

Accepted Manuscript

Research papers

Using multiple satellite-gauge merged precipitation products ensemble for hydrologic uncertainty analysis over the Huaihe River basin

Ruochen Sun, Huiling Yuan, Yize Yang

PII: S0022-1694(18)30712-1

DOI: <https://doi.org/10.1016/j.jhydrol.2018.09.024>

Reference: HYDROL 23121

To appear in: *Journal of Hydrology*

Received Date: 16 May 2018

Revised Date: 7 September 2018

Accepted Date: 10 September 2018



Please cite this article as: Sun, R., Yuan, H., Yang, Y., Using multiple satellite-gauge merged precipitation products ensemble for hydrologic uncertainty analysis over the Huaihe River basin, *Journal of Hydrology* (2018), doi: <https://doi.org/10.1016/j.jhydrol.2018.09.024>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1 Using multiple satellite-gauge merged precipitation products ensemble for
2 hydrologic uncertainty analysis over the Huaihe River basin

3 Ruochen Sun^a, Huiling Yuan^{a, b, *}, Yize Yang^a

4

5 ^a School of Atmospheric Sciences and Key Laboratory of Mesoscale
6 Severe Weather/Ministry of Education, Nanjing University, Nanjing,
7 China

8 ^b Joint Center for Atmospheric Radar Research of CMA/NJU, China

9 Corresponding author at: School of Atmospheric Sciences and Key
10 Laboratory of Mesoscale Severe Weather/Ministry of Education, Nanjing
11 University, Nanjing, 210023, China. Tel.: +86 25 89685016.

12 E-mail address: yuanhl@nju.edu.cn (H. Yuan).

Abstract

Global satellite–gauge merged precipitation (SGMP) products combine the advantages of satellite precipitation estimates with rain gauge data, providing great potential to hydrological applications. However, the inaccuracies of the precipitation products together with hydrologic model limitations, could cause great uncertainty in streamflow predictions. Therefore, this study investigates the hydrological value of three mainstream global SGMP products, including the Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA) 3B42V7 product, the Climate Prediction Center (CPC) MORPHing technique (CMORPH) satellite–gauge merged product (CMORPH BLD), the Global Satellite Mapping of Precipitation (GSMaP) Gauge-calibrated product (GSMaP Gauge). They are used as the precipitation input of the Variable Infiltration Capacity (VIC) hydrologic model over the Huaihe River basin in China. To better quantify their effects on parameter calibration and streamflow predictions, a newly developed residual error model accompanied with the Bayesian uncertainty analysis are performed. CMORPH satellite-gauge merged precipitation product, recently developed by the China Meteorological Administration (CMA) (CMORPH CMA), is a high-quality regional precipitation product. Thus, this study applies the CMORPH CMA within the same framework to

provide a benchmark. The results show that the parameter uncertainty are influenced significantly by the input of various precipitation products. There is a tradeoff between the deterministic streamflow performance and the probabilistic predictions for selecting the best input among the three global precipitation products. The streamflow uncertainty intervals of the three global precipitation products are then merged using the Bayesian Model Averaging (BMA) method. The BMA results show satisfying hydrological performance in terms of deterministic streamflow predictions, with the largest Nash-Sutcliffe coefficient of Efficiency (NSCE) values of 0.86 and 0.64, and the smallest absolute relative error (RE) values of 0% and 10.2% in the calibration and validation periods, respectively. In addition, the BMA results also produce much more reliable probabilistic predictions, which even outperform the outcomes of the high-quality CMORPH CMA. Our study demonstrates the potential uncertainty of various SGMP products for model calibration and streamflow predictions. The hydrologic ensemble using multiple global SGMP products provides a promising and advantageous approach to support water management and decision making, especially in ungauged basins.

Key words: Satellite–gauge; Residual error model; Parameter uncertainty; Streamflow prediction; Bayesian Model Averaging

1. Introduction

As a fundamental component of the global water cycle, precipitation often shows notable variability in both space and time (Kidd and Huffman, 2011), thus significantly affecting the land surface hydrological process and playing a key role in water resource management. Therefore, obtaining accurate precipitation data is of great importance for hydrological simulation (McMillan et al., 2011; Yu et al. 2011).

Satellite-based precipitation estimates have become important resources and have been used worldwide for different hydrological applications (Vergara et al., 2014; Sun et al., 2016), because many products have the advantage to provide global coverage of precipitation observation (Kucera et al., 2013). Currently a lot of quasi-global satellite-only precipitation estimates are widely used, such as the raw version of Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks (PERSIANN; Sorooshian et al., 2000), the National Oceanic and Atmospheric Administration (NOAA)'s Climate Prediction Center (CPC) morphing technique (CMORPH; Joyce et al., 2004) rainfall estimates, the near real-time product of the Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA; Huffman et al. 2007), the satellite rainfall estimates of the Global Satellite Mapping of Precipitation (GSMaP; Kubota et al., 2007) project. In addition,

a number of SGMP products have been generated, considering the pros and cons of satellite-only rainfall estimates and gauge observations. These include the PERSIANN-Climate Data Record (PERSIANN-CDR; Ashouri et al., 2015), the TMPA 3B42 version7 product (TMPA hereafter), the CMORPH bias-corrected product (CMORPH CRT), the CMORPH gauge-satellite blended precipitation product (CMORPH BLD; Xie et al., 2013), the CMORPH satellite–gauge merged product developed by the China Meteorological Administration (CMA) (CMORPH CMA; Yu et al., 2015), the GSMaP Gauge-calibrated product (GSMaP Gauge; Mega et al., 2014). In 2014, the successor of TRMM, the Global Precipitation Measurement (GPM) mission released the Integrated Multi-satellitE Retrievals for GPM (IMERG) with temporal resolution of 30 min and spatial resolution of 0.1° (Hou et al., 2014). With increasing spatial and temporal resolution, the use of satellite precipitation products in hydrological applications would provide wider venues to support water management (Maggioni and Massari, 2018).

