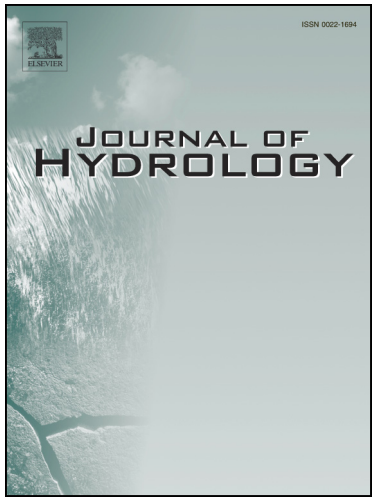
Accepted Manuscript



Research papers

An ensemble-based dynamic Bayesian averaging approach for discharge simulations using multiple global precipitation products and hydrological models

Wei Qi, Junguo Liu, Hong Yang, Chris Sweetapple

PII: S0022-1694(18)30025-8
DOI: https://doi.org/10.1016/j.jhydrol.2018.01.026
Reference: HYDROL 22508

To appear in: *Journal of Hydrology*

Received Date: 27 May 2017
Revised Date: 1 December 2017
Accepted Date: 9 January 2018

Please cite this article as: Qi, W., Liu, J., Yang, H., Sweetapple, C., An ensemble-based dynamic Bayesian averaging approach for discharge simulations using multiple global precipitation products and hydrological models, *Journal of Hydrology* (2018), doi: https://doi.org/10.1016/j.jhydrol.2018.01.026

An ensemble-based dynamic Bayesian averaging approach for discharge simulations using multiple global precipitation products and hydrological models

Wei Qi[1, 2*], Junguo Liu[1*], Hong Yang[3, 4] and Chris Sweetapple[5]

[1] School of Environmental Science and Engineering, South University of Science and Technology of China, Xueyuan Road 1088, Nanshan Districit, Shenzhen, 518055, Shenzhen, China

[2] State Key Laboratory of Water Resource & Hydropower Engineering Science, Wuhan University, Wuhan, 430072, China

[3] Eawag, Swiss Federal Institute of Aquatic Science and Technology, Duebendorf, Switzerland

[4] Department of Environmental Science, University of Basel, Basel, Switzerland

[5] Centre for Water Systems, College of Engineering, Mathematics and Physical Sciences, University of Exeter, North Park Road, Exeter, Devon EX4 4QF, United Kingdom

*Email: qiwei_waterresources@hotmail.com; junguo.liu@gmail.com; liujg@sustc.edu.cn

Highlights:

An ensemble-based dynamic Bayesian averaging approach is proposed

Estimate joint probability of precipitation and hydrological models being correct

Estimate posterior distribution based on magnitude and timing of flows

Outstanding capability to estimate expected discharges

Abstract:

Global precipitation products are very important datasets in flow simulations, especially in poorly gauged regions. Uncertainties resulting from precipitation products, hydrological models and their combinations vary with time and data magnitude, and undermine their application to flow simulations. However, previous studies have not quantified these uncertainties individually and explicitly. This study developed an ensemble-based dynamic Bayesian averaging approach (e-Bay) for deterministic discharge simulations using multiple global precipitation products and hydrological models. In this approach, the joint probability of precipitation products and hydrological models being correct is quantified based on uncertainties in maximum and mean estimation, posterior probability is quantified as functions of the magnitude and timing of discharges, and the law of total probability is implemented to calculate expected discharges. Six global fine-resolution precipitation products and two hydrological models of different complexities are included in an illustrative application. e-Bay can effectively quantify uncertainties and therefore generate better deterministic discharges than traditional approaches (weighted average methods with equal and varying weights and maximum likelihood approach). The mean Nash-Sutcliffe Efficiency values of e-Bay are up to 0.97 and 0.85 in training and validation periods respectively, which are at least 0.06 and 0.13 higher than traditional approaches. In addition, with increased training data, assessment criteria values of e-Bay show smaller fluctuations than traditional approaches and its performance becomes outstanding. The proposed e-Bay approach bridges the gap between global precipitation products and their pragmatic applications to discharge simulations, and is beneficial to water resources management in ungauged or poorly gauged regions across the world.

Keywords: Ensemble simulation; Hydrological modelling; Satellite precipitation products; Uncertainty; Water resources

1 Introduction

Knowledge of river flow plays an important role in understanding the water cycle and, therefore, in water resources management (Sellers, 1997; Sorooshian et al., 2005; Tapiador et al., 2012; Gao and Liu, 2013). However, flow data are not available in many regions, particularly in mountainous districts and rural areas in developing countries. In these regions, due to insufficient gauge observations, alternative approaches, such as hydrological model simulations, are required for efficient water resources assessment (Voisin et al., 2008; Yong et al., 2010; Ward et al., 2011; Beck et al., 2016; Qi et al., 2016a).

In hydrological simulations, precipitation data are very important inputs. There are a few data sources that are available to use: for example, precipitation gauge based datasets, remote sensing based data and reanalysis precipitation products. In these datasets, fine resolution precipitation products have been developed on a global scale, e.g. Tropical Rainfall Measuring Mission (TRMM) products (Huffman et al., 2007), Global Land Data Assimilation System (GLDAS) precipitation products (Kato et al., 2007), Asian Precipitation - Highly-Resolved Observational Data Integration Towards Evaluation of Water Resources (APHRODITE) (Xie et al., 2007; Yatagai et al., 2012), Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) (Sorooshian et al., 2000; Sorooshian et al., 2002), and Global Satellite Mapping of Precipitation product (GSMAP) (Kubota et al., 2007; Aonashi et al., 2009). The uncertainty of precipitation products varies between different months and data magnitudes: for example, Qi et al. (2016a) showed that, in a region in northeast China, TRMM3B42 and GLDAS-1 underestimate high precipitation intensities in summer periods and perform better in spring and autumn; Ochoa et al. (2014) found that the uncertainty of TRMM3B42 varies in different months in a region in western South American. Many studies in other regions across the world revealed similar results, for example, the study on TRMM3B42

in the whole of China by Chen et al. (2013), and the investigation on GSMAP-MVK+ in the United States by Tian et al. (2010).

Many researchers have implemented precipitation products in hydrological simulations (Li et al., 2013; Maggioni et al., 2013; Müller and Thompson, 2013; Xue et al., 2013; Qi et al., 2016a; Wong et al., 2017). In hydrological simulations using precipitation products, uncertainties in simulated discharges have been reported. There are three important uncertainty sources in the simulated discharges: precipitation product uncertainty (Nikolopoulos et al., 2010; Pan et al., 2010; Yong et al., 2010; Yong et al., 2012; Chen et al., 2013; Yong et al., 2014), hydrological model uncertainty (Beven and Binley, 1992; Beven and Freer, 2001; Beven et al., 2007) and uncertainty resulting from their combinations (Qi et al., 2016a). Here, the combination uncertainty represents the nonlinear propagation processes from precipitation product and hydrological model uncertainties to discharge simulation uncertainty. Most recently, Qi et al. (2016a) quantified the influence of the combination uncertainty on discharge simulations and showed that combination uncertainty can contribute a similar magnitude of uncertainty to discharges as hydrological models.

Many uncertainty analysis approaches have been introduced to quantify hydrological simulation uncertainty (Beven and Binley, 1992; Beven and Freer, 2001; Blasone et al., 2008; Vrugt et al., 2009; Nott et al., 2012). In these approaches, one of the commonly used methods is the generalized likelihood uncertainty estimation (GLUE) technique, introduced by Beven and Binley (1992). This approach outputs probability distribution of model parameters conditioned on input data, and the uncertainties in model input are represented by uncertain parameters. Similar to GLUE, Hong et al. (2006) proposed a Monte Carlo based method to quantify uncertainty in hydrological simulations using satellite precipitation data, in which flow

simulation uncertainty is represented by ensemble simulation results. However, the previous studies have not explicitly considered the respective uncertainty contributions of precipitation products, hydrological models and their combinations at different times and for different data magnitudes.

Ensemble discharge simulation is a very important approach to representing the overall uncertainty and should be handled in an appropriate way to generate a deterministic output (Wang et al., 2012). Several approaches can be applied to this process, for example, weighted average approaches, maximum likelihood approaches and Bayesian averaging approaches. In these methods, the Bayesian averaging approaches can combine multiple models and/or inputs based on their probabilities and have been implemented in many studies. For example, Vrugt and Robinson (2007) compared data assimilation and Bayesian averaging approaches, and indicated that the data assimilation methods cannot handle the cases where multiple inputs and models are involved. Also, Duan et al. (2007) developed a probabilistic hydrologic prediction approach by combining a few hydrological models using the Bayesian averaging approaches. In recent years, Madadgar and Moradkhani (2014) developed an improved Bayesian averaging approach by incorporating joint probability functions and then used the method to combine simulation results from several models; Madadgar et al. (2016) developed an approach based on Bayesian averaging to predict seasonal precipitation on the basis of precipitation simulation ensemble; and Roy et al. (2017) proposed an approach for real-time streamflow forecasting using three hydrological models and bias corrected satellite data based on a Bayesian averaging approach. However, to the best of the authors' knowledge, no research has been carried out to generate deterministic outputs from multiple precipitation products and hydrological models explicitly considering their joint probability and dynamic posterior probability at different times and for different data magnitudes.

The overall objective of this paper, therefore, is to develop an ensemble-based dynamic Bayesian averaging approach (e-Bay) for deterministic discharge simulations using multiple global precipitation products and hydrological models. The developed e-Bay approach is generic and is independent of geographical area, precipitation products and hydrological models: hence, it is widely applicable. In this approach, multiple precipitation products and hydrological models are used to consider their respective uncertainties. The joint probability of precipitation products and hydrological models being correct is quantified based on a two-objective likelihood function: uncertainties in maximum and mean estimation; dynamic posterior probability is quantified as a function of the magnitude and timing of estimated discharges; the law of total probability is implemented to average the ensemble discharge simulations. Six global fine-resolution precipitation products and two hydrological models of different complexities are used in this research to illustrate the application of the proposed approach in a river basin in northeast China. e-Bay is compared with three other approaches: the weighted average method with equal weights, weighted average method with varying weights, and the maximum likelihood approach. This study provides three advances from the present state of knowledge. First, an ensemble-based dynamic Bayesian averaging approach is developed, and it has outstanding capability to estimate expected discharges. Second, an approach is developed to estimate the joint probability of precipitation products and hydrological models being correct based on uncertainties of maximum and mean estimates. Third, an approach is proposed to estimate posterior probability based on the magnitude and timing of simulated discharges. The developed e-Bay approach bridges the gap between global precipitation products and their pragmatic applications to river discharge simulations, and could be beneficial to hydrological simulations and water resources management in ungauged or poorly gauged regions across the world.

The paper is organized as follows. Section 2 introduces the e-Bay approach. Section 3 presents the study area, data sets, precipitation products, hydrological models and criteria used. Section 4 shows the results of applying e-Bay to the study area. Discussion and conclusions are presented in Section 5 and Section 6.

2 The e-Bay approach

There are four main steps in the proposed e-Bay approach. First，ensemble discharge simulations using selected global precipitation products and hydrological models. Second, quantification of joint probability of the precipitation products and hydrological models being correct, which includes calculation of respective uncertainties of precipitation products, hydrological models and their combinations. Third, quantification of dynamic posterior probability, which depends on the timing and magnitude of simulated discharges. Fourth, calculation of expected discharges from the discharge ensemble on the basis of the law of total probability. The implementation of the proposed approach in this study splits datasets used into two periods: training period and prediction period. The training period is used to quantify the probabilities, and the prediction period used to validate the proposed approach. In this study, a series of training periods ranging from two to eight years are used, which results in prediction periods ranging from six-year to one-year. The flowchart of e-Bay is shown in Fig. 1.

< Figure 1 here please >

2.1 Quantification of joint probabilities of precipitation products and hydrological models

Assuming hydrological model M_i and precipitation product R_j are used, the probability of them being correct $p\left(M_i, R_j\right)$ can be assessed by the performances of M_i, R_j and their combination in replicating observed discharge data $\mathbf{Q}_{\text{obs}}$ (Madadgar and Moradkhani, 2014):

$$p\left(M_i, R_j \mid \mathbf{Q}_{\text{obs}}\right) = p\left(M_i \mid \mathbf{Q}_{M_i}, \mathbf{Q}_{\text{obs}}\right) \cdot p\left(\mathbf{R}_j \mid \mathbf{Rain}_{R_j}, \mathbf{Rain}_{\text{obs}}\right) \cdot c\left(M_i, R_j \mid \mathbf{Q}_{M_i,R_j}, \mathbf{Q}_{\text{obs}}\right) \quad (1)$$

where *i* and *j* represent the *i*th hydrological model and the *j*th precipitation product respectively; $p\left(M_i \mid \mathbf{Q}_{M_i}, \mathbf{Q}_{\text{obs}}\right)$ and $p\left(R_j \mid \mathbf{Rain}_{R_j}, \mathbf{Rain}_{\text{obs}}\right)$ represent the probability of model M_i and precipitation R_j respectively being correct; $c\left(M_i, R_j \mid \mathbf{Q}_{M_i,R_j}, \mathbf{Q}_{\text{obs}}\right)$ represents the probability of the combination of M_i and R_j being correct; $\mathbf{Q}_{\text{obs}}$ represents observed discharges; $\mathbf{Rain}_{\text{obs}}$ represents observed precipitation depth; $\mathbf{Q}_{M_i}$ represents simulated discharges of M_i using observed data as input; $\mathbf{Rain}_{R_j}$ represents precipitation depth of R_j; $\mathbf{Q}_{M_i,R_j}$ represents simulated discharges of M_i using $\mathbf{Rain}_{R_j}$ as input.

