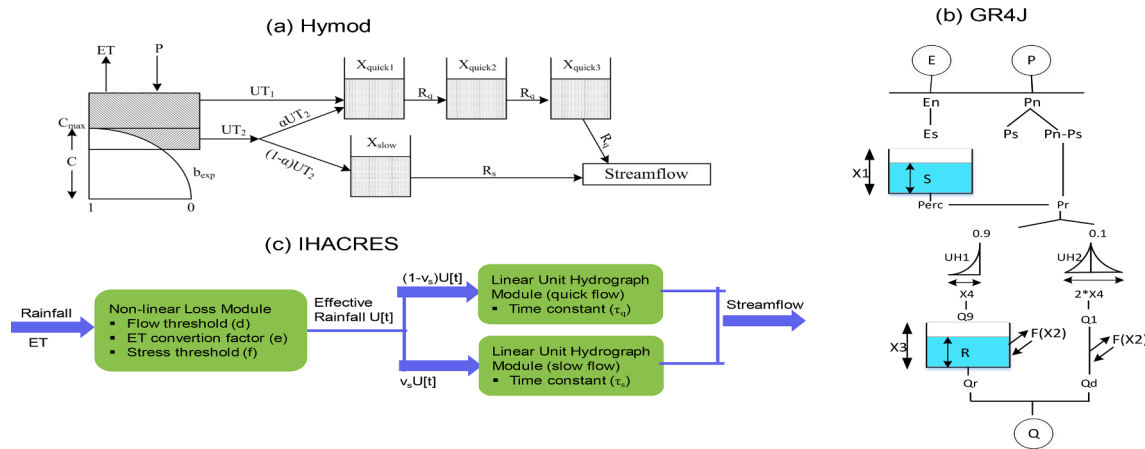


Fig. 2. Frameworks of Hymod, GR4J and IHACRES.

generated through the copula model and  $r = 0$  indicates the Gaussian perturbation scheme is merely employed. In this study,  $r = 0.5$  would be employed.

The Metropolis acceptance ratio is then employed to determine the acceptance of the candidate parameter samples and generate the posterior distributions for both model parameters and states as expressed by Eqs. (10) – (13). The prior weights and parameter samples for the next time step are then characterized by Eq. (14).

Fig. 1 presents the detailed procedure for the proposed PCMH method. There are underlying assumptions for the PCMH method: i) the parameters in a hydrological model are correlated among each other; ii) those correlations may not be well quantified through the multivariate Gaussian distribution and thus the multivariate copula need to be employed to reflect those interdependence; iii) outlier samples may be obtained from either the Gaussian distribution and the copula model and thus the Metropolis acceptance criterion is adopted to eliminate those unacceptable samples. Particularly for the second assumption, the multivariate Gaussian distribution, used to generate the candidate samples as presented in Eq. (9), implies that all the posterior distributions for individual parameters would be normally distributed. However, such an implication may not be valid in the data assimilation process since the posterior samples after assimilating observations may have arbitrary random features. Consequently, the Gaussian distribution cannot well quantify parameter posteriors and their interdependence in the data assimilation process.

There are a great number of copula functions available to quantify complex dependence structures among correlated variables (or parameters) such as Gaussian, Student-t, elliptical copula families (e.g., Clayton, Gumbel, Frank) (Brechmann and Schepsmeier, 2013). In this study, the vine copula model consisting of Frank copula will be used in the developed PCMH model due to the two following reasons: i) the Frank copula can reflect both positive and negative interdependence; ii) the bivariate copula structure is fixed in the vine copula model can accelerate computation since the selection procedure in establishing the vine model is neglected. Moreover, we argue that the vine model with fixed bivariate copulas is acceptable in PCHM since Metropolis acceptance criterion is included to exclude unacceptable samples from the vine model.

### 3. Synthetic case study

#### 3.1. Experiment setup

The applicability of the PCMH method will be firstly demonstrated through a conceptual model Hymod, as shown in Fig. 2(a). Hymod is a probability-distributed model, and consists of three components including a Pareto distribution for reflecting soil storage capacity, a

Table 1 Model parameters, their fluctuating ranges and predefined values in the synthetic experiment for Hymod.

Parameter	Description	Range	Parameter values in the synthetic experiment
$C_{max}$ (mm)	Maximum storage capacity of watershed	[200, 700]	402.2
$b_{exp}$	Spatial variability of soil moisture capacity	[0.5, 6.5]	4.66
$\alpha$	Factor distributing flow to the quick-flow tank	[0.1, 0.9]	0.76
$R_s$ (1/day)	Residence time of the slow-flow tank	[0.001, 0.2]	0.089
$R_q$ (1/day)	Residence time of the quick-flow tank	[0.1, 0.9]	0.52

slow-flow tank for routing groundwater flow, and three identical quick-flow tanks for routing surface flow (Moor, 1985; Fan et al., 2017b). Five parameters need to be calibrated/quantified as listed in Table 1. Precipitation ( $P$  (mm/day)) and potential evapotranspiration ( $ET$  (mm/day)) are the two inputs required by Hymod. State variables in Hymod include: storage in the nonlinear tank representing the watershed soil moisture content, the three quick-flow tank storages representing the temporary (short-time) detentions, and the slow-flow tank storage (subsurface storage) (Moradkhani et al., 2005a).

The experiment is conducted to demonstrate the capability of the PCMH method in quantifying parameter and state uncertainties in Hymod. The 1-year synthetic streamflow data are generated by Hymod with predefined parameters listed in Table 1, with the real inputs for precipitation and potential evapotranspiration. These synthetic streamflow data would be considered as the observations in data assimilation. Moreover, uncertainties in model inputs (i.e. precipitation and potential evapotranspiration) and outputs are characterized by random perturbations. For instance, consider an input at time  $t$  denoted as  $u_t$ , it can be perturbed as  $u_t^i = u_t + \xi_t^i$  where  $\xi_t^i$  is the noise added to the forcing data to generate the  $i$ th sample member. Normal distributed errors with standard deviations being 15% of the corresponding measurements will be added to synthetic streamflow and potential evapotranspiration data (e.g.,  $\xi_t^i \sim N(0, \sum_t^u)$  and  $\sum_t^u = \gamma u_t$  for the forcing data and  $\gamma = 0.15$ ), while a log-normal distributed error, with a proportionality factor of 0.15 will be added the precipitation (i.e.,  $\xi_t^i \sim \log N(0, \sum_t^u)$  and  $\sum_t^u = \gamma u_t$ ).

#### 3.2. Evaluation criteria

To evaluate the performance of different data assimilation methods,

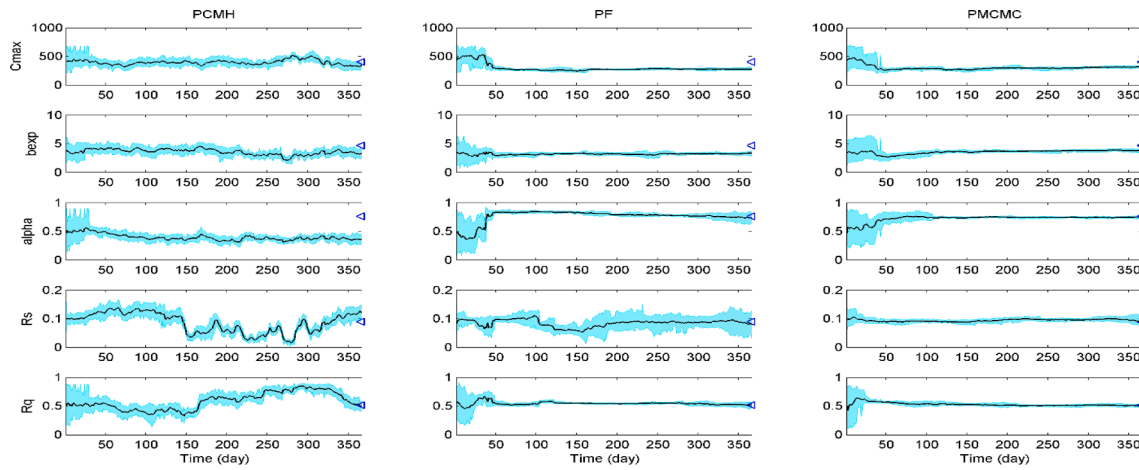


Fig. 3. Parameter Evolution through different data assimilation methods: The blue triangles are the true parameter values used in the synthetic experiment; the black solid lines are the mean estimations and the cyan regions show the 90% predictive intervals (bracketed by the 5% and 95% quantiles) for different parameters.

the root-mean-square error (RMSE), and Nash-Sutcliffe efficiency (NSE) coefficient and continuous ranked probability score (CRPS) will be used in this study, which are respectively formulated as follows (Murphy and Winkler, 1987; Wang et al., 2015; Wang and Wang, 2019):

$$NSE = 1 - \frac{\sum_{i=1}^N (Q_{obs,i} - Q_{sim,i})^2}{\sum_{i=1}^N (Q_{obs,i} - \bar{Q}_{obs})^2} \quad (20)$$

$$CRPS = \int_{-\infty}^{+\infty} [F^s(x) - F^o(x)]^2 dx \quad (21)$$

where  $N$  is the total number of streamflow measurements (or

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (Q_{obs,i} - Q_{sim,i})^2} \quad (19)$$