Numerous applications of satellite precipitation products have indicated their potentials and values in hydrological simulation and water resources management (Behrangi et al., 2011; Bitew and Gebremichael, 2011; Li et al., 2012; Li et al., 2015; Tong et al., 2014; Wang et al., 2017; Zubieta et al., 2015). These studies also pointed out that different satellite

precipitation products have variable accuracy and distinct hydrological performance in different basins. Because of various uncertainties in the data sources, retrieval algorithms and bias correction processes of satellite precipitation products (Dinku et al., 2010; Kidd et al. 2003; Tian et al., 2010), the quality of these products would affect hydrological simulations through the rainfall-runoff processes in hydrologic models (Nikolopoulos et al., 2012; Vergara et al., 2014; Zubieta et al., 2015). In another word, the uncertainty in satellite-based precipitation data, together with parameter uncertainty and structural uncertainty of hydrologic models, will result in uncertainty in streamflow predictions.

Although lots of studies have applied different satellite precipitation products to streamflow or flood modeling, most of them focused on evaluating the performance of deterministic simulations without uncertainty estimation. There are also some proposed methods to account for the satellite precipitation uncertainties in hydrological simulations. Hong et al. (2006) assessed the influence of satellite-based precipitation estimation error on the uncertainty of hydrological response using Monte Carlo simulation, by constructing an error function associated with the precipitation estimates. The two dimensional satellite rainfall error model (SREM2D; Hossain and Anagnostou, 2006) was used to generate ensembles of satellite rain fields for investigating the error propagation

from satellite rainfall to streamflow (Falck et al., 2015; Maggioni et al., 2013). However, these studies did not consider the parameter and model structure uncertainty when obtaining the streamflow ensemble, which could probably lead to unrealistic uncertainty bounds (Ajami et al., 2007). In addition, these studies generated ensemble streamflow simulations based on a single satellite precipitation product, thus the corresponding performance could be highly affected by the error characteristics of the selected satellite precipitation product in the region.

There is limited research on the impacts of different satellite precipitation products on model parameter uncertainty and hydrologic predictive uncertainty estimation, which are essential parts of hydrologic study (Kavetski, 2018; Stedinger et al., 2008). For uncertainty analysis of streamflow prediction, the errors should be described adequately with respect to their characteristics (Kuczera et al., 2010; Schoups and Vrugt, 2010; Sikorska and Seibert, 2016). A common method is to use residual error models to treat total uncertainty in a lumped manner (Bates and Campbell, 2001; Evin et al., 2013, 2014; Schoups and Vrugt, 2010; Sorooshian and Dracup, 1980; Sun et al., 2017). The residual errors typically consist of a combination of input, model structural and parameter errors (Schoups and Vrugt, 2010). Within the Bayesian framework, parameter posterior distributions are obtained based on a likelihood

function derived from the residual error model (Box and Tiao, 1992). Accordingly, posterior predictive distribution of streamflow can be created by adding residual errors to model outputs. Sikorska and Seibert (2016) performed a Bayesian uncertainty analysis with an improved characterization of model residual errors to quantify the effects of station-based and radar-based precipitation on model calibration and flood predictions. To date, similar studies have rarely been applied to satellite-based precipitation.

Ensemble modeling techniques based on different models or datasets have the potential advantage to improve uncertainty estimation in meteorological and hydrological modeling (Duan et al., 2007; Raftery et al., 2005; Ma et al., 2018a, 2018b). Bayesian model averaging (BMA) is one of the most widely used ensemble approaches, which derives the consensus prediction from multiple competing predictions (Hoeting et al., 1999) and generates a better merged forecast by exploiting the strengths of the individual predictions. The BMA method assigns weight to each member based on its corresponding predictive performance within the observations. In hydrological field, BMA has been successfully applied to provide improved and more reliable streamflow predictions with corresponding uncertainty measures. Most of the studies used the BMA method to combine streamflow simulations from different hydrologic

models for better accounting for model structure uncertainty (Ajami et al., 2007; Duan et al., 2007; He et al., 2018; Jiang et al., 2018; Liang et al., 2013; Roy et al., 2017). Through the use of multiple different hydrologic models, the model error is implied in the variability of different model predictions. This method of using BMA assumes there is enough variability in model structures to account for the uncertainties in the general model framework (DeChant and Moradkhani, 2014). Similarly, the framework has been transferred to account for precipitation input uncertainty of hydrologic models. Strauch et al. (2012) used the BMA method to account for precipitation uncertainty by merging streamflow simulations forced by different types of gauge-based precipitation datasets. However, studies focusing on satellite precipitation uncertainty in streamflow simulations are relatively few (Jiang et al., 2012, 2014).

A lot of studies have demonstrated that SGMP products which use gauge observations to correct the errors of original satellite rainfall estimates yield improved streamflow performance (Maggioni and Massari, 2018; Mei et al., 2016; Wang et al., 2015). The first objective of this study is therefore to investigate the hydrologic value of three mainstream global SGMP products (e.g., TMPA, GSMaP Gauge and CMORPH BLD) for hydrological simulations over the Huaihe River basin of China with the involvement of the Bayesian uncertainty analysis. Sun et al. (2017)

proposed a novel residual error model to address the strong and complex heteroscedasticity of residual errors showed in the Huaihe River basin. By means of this newly developed residual error model, the parameter and model structural uncertainty could also be properly considered in addition to the precipitation uncertainty. Through the Bayesian uncertainty analysis, the impact of different SGMP products on parameter uncertainty and streamflow predictive uncertainty estimation is explored. The second objective is to improve the streamflow prediction skills and provide more reliable uncertainty estimations by merging the individual model outputs of the three SGMP products using the BMA method. Sun et al. (2016) found out that the regional SGMP product CMORPH CMA produced the best streamflow simulations, and could serve as the high quality precipitation input data in China. To evaluate the performance of the streamflow predictions forced by the three global SGMP products and the BMA merged streamflow prediction, the high-quality CMORPH CMA forced streamflow prediction based on the same error model (Sun et al., 2017) is used as benchmark. The comparison could provide quantitative evaluation of the usage of multiple global satellite precipitation products within the BMA framework, thus indicating its application prospect in poorly gauged basins.

The rest part of the paper is organized as follows. Section 2 describes

the study area and various data used. Section 3 describes the detailed methodology including the hydrologic model calibration method, the residual error model, the BMA method and verification methods. Section 4 provides detailed results. Section 5 presents discussion, followed by conclusions in Section 6.