$p\left(M_i \mid \mathbf{Q}_{M_i}, \mathbf{Q}_{\text{obs}}\right)$ can be calculated using Bayesian theory as below:

$$p\left(M_i \mid \mathbf{Q}_{M_i}, \mathbf{Q}_{\text{obs}}\right) = \frac{l\left(\mathbf{Q}_{M_i} \mid M_i, \mathbf{Q}_{\text{obs}}\right) \cdot \pi\left(M_i\right)}{\sum_{i=1}^{m} l\left(\mathbf{Q}_{M_i} \mid M_i, \mathbf{Q}_{\text{obs}}\right) \cdot \pi\left(M_i\right)} \quad (2)$$

where $\pi\left(M_i\right)$ represents prior probability of M_i being correct; $l\left(\mathbf{Q}_{M_i} \mid M_i, \mathbf{Q}_{\text{obs}}\right)$ represents likelihood function conditioning on the observed discharge data $\mathbf{Q}_{\text{obs}}$. In this

research, $l\left(\mathbf{Q}_{M_i} | M_i, \mathbf{Q}_{\text{obs}}\right)$ is calculated on the basis of two uncertainty criteria: uncertainty of maximum discharge (f_1) and uncertainty of mean discharge (f_2):

$$l\left(\mathbf{Q}_{M_i} | M_i, \mathbf{Q}_{\text{obs}}\right) = \frac{1}{2} \cdot \left(f_1 + f_2\right) \tag{3}$$

f_1 can be calculated as below

$$f_1 = 1 - \frac{\left|\max\left(\mathbf{Q}_{M_i}\right) - \max\left(\mathbf{Q}_{\text{obs}}\right)\right|}{\max\left[maxQ, \max\left(\mathbf{Q}_{\text{obs}}\right)\right]} \tag{4}$$

where *max* represents the maximum values; $maxQ$ represents the maximum discharge among *m* hydrological model simulations using observed precipitation as input

$$maxQ = \max\left[\max\left(\mathbf{Q}_{M1}\right), \max\left(\mathbf{Q}_{M2}\right), \cdots, \max\left(\mathbf{Q}_{Mm}\right)\right] \tag{5}$$

f_2 can be calculated as below

$$f_2 = 1 - \frac{\left|\text{mean}\left(\mathbf{Q}_{M_i}\right) - \text{mean}\left(\mathbf{Q}_{\text{obs}}\right)\right|}{\max\left[maxMean, \text{mean}\left(\mathbf{Q}_{\text{obs}}\right)\right]} \tag{6}$$

where *mean* represents the mean values; $maxMean$ represents the maximum mean discharge among *m* hydrological model simulations using observed precipitation as input:

$$maxMean = \max\left[mean\left(\mathbf{Q}_{M1}\right), mean\left(\mathbf{Q}_{M2}\right), \cdots, mean\left(\mathbf{Q}_{Mm}\right)\right] \tag{7}$$

It should be noted that the likelihood function is formed by aggression of the two criteria, and requires the two criteria have the same magnitude (Khu and Madsen, 2005; Ngo et al., 2007). Eqs. (4) and (6) show that f_1 and f_2 have values between zero and one, and thus the constructed

likelihood function is appropriate. The larger of the likelihood function, the higher of the probability of being correct.

For precipitation products, the probability of being correct can be calculated as

$$p\left(R_j \mid \mathbf{Rain}_{R_j}, \mathbf{Rain}_{\text{obs}}\right) = \frac{l\left(\mathbf{Rain}_{R_j} \mid R_j, \mathbf{Rain}_{\text{obs}}\right) \cdot \pi\left(R_j\right)}{\sum_{j=1}^{r} l\left(\mathbf{Rain}_{R_j} \mid R_j, \mathbf{Rain}_{\text{obs}}\right) \cdot \pi\left(R_j\right)} \tag{8}$$

where $\pi\left(R_j\right)$ represents prior probability of R_j being corrrect; $l\left(\mathbf{Rain}_{R_j} \mid R_j, \mathbf{Rain}_{\text{obs}}\right)$ represents likelihood function conditioning on observed precipitation depth $\mathbf{Rain}_{\text{obs}}$. Similar to likelihood function calculation of discharges, $l\left(\mathbf{Rain}_{R_j} \mid R_j, \mathbf{Rain}_{\text{obs}}\right)$ is also calculated on the basis of two uncertainty criteria: uncertainty of maximum precipitation depth (f_1) and uncertainty of mean precipitation depth (f_2):

$$l\left(\mathbf{Rain}_{R_j} \mid R_j, \mathbf{Rain}_{\text{obs}}\right) = \frac{1}{2} \cdot \left(f_1 + f_2\right) \tag{9}$$

f_1 is calculated as

$$f_1 = 1 - \frac{\left|\max\left(\mathbf{Rain}_{R_j}\right) - \max\left(\mathbf{Rain}_{\text{obs}}\right)\right|}{\max\left[maxRain, \max\left(\mathbf{Rain}_{\text{obs}}\right)\right]} \tag{10}$$

where *maxRain* represents the maximum precipitation depth among *r* precipitation products

$$maxRain = \max\left[\max\left(\mathbf{Rain}_{R_1}\right), \max\left(\mathbf{Rain}_{R_2}\right), \cdots, \max\left(\mathbf{Rain}_{R_r}\right)\right] \tag{11}$$

f_2 is calculated as

$$f_2 = 1 - \frac{\left|\text{mean}\left(\mathbf{Rain}_{R_j}\right) - \text{mean}\left(\mathbf{Rain}_{\text{obs}}\right)\right|}{\max\left(maxMean, \text{mean}\left(\mathbf{Rain}_{\text{obs}}\right)\right)} \tag{12}$$

where $maxMean$ represents the maximum mean precipitation depth among *r* precipitation products

$$maxMean = \max\left[\text{mean}\left(\mathbf{Rain}_{R_1}\right), \text{mean}\left(\mathbf{Rain}_{R_2}\right), \cdots, \text{mean}\left(\mathbf{Rain}_{R_r}\right)\right] \tag{13}$$

For the combinations of precipitation products and hydrological models, the probabilities of being correct are

$$c\left(M_i, R_j \mid \mathbf{Q}_{M_i,R_j}, \mathbf{Q}_{\text{obs}}\right) = \frac{l\left(\mathbf{Q}_{M_i,R_j} \mid M_i, R_j, \mathbf{Q}_{\text{obs}}\right) \cdot \pi\left(M_i, R_j\right)}{\sum_{i=1}^{m}\sum_{j=1}^{r} l\left(\mathbf{Q}_{M_i,R_j} \mid M_i, R_j, \mathbf{Q}_{\text{obs}}\right) \cdot \pi\left(M_i, R_j\right)} \tag{14}$$

where $\pi\left(M_i, R_j\right)$ represents prior probability of the combination of R_j and M_i being correct; $l\left(\mathbf{Q}_{M_i,R_j} \mid M_i, R_j, \mathbf{Q}_{\text{obs}}\right)$ represents likelihood function conditioning on observed discharge data $\mathbf{Q}_{\text{obs}}$. Similar to likelihood function calculations of hydrological models and precipitation products, $l\left(\mathbf{Q}_{M_i,R_j} \mid M_i, R_j, \mathbf{Q}_{\text{obs}}\right)$ is also calculated on the basis of two uncertainty criteria: uncertainty of maximum value of $\mathbf{Q}_{M_i,R_j}$ (f_1) and uncertainty of mean value of $\mathbf{Q}_{M_i,R_j}$ (f_2):

$$l\left(\mathbf{Q}_{M_i,R_j} \mid M_i, R_j, \mathbf{Q}_{\text{obs}}\right) = \frac{1}{2} \cdot \left(f_1 + f_2\right) \tag{15}$$

f_1 is calculated as

$$f_1 = 1 - \frac{\left|\max\left(\mathbf{Q}_{M_i,R_j}\right) - \max\left(\mathbf{Q}_{\text{obs}}\right)\right|}{\max\left(maxQMR, \max\left(\mathbf{Q}_{\text{obs}}\right)\right)} \tag{16}$$

where $maxQMR$ represents the maximum discharge among the $m \times r$ combinations of hydrological models and precipitation products.

$$maxQMR = \max\left\lfloor \max\left(\mathbf{Q}_{M_1,R_1}\right), \max\left(\mathbf{Q}_{M_1,R_2}\right), \cdots, \max\left(\mathbf{Q}_{M_m,R_r}\right)\right\rfloor \quad (17)$$

f_2 is calculated as

$$f_2 = 1 - \frac{\left|\text{mean}\left(\mathbf{Q}_{M_i,R_j}\right) - \text{mean}\left(\mathbf{Q}_{\text{obs}}\right)\right|}{\max\left(maxmeanQMR, \text{mean}\left(\mathbf{Q}_{\text{obs}}\right)\right)} \quad (18)$$

where $maxmeanQMR$ represents the maximum of the mean discharges among the $m \times r$ combinations of hydrological models and precipitation products

$$maxmeanQMR = \max\left\lfloor \text{mean}\left(\mathbf{Q}_{M_1,R_1}\right), \text{mean}\left(\mathbf{Q}_{M_1,R_2}\right), \cdots, \text{mean}\left(\mathbf{Q}_{M_m,R_r}\right)\right\rfloor \quad (19)$$

2.2 Quantification of posterior probabilities

The probability of simulated discharge q being correct at time *k* can be quantified by performance of the combination of M_i and R_j in replicating observed data q_{obs}^k

$$p\left(q^k \mid M_i, R_j, q_{\text{obs}}^k\right) = p\left(q_{M_i,R_j}^k \middle| q_{\text{obs}}^k\right) = \frac{l\left(q_{M_i,R_j}^k, q_{\text{obs}}^k\right) \cdot \pi\left(q_{M_i,R_j}^k\right)}{\sum_{i=1}^{m}\sum_{j=1}^{r} l\left(q_{M_i,R_j}^k, q_{\text{obs}}^k\right) \cdot \pi\left(q_{M_i,R_j}^k\right)} \quad (20)$$

$p\left(q^k \mid M_i, R_j, q_{\text{obs}}^k\right)$ represents the posterior probability of q^k being correct; $\pi\left(q_{M_i,R_j}^k\right)$ represents the prior probability of q^k being correct; $l\left(q_{M_i,R_j}^k, q_{\text{obs}}^k\right)$ represents a likelihood function conditioning on observed discharge q_{obs}^k as below

$$\begin{cases} l\left(q^k_{M_i,R_j} \mid q^k_{\text{obs}}\right) = \dfrac{1}{\left(\left|q^k_{M_i,R_j} - q^k_{\text{obs}}\right|\right)^n} \\ if \left|q^k_{M_i,R_j} - q^k_{\text{obs}}\right| = 0,\ then\ l\left(q^k_{M_i,R_j} \mid q^k_{\text{obs}}\right) = Inf \end{cases} \tag{21}$$

where n represents a parameter to scale up or down the uncertainty and should be calibrated in the training period; Inf represents a punishment value and should be a large real number. From (21), it can be seen that the closer of simulated discharge $q^k_{M_i,R_j}$ to observed data q^k_{obs}, the larger the likelihood function values, implying higher probability of being correct.

After quantifying the posterior probability of being correct in the training period, the probability of predicted discharges being correct in the prediction period can be calculated using an interpolation approach. In this interpolation approach, it is assumed that the simulated discharges in the prediction periods have the same probability distribution as in the training periods, and the probability of predicted discharges being correct can be linearly interpolated as below

$$\begin{cases} p\left(q^k_{M_i,R_j}\right) = p\left(q^{k,\wedge}_{M_i,R_j}\right) + \dfrac{q^k_{M_i,R_j} - q^{k,\wedge}_{M_i,R_j}}{q^{k,\vee}_{M_i,R_j} - q^{k,\wedge}_{M_i,R_j}} \cdot \left\{p\left(q^{k,\vee}_{M_i,R_j}\right) - p\left(q^{k,\wedge}_{M_i,R_j}\right)\right\} \\ q^{k,\wedge}_{M_i,R_j} \le q^k_{M_i,R_j} < q^{k,\vee}_{M_i,R_j} \end{cases} \tag{22}$$

where $q^{k,\wedge}_{M_i,R_j}$ and $q^{k,\vee}_{M_i,R_j}$ represents simulated discharges in the training period at time *k* by the combination of M_i and R_j with a relationship of $q^{k,\wedge}_{M_i,R_j} < q^{k,\vee}_{M_i,R_j}$; $p\left(q^{k,\vee}_{M_i,R_j}\right)$ and $p\left(q^{k,\wedge}_{M_i,R_j}\right)$ represents probabilities calculated in the training period; $q^k_{M_i,R_j}$ and

$p\left(q^k_{M_i,R_j}\right)$ represents predicted discharges and their probability of being correct in the prediction period.

When predicted discharges are smaller than the smallest discharges in the training period, the probability of predicted discharges can be interpolated using Eq. (23), assuming that the probability of being correct is zero when simulated discharge values are zero

$$p\left(q^k_{M_i,R_j}\right)=p\left(\underline{q^k_{M_i,R_j}}\right)-\frac{\underline{q^k_{M_i,R_j}}-q^k_{M_i,R_j}}{\underline{q_{q^k_{M_i,R_j}}}-0}\cdot\left\{p\left(\underline{q^k_{M_i,R_j}}\right)-p(0)\right\} \tag{23}$$

where $\underline{q^k_{M_i,R_j}}$ and $p\left(\underline{q^k_{M_i,R_j}}\right)$ represent the smallest discharges and its probability of being correct in the training period at time *k*; $p(0)=0$ represents the probability of being correct when simulated discharge values equal zero.