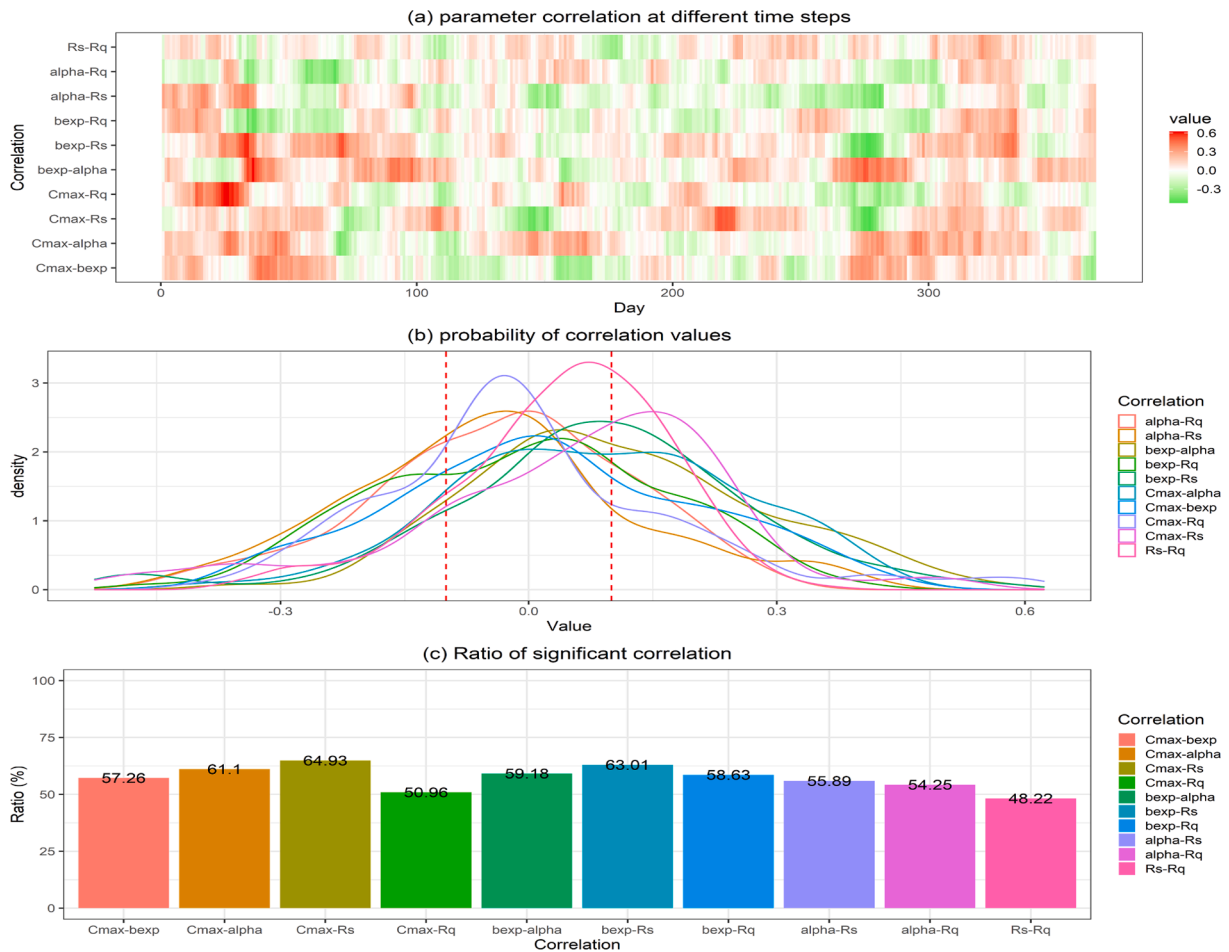
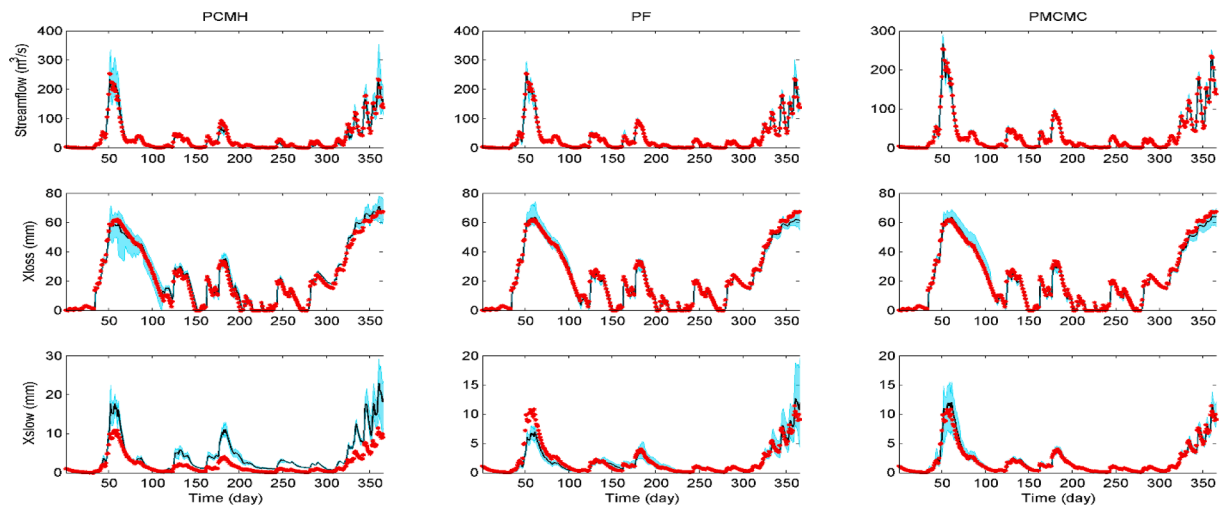


Fig. 4. Parameter correlation in the data assimilation process: (a) parameter correlation at different time steps; (b) the probability of correlation values over the whole data assimilation process.



**Fig. 5.** Comparison between synthetic values and predictions from PCMH, PF and PMCMC: the red stars show the synthetic values for streamflow, nonlinear tank storage (i.e.  $X_{loss}$ ) and slow tank storage ( $X_{slow}$ ); the black solid lines are the predictive means and the cyan regions exhibit the 90% predictive intervals (bracketed by the 5% and 95% quantiles).

predictions),  $Q_{obs}$  and  $Q_{sim}$  are streamflow observations and model predictions, respectively, and  $\overline{Q_{obs}}$  is the mean value of observations.  $F^o$  and  $F^f$  are cumulative distribution functions for observations and model predictions, respectively. The RMSE and NSE have been widely employed for evaluating deterministic predictions from a model, while CRPS is a measurement for probabilistic predictions.

### 3.3. Results analysis for the synthetic experiment

The synthetic experiment is initially conducted by PCMH with a sample size of 50. In addition, both PF and PMCMC, which have been widely adopted in hydrologic data assimilation, have also applied to the synthetic experiment as the benchmark approaches. Fig. 3 presents the sample evolution of model parameters during the data assimilation process by different approaches. It can be noticed that all parameters in Hymod would be well identified by PCMH with reasonable fluctuation ranges. In comparison, both PF and PMCMC seems to lead to over convergence for most model parameters. However, the parameter  $R_s$  is not identifiable by PF since it generally has significant variation ranges in the data assimilation process via PF. This may affect the robustness of PF and PMCMC in long-term data assimilation process since the parameters can hardly be evolved within rational spaces. However, the developed PCMH method can effectively rejuvenate model parameters and mitigate sample impoverishment within larger spaces than those from PF and PMCMC, and thus can help provide reliable results even for long-term data assimilation processes.

In the proposed PCMH method, the copula functions are employed to quantify interdependence among model parameters and then generate parameter samples under such correlation consideration. These samples are mixed with those from Gaussian perturbation to generate the candidate parameter particles. The above processes imply that the parameter samples in a hydrological model are correlated during their evolution in data assimilation. To verify this assumption, the parameter correlation at different time steps in data assimilation is presented in Fig. 4(a). It can be observed that as the data assimilation proceeds, the five parameters (i.e.,  $C_{max}$ ,  $b_{exp}$ ,  $\alpha$ ,  $R_s$ ,  $R_q$ ) in Hymod would show visible correlation with some of them highly correlated at some time steps. Also, the correlation values among those parameters are varied as the data assimilation process continues, which may change from positive correlation to negative correlation. Fig. 4(b) presents the probabilities of correlation values among those five parameters in the data assimilation process in which the dashed red lines show the thresholds of  $-0.1$  and  $0.1$ . Moreover, Fig. 4(c) present the ratios of the parameter correlations

larger than 0.1 or smaller than  $-0.1$  over the data assimilation process. The results indicate that, during the data assimilation process, the model parameters would be significantly correlated among each other at most time steps. The highest ratio (larger than 0.1 or smaller than  $-0.1$ ) of significant correlation appears between  $C_{max}$  and  $\alpha$  with a ratio of about 65%, while the lowest ratio of significant parameter correlation is observed between  $R_s$  and  $R_q$  but such a ratio still be more than 48%. Consequently, parameter correlation needs to be considered in the data assimilation process. Nevertheless, the parameter posteriors can hardly be normally distributed after data assimilation. Fig. S1 presents the histograms of parameter posteriors at different time periods. The results indicate that the parameter posteriors would be generally skewed distributed which would be unable to be modelled through Gaussian and multivariate Gaussian distributions. Consequently, it is desired to quantify those parameter correlation by the copula functions during the data assimilation process.

For the performances of the three data assimilation approaches (i.e. PCMH, PF and PMCMC) in the synthetic experiment, Fig. 5 show comparisons between predictions of streamflow and two state variables (i.e.  $X_{loss}$  and  $X_{slow}$ ) from different approaches with the corresponding true values. The results indicate that the predictive means for streamflow from all the three approaches can well track the variations of the synthetic observations. Moreover, the predictive intervals bounded by 5% and 95% quantiles can bracket most observations. The PCMH approach may produce relatively larger predictive intervals than those from PF and PMCMC during high-flow periods. This can be generally attribute to the larger parameter ranges obtained by PCMH, as shown in Fig. 3. For the nonlinear tank storage (i.e.  $X_{loss}$ ) which represents the soil moisture content, both the predictive means and 90% predictive intervals agree well with the corresponding true values in the synthetic experiment. Also, PCMH would generate relatively larger predictive intervals due to wider parameter ranges than those from PF and PMCMC. For the slow tank storage (i.e.  $X_{slow}$ ), it can be observed that the PCMH performed slightly worse than PF and PMCMC, which lead to underestimation for high slow tank storages. One possible reason is that the state variable  $X_{slow}$  may not be highly correlated with the streamflow rate in the hydrology model (i.e., Hymod) and thus the accurate streamflow prediction does not necessary lead to accurate quantification for  $X_{slow}$ . Fig. 6 compared the prediction accuracy evaluated by NSE between the  $X_{slow}$  and streamflow in the synthetic experiment in 20 replications. It can be seen that the prediction for  $X_{slow}$  is most likely unacceptable even though  $NSE$  value for streamflow prediction higher than 0.9. Consequently, the worse  $X_{slow}$  prediction from PCMH as presented in