2. Study area and data

2.1 Study area

The Huaihe River basin ($30^{\circ}55'-36^{\circ}36'N$, $111^{\circ}55'-121^{\circ}25'E$), located in eastern China between the Yangtze and Yellow River basins is the most densely inhabited river basin and the sixth largest river basin in China. The basin situates within the China's climate transition zone of subtropical moist and semi-moist monsoon climate. The average annual precipitation is approximately 900 mm, of which 50–75% falls during the summer monsoon season. The study area (Fig .1) covers the upper region of Bengbu hydrologic station over the Huaihe River basin, same as the one in Sun et al. (2016, 2017).

2.2 Data

This study aims to evaluate the effect of three global SGMP products on streamflow prediction uncertainty over the Huaihe River basin, including TMPA 3B42V7 (TMPA), CMORPH satellite-gauge merged

product (CMORPH BLD) developed at NOAA/CPC, GSMaP Gauge-calibrated Rainfall Product (GSMaP Gauge). GSMaP integrates passive microwave (PMW) retrievals and infrared (IR) radiometer data using a PMW–IR merged algorithm, which uses the Kalman filter to refine rainfall rate estimated from the two-way morphing technique from IR images (Joyce et al. 2004), to produce the GSMaP Moving Vector with Kalman-filter (GSMaP_MVK) product (Ushio et al., 2009). To develop GSMaP Gauge, NOAA CPC global rain gauge dataset is used to remove the bias of GSMaP_MVK (Mega et al., 2014). The GSMaP Gauge product used in this study is daily and covers 60°S–60°N, 180°W–180°E, with a spatial resolution of 0.1°. For TMPA and CMORPH BLD, the authors have made very detailed introductions in Sun et al. (2016), interested readers can refer to the paper for more information.

According to Sun et al. (2016), the recent CMORPH SGMP product developed by CMA (CMORPH CMA) showed even higher accuracy in rainfall monitoring and streamflow simulations than high density ground rain gauge data. Therefore, CMORPH CMA can serve as an alternative high quality precipitation product in China. The streamflow prediction interval forced by CMORPH CMA was adopted as a reference in this study.

The four SGMP products were aggregated into the uniform $0.25^{\circ} \times 0.25^{\circ}$ spatial grid and daily (00 UTC–00 UTC) resolution during the study

period of 2002–2012. The annual basin average precipitation of the four satellite-gauge products is shown in Table 1, which is characterized by consistent performance and slight interannual variability for the four products.

The daily discharge observations of the Bengbu hydrologic station were collected from the Ministry of Water Resources of the People's Republic of China. The collected discharge dataset contained the period of 2003–2007 for calibration, the spinup period of 2002, and the remaining dataset (2008–2012) for validation.

3. Methodology

3.1 Hydrologic model

In this study, the physically-based, distributed VIC hydrologic model was selected to simulate the rainfall–runoff processes. The VIC model is based on a soil vegetation atmospheric transfer scheme that considers both energy and water balances (Liang et al., 1994, 1996). The 3-Layer model also takes into account the spatial variation of vegetation cover, topography, precipitation, soil properties and soil moisture. Many of the VIC parameters are assigned according to the vegetation type and soil texture. Although most parameters can be directly estimated from the land surface database, in this study a total of six hydrologic parameters were optimized through calibration, including the exponent of variable infiltration capacity

curve (b), the maximum baseflow that can occur from the lowest soil layer (D_{smax}), the fraction of maximum base flow (D_s), the fraction of maximum soil moisture content of the third layer (W_s), the second and the third soil-layer thicknesses (d₂ and d₃).

3.2 Bayesian uncertainty analysis

Consider that a hydrologic model simulates streamflow Q with the forcing data X (e.g., precipitation, temperature) and a set of model parameters θ_H . A residual error model is used to describe the additive errors e , defined at time step t as:

$$e_t = \tilde{Q}_t - Q_t(\tilde{X}_{1:t}, \theta_H) \quad (1)$$

where \tilde{Q}_t is the observed streamflow and $Q_t(\tilde{X}_{1:t}, \theta_H)$ is the simulated streamflow with hydrologic model parameters θ_H and forcing data over time steps 1 to t , $\tilde{X}_{1:t}$.

According to the Bayesian inference, the posterior distribution of parameters is

$$p(\theta_H, \theta_e | \tilde{X}, \tilde{Q}) \propto p(\tilde{Q} | \theta_H, \theta_e, \tilde{X}) p(\theta_H, \theta_e) \quad (2)$$

where θ_e denotes the residual error model parameters, $p(\tilde{Q} | \theta_H, \theta_e, \tilde{X})$ is the likelihood function and $p(\theta_H, \theta_e)$ is the prior distribution of hydrologic and residual error model parameters. The likelihood function can be represented in the form of the joint probability density function (PDF) of the residual errors

$$p(\tilde{Q} | \theta_H, \theta_e, \tilde{X}) = p(e[\theta_H, \tilde{X}, \tilde{Q}] | \theta_e) \quad (3)$$

where $e[\theta_H, \tilde{X}, \tilde{Q}]$ is the vector of residual errors obtained over the calibration period.

The successful application of Bayesian inferential approaches is typically based on the formal likelihood function properly characterizing the form of the residual errors (Mantovan and Todini, 2006). Residual errors often show characteristics of autocorrelation, nonnormality and heteroscedasticity and a number of strategies have been proposed to handle these features. Recently, Sun et al. (2017) proposed a combined approach (CA) combining the advantages of the linear modeling and Box-Cox transformation to deal with the strong and complex heteroscedasticity showed in the Huaihe River basin. Together with the first order autoregressive model and the skew exponential power (SEP) distribution, the residual error model, CA-SEP was generated. The expression of the CA-SEP is shown below:

$$\eta_t = \frac{e_t^*}{\sigma_t}; \quad \eta_t = \varphi_1 \eta_{t-1} + a_t \quad \text{with} \quad a_t \sim SEP(0, 1, \xi, \beta); \quad \sigma_t = \sigma_0 + \sigma_1 Q_t^* \quad (4)$$

where e_t^* denotes the residual errors after the Box-Cox transformation, σ_t is a normalization term, φ_1 is the first-order autoregressive coefficient and a_t is the innovation described by the SEP distribution, with parameters ξ and β accounting for nonnormality. σ_t is conditioned on the simulated streamflow after Box-Cox transformation (Q_t^*), with

parameters σ_0 and σ_1 . The Box-Cox transformation is shown below:

$$f(y) = \begin{cases} \frac{(y + \lambda_2)^{\lambda_1} - 1}{\lambda_1} & (\lambda_1 \neq 0; y > -\lambda_2) \\ \ln(y + \lambda_2) & (\lambda_1 = 0; y > -\lambda_2) \end{cases}$$