When predicted discharges are larger than the largest discharges in the training period, the probability of the predicted discharges being correct can be derived from interpolation using Eq. (24), assuming that the probability is a linear function

$$p\left(q^k_{M_i,R_j}\right)=a\cdot q^k_{M_i,R_j}+b \tag{24}$$

where a and b are coefficients of a linear function representing a line of the best fit of the largest and second largest discharges and their corresponding probabilities in the training period.

2.3 Expected discharges based on the law of total probability

On the basis of the law of total probability, given a set of observations $\mathbf{Q}_{\text{obs}}$, the probability of simulated discharge *q* being correct can be calculated as Eq. (25)

$$p\left(q \mid \mathbf{M}, \mathbf{R}, \mathbf{Q}_{\text{obs}}\right) = \sum_{i=1}^{m} \sum_{j=1}^{r} p\left(M_i, R_j \mid \mathbf{Q}_{\text{obs}}\right) \cdot p\left(q \mid M_i, R_j, \mathbf{Q}_{\text{obs}}\right) \tag{25}$$

where p represents probability of q being correct; **M** and **R** represents a set of hydrological models and precipitation products respectively; $p\left(q \mid M_i, R_j, \mathbf{Q}_{\text{obs}}\right)$ represents probability of q simulated by M_i and R_j being correct; $p\left(M_i, R_j \mid \mathbf{Q}_{\text{obs}}\right)$ can be rewritten as below

$$w_{ij} = p\left(M_i, R_j \mid \mathbf{Q}_{\text{obs}}\right) = p\left(M_i\right) \cdot p\left(R_j\right) \cdot c\left(M_i, R_j\right) \tag{26}$$

and represents overall performance of M_i and R_j in replicating the observed discharge $\mathbf{Q}_{\text{obs}}$ in the training period. The sum of w_{ij} is one (Madadgar and Moradkhani, 2014). w_{ij} can be calculated using Eq. (1) and will be called joint probability in the rest of this paper.

When considering time variant discharge simulations, Eq. (25) can be rewritten as below (Madadgar and Moradkhani, 2014)

$$p\left(q^k \mid \mathbf{M}, \mathbf{R}, \mathbf{Q}_{\text{obs}}\right) = \sum_{i=1}^{m} \sum_{j=1}^{r} w_{ij} \cdot p\left(q^k \mid M_i, R_j, q_{\text{obs}}^k\right) \tag{27}$$

where k represents the timing of simulated discharge q; $p\left(q^k \mid M_i, R_j, q_{\text{obs}}^k\right)$ can be calculated using Eqs. (20)-(24). Thus, expected discharges on the basis of the law of total probability can be calculated as below

$$q^k = \sum_{i=1}^{m} \sum_{j=1}^{r} q_{M_i, R_j}^k \cdot w_{ij} \cdot p\left(q^k \mid M_i, R_j, q_{\text{obs}}^k\right) \tag{28}$$

2.4 Steps to implement the e-Bay method

This section is devoted to describing how the newly proposed e-Bay approach can be applied in practice. The numerical procedures are implemented according to the following main steps:

1. Collect available observation data, including precipitation, discharge and other inputs of

hydrological models.

2. Quantify the joint probabilities of precipitation products and hydrological models being correct and the dynamic posterior probability, including calibration of parameter *n* in Eq. (21).
3. Validate the performance of e-Bay using observed discharges and go to step 2 if the performance is not desirable.
4. Predict future discharges using e-Bay, in which the global precipitation products should be used as input to hydrological models.

3 Real world study area, data sets, precipitation products, models and criteria

3.1 The Biliu river basin

The Biliu basin (2814 km^2) is located in the coastal region between the China Bohai Sea and the China Huanghai Sea, and is characterized by a dry winter - hot summer climate (Koppen climate classification). There are 11 precipitation stations and one discharge gauge which have historical data from January 2000 to December 2007. The Biliu reservoir is a small reservoir, with a maximum capacity of 934 million cubic meter, and is managed to supply water to nearby cities. The flow data used in this research is the inflow to the Biliu reservoir, which is not influenced by reservoir management. The gauge distribution in the Biliu basin is shown in Fig. 2. The elevation of Biliu river basin ranges from 4.5 m to 984.9 m. Land-use data are obtained from the USGS (http://edc2.usgs.gov/glcc/glcc.php). The land-use types have been reclassified to Simple Biosphere scheme – version 2 (SiB2) land-use types for this study (Sellers et al., 1996). Soil data are obtained from the Food and Agriculture Organization (FAO) (2003) Global data product.

< Figure 2 here please >

3.2 Data sets

The weather stations provide sunshine duration, air temperature, humidity and wind speed data, which can be obtained from the China Meteorological Data Sharing Service System (downloaded from http://cdc.cma.gov.cn/home.do). Precipitation data were obtained from the 11 rain gauges. Downward solar radiation is estimated from sunshine duration, temperature and humidity using a hybrid model (Yang et al., 2006). Long wave radiation is obtained from the GLDAS/Noah (Rodell et al., 2004). Air pressure is estimated according to elevation (Yang et al., 2006; McVicar and Körner, 2013). These meteorological data are then interpolated to 300 m × 300 m model cells through an inverse-distance weighting approach. Because of the elevation differences among model cells and meteorological gauges, the interpolated surface air temperatures are further modified with a lapse rate of 6.5K/km. Gauge precipitation data are also interpolated to 300 m × 300 m model cells and basin-averaged gauge precipitation data are calculated on the basis of the interpolation results. In addition to the above, the leaf area index and fraction of photosynthetically active radiation data are obtained from level-4 MODIS global products-MOD15A2. The Digital Elevation Model (DEM) is from the NASA SRTM (Shuttle Radar Topographic Mission) and has a resolution of 30 m × 30 m. This was resampled to a 300 m resolution in the model calculation to reduce computational cost, while the model processed the finer DEM (30 m grid) to generate sub-grid parameters (such as hillslope angle and length). These data were used as inputs for the two hydrological models to calibrate model parameters.

3.3 Precipitation products

The precipitation products used in this study are shown in Table 1. These data are all freely available. In these selected precipitation products, APHRODITE is wholly based on gauge data; TRMM3B42 is based on both remote satellite estimation and gauge data; GLDAS-1

(GLDAS_Noah025SUBP_3H) is generated based on Global Data Assimilation System data (which is a reanalysis data) and NOAA Climate Prediction Center Merged Analysis of Precipitation (which is based on gauge, reanalysis and satellite-derived precipitation estimates); while others are based on remote satellite estimation without gauge data corrections. Remote-based precipitation estimation has many weaknesses, e.g. microwave estimation could miss convective precipitation and typhoon rain because of its sparse time interval resolution, and infrared estimation has a higher temporal resolution but cannot penetrate clouds. Ground rain gauge-based interpolation products are limited by interpolation algorithms, gauge density and gauge data quality (Xie et al., 2007). The details of data sources used in each precipitation product can be found in Table 1.

< Table 1 here please >

3.4 Hydrological models

Two hydrological models with different complexities – WEB-DHM and TOPMODEL – are used in this study, since previous research showed that they can simulate the hydrological processes very well in the case study region used in this paper (Qi et al., 2015; Qi et al., 2016a; Qi et al., 2016b; Qi et al., 2016c).

3.4.1 WEB-DHM

The distributed biosphere hydrological model – Water and Energy Budget-based Distributed Hydrological Model (WEB-DHM) (Wang et al., 2009a; Wang et al., 2009b; Wang et al., 2009c) – was developed by coupling SiB2 (Sellers et al., 1986) with a geomorphology-based hydrological model (Yang, 1998) to describe water, energy and CO_2 fluxes at a basin scale. WEB-DHM has been used in several evaluations and applications (Wang et al., 2010a; Wang et

al., 2010b; Qi et al., 2015; Qi et al., 2016a). Input data of WEB-DHM include precipitation, temperature, downward solar radiation, long wave radiation, air pressure, wind speed and humidity.

3.4.2 TOPMODEL

TOPMODEL is a physically-based, variable contributing area model of basin hydrology which attempts to combine the advantages of a simple lumped parameter model with distributed effects (Beven and Kirkby, 1979). More detailed descriptions of TOPMODEL and its mathematical formulation can be found in Beven et al. (1979). TOPMODEL has been commonly utilized in research across the world (Blazkova and Beven, 1997; Cameron et al., 1999; Hossain and Anagnostou, 2005; Bastola et al., 2008; Gallart et al., 2008; Bouilloud et al., 2010; Qi et al., 2016b; Qi et al., 2016c) because of its relatively simple model structure. The input data of TOPMODEL include basin averaged precipitation and topographic data which can be estimated from digital elevation models (DEM).

3.5 Criteria

In model performance assessment, NSE and relative bias (RB) are used as error metrics:

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{n}\left(Q_{pi} - Q_{ti}\right)^2}{\sum_{i=1}^{n}\left(Q_{ti} - \overline{Q_t}\right)^2} \tag{29}$$

$$\mathrm{RB} = \frac{\sum_{i=1}^{n} Q_{pi} - \sum_{i=1}^{n} Q_{ti}}{\sum_{i=1}^{n} Q_{ti}} \times 100\% \tag{30}$$

where Q_{pi} and Q_{ti} are a simulated and an observed flow respectively at time i; $\overline{Q_t}$ is the average of observed flows. In addition, a comprehensive criterion F is also used

$$\mathrm{F} = 1 - \mathrm{NSE} + \left|\mathrm{RB}\right| \tag{31}$$

The lower the F values, the better the results.

4 Application to the Biliu river basin

This section intends to present the detailed results of the application of e-Bay to the Biliu river basin, in which the advantages of e-Bay are shown.

4.1 Hydrological model calibration, validation and simulated discharge ensemble

For WEB-DHM, six main parameters were selected for calibration using a trial and error approach due to the model's computational burden (it takes two hours to simulate one year of discharge in a personal computer with two 3.41GHz processors). Model parameter multipliers were calibrated, similar to the studies by Wang et al. (2011), Qi et al. (2015) and Qi et al. (2016a). The 'trial and error' approach has two steps. First, all the multiplier values are set to 1, which represents the default parameter values from Food and Agriculture Organization (FAO) (2003) and SiB2 model. Second, multiplier values are varied until acceptable discharge simulation accuracy is obtained. The calibrated parameter values are listed in Table 2. The simulated daily, monthly and inter-annual results are shown in Figs. 3a, 3c and 3e.

< Table 2 here please >

< Figure 3 here please >

TOPMODEL uses basin-averaged parameter values, and these parameter values are estimated by experience or observation. However, these methods do not give precise parameter values. Therefore, the parameter values are considered as uncertain and provided with ranges based on previous studies in this region (Qi et al., 2016a; Qi et al., 2016b; Qi et al., 2016c). Six parameters of TOPMODEL were calibrated using the dynamically dimensioned search algorithm (Tolson and

Shoemaker, 2007), and the results are given in Table 3. The simulated daily, monthly and inter-annual results are shown in Figs. 3b, 3d and 3f.

< Table 3 here please >

Note that the parameters of TOPMODEL and WEB-DHM were calibrated using observed precipitation data, and the accuracy of simulated discharges was validated using gauge observations. These parameter values were employed when using the global precipitation products as model inputs. The discharge simulation ensemble using WEB-DHM, TOPMODEL and the six global precipitation products is shown in Fig. 4.

< Figure 4 here please >

None of the simulated discharges accurately replicate the variations in observed discharge. For example, the combination of WEB-DHM and PERSIANN seriously overestimates discharges across the whole year, and all the other hydrological model and precipitation product combinations underestimate the discharge in August. This information implies that it is difficult to accurately simulate discharge processes using any of the combinations, and therefore it is necessary to generate a synthetic discharge process from the ensemble considering all the uncertainties.

4.2 Probability quantification

Fig. 5 shows the joint probability density in the training period, based on Eq. (1). These results are calculated using data duration ranging from two years to eight years. Each hydrological model and precipitation product combination is represented by a different color.

< Figure 5 here please >

In the case of 2-year training data, the joint probability of PERSIANN and TOPMODEL is the largest, with a value of 0.13; TRMM3B42 and WEB-DHM have the second largest joint probability (0.12); GLDAS and TOPMODEL have the lowest probability, with a value of just 0.05. Results are different in the case of 3-year training data: for example, the joint probability of TRMM3B42 and WEB-DHM is the largest, with a value of 0.13; TRMM3B42 and TOPMODEL have the second largest value (0.12). These differences may result from the changes in the data implemented, and implies that the training data influences the quantification of joint probability. Similar to 3-year training data case, in the cases of 4-year and 5-year training data the joint probability of TRMM3B42 and WEB-DHM is the largest; TRMM3B42 and TOPMODEL have the second largest value; GLDAS and TOPMODEL have the lowest value. However, in the cases of 6-year, 7-year and 8-year training data, the results are a little different: for example, TRMM3B42RT and TOPMODEL have the lowest joint probability; TRMM3B42 and WEB-DHM have almost the same joint probability as PERSIANN and TOPMODEL. These differences are also due to the increase in training data, which implies that the calculation of the joint probability should use as much available data as possible.

Fig. 6 shows the dynamic posterior probability of simulated discharges across twelve months. The duration of training data ranges from two to eight years. In this study, the scale factor n and Inf in Eq. (21) equal 4 and 1000 respectively; these values were obtained through a trial and error approach and found to provide the best replication of discharges observed during the training period. The results in the following sections show this scale factor value is appropriate, replicating the observed data very well.