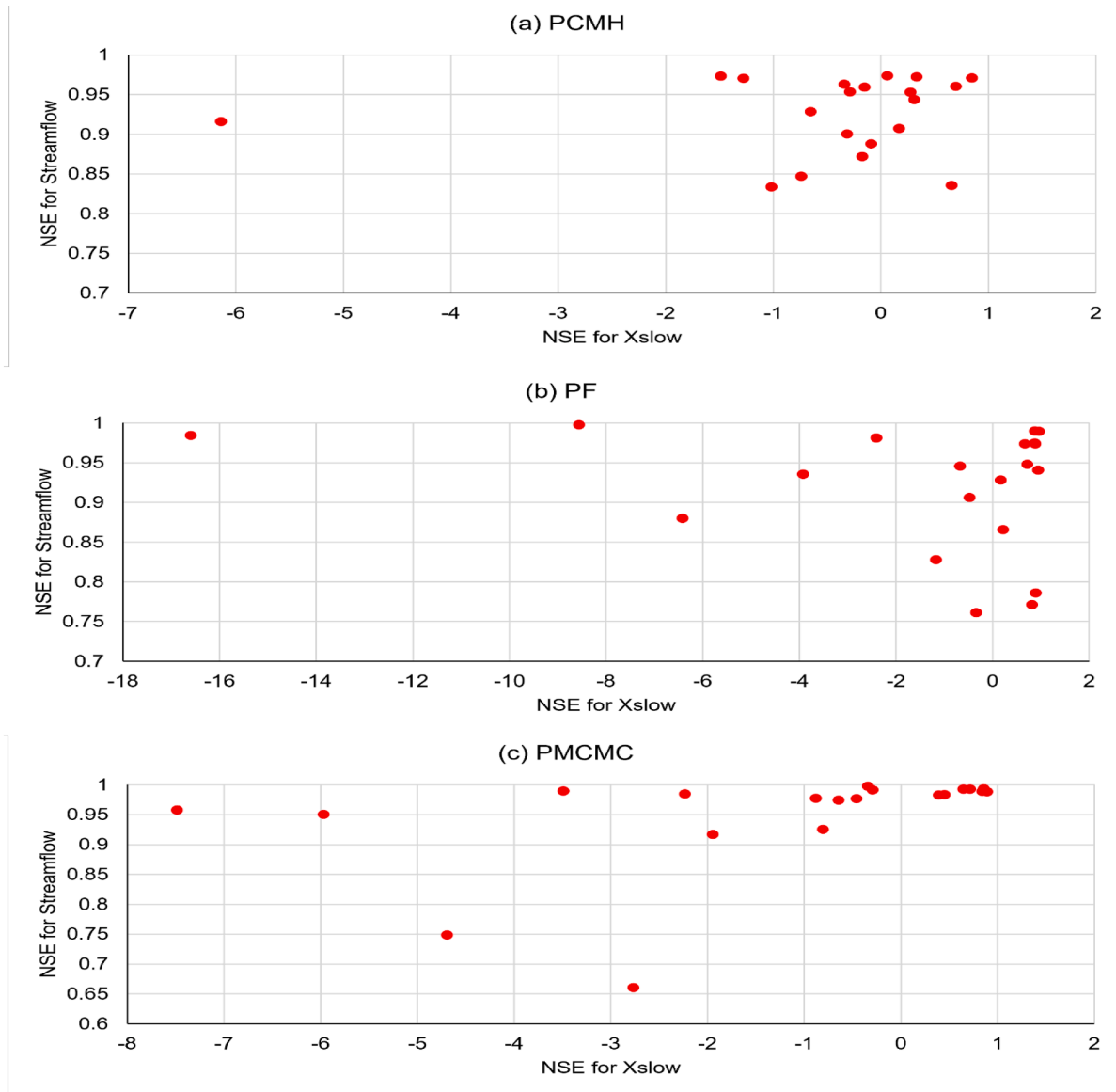


Fig. 6. Comparison of prediction accuracy between the slow tank storage (Xslow) and the streamflow under different data assimilation schemes.

Fig. 5 may possibly be resulted from the poor correlation between streamflow rate and the state variable in Hymod.

To further demonstrate the applicability of the proposed PCMH approach, we have run PCMH, as well as PF and PMCMC for the synthetic experiment under different sample size scenarios. In details, sample size scenarios of {20, 50, 100, 200, 500} are adopted with each scenario being run 20 times by PCMH, PF and PMCMC. This means that for one sample size scenario, all the three data assimilation approaches would be run independently for 20 times (i.e., generate different particles in different runs). The purpose for running multiple times for different approaches is to evaluate the robustness for those methods in uncertainty quantification. The boxplots in Fig. 7 exhibit the performance variations of NSE, RMSE and CRPS for PCMH, PF and PMCMC for different runs under different sample sizes. In this figure, the x-axis indicates the sample scenarios whilst the y-axis shows the values of NSE, RMSE or CRPS. The indices of NSE and RMSE evaluate the accuracy for deterministic predictions and thus the mean prediction at each time step (i.e.,  $Q_{sim}$  in Eqs. (19) and (20)) would be adopted to calculate these two indices. In comparison, the CRPS reflects the accuracy in probabilistic predictions. Therefore, the ensemble predictions in each time step would be used to calculate  $F^0$  and the perturbed observations are employed to

generate the value of  $F^0$ . Based on the above procedures, each run would generate one set values for NSE, RMSE and CRPS and 20 sets would then be obtained for each sample scenario. The boxplot is adopted to display the minimum, maximum, median, first and third quantiles for the 20 values for each index.

Based on the results in Fig. 6, we can conclude that the performances of PCMH show less variation than PF and PMCMC under small sample sizes. Moreover, Table S1 presents the performances of the three data assimilation approaches evaluated by NSE, RMSE and CRPS under different sample sizes. The results suggest that, for large sample size scenarios such as 200 and 500, all the methods performed quite well with high NSE values and low RMSE and CRPS values. These performances would not vary significantly for different runs. For small sample size scenarios such as 20 and 50, PF and PMCMC can generate accurate results with NSE larger than 0.95, but they may also produce unacceptable results with NSE values less than 0.5. In comparison, the developed PCMH method can produce reliable predictions even for small sample sizes. For instance, NSE values for PCMH range from 0.779 to 0.976 for different runs under the sample size of 20, while the NSE values for PF and PMCMC vary within [0.276, 0.997] and [0.400, 0.994], respectively. This demonstrates that the proposed PCMH



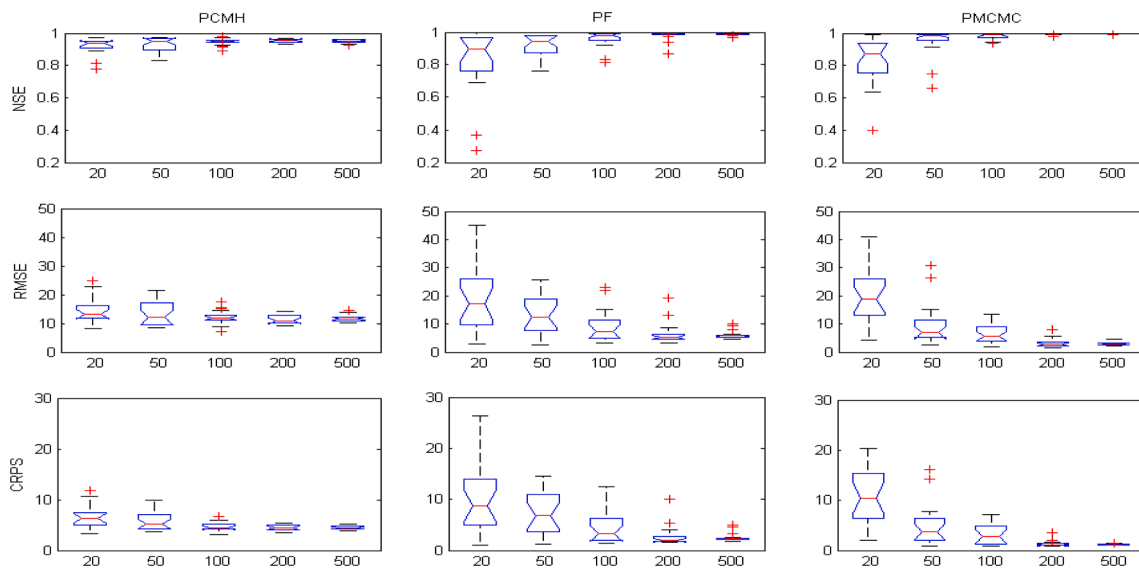


Fig. 7. Performance variations for PCMH, PF and PMCMC for different runs under different sample size scenarios.

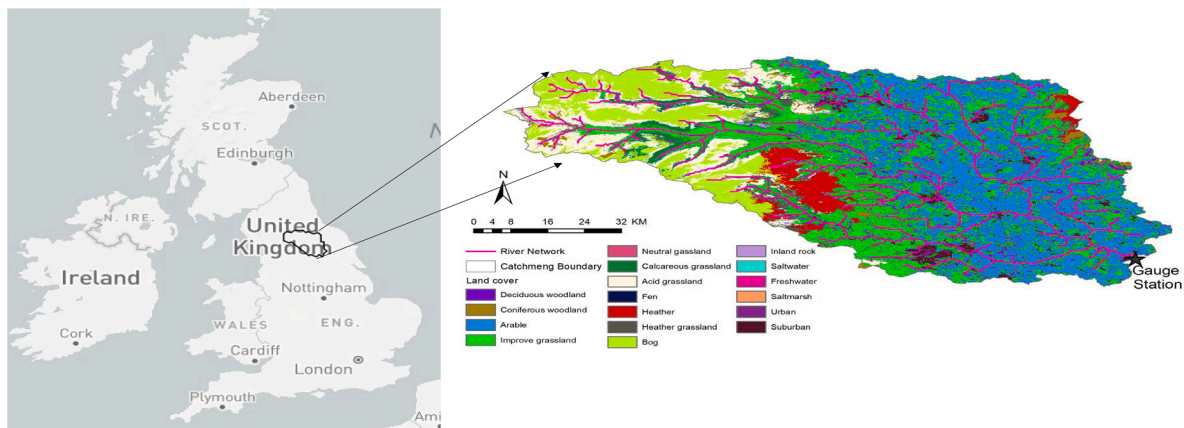


Fig. 8. Location of the River Ouse (Source: Center of Ecology and Hydrology, <https://nrfa.ceh.ac.uk/data/station/info/27009>).

method shows more robustness than PF and PMCMC, especially for small sample size scenarios.