(5)

where y is observed or simulated streamflow, λ_1 and λ_2 are the Box-Cox parameters. We fixed the value of λ_2 to 1 to make the original zero flows still zero after the transformation. The corresponding likelihood function derived from the residual error model can be expressed using the generalized likelihood (GL) function (Schoups and Vrugt, 2010). The expression for the log-likelihood function is:

$$\log L(\tilde{Q} | \theta_H, \theta_e) = n \log \left(\omega_\beta \frac{2\sigma_\xi}{(\xi + \xi^{-1})} \right) - \sum_{t=1}^n \{ \log(\sigma_t) \} - c_\beta \sum_{t=1}^n |a_{\xi,t}|^{2/(1+\beta)} + (\lambda_1 - 1) \sum_{t=1}^n \log(\tilde{Q}_t + 1)$$

(6)

where n is the number of observations, $a_{\xi,t} = \xi^{-\text{sign}(\mu_\xi + \sigma_\xi a_t)}$ and values for μ_ξ , σ_β , ω_β and c_β are all functions of the skewness parameter ξ and the kurtosis parameter β (see Schoups and Vrugt (2010) for details).

In sun et al. (2017), the residual error model CA-SEP was developed using the high quality precipitation product CMORPH CMA as the input and the obtained streamflow prediction interval was reliable. To investigate the effects of imprecise precipitation input on streamflow predictions, we

performed a comparative Bayesian uncertainty analysis for the three global SGMP products, in which the parameters of the hydrologic model (VIC) and residual error model (CA-SEP) were calibrated jointly with each of the three precipitation products as model input. This led to three different parameter posterior distributions:

$$p(\theta_H, \theta_e | \tilde{X}^i, \tilde{Q}) \propto p(\tilde{Q} | \theta_H, \theta_e, \tilde{X}^i) p(\theta_H, \theta_e) \quad (7)$$

where \tilde{X}^i is the forcing data containing the i-th SGMP product, $p(\theta_H, \theta_e | \tilde{X}^i, \tilde{Q})$ is the i-th parameter posterior distribution. The i-th posterior distribution of the hydrologic model parameters θ_H was then used to simulate the streamflow with the i-th precipitation product, producing the i-th streamflow uncertainty interval. Moreover, by adding residual errors to the i-th streamflow uncertainty interval based on the i-th posterior distribution of the residual error model parameters θ_e , the i-th streamflow prediction interval was obtained. For better comparison, the parameter posterior distributions and the streamflow prediction interval calibrated with CMORPH CMA were used as benchmark.

3.3 Calibration method

A state-of-the-art Markov chain Monte Carlo (MCMC) algorithm called DREAM_(ZS) was used in this study to perform the Bayesian calibration and uncertainty estimation. DREAM_(ZS) is based on the original

DREAM (Differential Evolution Adaptive Metropolis) algorithm (Vrugt et al. 2008a, 2009), which runs multiple Markov chains simultaneously. In this study, five parallel chains were used and each chain contained 14,000 iterations. The last 1000 samples after each chain's convergence (altogether 5000 samples) were retained to estimate parameter posterior distribution and predictive performance. Both visual observations and R statistic calculations (Gelman and Rubin, 1992) were adopted to check the chains' convergence.

3.4 Bayesian model averaging (BMA)

Bayesian model averaging is a probabilistic scheme for model combination. A brief description of the essence of BMA is shown below. Consider a quantity y as the forecasted variable (here streamflow), D as the observations of the forecasted variable, and the $f = [f_1, f_2, \dots, f_K]$ as the K sets of considered predictions. According to the law of total probability, the posterior distribution of the BMA prediction can be represented as:

$$p(y | D) = \sum_{k=1}^K p(f_k | D) \cdot p_k(y | f_k, D) \quad (8)$$

where $p(f_k | D)$ is the posterior probability of prediction f_k , also known as the likelihood of prediction f_k being the correct prediction given the observation data, D . If we denote $w_k = p(f_k | D)$, w_k can be considered as

weight and $\sum_{k=1}^K w_k = 1$. $p_k(y|f_k, D)$ is the posterior distribution of y given prediction f_k and observations D , and is usually approximated by a normal distribution with mean $a_k + b_k f_k$ and variance σ^2 , where a_k and b_k are regression coefficients obtained through simple linear regression of y on f_k using the training data. The estimation of a_k and b_k can be viewed as a simple bias-correction process (Raftery et al., 2005).

The BMA weights w_k and variance σ^2 were estimated by maximizing the likelihood of occurrence of the observed data, y . It is easier to maximize the logarithm of likelihood function, which is defined as follows:

$$L(w_1, w_2, \dots, w_K, \sigma^2) = \sum_{t=1}^T \log \left(\sum_{k=1}^K w_k \cdot p(y_t | f_{kt}) \right) \quad (9)$$

In this study we assumed a heteroscedastic variance of the conditional forecast distribution as $\sigma_{kt}^2 = c \cdot f_{kt}$. The multiplier c applied to all predictions of the ensemble. The value of c along with the BMA weights were estimated by maximizing equation (9) using Markov chain Monte Carlo (MCMC) simulation with DREAM (Vrugt et al., 2008b). The BMA predictive mean is then given by

$$E(y | f_1, f_2, \dots, f_K, D) = \sum_{k=1}^K w_k (a_k + b_k f_k) \quad (10)$$

Finally, the BMA probabilistic ensemble predictions were generated following Raftery et al. (2005). The procedure involves (i) selecting the ensemble size M , (ii) generating a value of k from the numbers $[1, 2, \dots, K]$ with the probabilities $[w_1, w_2, \dots, w_k]$, (iii) drawing a replication of y from the conditional forecast distribution $p(y|f_k)$, and (iiii) repeating steps (ii) and (iii) M times to obtain M values of y for each time step. In this study, the generated BMA probabilistic prediction consisted of 5000 ensemble streamflow simulations, which had the same sample size as the single precipitation product forced streamflow prediction interval.