< Figure 6 here please >

In the case of 2-year training data in January (Fig. 6(1)), the CDF values vary as discharge varies. Compared with Fig. 6(1), Figs. 6(2)-(7) show different CDF curves. These differences result from the increase in training data. Overall, these results reveal that the probability of simulated discharges being correct is dynamic: the probability depends on the magnitude of discharge. Compared with the results for January (Figs. 6(1)-6(7)), the panels for February show different patterns in the variation of CDFs (Figs. 6(8)-6(14)). For example, Fig. 6(14) shows that the CDF value is 0.61 when discharges are smaller than 2.70 m^3/s, whereas in Fig. 6(7) the CDF value is only 0.29. These differences between January and February indicate that the probability of simulated discharges being correct also varies with times. Similar to the cases in January and February, other months also show the dynamic posterior probability of simulated discharges.

4.3 Expected discharges in the training period

The expected discharges in the training period can be obtained after calculation of the joint probabilities and the dynamic posterior probabilities. Fig. 7 shows the comparison of simulated discharges calculated from four approaches. 'Mean' represents the mean values of the discharge simulation ensemble; 'Maximum NSE' represents the result of a combination of precipitation product and hydrological model with maximum NSE. 'Weighted average' represents the results that are averaged on the basis of the performance of hydrological model and precipitation product combinations, with the weight calculated from Eq. (1).

< Figure 7 here please >

In the case of 2-year training data, with respect to e-Bay, the calculated discharge values are very similar to the observations (Fig. 7(1)), while the results of other approaches are a little worse. For example, Fig. 7(2) shows the calculated results have large uncertainty in the peak discharge; Figs. 7(3) - (4) show the results have larger uncertainty in smaller discharges. These results are quantitatively verified in Table 4 in terms of three criteria: NSE, RB and F. The best criterion values are in bold fonts. In the case of 2-year training data, e-Bay has the best performance in terms of all the three criteria: NSE, RB and F values are 0.96, -0.14 and 0.18 respectively.

Compared to the 2-year training case, the results are different with 3 years of training data. For example, the 'Maximum NSE' approach shows better performance (Figs. 7(7)) than e-Bay (Fig. 7(5)). The differences are also shown in Table 4, from which it can be seen that, although e-Bay has a NSE value of up to 0.96, which is the same as the 'Maximum NSE' approach, its RB value shows larger error, resulting in larger F value. The differences between 2- and 3-year training data result from the changes in the training data and the joint probability (recall the results in section 4.2).

To consider the impacts of increasing training data, the trends of the performance criteria are analyzed. Fig. 8 shows the trends of the three performance criteria, which are represented in different colors. As the number of years of training data increases, the performance of e-Bay continually improves. For example, when the training data increase from 2 to 8 years, the NSE values increase from 0.96 to 0.99; absolute RB values decrease from 0.14 to 0.05; F values decrease from 0.18 to 0.06. It should be noted that, although the absolute RB and F values of the 6-year training data are 0.01 larger than the ones of the 5-year training data, this difference is very small. Compared to e-Bay, the other three approaches show large variations in NSE, RB

and F, demonstrating that they cannot guarantee better results when new data are added to the calculation. This implies that they would devalue the efforts to improve discharge simulation accuracy by collecting new data. In contrast, with an increase in the training data, the criteria values of e-Bay show smaller fluctuations and the performance of e-Bay becomes outstanding. From Table 4, it can also be seen that the average NSE and F values of e-Bay over all the training years are up to 0.97 and 0.12 respectively, whereas their best values among the other three approaches are only 0.91 and 0.18. Thus, the overall performance of e-Bay is better.

< Figure 8 here please >

< Table 4 here please >

4.4 Expected discharges in the validation period

Fig. 9 shows the comparison of simulated discharge values among the four approaches in the validation period.

< Figure 9 here please >

In the case of 6-year validation data, e-Bay seriously overestimates discharges in June and July, but simulates the peak value (in August) very well. In contrast, the other three approaches seriously underestimate peak discharges but their simulated values in June and July are very close to observations. The quantitative comparison is shown in Table 5, from which it can be seen that the other three approaches outperform e-Bay in terms of the three criteria. This result differs from section 4.3, in which e-Bay is better than the other three approaches according to all the three criteria when 2 years of training data is used. This difference results from the influence of

the data in the training period: the calculated probability is biased by the small duration of sample data in the training period.

< Table 5 here please >

According to the results in section 4.2, the performance of e-Bay becomes better as the training data duration increases. Therefore, the performance of e-Bay in the validation period should be better than other approaches when the duration of training data is large. This hypothesis is verified according to the quantitative criteria in Table 5. e-Bay outperforms the other three approaches when the duration of training data is large: in terms of NSE, e-Bay is the best when 5 or fewer years are used for validation (corresponding to training data duration of at least 3 years); in terms of the overall performance criterion F, e-Bay has better performance when 1, 2 or 3 years are used for validation (corresponding to 5, 6 and 7-year training data). Especially when the number of the validation years equals 1, e-Bay successfully replicates the observations, with NSE and RB values of 0.95 and 0.01, but the other three approaches have very poor performance. For example, the 'Mean' and 'Weighted average' approaches have NSE values of 0.65 and 0.68, respectively; 'Maximum NSE' has NSE and RB values of -0.02 and -0.42. In addition, from Table 5, it can be seen that the average NSE and F values of e-Bay over all the different years of validations are up to 0.85 and 0.32, whereas their best values among the other three approaches are only 0.72 and 0.37. Thus, e-Bay is better than the other three approaches when an extended period of data is used for training.

It should be noted that the increase of the performance of e-Bay is not continuous, for example, the NSE, RB and F values are 0.87, 0.08 and 0.21 in 5-year validation, but they are 0.85, -0.34 and 0.49 in 4-year validation. This could result from the influence of the changes of validation

data. Thus, the comparisons of the four approaches are carried out in each of the validation cases.

5 Discussion

Compared with previous Bayesian model averaging approaches (e.g., Box and Tiao (1992); Duan et al. (2007); Roy et al. (2017)), there are two main differences in e-Bay: (i) e-Bay explicitly quantifies the probabilities of being correct following the combinations of precipitation products and hydrological models to replicate observed flows; (ii) e-Bay explicitly quantifies the posterior probabilities at different times and under different flow magnitudes. A function describing the combination behavior of uncertainty sources at different times and flow magnitudes is effective in removing the output bias (Madadgar and Moradkhani, 2014). The traditional approaches ('Mean', 'Maximum NSE' and 'Weighted average') do not explicitly consider the combination effects. For example, the 'Mean' approach assumes every ensemble member has the same probability of being correct; the 'Maximum NSE' approach assumes that the probability of the ensemble member with a highest NSE value being correct is one; although 'Weighted average' considers the varying weights, these weights represents the overall joint performance of hydrological models and precipitation products, and therefore the probability of the combinations being correct at different times and discharge magnitudes is not explicitly considered. However, in this research, the uncertainties at different times and discharge magnitudes are explicitly quantified. This explicit quantification makes the e-Bay approach capable of coping with detailed uncertainties, which is beyond the capability of other traditional approaches, and therefore e-Bay generates better discharge simulation results.

In standard Bayesian averaging approaches, error distributions are assumed to be normal or gamma distributions, which are appropriate for several variables such as temperatures and sea

level pressure (Duan et al., 2007; Refsgaard et al., 2012; Madadgar and Moradkhani, 2014; Roy et al., 2017). In addition, non-parametric approaches, for example, the ensemble based approaches, can also be utilized to simulate complex error distributions (Hesterberg, 2002; Duan et al., 2007). In this study, the latter approach is utilized since the likelihood functions used (Eqs. (3), (9), (15) and (21)) consider uncertainties in maximum and mean estimation and also change with timing and flow magnitudes, and it is difficult to derive an analytical error distribution.

The scale factor in Eq. (21) has influence on the performance of e-Bay and should be calculated to maximize the probability of simulated discharges being correct. Based on the results in section 4, it can be seen that the scale factor used in this study is appropriate. Although a trial and error approach was used in this research to obtain the scale factor value, other approaches such as optimization algorithms and maximum likelihood estimate approaches may also be applied. In addition, the linear assumptions used to predict the probability of simulated discharges being correct in the future (Eqs. (23) and (24)) may have impacts on the performance of e-Bay. However, the performances of e-Bay in both calibration and validation periods in this study are very good, and are better than the other approaches if given enough training data. Thus, the linear assumptions are acceptable.

The overall objective of this study is to introduce the developed e-Bay approach, including its theoretical foundations and formulations, which are independent of the study area, precipitation products and hydrological models. Therefore, e-Bay is not restricted to the six products and two models used in this study and more precipitation products and hydrological models could be used, including, for example, a recent new precipitation product - MSWEP (Beck et al., 2017a). e-Bay provides an approach for deterministic discharge simulations using multiple global precipitation products since all the precipitation products (including MSWEP) have uncertainties

to some degree, as highlighted recently by Beck et al. (2017b). The application of e-Bay provides very good discharge estimation, indicating that the use of six products and two models in this research is acceptable. When using other precipitation products and hydrological models in other regions, the uncertainties of global precipitation products and hydrological models may vary, and the probabilities and parameters quantified in this study will change. However, this does not affect the applicability of e-Bay to other river basins.

In addition to precipitation, hydrological models may also need other input data, for example, wind speed, temperature and net radiation, and the used temperature lapse rates may vary in different months (McVicar et al., 2007). In this study, uncertainties in these data are not included since the overall objective of this study is to quantify probability of precipitation product uncertainties. The hydrological model parameters are calibrated using in situ gauge observation data in this study. While in situ data do not always yield the 'best' parameters, in our case, validation using observed flows shows that the performance of hydrological models is acceptable, with NSE and absolute values of RB being above 0.90 and below 1% respectively (recall results in Section 4.1). Therefore, the hydrological model parameter uncertainty was not considered in this study. The duration of precipitation and discharge data used can influence the performance of e-Bay: the longer the training data series, the better the performance of e-Bay. In this study, the WEB-DHM model needs to use a few satellite-based data sets as inputs, for example, the LAI and FPAR data from MODIS global products-MOD15A2, which can provide data only after 18 February 2000. In addition, the in situ gauge observations and APHRODITE precipitation product are only available until 31 December 2007. Thus, an eight-year data period is used in this study. The results obtained (recall section 4) show that this is sufficient to demonstrate the advantages of e-Bay and, therefore, the use of an eight-year data period is considered acceptable.

The e-Bay approach is developed on an inter-annual scale in this study, which represents a state-of-the-art process in ensemble-based Bayesian averaging estimation for discharges using multiple global precipitation products and hydrological models. Developing such approaches for daily and monthly scales are more challenging works, and future research is encouraged. Although the e-Bay approach is developed and evaluated on an inter-annual scale in this study, it shows better results than the other three approaches.

The spatial variations in precipitation are not considered in this study. Sapriza-Azuri et al. (2015) concluded that the spatial variability of precipitation has little influence on rapidly responding river discharges. This is the case in this study area because the flow transport time from the uppermost part of the Biliu basin to the downstream discharge gauge is approximately 6 hours, which is shorter than the monthly time steps of discharges investigated. Therefore, e-Bay is suitable for rapidly responding river basins (i.e. those in which the runoff concentration time is less than the discharge time step investigated, as in this case study). To extend e-Bay to a global scale when using global land surface/hydrological models, two approaches could be utilized: (i) divide river basins into sub-basins, apply e-Bay in each sub-basin and then route simulated flow to basin outlets; (ii) apply e-Bay in each model grid and then route simulated flow to basin outlets.

The e-Bay approach provides a useful tool to average ensemble discharge simulations using global precipitation products and hydrological models. Future research is encouraged to apply the approach in river basins world-wide using more land surface/hydrological models and precipitation products, and to compare e-Bay with traditional approaches comprehensively. e-Bay quantifies the probability of being correct based on combinations of precipitation products and hydrological models at different times and flow magnitudes, which is more sophisticated than the

traditional approaches. Therefore, the quantified probability of being correct using e-Bay may not be zero even when the probability values are zero based on traditional Bayesian approaches, which may extend knowledge beyond what is known from traditional Bayesian approaches. In addition, e-Bay could be applied to generate deterministic results from ensembles of temperatures, humidity, radiation, etc. Because e-Bay can cope with complex uncertainty, and the generated deterministic results could be compared with results using traditional Bayesian averaging approaches.

6 Conclusion

In discharge simulations, it is a challenge to consider the uncertainties of global precipitation products, hydrological models and their combinations individually, especially at different times and for different discharge magnitudes. This study presents a state-of-the-art methodology to address this issue. Six global fine-resolution precipitation products and two hydrological models of different complexities are used to illustrate the methodology in a river basin in northeast China. The major contributions are as follows.

First, an ensemble-based dynamic Bayesian averaging approach (e-Bay) is developed. e-Bay utilizes the law of total probability to calculate expected discharges using multiple global precipitation products and hydrological models, based on the total probabilities: precipitation product probability, hydrological model probability, their combination probability and the dynamic posterior probability.

Second, an approach is developed to estimate the joint probability of precipitation products and hydrological models based on uncertainties in maximum and mean estimation, and a method is

proposed to estimate dynamic posterior probability based on the magnitude and timing of simulated discharges.

Third, e-Bay has outstanding capability to estimate expected discharges. The Nash-Sutcliffe Efficiency values of e-Bay simulations are up to 0.97 and 0.85 in training and validation periods respectively. With an increase in the training data, the advantages of e-Bay are unveiled, and its performance becomes outstanding.