Moreover, it is visible from Fig. 7 and Table S1 that the performance of PCMH may be slightly worse than PF and PMCMC for some larger sample scenarios. The performance of PCMH is affected by different factors such as the copula function in the vine model as well as the mixing rate (i.e.,  $r$  in Eq. (18)). We initially used the Frank copula in the vine model to quantify the interdependence among model parameters and set  $r = 0.5$  to mix the samples from copula and Gaussian perturbation. Such a setting may not lead to the better predictions from PCMH. Consequently, the PCMH with different settings on copula structure and sample mixing rate would be further tested and compared with PF and PMCMC to explore the impacts of these two factors. In detail, for the copula structure, the PCMH can use i) fixed copula functions in the vine model (e.g., Frank copula) or ii) the best copula function selected through the Akaike information criterion. The mixing rate  $r$  in Eq. (18) would be specified as i)  $r = 0.5$  for the whole data assimilation process or ii)  $r = 0.5$  when all correlation coefficients among parameters are larger than 0.05 and otherwise  $r = 0.05$ . Figs. S2 – S4 exhibit the performances of PCMH under different settings and their comparisons with the results from PF and PMCMC. It is visible that the PCMH can produce better streamflow predictions than those from PF and PMCMC with small sample size scenarios. As the sample size increases, the PCMH with copula selection but a fixed mixing rate, as shown in Fig. S3, may not

perform as good as PF and PMCMC. However, the PCMH with varied sample mixing rate (i.e.,  $r = 0.05$  or  $0.5$ ) can generate comparable predictions with those from PF and PMCMC for large sample sizes, as presented in Figs. S2 and S4. This also implies that the copula structure in the vine model may not have a significant effect on the performance of PCMH. Therefore, the fixed copula function (e.g., Frank) is recommended in the vine model to quantify parameter interdependence in the data assimilation process.

#### 4. Real case study

##### 4.1. Overview of the catchment

The applicability of PCMH would be further demonstrated through a long-term data assimilation problem at the River Ouse. The River Ouse is located in North Yorkshire, which has a mainstream length of 208 km, making it the sixth longest river of the United Kingdom and the longest to flow entirely in one county. Fig. 8 shows the location of the River Ouse. The River Ouse flows from northwest to southeast, with the altitude ranging from 5.6 m to 714.3 m. The catchment is mostly covered by grassland (43.92%) and arable/horticultural crops (31.34%), with some parts covered by mountain, heath and bog plants as well as woodland. This catchment has mixed geological structures, mainly consisting of moderate permeability bedrocks and mixed permeability superficial

**Table 2**  
Model parameters for GR4J and IHACRES.

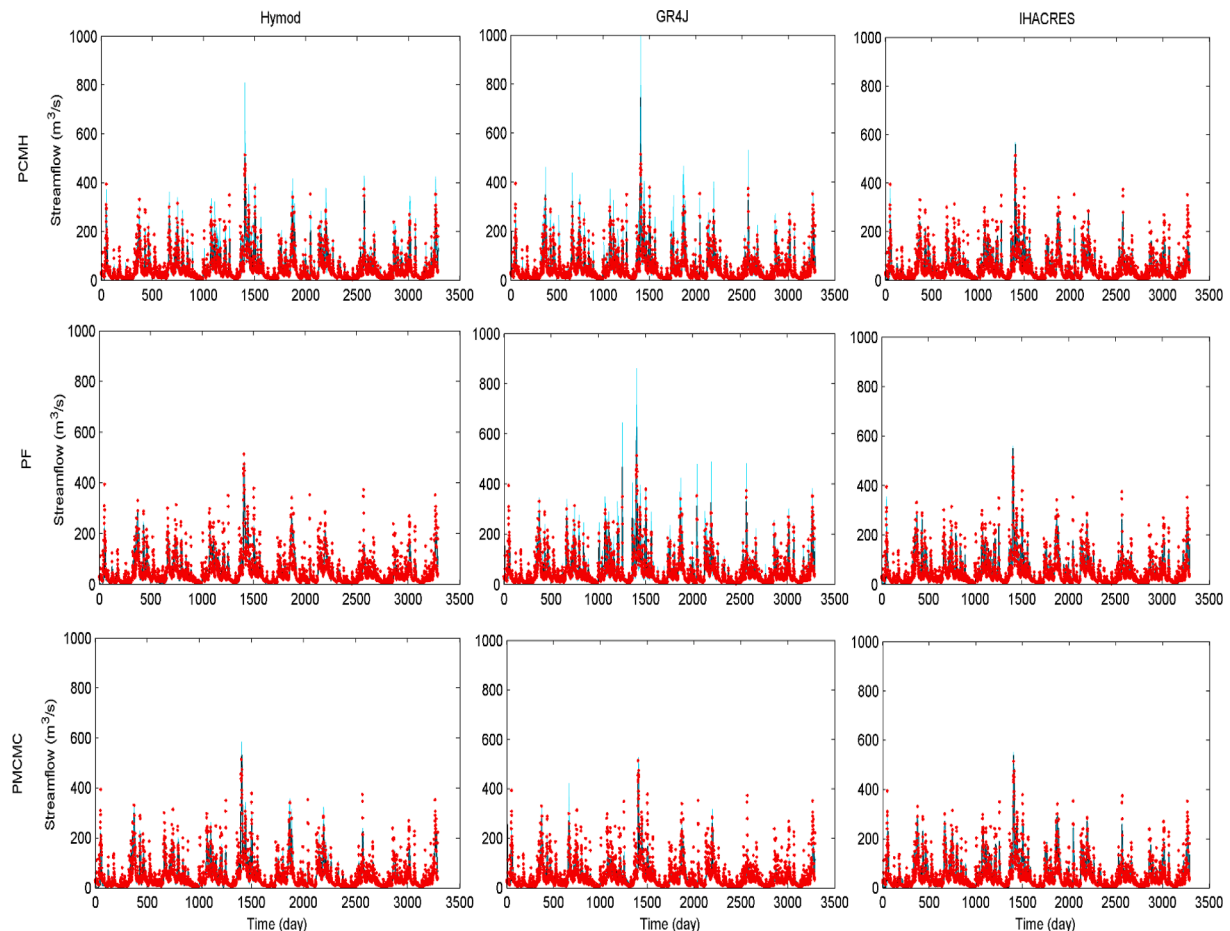
Models	Parameter	Description	Range
GR4J	$X_1$ (mm)	Capacity of the production soil (SMA) store	[100, 1200]
	$X_2$ (mm)	Groundwater exchange coefficient	[-5, 3]
	$X_3$ (mm)	Capacity of the routing store	[20, 300]
	$X_4$ (day)	Time parameter for unit hydrographs	[0.5, 4]
IHACRES	$\tau_q$ (day)	Time constant governing the rate of recession of quickflow	[0.5, 10]
	$\tau_s$ (day)	Time constant governing rate of recession of slowflow, $\tau_q < \tau_s$	[10, 350]
	$V_s$	The proportion of slow flow to total flow	[0, 1]
	$d$ (mm)	CMD threshold for producing flow	[50, 550]
	$e$	Conversion parameter from potential evapotranspiration to actual evapotranspiration	[0.01, 1.5]
	$f$	CMD stress threshold as a proportion of $d$	[0.01, 3]

deposits, as well as highly permeable bedrocks in lower reaches. The streamflow records from 1997 to 2006 at the Gauge station in Skelton are adopted in this study. The associated precipitation data are also obtained from the National River Flow Archive (<https://nrfa.ceh.ac.uk/>) and the potential evapotranspiration data are obtained through averaging the historic gridded potential evapotranspiration developed by Tanguy et al. (2017).

#### 4.2. Hydrologic models

In addition to the Hymod used in the synthetic experiment, two more hydrologic models, GR4J and IHACRES will also be employed to comprehensively demonstrate the applicability of the developed PCMH method. The GR4J model is a lumped hydrologic model with four parameters and can be run in daily scale. The structure of GR4J is shown in Fig. 2(b), in which it consists of an interception reservoir is employed to intercept rainfall and potential evapotranspiration, a soil moisture accounting procedure is adopted to generate effective rainfall, and a water exchange term used to model water losses to or gains from deep aquifers. The GR4J model also has a routing module including two flow components with constant volumetric split (10–90%), two unit hydrographs, and a non-linear routing store (as shown in Fig. 2(b)). More details about GR4J can be found in some literates (e.g. Perrin et al., 2003; Westra et al., 2014; Smith et al., n.d.).

The IHACRES model (Jakeman et al., 1990; Jakeman and Hornberger, 1993) is another conceptual rainfall-runoff model to be employed in this study. It typically has unknown parameters ranging between 5 and 7 parameters. The IHACRES model has many versions, developed by a number of studies (e.g. Viney et al. 2009; Vaze et al. 2010; Shin et al., 2013; Shin et al., 2015). In this study, the catchment moisture deficit (CMD) version developed by Croke and Jakeman (2004) as it has been recognized to have more physical meaning for the parameters (Shin and Kim, 2017). As shown in Fig. 2(c), the CMD-based IHACRES model is composed of two modules, in which a nonlinear module is adopted to generate effective rainfall as well as the CMD output, and the linear module is employed to translate effective rainfall into streamflow by routing it through two parallel linear stores based on



**Fig. 9.** Comparison between observations and predictions from different hydrologic models quantified by different data assimilation techniques: the red points show the real observations, the black solid lines indicate the predictive means and the cyan belts exhibit the 90% predictive intervals.

**Table 3**  
Uncertainty scenarios in model inputs, outputs and sample sizes.

Scenarios	Proportional factor for PET	Proportional factor for precipitation	Proportional factor of observation	Sample size
1	0.15	0.15	0.15	50
2	0.15	0.15	0.15	100
3	0.15	0.15	0.15	200
4	0.25	0.25	0.25	50
5	0.25	0.25	0.25	100
6	0.25	0.25	0.25	200
7	0.35	0.35	0.35	50
8	0.35	0.35	0.35	100
9	0.35	0.35	0.35	200

the instantaneous unit hydrograph theory (Croke and Jakeman, 2007).