Inspired by the integrated Bayesian uncertainty estimator (IBUNE; Ajami et al., 2007) framework, this study merged the three streamflow uncertainty intervals using BMA to fully consider the input, parameter estimation, and model structural uncertainty. The detailed procedure is described as follows:

(1) Obtain posterior distribution of hydrologic model and residual error model parameters forced by each global SGMP product using DREAM_(zs).

(2) Select three streamflow ensembles from the streamflow uncertainty interval forced by each SGMP product at 5%, 50% and 95% quantiles, respectively, thus generating a nine member ensemble.

(3) Estimate the model weight and variance multiplier of each

ensemble member.

(4) Compute the final weight of each precipitation product by summing the weights for all ensemble members generated from the corresponding streamflow uncertainty interval.

(5) Generate the BMA probabilistic prediction of 5000 ensemble streamflow simulations.

(6) Assess the performance of BMA predictive mean and predictive uncertainty.

3.5 Verification methods

Both deterministic and probabilistic verification methods were applied in this study to comprehensively assess the predictive performance. Deterministic verification methods included Nash–Sutcliffe Coefficient of Efficiency (NSCE) and relative error (RE). Probabilistic verification metrics used to quantify reliability and forecast skill included probability integral transform (PIT) histogram, quantile-quantile (QQ) plot with the corresponding quantified metric π_{reliab} , the Brier skill score (BSS), and the continuous ranked probability skill score (CRPSS). On the other hand, the methods for quantifying sharpness included the cover rate (CR), and the d-factor. The detailed explanations and equations of the verification methods can be found in Appendix A.

4. Results

4.1 Effects of different precipitation products on calibration procedure

The final parameter ranges calibrated with the four SGMP products are shown in Fig. 2. It is obvious that using different precipitation products to calibrate the VIC model results in different parameter values. It is shown that GSMaP Gauge and CMORPH BLD yield the largest parameter uncertainty. For example, the interquartile ranges of the final parameter distributions exceed 25% and 15% of the initial parameter range for Dsmax and d3, respectively. In contrast, the final parameter distributions for TMPA and CMORPH CMA (except for the parameter Ds) are much narrower. Sun et al. (2016) has reported that CMORPH CMA is the most accurate among the four precipitation products and sometimes even performs better than the high quality ground gauge rainfall product. Correspondingly, less parameter uncertainty is associated with less error in precipitation for CMORPH CMA. The reason that CMORPH CMA produces large uncertainty for Ds may be partly explained by the residual error model used in this study. Because the parameter Ds is insensitive and has much less impact on the streamflow simulation than the other five parameters in the VIC model, it is the most likely to be affected by the interaction with residual error model parameters, which was also observed in Sun et al. (2017). As for GSMaP Gauge and CMORPH BLD, the errors

in precipitation are propagated into simulated streamflow and amplified through the interaction with the hydrologic processes. When these two are used to calibrate the VIC model, the parameter uncertainty increases. Interestingly, final parameter ranges calibrated with TMPA are narrow. The additive form of the model residual errors represent the model structural uncertainty, the model parameter uncertainty and the input uncertainty. Because of the interaction of different sources of uncertainty during the calibration process, the informativeness of different precipitation datasets for parameter estimation is varied. Therefore, it's inferred that TMPA contains more information to inform model parameters with the residual error model used in this study.

It is noted that the final parameter ranges vary significantly among input precipitation products except the parameter b . Both GSMaP Gauge and CMORPH BLD produce higher estimations of parameter D_{max} , with median values of 23 and 25, compared to the other two precipitation products (with median values of 5 and 3.7). Large D_{max} tends to cause a high baseflow produced by the bottom soil layer, thus having an impact on the simulated streamflow. In addition, GSMaP Gauge and CMORPH BLD provide lower estimation of d_2 and higher estimation of d_3 than the other two. The thicker bottom soil layer may lead to larger soil moisture capacity and hence increase baseflow, which has consistent effect with the large

estimation of D_{\max} . At the same time, thinner second soil layer results in smaller soil moisture capacity in the upper layer, and thus reduces the infiltration when rainfall falls on the surface.

The different calibrated parameter values will directly affect the streamflow predictions. As shown in Fig. 3, in general, the 90% uncertainty interval of CMORPH BLD overestimates the baseflow and produces higher peak flows than others because CMORPH BLD has higher estimation of D_{\max} and d_3 as well as lower estimation of d_2 . Based on the definition of the residual error model (Eq. (4)), the higher flow corresponds to higher residual error. In addition, the estimated residual error model parameters of CMORPH BLD also have the largest uncertainty. Therefore, the prediction band of CMORPH BLD is the widest after adding the additive residual errors. Similarly, the prediction band of GSMaP Gauge is also wider than those of TMPA and CMORPH CMA, especially at high flows. The prediction interval of TMPA displays similar magnitudes to that of CMORPH CMA, while the latter one is narrower. Sun et al. (2017) has demonstrated that the newly developed residual error model CA-SEP is capable of addressing the complicated heteroscedasticity over the Huaihe River basin. When CMORPH CMA was used as the input, CA-SEP produced accurate predictions with high reliability and effectively avoided the negative flows. It is worth mentioning that the prediction intervals

calibrated with the three global SGMP products could effectively avoid the negative flows, which further confirms the stability of CA-SEP over this basin.

As summarized in Table 2, the maximum-likelihood simulated streamflow calibrated with CMORPH CMA has the highest NSCE values of 0.802 and 0.51 in the calibration and validation periods, respectively. Although the calibrated parameters have relatively large uncertainty for CMORPH BLD, its maximum-likelihood simulated streamflow has the second highest NSCE and the best RE in the calibration period. Because CMORPH BLD overestimates the baseflow and its underestimation of peak flows is less severe than TMPA and CMORPH CMA, the overall average biases are smaller. GSMap Gauge forced simulation has larger NSCE than that of TMPA in the calibration period, but their difference decreases in the validation period.