Acknowledgements:

This study was supported by the National Science Fund for Distinguished Youth Scholars (41625001), the National Natural Science Foundation of China (41571022), and Beijing Natural Science Foundation Grant (8151002). Additional support was provided by the Southern University of Science and Technology (Grant no. G01296001). The authors are deeply indebted to the editors and two anonymous reviewers for their valuable time and constructive suggestions that greatly improved the quality of this paper. The APHRODITE data was downloaded from http://www.chikyu.ac.jp/precip/pro-ducts/index.html. The TRMM3B42 data was downloaded from http://mirador.gsfc.nasa.gov/cgi-bin/mirador/presentNavigation.pl?tree=project&project=TRMM&dataGroup=Gridded. PERSIANN data was downloaded from http://chrs.web.uci.edu/persiann/data.html. GSMAP−MVK+ data was downloaded from http://sharaku.eorc.jaxa.jp/GSMaP_crest/. The GLDAS data was downloaded from http://mirador.gsfc.nasa.gov/cgi-bin/mirador/homepageAlt.pl?keyword=GLDAS_NOAH025SUBP_3H. The data of Biliu basin was obtained from the Biliu reservoir administration.

References:

Aonashi, K. et al., 2009. GSMaP Passive Microwave Precipitation Retrieval Algorithm: Algorithm Description and Validation. Journal of the Meteorological Society of Japan, 87A: 119-136.

Bastola, S., Ishidaira, H., Takeuchi, K., 2008. Regionalisation of hydrological model parameters under parameter uncertainty: A case study involving TOPMODEL and basins across the globe. Journal of Hydrology, 357(3-4): 188-206.

Beck, H.E., van Dijk, A.I.J.M., de Roo, A., Miralles, D.G., McVicar, T.R., Schellekens, J., Bruijnzeel, A.L., 2016. Global- scale regionalization of hydrologic model parameters. Water Resources Research, 52(5): 3599-3622.

Beck, H.E., van Dijk, A.I.J.M., Levizzani, V., Schellekens, J., Miralles, D.G., Martens, B., de Roo, A., 2017a. MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. Hydrology and Earth System Sciences, 21(1): 589-615.

Beck, H.E. et al., 2017b. Global-scale evaluation of 23 precipitation datasets using gauge observations and hydrological modeling. Hydrology and Earth System Sciences Discussions: 1-23.

Beven, K.J., Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology. Hydrological Sciences Bulletin, 24(1): 43-69.

Beven, K.J., Binley, A., 1992. The future of distributed models: Model calibration and uncertainty prediction. Hydrological Processes, 6(3): 279-298.

Beven, K.J., Freer, J.E., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. Journal of Hydrology, 249(1-4): 11–29.

Beven, K.J., Smith, P.J., Freer, J.E., 2007. Comment on "Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology" by Pietro Mantovan and Ezio Todini. Journal of Hydrology, 338(3-4): 315-318.

Blasone, R.-S., Vrugt, J.A., Madsen, H., Rosbjerg, D., Robinson, B.A., Zyvoloski, G.A., 2008. Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov Chain Monte Carlo sampling. Advances in Water Resources, 31(4): 630-648.

Blazkova, S., Beven, K., 1997. Flood frequency prediction for data limited catchments in the Czech Republic using a stochastic rainfall model and TOPMODEL. Journal of Hydrology, 195(1-4): 256-278.

Bouilloud, L., Chancibault, K., Vincendon, B., Ducrocq, V., Habets, F., Saulnier, G.-M., Anquetin, S., Martin, E., Noilhan, J., 2010. Coupling the ISBA Land Surface Model and the TOPMODEL Hydrological Model for Mediterranean Flash-Flood Forecasting: Description, Calibration, and Validation. Journal of Hydrometeorology, 11(2): 315-333.

Box, G.E.P., Tiao, G.C., 1992. Bayesian Inference in Statistical Analysis. John Wiley & Sons, Inc., Hoboken, NJ, USA.

Cameron, D.S., Beven, K.J., Tawn, J., Blazkova, S., Naden, P., 1999. Flood frequency estimation by continuous simulation for a gauged upland catchment (with uncertainty). Journal of Hydrology, 219(3-4): 169-187.

Chen, S. et al., 2013. Similarity and difference of the two successive V6 and V7 TRMM multisatellite precipitation analysis performance over China. Journal of Geophysical Research: Atmospheres, 118(23): 13,060-13,074.

Duan, Q., Ajami, N.K., Gao, X., Sorooshian, S., 2007. Multi-model ensemble hydrologic prediction using Bayesian model averaging. Advances in Water Resources, 30(5): 1371-1386.

Food and Agriculture Organization (FAO), 2003. Digital soil map of the world and derived soil properties, land and water digital media series [CD-ROM], Rome, Italy.

Gallart, F., Latron, J., Llorens, P., Beven, K.J., 2008. Upscaling discrete internal observations for obtaining catchment-averaged TOPMODEL parameters in a small Mediterranean mountain basin. Physics and Chemistry of the Earth, 33(17-18): 1090-1094.

Gao, Y.C., Liu, M.F., 2013. Evaluation of high-resolution satellite precipitation products using rain gauge observations over the Tibetan Plateau. Hydrology and Earth System Sciences, 17(2): 837-849.

Hesterberg, T., 2002. Monte Carlo Strategies in Scientific Computing, Monte Carlo Strategies in Scientific Computing. Taylor & Francis.

Hong, Y., Hsu, K.-l., Moradkhani, H., Sorooshian, S., 2006. Uncertainty quantification of satellite precipitation estimation and Monte Carlo assessment of the error propagation into hydrologic response. Water Resources Research, 42(8).

Hossain, F., Anagnostou, E.N., 2005. Assessment of a stochastic interpolation based parameter sampling scheme for efficient uncertainty analyses of hydrologic models. Computers & Geosciences, 31(4): 497-512.

Huffman, G.J., Bolvin, D.T., Nelkin, E.J., Wolff, D.B., Adler, R.F., Gu, G., Hong, Y., Bowman, K.P., Stocker, E.F., 2007. The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales. Journal of Hydrometeorology, 8(1): 38-55.

Kato, H., Rodell, M., Beyrich, F., Cleugh, H., Gorsel, E.v., Liu, H., Meyers, T.P., 2007. Sensitivity of Land Surface Simulations to Model Physics, Land Characteristics, and Forcings, at Four CEOP Sites. Journal of the Meteorological Society of Japan, 85A: 187-204.

Khu, S., Madsen, H., 2005. Multiobjective calibration with Pareto preference ordering: An application to rainfall- runoff model calibration. Water Resources Research, 41(3).

Kubota, T. et al., 2007. Global precipitation map using satellite-borne microwave radiometers by the GSMaP project: Production and validation. Ieee Transactions on Geoscience and Remote Sensing, 45(7): 2259-2275.

Li, Z., Yang, D., Hong, Y., 2013. Multi-scale evaluation of high-resolution multi-sensor blended global precipitation products over the Yangtze River. Journal of Hydrology, 500: 157-169.

Madadgar, S., Moradkhani, H., 2014. Improved Bayesian multimodeling: Integration of copulas and Bayesian model averaging. Water Resources Research, 50(12): 9586-9603.

Madadgar, S., AghaKouchak, A., Shukla, S., Wood, A.W., Cheng, L., Hsu, K., Svoboda, M., 2016. A hybrid statistical dynamical framework for meteorological drought prediction:

Application to the southwestern United States. Water Resources Research, 52(7): 5095-5110.

Maggioni, V., Vergara, H.J., Anagnostou, E.N., Gourley, J.J., Hong, Y., Stampoulis, D., 2013. Investigating the Applicability of Error Correction Ensembles of Satellite Rainfall Products in River Flow Simulations. Journal of Hydrometeorology, 14(4): 1194-1211.

McVicar, T.R., Niel, T.G., Li, L., Hutchinson, M.F., Mu, X., Liu, Z., 2007. Spatially distributing monthly reference evapotranspiration and pan evaporation considering topographic influences. Journal of Hydrology, 338(3-4): 196-220.

McVicar, T.R., Körner, C., 2013. On the use of elevation, altitude, and height in the ecological and climatological literature. Oecologia, 171(2): 335-337.

Müller, M.F., Thompson, S.E., 2013. Bias adjustment of satellite rainfall data through stochastic modeling: Methods development and application to Nepal. Advances in Water Resources, 60: 121-134.

Ngo, L.L., Madsen, H., Rosbjerg, D., 2007. Simulation and optimisation modelling approach for operation of the Hoa Binh reservoir, Vietnam. Journal of Hydrology, 336(3-4): 269-281.

Nikolopoulos, E.I., Anagnostou, E.N., Hossain, F., Gebremichael, M., Borga, M., 2010. Understanding the Scale Relationships of Uncertainty Propagation of Satellite Rainfall through a Distributed Hydrologic Model. Journal of Hydrometeorology, 11(2): 520-532.

Nott, D.J., Marshall, L., Brown, J., 2012. Generalized likelihood uncertainty estimation (GLUE) and approximate Bayesian computation: What's the connection? Water Resources Research, 48(12): W12602.

Ochoa, A., Pineda, L., Crespo, P., Willems, P., 2014. Evaluation of TRMM 3B42 precipitation estimates and WRF retrospective precipitation simulation over the Pacific–Andean region of Ecuador and Peru. Hydrology and Earth System Sciences, 18(8): 3179-3193.

Pan, M., Li, H., Wood, E., 2010. Assessing the skill of satellite-based precipitation estimates in hydrologic applications. Water Resources Research, 46(9).

Qi, W., Zhang, C., Fu, G., Zhou, H., 2015. Global Land Data Assimilation System data assessment using a distributed biosphere hydrological model. Journal of Hydrology, 528: 652-667.

Qi, W., Zhang, C., Fu, G., Sweetapple, C., Zhou, H., 2016a. Evaluation of global fine-resolution precipitation products and their uncertainty quantification in ensemble discharge simulations. Hydrology and Earth System Sciences, 20(2): 903-920.

Qi, W., Zhang, C., Fu, G., Zhou, H., 2016b. Quantifying dynamic sensitivity of optimization algorithm parameters to improve hydrological model calibration. Journal of Hydrology, 533: 213-223.

Qi, W., Zhang, C., Fu, G., Zhou, H., Liu, J., 2016c. Quantifying uncertainties in extreme flood predictions under climate change for a medium-sized basin in northeast China. Journal of Hydrometeorology(17): 3099–3112.

Refsgaard, J., Christensen, S., Sonnenborg, T.O., Seifert, D., Højberg, A., Troldborg, L., 2012. Review of strategies for handling geological uncertainty in groundwater flow and transport modeling. Advances in Water Resources, 36: 36-50.

Rodell, M. et al., 2004. The Global Land Data Assimilation System. Bulletin of the American Meteorological Society, 85(3): 381-394.

Roy, T., Serrat- Capdevila, A., Gupta, H., Valdes, J., 2017. A platform for probabilistic Multimodel and Multiproduct Streamflow Forecasting. Water Resources Research, 53(1): 376-399.

Sapriza-Azuri, G., Jódar, J., Navarro, V., Slooten, L.J., Carrera, J., Gupta, H.V., 2015. Impacts of rainfall spatial variability on hydrogeological response. Water Resources Research, 51(2): 1300-1314.

Sellers, P.J., Mintz, Y., Sud, Y.C., Dalcher, A., 1986. A Simple Biosphere Model (SIB) for Use within General Circulation Models. Journal of the Atmospheric Sciences, 43(6): 505-531.

Sellers, P.J., Randall, D.A., Collatz, G.J., Berry, J.A., Field, C.B., Dazlich, D.A., Zhang, C., Collelo, G.D., Bounoua, L., 1996. A Revised Land Surface Parameterization (SiB2) for Atmospheric GCMS. Part I: Model Formulation. Journal of Climate, 9(4): 676-705.

Sellers, P.J., 1997. Modeling the Exchanges of Energy, Water, and Carbon Between Continents and the Atmosphere. Science, 275(5299): 502-509.

Sorooshian, S., Hsu, K.L., Gao, X., Gupta, H.V., Imam, B., Braithwaite, D., 2000. Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. Bulletin of the American Meteorological Society, 81(9): 2035-2046.

Sorooshian, S., Gao, X., Hsu, K., Maddox, R.A., Hong, Y., Gupta, H.V., Imam, B., 2002. Diurnal Variability of Tropical Rainfall Retrieved from Combined GOES and TRMM Satellite Information. Journal of Climate, 15(9): 983-1001.

Sorooshian, S., Lawford, R.G., Try, P., Rossow, W., Roads, J., Polcher, J., Sommeria, G., Schifer, R., 2005. Water and energy cycles: Investigating the links. WMO Bull. , 54(2).

Tapiador, F.J. et al., 2012. Global precipitation measurement: Methods, datasets and applications. Atmospheric Research, 104-105: 70-97.

Tian, Y., Peters-Lidard, C.D., Adler, R.F., Kubota, T., Ushio, T., 2010. Evaluation of GSMaP Precipitation Estimates over the Contiguous United States. Journal of Hydrometeorology, 11(2): 566574.

Tolson, B.A., Shoemaker, C.A., 2007. Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. Water Resources Research, 43(1).

Voisin, N., Wood, A.W., Lettenmaier, D.P., 2008. Evaluation of precipitation products for global hydrological prediction. Journal of Hydrometeorology, 9(3): 388-407.

Vrugt, J.A., Robinson, B.A., 2007. Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging. Water Resources Research, 43(1): W01411.

Vrugt, J.A., ter Braak, C.J.F., Diks, C.G.H., Robinson, B.A., Hyman, J.M., Higdon, D., 2009. Accelerating Markov Chain Monte Carlo Simulation by Differential Evolution with Self-Adaptive Randomized Subspace Sampling, International Journal of Nonlinear Sciences and Numerical Simulation, pp. 273.

Wang, F., Wang, L., Koike, T., Zhou, H., Yang, K., Wang, A., Li, W., 2011. Evaluation and application of a fine-resolution global data set in a semiarid mesoscale river basin with a distributed biosphere hydrological model. Journal of Geophysical Research, 116(D21).