Table 2 presents the model parameters in GR4J and IHACRES. There are four unknown model parameters (defined as  $X_1 - X_4$ ) in GR4J model. There are also other modified versions of GR4J with five and six parameters (i.e. GR5J and GR6J), but the four-parameter model (i.e. GR4J) will be adopted in this study. The precipitation and potential evapotranspiration with a daily scale will be used as the inputs for GR4J. The IHACRES model would have different parameters for different versions. In current study, six parameters will need to be estimated for the CMD-based IHACRES model, as presented in Table 2. The IHACRES model can be run based on different inputs such as precipitation, temperature and evapotranspiration. Similar with GR4J and Hymod, the precipitation and potential evapotranspiration will be adopted to force the IHACRES model.

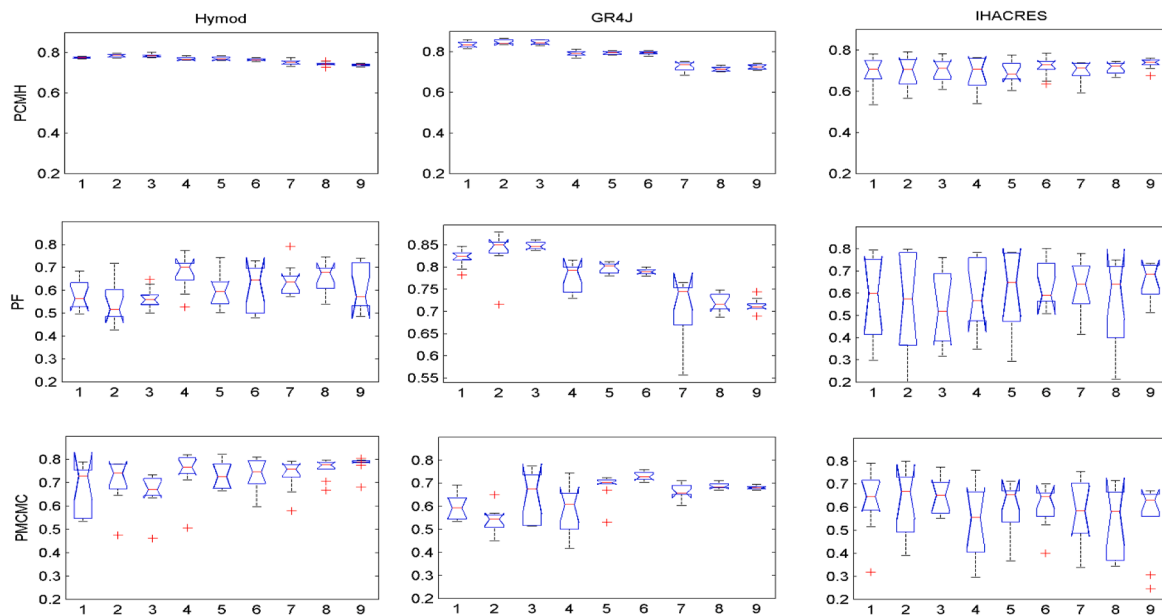
### 4.3. Results analysis for the real-case study

The developed PCMH method, as well as PF and PMCMC are employed to quantify uncertainties in the hydrologic models of Hymod, GR4J and IHACRES on streamflow predictions for the River Ouse. A proportional factor of 0.15 is adopted for the random errors to be added to inputs and outputs to reflect their inherent uncertainties, in which the normal distributed errors are used for potential evapotranspiration and streamflow whilst the log-normal distributed error is used for

precipitation. Moreover, a total number of 10-year data (1997–2006) are employed to demonstrate the applicability of PCMH in hydrological data assimilation processes. All the three data assimilation methods are run with a sample size of 100. Fig. S5 presents the flowchart for the real-case experiment.

Fig. 9 exhibits the comparison between observations and predictions from different hydrologic models quantified by different data assimilation techniques, in which the solid lines indicate the mean prediction at each time step from data assimilation and the cyan region shows the corresponding 90% predictive interval bounded by the 5% and 95% quantiles of the streamflow prediction samples. The first column shows the performances of PCMH, PF and PMCMC on the Hymod. The results indicate that the predictive means from PCMH can well track the variations of real streamflow records, while in comparison, both PF and PMCMC have led to underestimation for many high flow records. Moreover, the 90% predictive intervals from PCMH can well bracket the real observations even though large predictive intervals can be observed at some high flow periods. Similar with the deterministic predictions, PF and PMCMC also generate underestimation for probabilistic predictions at high flow periods, which cannot cover the corresponding observations. For the GR4J model, predictions from both PCMH and PF agree well with most observations except overestimations from high flows. But PCMH seems to produce slightly higher overestimations than PF. However, the PMCMC method would produce underestimation for both deterministic and probabilistic predictions at many high flow periods, which is similar with results from Hymod. For the IHACRES model as shown in the last column of Fig. 9, it can be observed that all the three approaches perform well without noticeable discrepancy between predictions and observations. Also, the 90% predictive intervals from PCMH, PF and PMCMC would well bracket the corresponding observations. In general, the developed PCMH method performed well on all the three hydrologic models, while the PF method produced underestimations for Hymod at some high flow periods and PMCMC also generated underestimations for Hymod and IHACRES at some high flow periods. This can demonstrate the robustness of the developed PCMH method on applications to different hydrologic models.

To further demonstrate the reliability of PCMH, we have run PCMH, PF and PMCMC for different hydrological models with different input



**Fig. 10.** Variations of NSE coefficients for PCMH, PF and PMCMC on different models under different uncertainty scenarios: {1, 2, ..., 9} on x-axis indicate the uncertainty scenarios presented in Table 3.

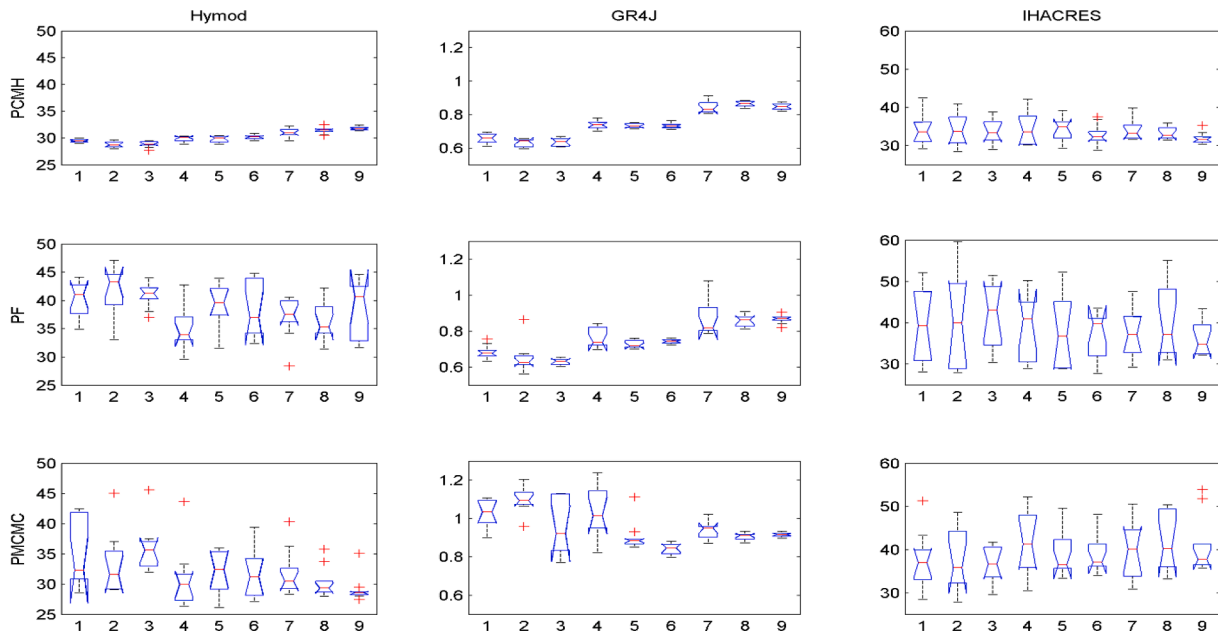


Fig. 11. Variations of RMSE for PCMH, PF and PMCMC on different models under different uncertainty scenarios: {1, 2, ..., 9} on x-axis indicate the uncertainty scenarios presented in Table 3.

and output uncertainties, as well as different sample sizes. In detail, a total number of 9 scenarios, as presented in Table 3, have been designed with different uncertainties in model inputs, outputs and sample sizes. Moreover, 10 runs will be conducted for each scenario to reveal the robustness of PCMH, PF and PMCMC.

The boxplot shown in Fig. 10 exhibit the variations of NSE coefficients for PCMH, PF and PMCMC on different models under different uncertainty scenarios. Here the NSE value is generated by comparing the mean predictions and observations through Equation (20). Under each

scenario, 10 runs were performed for a data assimilation (i.e., PCHM, PF or PMCMC) method on a hydrological model (i.e., Hymod, GR4J, or IHACRES) and thus 10 NSE values would be correspondingly obtained. It can be concluded that the PCMH would generate reliable predictions with acceptable variations even for the IHACRES model with small sample sizes (e.g. Scenario 1, 4 and 7). In comparison, both PF and PMCMC show remarkable variations for their performances on all the three hydrological models, especially under small sample size scenarios. Table S2 presents the performances of PCMH, PF and PMCMC on