4.2 Deterministic predictive performance

The daily mean streamflow of the posterior streamflow distribution forced by the different precipitation products and the BMA predictive mean are shown in Fig. 4. The mean flows for the four precipitation input have the same trend as their 90% streamflow uncertainty intervals shown in Fig. 3. The detailed evaluation scores for the mean flows are shown in Table 3. CMORPH BLD has the highest NSCE and best RE among the three global

SGMP products in the calibration period. Because BMA weights reflect the ensemble members' deterministic predictive performance over the training period (Duan et al., 2007; Raftery et al., 2005), CMORPH BLD receives the largest weight (Fig. 5). TMPA, however, receives the second largest weight despite having lower NSCE than GSMaP Gauge. The parameter posterior distributions of GSMaP Gauge and CMORPH BLD are very similar, their corresponding streamflow uncertainty intervals also have similar shape. Since GSMaP Gauge performs worse than CMORPH BLD, it could contribute little additional information for BMA predictions and receives the lowest weight. Similarly, the two members selected from each streamflow uncertainty interval at the 5% and 50% quantiles receive significantly lower weights than the member selected at the 95% quantile (Table 4).

In terms of NSCE and RE, the BMA predictive mean performs better than any individual deterministic prediction during both calibration and validation periods (Table 3). The NSCE values of the BMA predictive mean are even higher than those of the maximum-likelihood simulated streamflow calibrated with CMORPH CMA (Table 2). It can also be seen that the BMA predictive mean has the consistent trend with the observed flow and can reproduce both the peak flow and low flow very well (Fig. 4). The high agreement between the BMA predictive mean and the observed

flow is also reflected by the best RE, which is even zero in the calibration period.

4.3 Predictive uncertainty

Fig. 6 displays the 90% BMA probabilistic prediction intervals in the calibration and validation periods. Compared to the most reliable prediction band using CMORPH CMA as input, the BMA prediction band is narrower under low flow conditions, but a little bit wider at peak flows in the calibration period. In the validation period, it seems that the BMA prediction band becomes narrower at high flows and some flood peaks are not covered. To further evaluate the predictive performance, a set of probabilistic verification metrics are then applied.

Fig. 7 shows the PIT histograms for the different prediction intervals. It can be seen that the PIT histograms for the same probabilistic prediction in the calibration and validation periods show similar shape. The shape of PIT histograms for TMPA indicates their overestimation of the prediction uncertainty. For GSMaP Gauge and CMORPH BLD, more observations fall into the low probability bins than are expected by the forecasted distributions, indicating a positive bias for its prediction interval, which is consistent with the former result that GSMaP Gauge and CMORPH BLD tends to overestimate the streamflow. CMORPH CMA forced prediction interval also shows a little over dispersive in the calibration period, but gets

552 better in the validation period. The prediction interval for BMA shows the
553 most reliable performance with the flattest PIT histogram in the calibration
554 period. In the validation period, the more obvious U-shaped histogram for
555 BMA is an indication of under dispersion.

556 As expected, the quantile-quantile (QQ) plots (Fig. 8) for BMA are
557 close to the diagonal line, indicating its prediction intervals are reliable.
558 Despite a slight underestimation of the prediction uncertainty, the
559 performance of BMA is further improved in the validation period. The S-
560 shaped curvature of the QQ plots for TMPA and GSMaP Gauge suggests a
561 consistent overestimation of the prediction uncertainty. While the S-shaped
562 feature for TMPA remains unchanged, the QQ plot for GSMaP Gauge
563 shows a positive bias of prediction in the validation period. The QQ plots
564 for CMORPH CMA also feature systematic under-prediction, but less
565 pronounced in the validation period. As for CMORPH BLD, its QQ plots
566 indicate evident systematic over-prediction, which is more pronounced in
567 the validation period. It is noticed that the poor performance of GSMaP
568 Gauge and CMORPH BLD becomes worse in the validation period, when
569 the streamflow record is relatively dominated by low flows, interspersed
570 with high flow events. Therefore, the weakness of GSMaP Gauge and
571 CMORPH BLD for producing overlarge predictions is again highlighted.
572 To quantitatively analyze the departure of the QQ plot from the diagonal,

Table 5 reports the corresponding statistical metric π_{reliab} . Although the deterministic prediction of TMPA is inferior to the other two global SGMP products, the probabilistic prediction of TMPA is much better with the smallest π_{reliab} of 0.095 and 0.084 in the calibration and validation periods, respectively. BMA also shows high reliability of predictive distributions with small values of π_{reliab} , which is slightly lower than those of CMORPH CMA.

To further compare the skills of probabilistic predictions, Fig. 9 shows the BSS at different streamflow thresholds in both calibration and validation periods. In the calibration period, CMORPH CMA has the highest scores at low thresholds of 20% and 40%. The BSSs of TMPA and GSMaP Gauge also outperform that of BMA at the low threshold of 20%. BMA shows superior predictive skills over CMORPH CMA at high thresholds of 60% and 80%. The BSSs of the four precipitation products have an obvious deterioration in the validation period, especially at 20% threshold. The result of BMA, however, shows consistently high skills. CMORPH CMA outperforms BMA at 40% threshold though, BMA is evidently superior to the results of the four precipitation products at other thresholds. In addition, the BSSs of CMORPH BLD in both periods reinforce its lower predictive skills than others. Additionally, CRPSS shown in Table 5 can also reflect the reliability of the probabilistic

prediction. BMA has the highest CRPSSs in both calibration and validation periods, further indicating that combining individual predictions simulated from the global SGMP products through BMA is an efficient way to produce reliable probabilistic predictions.