Wang, F., Wang, L., Zhou, H., Saavedra Valeriano, O.C., Koike, T., Li, W., 2012. Ensemble hydrological prediction-based real-time optimization of a multiobjective reservoir during flood season in a semiarid basin with global numerical weather predictions. Water Resources Research, 48(7): W07520.

Wang, L., Koike, T., Yang, D.W., Yang, K., 2009a. Improving the hydrology of the Simple Biosphere Model 2 and its evaluation within the framework of a distributed hydrological model. Hydrological Sciences Journal, 54(6): 989-1006.

Wang, L., Koike, T., Yang, K., Jackson, T.J., Bindlish, R., Yang, D., 2009b. Development of a distributed biosphere hydrological model and its evaluation with the Southern Great Plains Experiments (SGP97 and SGP99). Journal of Geophysical Research, 114(D8).

Wang, L., Koike, T., Yang, K., Yeh, P.J.-F., 2009c. Assessment of a distributed biosphere hydrological model against streamflow and MODIS land surface temperature in the upper Tone River Basin. Journal of Hydrology, 377(1-2): 21-34.

Wang, L., Koike, T., Yang, K., Jin, R., Li, H., 2010a. Frozen soil parameterization in a distributed biosphere hydrological model. Hydrology and Earth System Sciences, 14(3): 557-571.

Wang, L., Wang, Z., Koike, T., Yin, H., Yang, D., He, S., 2010b. The assessment of surface water resources for the semi-arid Yongding River Basin from 1956 to 2000 and the impact of land use change. Hydrological Processes, 24(9): 1123-1132.

Ward, E., Buytaert, W., Peaver, L., Wheater, H., 2011. Evaluation of precipitation products over complex mountainous terrain: A water resources perspective. Advances in Water Resources, 34(10): 1222-1231.

Wong, J.S., Razavi, S., Bonsal, B.R., Wheater, H.S., Asong, Z.E., 2017. Inter-comparison of daily precipitation products for large-scale hydro-climatic applications over Canada. Hydrology and Earth System Sciences, 21(4): 2163-2185.

Xie, P., Chen, M., Yang, S., Yatagai, A., Hayasaka, T., Fukushima, Y., Liu, C., 2007. A Gauge-Based Analysis of Daily Precipitation over East Asia. Journal of Hydrometeorology, 8(3): 607-626.

Xue, X., Hong, Y., Limaye, A.S., Gourley, J.J., Huffman, G.J., Khan, S.I., Dorji, C., Chen, S., 2013. Statistical and hydrological evaluation of TRMM-based Multi-satellite Precipitation Analysis over the Wangchu Basin of Bhutan: Are the latest satellite precipitation products 3B42V7 ready for use in ungauged basins? Journal of Hydrology, 499: 91-99.

Yang, D., 1998. Distributed hydrological model using hillslope discretization based on catchment area function: development and applications, University of Tokyo, Tokyo.

Yang, K., Koike, T., Ye, B., 2006. Improving estimation of hourly, daily, and monthly solar radiation by importing global data sets. Agricultural and Forest Meteorology, 137(1-2): 43-55.

Yatagai, A., Kamiguchi, K., Arakawa, O., Hamada, A., Yasutomi, N., Kitoh, A., 2012. APHRODITE: Constructing a Long-Term Daily Gridded Precipitation Dataset for Asia Based on a Dense Network of Rain Gauges. Bulletin of the American Meteorological Society, 93(9): 1401-1415.

Yong, B., Ren, L.-L., Hong, Y., Wang, J.-H., Gourley, J.J., Jiang, S.-H., Chen, X., Wang, W., 2010. Hydrologic evaluation of Multisatellite Precipitation Analysis standard precipitation

products in basins beyond its inclined latitude band: A case study in Laohahe basin, China. Water Resources Research, 46(7).

Yong, B., Hong, Y., Ren, L.-L., Gourley, J.J., Huffman, G.J., Chen, X., Wang, W., Khan, S.I., 2012. Assessment of evolving TRMM-based multisatellite real-time precipitation estimation methods and their impacts on hydrologic prediction in a high latitude basin. Journal of Geophysical Research, 117(D9).

Yong, B., Chen, B., Gourley, J.J., Ren, L., Hong, Y., Chen, X., Wang, W., Chen, S., Gong, L., 2014. Intercomparison of the Version-6 and Version-7 TMPA precipitation products over high and low latitudes basins with independent gauge networks: Is the newer version better in both real-time and post-real-time analysis for water resources and hydrologic extremes? Journal of Hydrology, 508: 77-87.

Table 1 Precipitation products

Product	Spatial resolution	Temporal resolution	Areal coverage	Temporal coverage	Data sources
TRMM3B42 V7	0.25°	3h	Global 50°N-S	1 Jan 1998-present	PR+IR+MW +gauge
TRMM3B42RT V7	0.25°	3h	Global 50°N-S	1 Mar 2000-present	IR+MW
GLDAS-1	0.25°	3h	Global 90°N-60°S	24 Feb 2000-present	IR+MW+gauge+R
GSMAP-MVK+ V6	0.1°	1h	Global 60°N-S	1 Mar 2000-present	IR+MW+CMV
PERSIANN	0.25°	3h	Global 60°N-S	1 Mar 2000-present	PR+IR+ANN
APHRODITE V1101R1	0.25°	1day	60°E-150°E, 15°S-55°N	1 Jan 1961 to 2007	gauge

PR: precipitation radar; IR: infrared estimation; MW: microwave estimation; CMV: cloud motion vectors; ANN: artificial neural network; R: reanalysis.

Table 2 WEB-DHM parameters

Symbol (unit)	Brief description	Basin-averaged value
KS (mm/h)	Saturated hydraulic conductivity for soil surface	26.43
Anik	Hydraulic conductivity anisotropy ratio	11.49
Sstmax (mm)	Maximum surface water storage	42.75
Kg (mm/h)	Hydraulic conductivity for groundwater	0.36
alpha	van Genuchten parameter	0.01
n	van Genuchten parameter	1.88

Table 3 TOPMODEL parameters

Name (unit)	Description	Lower bound	Upper bound	Calibration
SZM (m)	form of the exponential decline in conductivity	0.01	0.04	0.019
LNT0 (m^2 h^{-1})	log value of effective lateral saturated transmissivity	-25	1	-11.911
RV (m h^{-1})	hill slope routing velocity	2000	5000	2608.4
SR_{max} (m)	maximum root zone storage	0.001	0.01	0.006
SR_0 (m)	initial root zone deficit	0	0.01	0.005
TD (m h^{-1})	unsaturated zone time delay per unit deficit	2	4	2.885

Table 4 Comparison of NSE, RB and F for four approaches with training periods of 2 to 8 years

Training Years		NSE	RB	F
2	e-Bay	**0.96**	**-0.14**	**0.18**
	Mean	0.69	-0.22	0.53
	Maximum NSE	0.9	-0.19	0.29
	Weighted average	0.74	-0.18	0.44
3	e-Bay	**0.96**	-0.12	0.16
	Mean	0.78	-0.07	0.29
	Maximum NSE	**0.96**	**0.04**	**0.08**
	Weighted average	0.81	-0.05	0.24
4	e-Bay	**0.96**	-0.1	**0.14**
	Mean	0.81	0.2	0.39
	Maximum NSE	0.8	**-0.09**	0.29
	Weighted average	0.83	0.16	0.33
5	e-Bay	**0.98**	**-0.08**	**0.1**
	Mean	0.9	0.09	0.19
	Maximum NSE	0.92	**0.08**	0.16
	Weighted average	0.9	0.06	0.16
6	e-Bay	**0.98**	-0.09	0.11
	Mean	0.81	-0.06	0.25
	Maximum NSE	0.93	**-0.02**	**0.09**
	Weighted average	0.83	-0.05	0.22
7	e-Bay	**0.99**	**-0.06**	**0.07**
	Mean	0.8	-0.08	0.28
	Maximum NSE	0.95	0.1	0.15
	Weighted average	0.82	-0.07	0.25
8	e-Bay	**0.99**	-0.05	**0.06**
	Mean	0.78	-0.05	0.27
	Maximum NSE	0.88	-0.07	0.19
	Weighted average	0.79	**-0.04**	0.25
Average	e-Bay	**0.97**	-0.09	**0.12**
	Mean	0.80	-0.03	0.31
	Maximum NSE	0.91	**-0.02**	0.18
	Weighted average	0.82	**-0.02**	0.27

Note: The best criterion values are shown in bold fonts.

Table 5 Comparison of NSE, RB and F for four approaches with validation periods of 1 to 6 years

Validation years		NSE	RB	F
6	e-Bay	0.7	0.19	0.49
	Mean	0.79	**0**	0.21
	Maximum NSE	**0.85**	-0.03	**0.18**
	Weighted average	0.8	-0.03	0.23
5	e-Bay	**0.87**	0.08	**0.21**
	Mean	0.77	**-0.04**	0.27
	Maximum NSE	0.81	-0.1	0.29
	Weighted average	0.78	**-0.04**	0.26
4	e-Bay	**0.85**	-0.34	0.49
	Mean	0.74	**-0.16**	**0.42**
	Maximum NSE	0.57	-0.49	0.92
	Weighted average	0.72	-0.2	0.48
3	e-Bay	**0.79**	-0.21	**0.42**
	Mean	0.66	**-0.17**	0.51
	Maximum NSE	0.73	-0.19	0.46
	Weighted average	0.65	-0.19	0.54
2	e-Bay	**0.91**	0.17	**0.26**
	Mean	0.69	**-0.03**	0.34
	Maximum NSE	0.63	-0.16	0.53
	Weighted average	0.71	**-0.03**	0.32
1	e-Bay	**0.95**	**0.01**	**0.06**
	Mean	0.65	0.1	0.45
	Maximum NSE	-0.02	-0.42	1.44
	Weighted average	0.68	0.11	0.43
Average	e-Bay	**0.85**	**-0.02**	**0.32**
	Mean	0.72	-0.07	0.37
	Maximum NSE	0.60	-0.23	0.64
	Weighted average	0.72	-0.06	0.38

Note: The best results are shown in bold fonts.

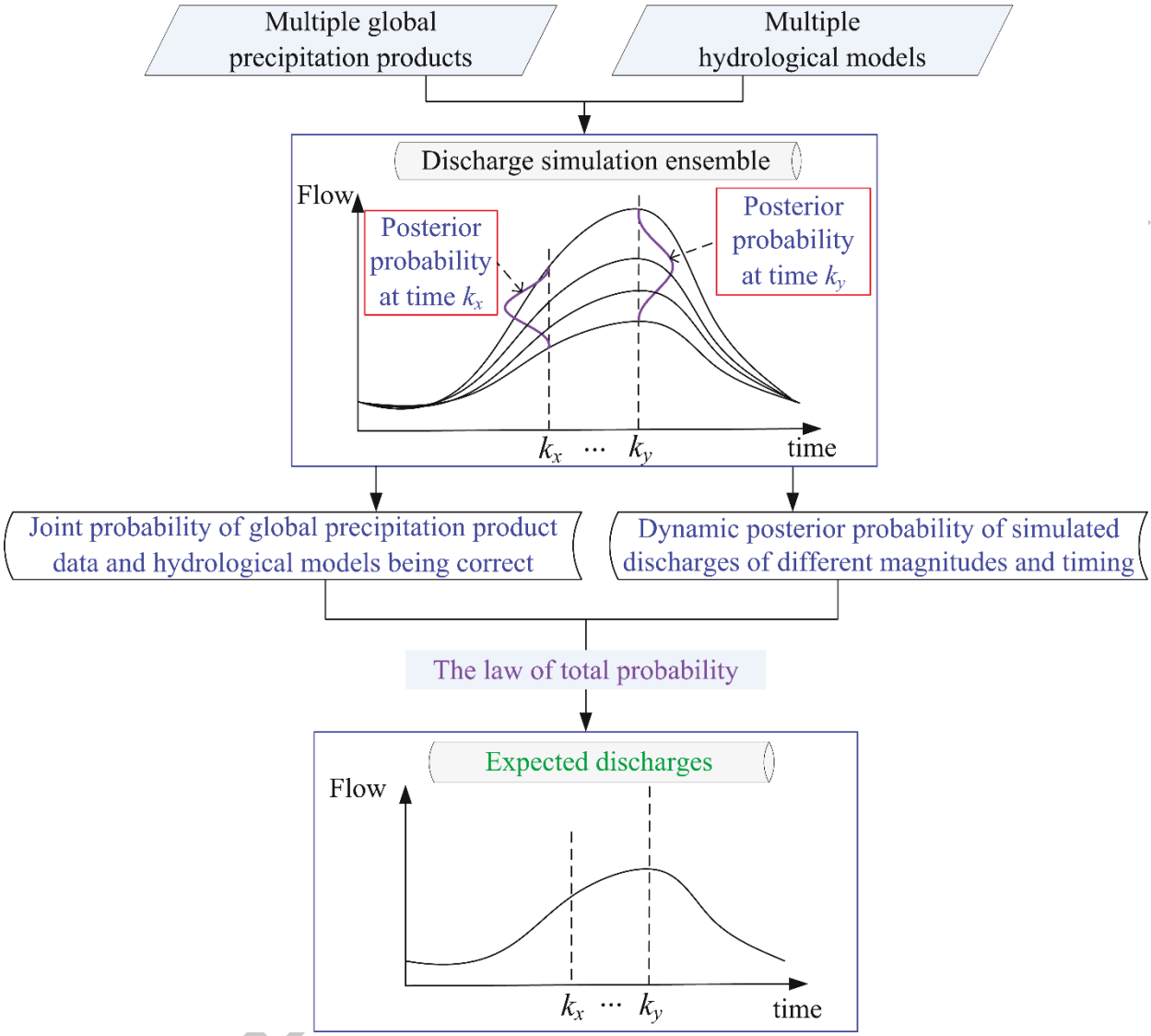


Fig. 1 Flowchart of the developed e-Bay approach.