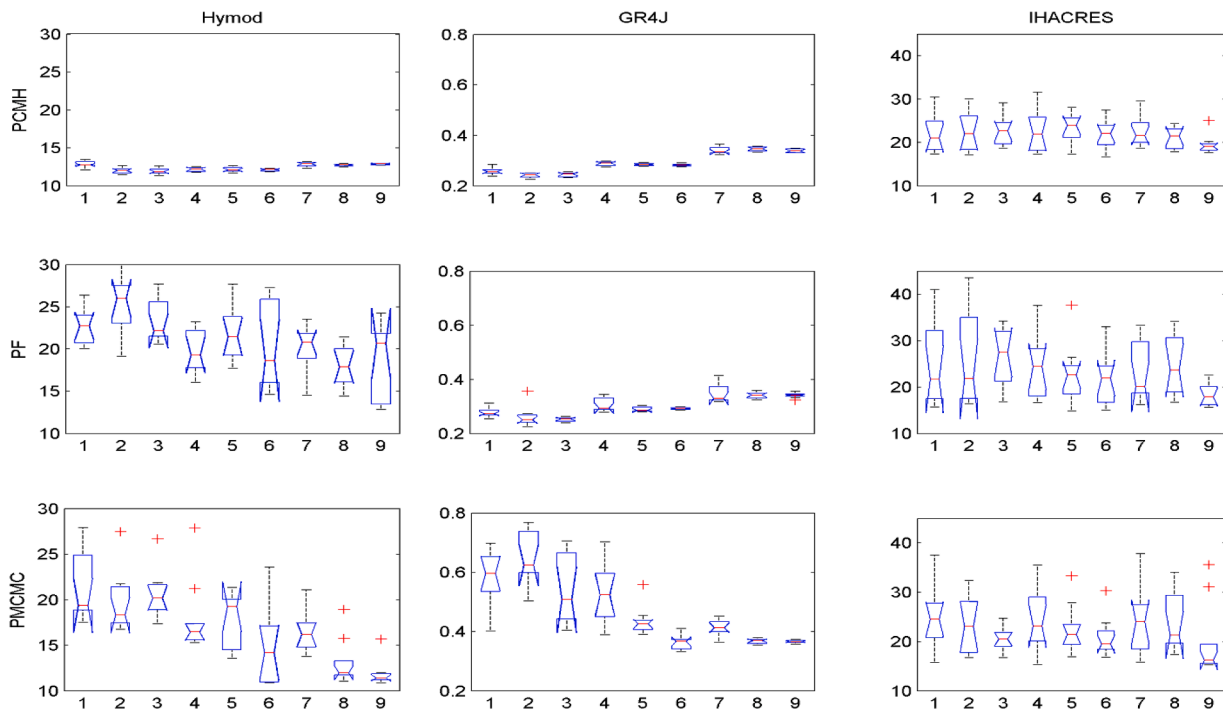
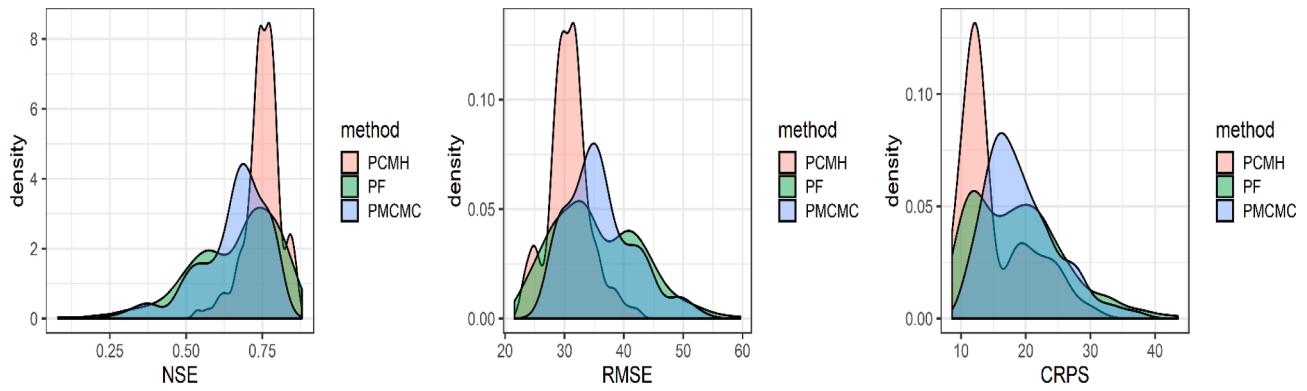


Fig. 12. Variations of CRPS for PCMH, PF and PMCMC on different models under different uncertainty scenarios: {1, 2, ..., 9} on x-axis indicate the uncertainty scenarios presented in Table 3.



**Fig. 13.** The overall performances of PCMH, PF and PMCMC: a total number of 270 replicates are conducted by each data assimilation (i.e.  $3 \times 9 \times 10$ , consisting of three hydrologic models, nine scenarios, and ten runs for each scenarios).

deterministic predictions evaluated by the NSE coefficient for different hydrologic models under different running scenarios. For one specific uncertainty scenario, the PCMH can produce acceptable predictions even for multiple replicates, indicating the reliability of the PCMH method. As presented in Table S2, the worst case of PCMH occurred for the scenarios of (0.15, 0.15, 0.15, 50). Nevertheless, the predictions from PCMH are still satisfactory with the NSE coefficient ranging between 0.534 and 0.781. In comparison, the PF and PMCMC would have chances to generate unacceptable predictions with the NSE coefficient being as low as around 0.3. Moreover, the proposed PCMH approach seems not to be influenced remarkably by uncertainties in inputs, outputs and sample sizes, with the NSE coefficient varying within [0.72, 0.80] for Hymod, [0.70, 0.87] for GR4J and [0.53, 0.79] for IHACRES. In comparison, the performances of PF and PMCMC would vary significantly for different uncertainty scenarios. Even though they may produce better results (e.g. an NSE coefficient larger than 0.82 for PMCMC on Hymod) than PCMH, their performances are not as robust as PCMH and they may also generate unacceptable predictions with the NSE coefficient as low as 0.2 for some cases. These conclusions are consistent with the results in Fig. 10.

Fig. 11 and Table S3 show the performances, evaluated by RMSE, of PCMH, PF and PMCMC on different hydrologic models under different uncertainty scenarios. Here the RMSE values for GR4J are significantly smaller than those for Hymod and IHACRES. This is because that the outputs of GR4J are denoted as streamflow depth in mm, where the outputs from Hymod and IHACRES are expressed as streamflow volume in  $m^3/s$ . We can observe low RMSE values with small variations for PCMH, especially on the hydrologic models of Hymod and GR4J. In comparison, the performances of PF and PMCMC would vary remarkably, which may lead to inaccurate predictions even for a small size of 200.

In addition to deterministic predictions, one of the attractive advantages for PCMH, PF and PMCMC is that they can also produce probabilistic predictions, which can be adapted for future risk inferences. The probabilistic predictions from the three data assimilation methods on different hydrologic models are evaluated by CRPS expressed as Eq. (21). Table S4 and Fig. 12 exhibit the performance variations of PCMH, PF and PMCMC for probabilistic predictions through different hydrologic models. Similar with RMSE values, the CRPS values from GR4J model is not comparable with those from Hymod and IHACRES since the outputs from GR4J are denoted as streamflow depth in mm. From Table S4 and Fig. 9, we can observe the robustness of PCMH on different hydrologic models for generating probabilistic predictions under different uncertainty scenarios. In detail, the CRPS values ranged between 11.2 and 13.5 for PCMH on Hymod, between 0.22 and 0.37 for GR4J, and between 16.6 and 31.6 for IHACRES. There are large discrepancies between the CRPS values from GR4J and those from Hymod and IHACRES. This is because that the

CRPS from GR4J is calculated based on the streamflow depth whilst this index for the other two models is generated based on streamflow rate. In comparison, as presented in Fig. 12, even though PF and PMCMC can produce reliable probabilistic predictions through GR4J model under some uncertainty scenarios (e.g. Scenarios 6 to 9 for PMCMC and Scenarios 3, 5, 6, 8, 9 for PF), they would have much more chances to generate discrepant results than PCMH.

To demonstrate the applicability of the proposed PCMH method, a total number of nine scenarios have been designed with each running 10 times through three hydrologic models (i.e. Hymod, GR4J, IHACRES). Consequently, we have run 270 replicates for PCMH, as well as PF and PMCMC. Fig. 13 exhibits the performance variations for all the replicates evaluated by NSE coefficient, RMSE and CRPS. Here, the RMSE and CRPS values for GR4J, evaluated in streamflow depth as shown in Tables S3 and S4 and Figs. 11 and 12, are converted into streamflow rate in order to make them comparable with those from Hymod and IHACRES. The results indicate that the proposed PCMH approach would lead to sharp distributions for its performance variations, while in comparison, the PF and PMCMC methods would have much flatter distributions. In detail, the mean values and the associated variation intervals of NSE over the total 270 runs for PCMH, PF and PMCMC are 0.752 (with a variation interval of [0.534, 0.866]), 0.661 (with a variation interval of [0.080, 0.879]), and 0.655 (with a variation interval of [0.247, 0.824]), respectively. Similarly, for the probabilistic predictions, the mean values and fluctuation ranges from PCMH, PF and PMCMC are respectively 15.215 ([8.624, 31.549]), 18.758 ([8.595, 43.536]), 19.308 ([10.848, 37.799]). These results suggest that the proposed PCMH method is more robust than PF and PMCMC in generating reliable hydrologic predictions and seems to be influenced less by the hydrologic models to be used, uncertainty scenarios, and its inherent replicates.

## 5. Discussion

In the proposed PCMH method, the new particles for model parameters are evolved through a mixed sampling scheme consisting of both Gaussian perturbation (i.e. Eq. (9)) and copula sampling (i.e. Eqs. (15) to (17)) methods. The vine copula approach is introduced into PCMH to reflect parameter interdependence among model parameters and the Gaussian perturbation approach aims to enhance the chosen chances for the particles around posterior samples. The proposed PCMH method can rejuvenate model parameters in wider spaces than those of PF and PMCMC, and effectively alleviate the sample impoverishment, especially in long-term data assimilation processes.

The CopPF method proposed by Fan et al. (2017b) has already introduced copula-based sample scheme into data assimilation through Particle Filter. However, the proposed PCMH is significantly different from CopPF. In PCMH, the candidate samples are generated by mixing the samples from Gaussian perturbation (i.e. Eq. (9)) and vine copula (i.

**Table 4**  
The performance comparison between PCMH and CopPF under different sample size scenarios.

Criteria	Sample size	PCMH			CopPF		
		Mean	Min	Max	Mean	Min	Max
NSE	50	0.775	0.768	0.782	0.774	0.740	0.804
	100	0.785	0.773	0.797	0.762	0.712	0.783
	200	0.784	0.775	0.801	0.750	0.671	0.789
	50	29.486	29.027	29.972	29.527	27.553	31.698
RMSE	100	28.803	28.025	29.635	30.342	28.956	33.362
	200	28.870	27.712	29.497	31.002	28.556	35.678
	50	12.838	12.096	13.453	15.825	15.077	17.747
CRPS	100	12.046	11.438	12.662	15.462	14.408	17.725
	200	11.919	11.298	12.625	15.327	13.412	19.567

e. Eqs. (15) to (17)) methods, and the Metropolis acceptance ratio is then employed to determine the acceptance of the candidate samples. Such a scheme can (i) balances possible extreme samples from the vine copula method, and (ii) remove unacceptable samples through the Metropolis ratio. In comparison, all new sample would be from either the copula method or the Gaussian perturbation method in CopPF, which suggests extreme samples from the outlier of the copula model (even though not too many) may exist in parameter evolution, especially for large sample sizes.