In addition to the above evaluations focusing on the reliability of the probabilistic predictions, Table 5 further shows the coverage rate (CR) and the d-factor for evaluating the sharpness of the probabilistic predictions. Perfect predictive uncertainty would expect the CR close to the assumed 90% prediction level and the d-factor close to 1. It is clear that BMA probabilistic predictions leads to accurate CR values, with only slight overestimation in the calibration period (+1%) and underestimation in the validation period (-4.5%). The d-factors of BMA are also the closest to 1 in both calibration and validation periods, indicating the intervals would correspond to the standard deviation of the observations with good sharpness. All the three global SGMP products, especially CMORPH BLD, produce pretty wide prediction bands with CRs larger than 97% and d-factors larger than 2. The d-factor of CMORPH CMA is 1.37 in the calibration period, indicating modest overestimation of the observed variance in streamflow. In the validation period, the predictive bounds of the four precipitation products become wider with increased d-factors. By contrast, the d-factor of BMA further decreases during the validation

period, indicating much more stable performance.

Generally speaking, the prediction band of BMA performs the best in both reliability and sharpness. It is noted that the BMA predictions even outperform the predictions forced by a high quality precipitation input. Therefore, using BMA for ensemble modeling with multiple global SGMP inputs can considerably increase both accuracy and reliability in streamflow simulation results, which provides a promising approach to support hydrological modeling and water management in regions with sparse precipitation data.

5. Discussion

This study adopted the Bayesian framework to assess three global SGMP products. This is a general framework for evaluating the value of the precipitation data with consideration of the parameter uncertainty and model structural uncertainty. It could easily be applied using different hydrologic models, different residual error models and other precipitation products.

Through the calibration, CMOPRH BLD obtained higher estimation of parameter D_{\max} and d_3 , but lower estimation of d_2 , thus producing higher baseflow and surface runoff than GSMaP Gauge and TMPA. Therefore, the peak flows of CMORPH BLD were closer to the observations than the other two, which helped CMORPH BLD to acquire

higher NSCE. This is because NSCE places more emphasis on high flows due to the squaring of the errors between the simulated and observed data (Krause et al., 2005). However, higher flows along with higher parameter uncertainty of CMORPH BLD also resulted in large predictive uncertainty and unreliable predictive performance. Due to the tradeoff between the deterministic performance and the probabilistic performance, sometimes using a single satellite precipitation product may not be enough in hydrological application. Therefore, multiple precipitation ensembles are suggested for distributed hydrologic models (such as VIC), by taking advantage of varied performance on model calibration and streamflow predictions using different precipitation inputs.

This study used the BMA method to account for precipitation uncertainty in streamflow simulations over the Huaihe River basin of China. Currently there are many kinds of mainstream global satellite precipitation products with high spatiotemporal resolutions, the selection of BMA input could affect the BMA results. Since many studies have demonstrated that SGMP products have better performance than original satellite rainfall estimates in streamflow simulations (Maggioni and Massari, 2018; Mei et al., 2016; Wang et al., 2015), this study selected three mainstream global SGMP products. Involving the poorly performing members in the BMA scheme could even worsen the BMA predictive

performance (Chen et al., 2015). Besides, the BMA method would assign low weights to the members with poor performance (Raftery et al., 2005), and thus these member would have little impact on the BMA results. In this study, we also did an experiment to include another popular SGMP product, namely PERSIANN-CDR, as BMA input. PERSIANN-CDR only uses the Global Precipitation Climatology Project (GPCP) monthly product to adjust the bias at 2.5 degree monthly scale. The relatively rough bias adjustment process causes its worse performance than other precipitation products over the Huaihe River basin. Consequently, the large bias identified for PERSIANN-CDR was followed by unrealistic parameter estimation and large predictive uncertainty when using CA-SEP in the Bayesian framework. The BMA results with PERSIANN-CDR showed slightly worse performance (not shown) and the weight of PERSIANN-CDR was smaller than 0.01.

When applying the BMA method, some previous studies (Ajami et al., 2007; Strauch et al., 2012) compared the BMA prediction intervals with the uncertainty intervals associated with parameter uncertainty. However, such comparisons were not strict, because the spread of the ensemble streamflow simulations highly depends on parameter uncertainties. In our study, all the SGMP products generated very narrow uncertainty intervals. Therefore, the uncertainty intervals can be easily outperformed by BMA

prediction intervals. To properly estimate the predictive uncertainty corresponding to each precipitation input, an explicit model of the residual error, say, CA-SEP in this study, is also recommended.

The classical BMA approach assumes that the conditional PDF of each ensemble member is described with a normal distribution. The assumption of normal distribution may be inappropriate for streamflow primarily driven by precipitation, while the gamma distribution seems more reasonable. However, when we examined the normal and gamma distribution, the assumption of normality obtained more improvements for streamflow predictions. Vrugt and Robinson (2007) also found an improvement of BMA method for streamflow forecasting, under the assumption of normal distribution instead of the gamma distribution. Many studies also recommended using the Box-Cox transformation for the nonnormal distribution prior to the BMA process (Duan et al., 2007; Ma et al., 2018a). However, the Box-Cox transformation would not always guarantee the dataset after transformation is normally distributed. Since a lot of studies applying BMA to hydrological predictions still directly adopted the normal distribution for computational simplicity (Ajami et al., 2007; Kim et al., 2015; Roy et al., 2017; Strauch et al., 2012; Yen et al., 2014), and the focus of this study is not on improving the BMA approach, the normal distribution is acceptable. In the future, we will test the effect

of using the Box-Cox transformation and other conditional distributions on BMA performance.

There is still room for improvement in the BMA application. For example, stationary weights were assigned to the ensemble members in this study. Some studies have found that dynamically adaptive weights based on the nature of the forecasts and/or catchment states may have advantages for improving predictive performance (Devineni et al., 2008; Ma et al., 2018a, 2018b). In addition, some technical improvements have also been made on the BMA method itself (Madadgar and Moradkhani, 2014; Rings et al., 2012). Future works are needed to further improve the performance by fully exploring the advantage of the BMA method.