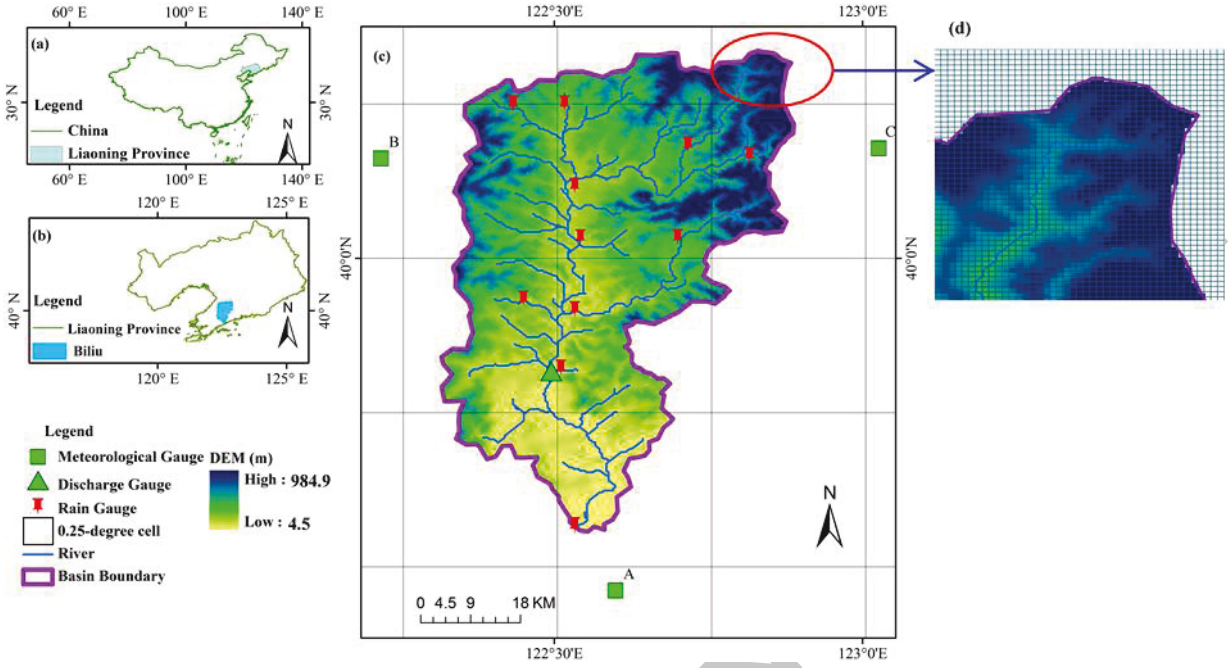


Fig. 2 The Biliu river basin: (a) location of Liaoning province within China; (b) location of the Biliu basin within Liaoning province; (c) distributions of precipitation gauges, discharge gauge, automatic weather stations, digital elevation model, and 0.25-degree precipitation cells which are represented by the horizontal and vertical lines together; (d) illustration of resampling into 300 m × 300 m cells in a precipitation cell that crosses a basin boundary to enable calculation of basin-averaged values.

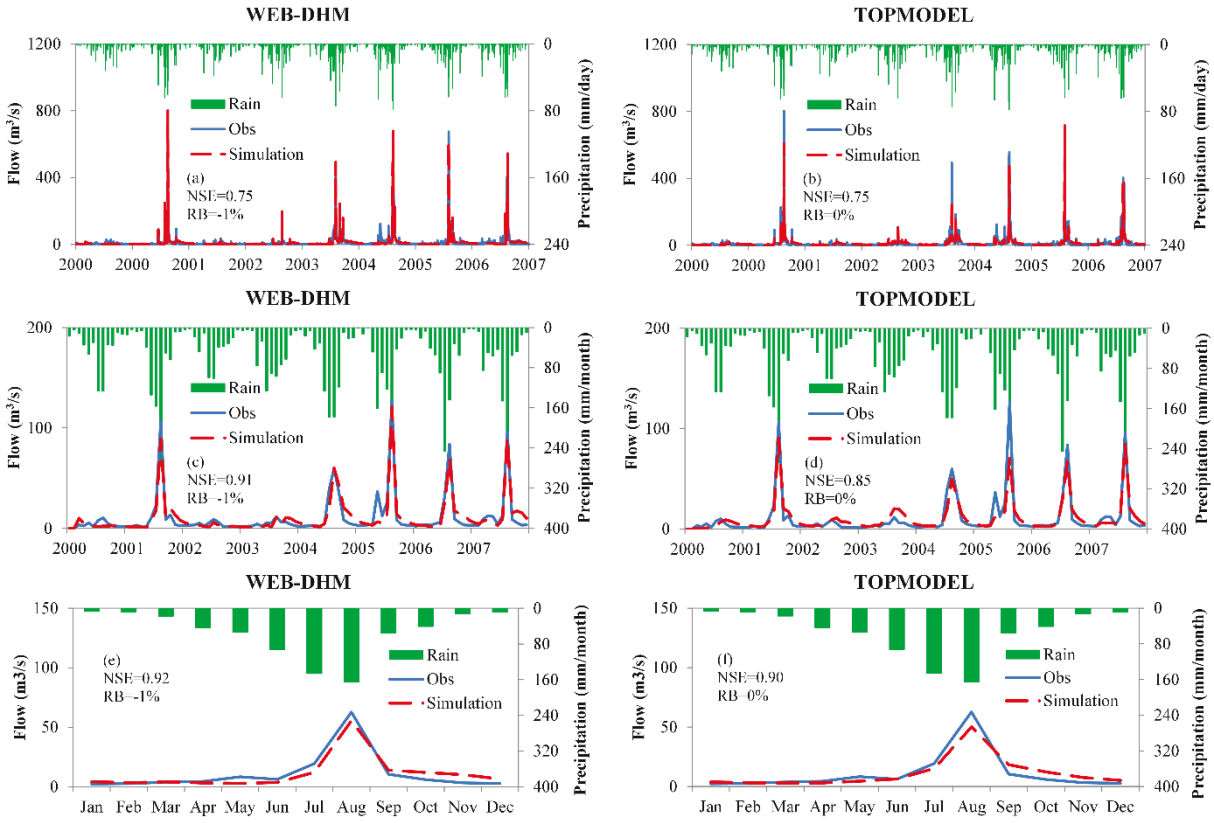


Fig. 3 Observed and simulated flows using WEB-DHM and TOPMODEL from 2000 to 2007: (a), (c) and (e) are daily, monthly and inter-annual simulations using WEB-DHM respectively; (b), (d) and (f) are daily, monthly and inter-annual simulations using TOPMODEL respectively.

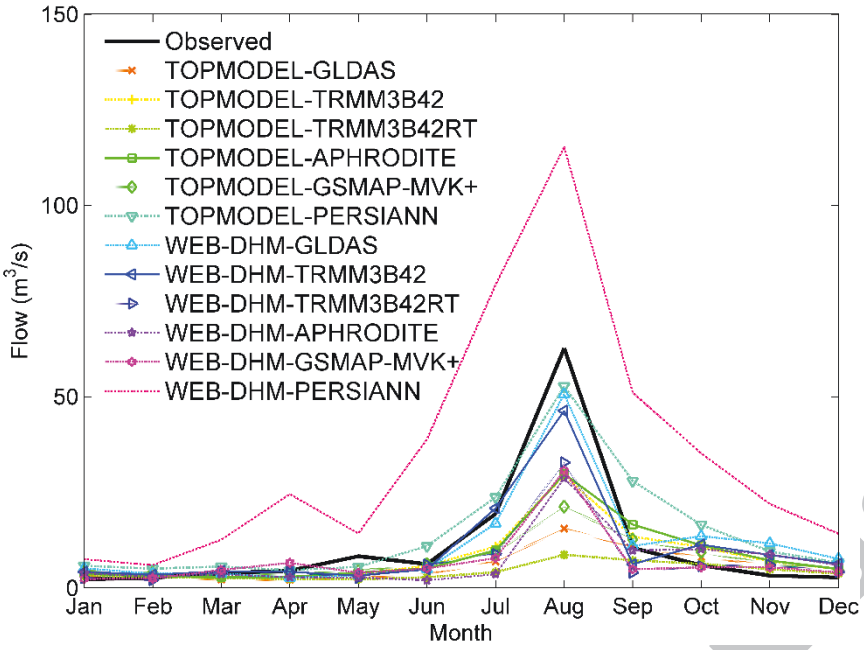


Fig. 4 The discharge simulation ensemble using two hydrological models (WEB-DHM and TOPMODEL) and six global precipitation products (TRMM3B42, TRMM3B42RT, GLDAS-1, APHRODITE, PERSIANN and GSMAP-MVK+).

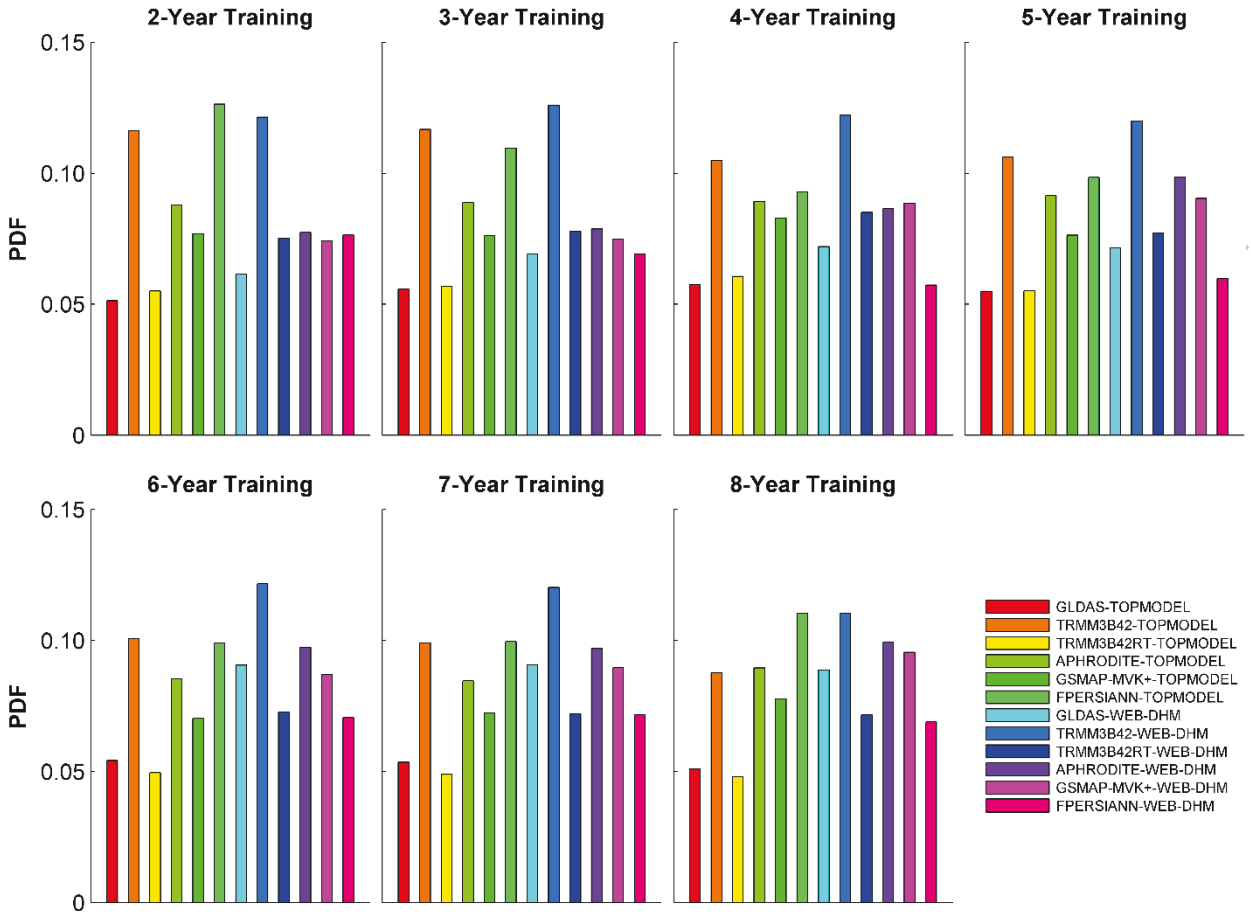


Fig. 5 Joint Probability density Distribution Functions (PDF) of precipitation products and hydrological models being correct in the training period. The duration of training data ranges from two to eight years. Each hydrological model and precipitation product combination is represented by a different color.