Table 4 present the performances of PCMH and CopPF through Hymod with a proportional factor of 0.15 for reflecting uncertainties in model inputs and outputs, and sample sizes of 50, 100 and 200. Moreover, 10 replicates have been conducted for CopPF to make the results comparable with those from PCMH. The results indicate that the CopPF is also able to generate reliable predictions under different sample size scenarios. However, compared with PCMH, the CopPF performed slightly worse in both deterministic and probabilistic predictions. Moreover, the CopPF seems to be less robust than PCMH, especially under large sample size scenarios. For instance, the NSE values from CopPF range between 0.740 and 0.804 for a sample size of 50, and between 0.671 and 0.789 for a sample size of 200, while in comparison, the proposed PCMH in this study would bring NSE values respectively ranging within [0.768, 0.782] and [0.775, 0.801]. This is mainly because that the CopPF would have more chances to involve abnormal (or extreme) samples generated by the copula sampling scheme (expressed by Eqs. (15) to (17)), and also it does not have a mechanism remove those abnormal samples.

## 6. Conclusions

In this study, a particle copula Metropolis-Hastings (PCMH) approach has been proposed to provide reliable predictions in hydrological data assimilation. The proposed PCMH approach can improve the hydrological data assimilation process through a mixed sample evolution scheme consisting of Gaussian evolution (expressed by Eq. (8)) and copula-based evolution schemes. The Metropolis ratio is then employed to determine the acceptance of the candidate samples. The developed PCMH method is able to rejuvenate model parameters in wider spaces than those of PF and PMCMC, and effectively alleviate the sample impoverishment, especially in long-term data assimilation processes.

The applicability of the PCMH approach was comprehensively demonstrated at the River Ouse in UK, through three hydrological models and different uncertainty settings in inputs, outputs and sample sizes. A total number of nine scenarios consisting of different uncertainty settings in model inputs, outputs and sample sizes are designed, with each scenario being replicating for 10 times for each hydrological model. The obtained results indicate that the PCMH approach can generally generate reliable predictions for all hydrological models, with the NSE values larger than 0.53 for all 270 runs. Particularly, the PCMH

method shows more robustness than traditional PF and PMCMC methods, which can produce better NSE, RMSE, and CRPS values with narrower fluctuation ranges than those from PF and PMCMC. Moreover, compared with the CopPF approach, the developed PCMH method can bring more robust results, especially under large sample size scenarios. This is mainly due to the capabilities of PCMH in balancing possible extreme samples from the copula sampling procedure by the mixed evolution scheme, and also removing unacceptable samples through the Metropolis ratio method.

This study has demonstrated that the proposed PCMH approach can generate more reliable predictions in hydrological data assimilation. Even though only the conceptual models have been tested for the PCMH method, this method can also be applicable for complex hydrological or land surface models. However, one challenge for the application of PCMH is that, when a large number of parameters are included in hydrological or land surface models, it may be complex and time consuming to quantify interdependence among those model parameters. Nevertheless, this issue can be resolved through the following ideas: i) parameter correlation screening is introduced, and the vine copula models are built for those highly correlated parameters whilst the others are sampled independently; ii) the vine copula models are used for the parameters in the same module/component or the same grid for complex models and thus different vine copula models will be used to quantify parameter interdependence in different modules or grids. Moreover, it can be observed from the results that, even for the same data assimilation method, it will produce predictions with different accuracies from different hydrological models under different uncertainties in inputs, outputs, and sample sizes. It is apparent that the predictions from data assimilation would be influenced by various factors such as the hydrological models and DA methods to be used, as well as uncertainties in inputs, outputs and sample sizes. However, currently it is not clear which factor would pose dominant effects on the predictability of the data assimilation. This issue is to be investigated in our companion paper (Fan et al., 2022).

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data will be made available on request.

## Acknowledgement

This research was supported by the Royal Society International Exchanges Program (No. IES\R2\202075).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jhydrol.2022.128163>.

## References

- Aas, K., Czado, C., Frigessi, A., Bakken, H., 2009. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics* 44, 182–198.
- Abbaszadeh, P., Moradkhani, H., Yan, H., 2018. Enhancing hydrologic data assimilation by evolutionary Particle Filter and Markov Chain Monte Carlo. *Advances in Water Resources* 111, 192–204. <https://doi.org/10.1016/j.advwatres.2017.11.011>.
- Abbaszadeh, P., Moradkhani, H., Daescu, D.N., 2019. The Quest for Model Uncertainty Quantification: A Hybrid Ensemble and Variational Data Assimilation Framework. *Water Resources Research* 55 (3), 2407–2431. <https://doi.org/10.1029/2018WR023629>.
- Arulampalam, M.S., Maskell, S., Gordon, N., Clapp, T., 2002. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing* 50, 174–188.
- Bárdossy, A., Hörning, S., 2016. Gaussian and non-Gaussian inverse modeling of groundwater flow using copulas and random mixing. *Water Resources Research* 52 (6), 4504–4526.
- Brechmann, E.C., Schepsmeier, U., 2013. Modeling Dependence with C- and D-Vine Copulas: The R Package CDvine. *Journal of Statistical Software* 52 (3), 1–27.
- Croke, B., Jakeman, A., 2007. Use of the IHACRES rainfall-runoff model in arid and semi-arid regions. In: Wheatler, H., Sorooshian, S., Sharma, K. (Eds.), *Hydrological Modelling in Arid and Semi-Arid Areas* (International Hydrology Series. Cambridge University Press, Cambridge, pp. 41–48. <https://doi.org/10.1017/CBO9780511535734.005>.
- Di, Z.H., Ye, A.Z., Duan, Q.Y., Wang, X.X., 2021. Assessment of Parametric Sensitivity Analysis Methods Based on a Quasi Two-Dimensional Groundwater Model. *Journal of Environmental Informatics* 37 (1), 62–78. <https://doi.org/10.3808/jei.201900413>.
- Evensen, G., 2003. The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics* 53, 343–367.
- Fan, Y.R., 2019. Uncertainty Quantification in Hydrologic Predictions: A Brief Review. *Journal of Environmental Informatics Letters* 2 (2), 48–56.
- Fan, Y.R., Huang, G.H., Baetz, B.W., Li, Y.P., Huang, K., Li, Z., Chen, X., Xiong, L.H., 2016. Parameter uncertainty and temporal dynamics of sensitivity for hydrologic models: A hybrid sequential data assimilation and probabilistic collocation method. *Environmental Modelling & Software* 86, 30–49.
- Fan, Y.R., Huang, G.H., Baetz, B.W., Li, Y.P., Huang, K., 2017a. Development of a Copula-based Particle Filter (CopPF) Approach for Hydrologic Data Assimilation under Consideration of Parameter Interdependence. *Water Resour. Res.* 53 (6), 4850–4875.
- Fan, Y.R., Huang, G.H., Baetz, B.W., Li, Y.P., Huang, K., Chen, X., Gao, M., 2017b. Development of Integrated Approaches for Hydrological Data Assimilation through Combination of Ensemble Kalman Filter and Particle Filter Methods. *Journal of Hydrology (Elsevier)* 550, 412–426.
- Fan, Y.R., Huang, G.H., Li, Y.P., Baetz, B.W., Huang, K., 2020. Uncertainty Characterization and Partition in Multivariate Risk Inference: A Factorial Bayesian Copula Framework. *Environmental Research* 183, 109215.
- Fan, Y.R., Huang, W.W., Li, Y.P., Huang, G.H., Huang, K., Li, Y.P., 2015. A coupled ensemble filtering and probabilistic collocation approach for uncertainty quantification of hydrological models. *Journal of Hydrology* 530, 255–272.
- Fan, Y.R., Shi, X., Duan, Q.Y., Yu, L., et al., 2022. Towards reliable uncertainty quantification for hydrologic predictions, Part II: Characterizing impacts of uncertain factors through an iterative factorial data assimilation framework. *Journal of Hydrology*. <https://doi.org/10.1016/j.jhydrol.2022.128136>.
- Favre, A.C., El Adlouni, S., Perreault, L., Thiémond, N., Bobée, B., 2004. Multivariate hydrological frequency analysis using copulas. *Water Resour. Res.* 40, W01101. <https://doi.org/10.1029/2003WR002456>.
- M. Ghaith Z. Li B.W. Baetz Uncertainty Analysis for Hydrological Models with Interdependent Parameters: An Improved Polynomial Chaos Expansion Approach. *Water Resources Research* 57 8 2021 e2020WR029149.
- Ghaith, M., Li, Z., 2020. Propagation of parameter uncertainty in SWAT: A probabilistic forecasting method based on polynomial chaos expansion and machine learning. *Journal of Hydrology* 586, 124854.
- Gordon, N.J., Salmond, D.J., Smith, A.F.M., 1993. Novel approach to nonlinear/nonGaussian Bayesian state estimation. *IEEE Proceedings F: Radar Signal Process.* 140 (2), 107–113.
- Hernández, F., Liang, X., 2018. Hybridizing Bayesian and variational data assimilation for robust high-resolution hydrologic forecasting. *Hydrology and Earth System Sciences* 22, 5759–5779. <https://doi.org/10.5194/hess-22-5759-2018>.
- Hu, Y.M., Liang, Z.M., Solomatine, D.P., Wang, H.M., Liu, T., 2021. Assessing the Impact of Precipitation Change on Design Annual Runoff in the Headwater Region of Yellow River. *Journal of Environmental Informatics* 37 (2), 122–129. <https://doi.org/10.3808/jei.201900421>.
- Huang, K., Fan, Y.R., 2021. Parameter Uncertainty and Sensitivity Evaluation of Copula-Based Multivariate Hydroclimatic Risk Assessment. *Journal of Environmental Informatics* 38 (2), 131–144. <https://doi.org/10.3808/jei.202100462>.
- Jakeman, A.J., Hornberger, G.M., 1993. How much complexity is warranted in a rainfall-runoff model? *Water Resour. Res.* 29 (8), 2637–2649.
- Jakeman, A.J., Littlewood, I.G., Whitehead, P.G., 1990. Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments. *J. Hydrol.* 117, 275–300.
- Khaki, M., Hoteit, I., Kuhn, M., Awange, J., Forootan, E., van Dijk, A.L.J.M., Schumacher, M., Pattiaratchi, C., 2017. Assessing sequential data assimilation techniques for integrating GRACE data into a hydrological model. *Advances in Water Resources* 107, 301–316.
- Liu, J.S., Chen, R., Logvinenko, T., et al., 2001. A theoretical framework for sequential importance sampling and resampling, in *Sequential Monte Carlo Methods in Practice*. Springer, New York.
- Liu, Y., Gupta, H.V., 2007. Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. *Water Resour. Res.* 43, W07401. <https://doi.org/10.1029/2006WR005756>.
- Liu, Y., Weerts, A.H., Clark, M., Hendricks Franssen, H.-J., Kumar, S., Moradkhani, H., Seo, D.-J., Schwanenber, D., Smith, P., van Dijk, A.L.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.
- Lyu, X.D., Fan, Y.R., 2021. Characterizing Impact Factors on the Performance of Data Assimilation for Hydroclimatic Predictions through Multilevel Factorial Analysis. *Journal of Environmental Informatics* 38 (1), 68–82. <https://doi.org/10.3808/jei.202100463>.
- Moor, R.J., 1985. The probability-distributed principle and runoff production at point and basin scales. *Hydrological Science Journal* 30, 273–297.
- Moradkhani, H., 2008. Hydrologic remote sensing and land surface data assimilation. *Sensors* 8 (5), 2986–3004.
- 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. <https://doi.org/10.1029/2012WR012144>.
- Moradkhani, H., Sorooshian, S., Gupta, H.V., Houser, P., 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), W05012. <https://doi.org/10.1029/2004WR003604>.
- Moradkhani, H., Nearing, G.S., Abbaszadeh, P., Pathiraja, S., 2019. Fundamentals of data assimilation and theoretical advances. In: Duan, Q., Pappenberger, F., Wood, A., Cloke, H.L., Schaake, J.C. (Eds.), *Handbook of Hydrometeorological Ensemble Forecasting*. Springer, Berlin. [https://doi.org/10.1007/978-3-642-39925-1\\_30](https://doi.org/10.1007/978-3-642-39925-1_30).
- Moradkhani, H., Sorooshian, S., 2008. General Review of Rainfall-Runoff Modeling: Model Calibration, Data Assimilation, and Uncertainty Analysis. In: Sorooshian, S. (Ed.), *Hydrological Modelling and the Water Cycle: Coupling the Atmospheric and Hydrological Models*. Springer, p. 2008.
- Murphy, A.H., Winkler, R.L., 1987. A general framework for forecast verification. *Mon. Weather Rev.* 115, 1330–1338.
- Nelsen, R.B., 2006. *An Introduction to Copulas*. Springer, New York.
- Perrin, C., Michel, C., Andréassian, V., 2003. Improvement of a parsimonious model for streamflow simulation. *J. Hydrol.* 279, 275–289. [https://doi.org/10.1016/s0022-1694\(03\)00225-7](https://doi.org/10.1016/s0022-1694(03)00225-7).
- Qing, Y., Wang, S., Ancell, B., Yang, Z.L., 2022. Accelerating flash droughts induced by the joint influence of soil moisture depletion and atmospheric aridity. *Nature Communications* 13, 1139.
- Renard, B., Kavetski, D., Kuczera, G., Thyer, M., Franks, S.W., 2010. Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. *Water Resour. Res.* 46, W05521. <https://doi.org/10.1029/2009WR008328>.
- Salamon, P., Feyen, L., 2010. Disentangling uncertainties in distributed hydrological modeling using multiplicative error models and sequential data assimilation. *Water Resources Research* 46, W12501. <https://doi.org/10.1029/2009WR009022>.
- Shen, Z., Tang, Y., 2015. A modified ensemble Kalman particle filter for non-Gaussian systems with nonlinear measurement functions. *J. Adv. Model. Earth Syst.* 7, 50–66. <https://doi.org/10.1002/2014MS000373>.
- Shin, M.-J., Guillaume, J.H.A., Croke, B.F.W., Jakeman, A.J., 2013. Addressing ten questions about conceptual rainfall-runoff models with global sensitivity analyses in R. *J. Hydrol.* 503, 135–152.
- Shin, M.-J., Guillaume, J.H.A., Croke, B.F.W., Jakeman, A.J., 2015. A review of foundational methods for checking the structural identifiability of models: results for rainfall-runoff. *J. Hydrol.* 520, 1–16.
- Shin, M.-J., Kim, C.-S., 2017. Assessment of the suitability of rainfall-runoff models by coupling performance statistics and sensitivity analysis. *Hydrology Research* 48 (5), 1192–1213.
- Smith K.A., Barker L.J., Tanguy M., Parry S., Harrigan S., Legg T.P., Prudhomme C., and Hannaford J., A multi-objective ensemble approach to hydrological modelling in the UK: an application to historic drought reconstruction, *Hydrol. Earth Syst. Sci.*, 23, 3247–3268.
- Sun, C.X., Huang, G.H., Fan, Y.R., Zhou, X., Lu, C., Wang, X.Q., 2019. Drought occurring with hot extremes: changes under future climate change on Loess Plateau, China. *Earth's Future* (AGU) 7, 587–604. <https://doi.org/10.1029/2018EF001103>.
- Tanguy, M., Prudhomme, C., Smith, K., Hannaford, J., 2017. Historic Gridded Potential Evapotranspiration (PET) based on temperature-based equation McGuinness-Bordne calibrated for the UK (1891–2015). NERC Environmental Information Data Centre. <https://doi.org/10.5285/17b9c4f7-1c30-4b6f-b2fe-f7780159939c>.

- Vaze, J., Post, D.A., Chiew, F.H.S., Perraud, J.M., Viney, N.R., Teng, J., 2010. Climate non-stationarity – validity of calibrated rainfall-runoff models for use in climate change studies. *J. Hydrol.* 394 (3–4), 447–457.
- Viney, N.R., Bormann, H., Breuer, L., Bronstert, A., Croke, B.F.W., Frede, H., Gräff, T., Hubrechts, L., Huisman, J.A., Jakeman, A.J., Kite, G.W., Lanini, J., Leavesley, G., Lettenmaier, D.P., Lindström, G., Seibert, J., Sivapalan, M., Willems, P., 2009. Assessing the impact of land use change on hydrology by ensemble modelling (LUCHEM) II: ensemble combinations and predictions. *Adv. Water Resour.* 32 (2), 147–158.
- Vrugt, J.A., ter Braak, C.J.F., Diks, C.G.H., Schoups, G., 2013. Hydrologic data assimilation using particle Markov chain Monte Carlo simulation: Theory, concepts and applications. *Advances in Water Resources* 51, 457–478.
- 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.
- Wang, S., Ancell, B.C., Huang, G.H., Baetz, B.W., 2018. Improving Robustness of Hydrologic Ensemble Predictions Through Probabilistic Pre- and Post-Processing in Sequential Data Assimilation. *Water Resources Research* 54 (3), 2129–2151.
- Wang, S., Huang, G.H., Baetz, B.W., Huang, W., 2015. A polynomial chaos ensemble hydrologic prediction system for efficient parameter inference and robust uncertainty assessment. *Journal of Hydrology* 530, 716–733.
- Wang, F., Huang, G.H., Fan, Y.R., Li, Y.P., 2022. Development of a disaggregated multi-level factorial hydrologic data assimilation model. *Journal of Hydrology* 610, 127802. <https://doi.org/10.1016/j.jhydrol.2022.127802>.
- Wang, S., Wang, Y., 2019. Improving probabilistic hydroclimatic projections through high-resolution convection-permitting climate modeling and Markov chain Monte Carlo simulations. *Clim Dyn* 53, 1613–1636.
- Westra, S., Thyer, M., Leonard, M., Kavetski, D., Lambert, M., 2014. A strategy for diagnosing and interpreting hydrological model nonstationarity. *Water Resour. Res.* 50, 5090–5113. <https://doi.org/10.1002/2013WR014719>.
- Wu, H., Chen, B., Snelgrove, K., Lye, L.M., 2019. Quantification of Uncertainty Propagation Effects during Statistical Downscaling of Precipitation and Temperature to Hydrological Modeling. *Journal of Environmental Informatics* 34 (2), 139–148.
- Yan, H., DeChant, C.M., Moradkhani, H., 2015. Improving soil moisture profile prediction with the particle filter-Markov chain Monte Carlo method. *IEEE Transactions on Geoscience and Remote Sensing.* 53, 6134–6147.
- Yan, H.X., Moradkhani, H., 2016. Combined assimilation of streamflow and satellite soil moisture with the particle filter and geostatistical modeling. *Advances in Water Resources* 94, 364–378.
- Yu, L., Li, Q.W., Jin, W.W., Chen, C., Li, Y.P., Fan, Y.R., Zuo, Q.T., 2020a. Coupling the two-level programming and copula for optimizing energy-water nexus system management—A case study of Henan Province. *Journal of Hydrology* 124832.
- Yu, L., Xiao, Y., Jiang, S., Li, Y.P., Fan, Y.R., Huang, G.H., Lv, J., Zuo, Q.T., Wang, F.Q., 2020b. A copula-based fuzzy interval-random programming approach for planning water-energy nexus system under uncertainty. *Energy* 196, 117063.
- Zhou, P., Li, C., Li, Z., Cai, Y., 2022. Assessing uncertainty propagation in hybrid models for daily streamflow simulation based on arbitrary polynomial chaos expansion. *Advances in Water Resources* 160, 104110.
- Zhang, B., Wang, S., Wang, Y., 2021. Probabilistic Projections of Multidimensional Flood Risks at a Convection-Permitting Scale. *Water Resources Research* 57 1, e2020WR028582. <https://doi.org/10.1029/2020WR028582>.
- Zhou, X., Huang, G.H., Piwowar, J., Fan, Y.R., Wang, X.Q., Li, Z., Cheng, G.H., 2018a. Hydrologic Impacts of Ensemble-RCM-Projected Climate Changes in the Athabasca River Basin, Canada. *Journal of Hydrometeorology (AMS)* 19 (12), 1953–1971.
- Zhou, X., Huang, G.H., Wang, X.Q., Fan, Y.R., Cheng, G.H., 2018b. A coupled dynamical-copula downscaling approach for temperature projections over the Canadian Prairies. *Climate Dynamics, (Springer)* 51 (7–8), 2413–2431.