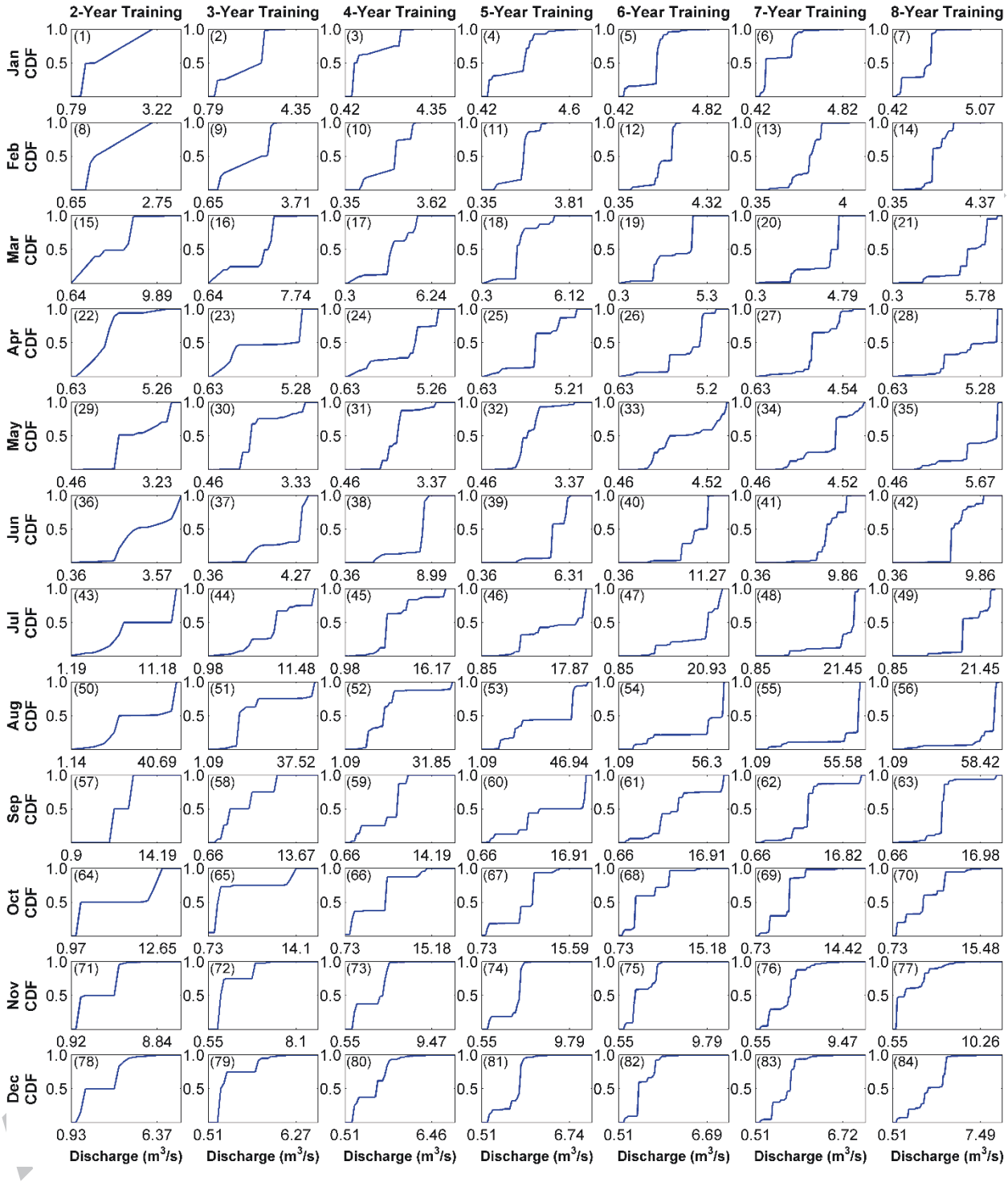


Fig. 6 Dynamic posterior probability (Cumulative probability Distribution Functions - CDF) of simulated discharges over twelve months. The duration of training data ranges from two to eight years. Each row shows the results of one month; each column represents the results of twelve months using the same years of training data. In each panel, *x*-axis represents the simulated discharges.

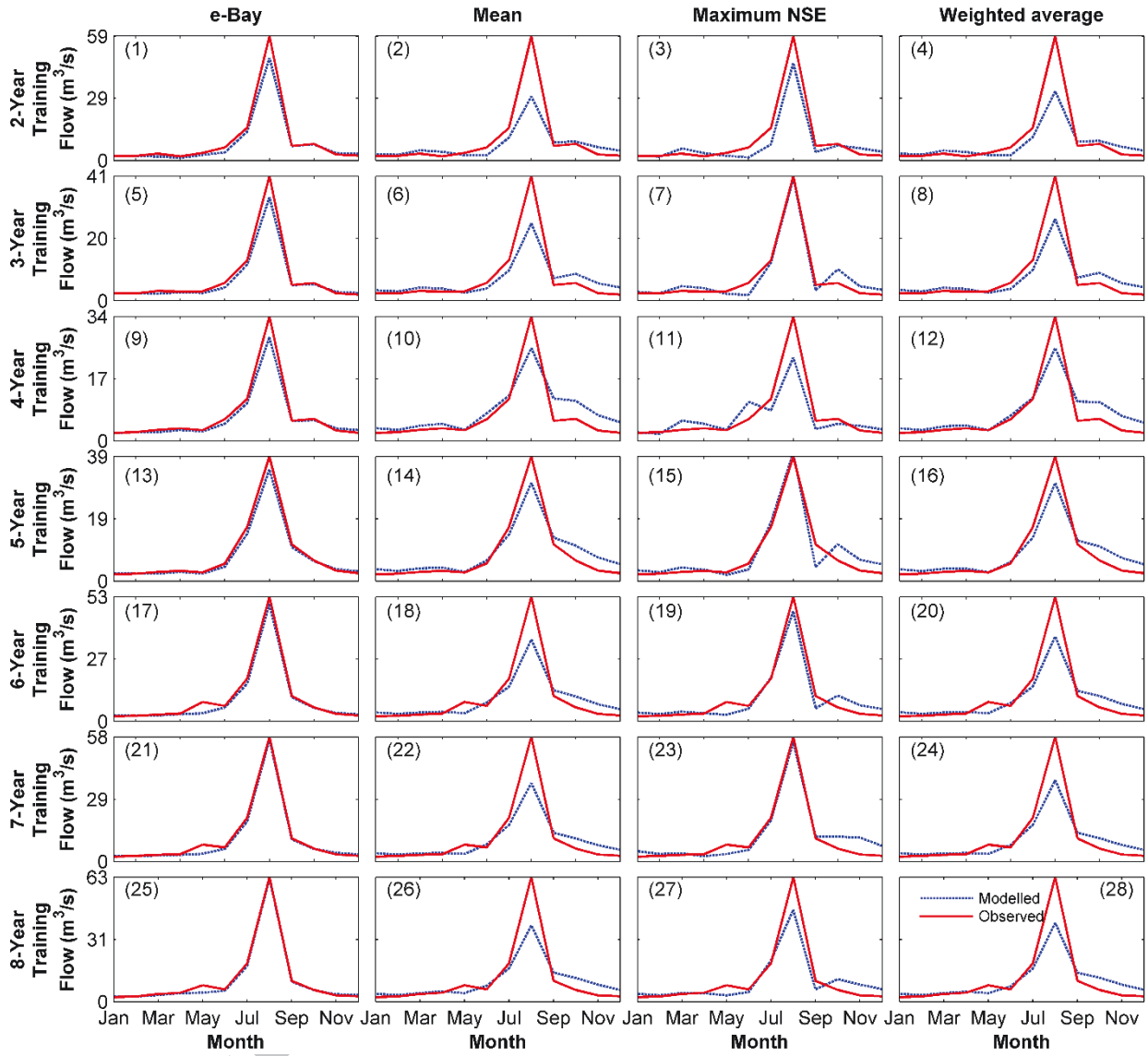


Fig. 7 Comparison of discharge values simulated using each approach in the training period. Each row shows the results of the four approaches using the same years of training data. Each column represents the results of the same approach using different years of training data.

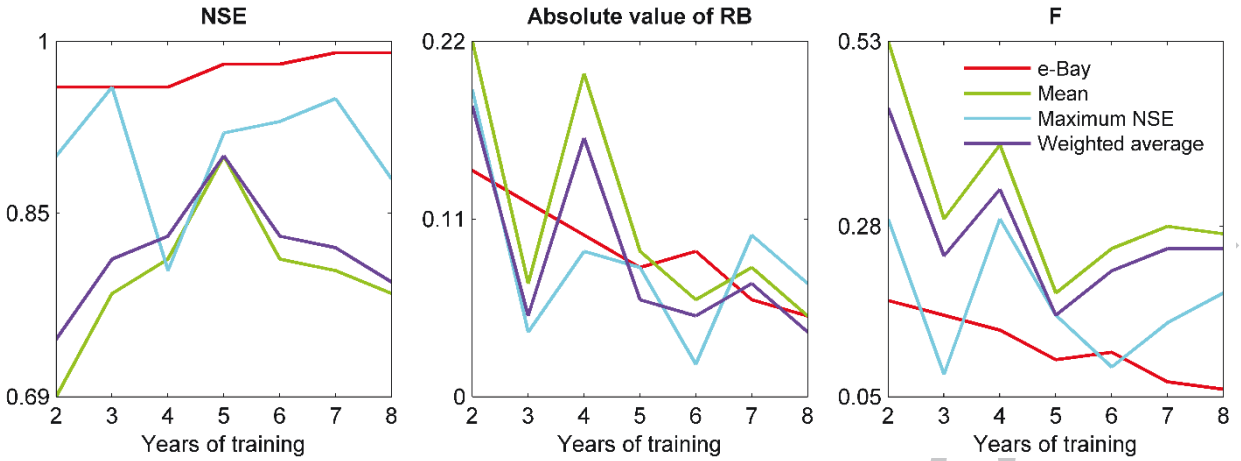


Fig. 8 Trends of the three performance criteria under the four approaches, each of which is represented by a different color. The *x*-axis shows the duration of training data in years. The *y*-axis shows the values of the performance criteria.

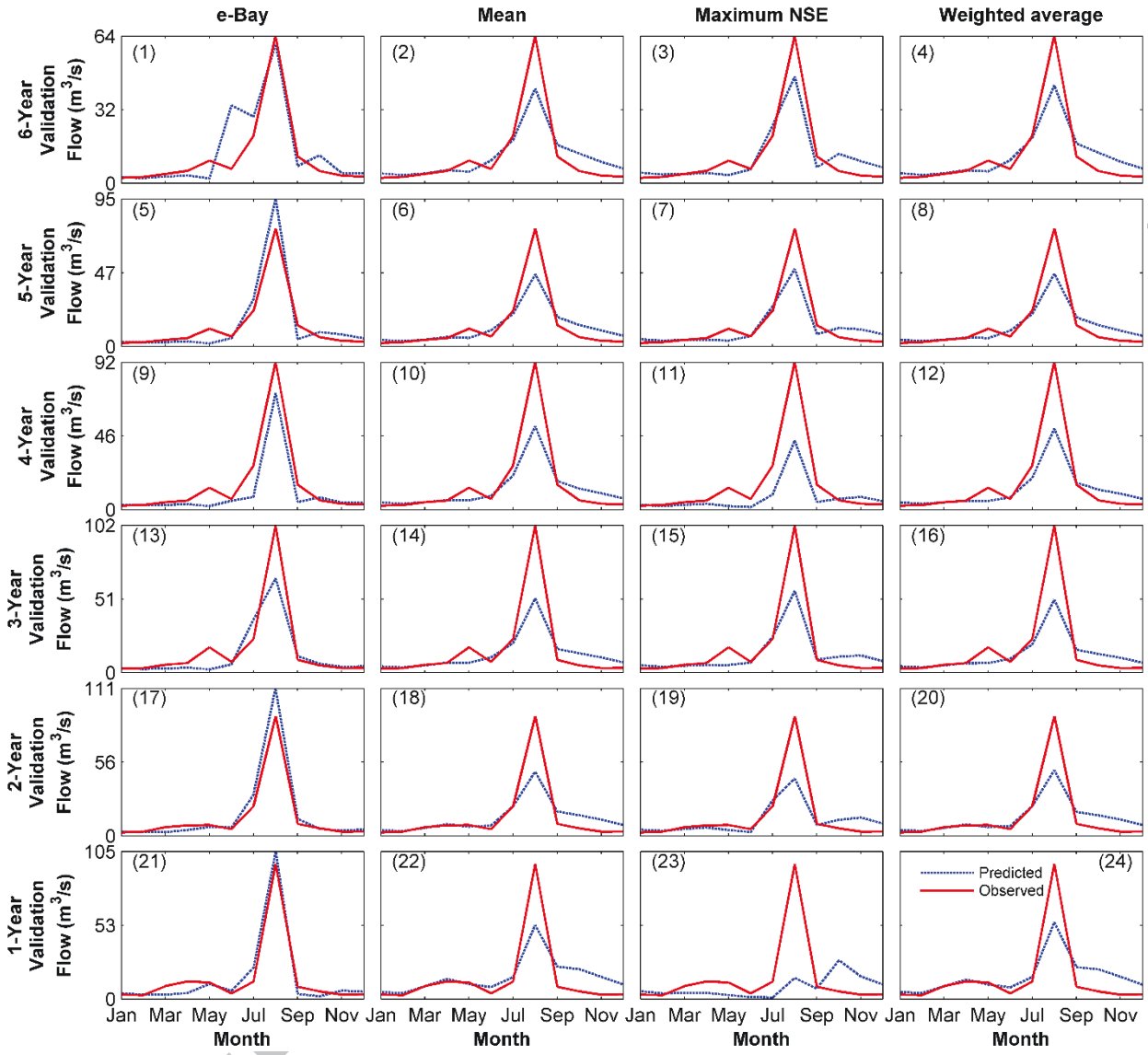


Fig. 9 Comparison of discharge values simulated using each approach in the validation period. Each row shows the results of the four approaches using the same years of data. Each column represents the results of the same approach using different years of data.

Highlights:

An ensemble-based dynamic Bayesian averaging approach is proposed

Estimate joint probability of precipitation and hydrological models being correct

Estimate posterior distribution based on magnitude and timing of flows

Outstanding capability to estimate expected discharges