6. Conclusions

This study investigated the hydrologic value of three global SGMP products, namely TMPA, GSMaP Gauge, CMORPH BLD, for hydrologic model parameter calibration and streamflow predictions over the Huaihe River basin in China. A newly developed residual error model CA-SEP (Sun et al., 2017) was used to represent total uncertainty in streamflow predictions, and the Bayesian uncertainty analysis was performed based on this residual error model. In addition, the individual streamflow simulations selected from each streamflow uncertainty interval were combined using BMA. The regional high quality SGMP product

CMORPH CMA was also applied within the same framework to provide a benchmark for evaluation. The primary conclusions are summarized as follows:

1. Using different precipitation products to calibrate the VIC model resulted in different parameter posterior values. The parameter posterior distributions calibrated with GSMaP Gauge and CMORPH BLD had larger uncertainty than those calibrated with TMPA. The final estimated parameter values directly affected the streamflow simulation results. The higher estimation of parameter Dsmax and d3, but lower estimation of d2 obtained by CMORPH BLD produced higher baseflow and surface runoff than GSMaP Gauge and TMPA, and further led to larger predictive uncertainty.
2. Among the three global SGMP products, CMORPH BLD forced deterministic simulations performed the best with the highest NSCE values of 0.685 and 0.436 accompanied with the smallest absolute RE values of 12% and 4.7%, in the calibration and validation periods, respectively. However, its corresponding probabilistic prediction bands were extremely wide with the worst reliability. On the contrary, TMPA forced deterministic simulations had the lowest NSCE values of 0.623 and 0.397 in the calibration and validation periods, respectively. However, its probabilistic

predictions were much better than the other two products.

3. The BMA predictive mean not only performed better than any individual deterministic prediction of the three global precipitation products, but also obtained higher NSCEs (0.859 and 0.639 for the calibration and validation periods, respectively) than those of the maximum-likelihood simulated streamflow calibrated with CMORPH CMA. It also fitted fairly well with the observed streamflow and captured most peak flows. In addition, BMA largely improved probabilistic streamflow predictions with the best reliability and sharpness compared to other prediction bands based on the various probabilistic verification methods.

Overall, the demonstrated advantages of using multiple global SGMP products ensemble should be transferable to other hydrologic models and other regions, especially in ungauged basins. BMA can be used as an efficient tool to improve the performance of both the deterministic and probabilistic streamflow predictions, hence better supporting water management and decision making.

Acknowledgments

The authors are very grateful to the Editors and two anonymous reviewers for their constructive comments which led to considerable improvements of the paper. The authors also would like to appreciate the

China Meteorological Administration (CMA) for providing CMORPH
CMA. The TMPA 3B42V7 data were obtained from
<https://pmm.nasa.gov/data-access/downloads/trmm>. The CMORPH BLD
data were obtained from
ftp://ftp.cpc.ncep.noaa.gov/precip/CMORPH_V1.0. The GSMaP Gauge
data were obtained from
<http://sharaku.eorc.jaxa.jp/GSMaP/guide.html#07>. This work is supported
by the National Key R&D Program of China (2017YFC1502703), the
National Natural Science Foundation of China (41675109), Jiangsu
Collaborative Innovation Center for Climate Change, and High
Performance Computing Center of Nanjing University.

Appendix A. Verification metrics for streamflow prediction

A1. Deterministic verification methods

Nash–Sutcliffe Coefficient of Efficiency (NSCE). NSCE measures
how well model simulations represent the observed data, with NSCE = 1
being the optimal value. The NSCE is expressed as:

$$NSCE = 1 - \frac{\sum_{i=1}^n (Q_i - \tilde{Q}_i)^2}{\sum_{i=1}^n (\tilde{Q}_i - \bar{Q})^2} \quad (A.1)$$

where n is the total number of events; Q_i and \tilde{Q}_i are the i th pairs of
simulated and observed streamflow; \bar{Q} is the mean value of the
observation.

Relative error (RE). RE measures the average tendency of the model simulations to over or under estimate the observed data, where low-magnitude values of RE are preferred. The RE is expressed as:

$$RE = \frac{\sum_{i=1}^n (Q_i - \tilde{Q}_i)}{\sum_{i=1}^n \tilde{Q}_i} \times 100\% \quad (\text{A.2})$$

A2. Probabilistic verification methods

Probability integral transform (PIT) histogram. The PIT histogram is analogous to the rank histogram. It's used to assess the consistency between the frequency of occurrence of observed values and the forecast probabilities (Bourdin et al., 2014). PIT histogram shows the distribution of PIT values over equally sized bins. PIT value is defined as:

$$PIT_t = F_t(O_t)$$

(A.3)

where O_t is the observation at time step t , F_t is the cumulative distribution function (CDF) of the probabilistic forecast at that time. In this study, we divide the interval $[0, 1]$ into 10 equally sized bins. For perfectly reliable forecasts, the PIT histogram will be approximately flat. If the PIT histogram is not flat, its shape can be used to diagnose problems with the predictive uncertainty. For example, dome-shaped PIT histogram indicates too large spread or uncertainty of the forecast. Note that a flat PIT histogram does not necessarily indicate a reliable forecast (Hamill, 2001).

Quantile-quantile (QQ) plot and π_{reliab} . The QQ plot describes how well probabilistic forecasts represent the uncertainty in observations (Laio and Tamea, 2007; Thyer et al., 2009). In the QQ plot, the set of cumulative distribution function (CDF) values of observed data within the predictive distribution is compared to the cumulative uniform distribution, $U[0, 1]$. If the curve matches the diagonal line, it means that the predictive distribution adequately captures the distributional properties of the observed data. Therefore, the quantitative reliability metric π_{reliab} can be derived by considering the difference between the QQ plot curve and the diagonal line. The π_{reliab} is defined as:

$$\pi_{reliab} = \frac{2}{n} \sum_{t=1}^n |F_U - F_{Q(t)}(\tilde{Q}_t)| \quad (A.4)$$

where F_U is the uniform CDF and $F_{Q(t)}$ is the predictive CDF at time step t . The π_{reliab} lies between 0 (perfect reliability) and 1 (worst reliability).

Brier skill score (BSS). The BSS measures the accuracy of probability forecasts relative to a climatological forecast given a threshold. It is calculated by comparing the Brier score (BS) of the probability forecasts to that of the climatological forecast. The BS is defined as the mean squared error between the predicted and observed probabilities for a set of events, which is defined as:

$$BS = \frac{1}{n} \sum_{j=1}^n (p_j - o_j)^2$$

(A.5)

where n is the number of events, $p_j, j=1,2,\dots,n$, is forecast probability, $o_j, j=1,2,\dots,n$, is the observed probability, while o_j equals 1 if the observation exceeds a selected threshold and 0 otherwise.

BSS of 0 indicates a forecast with skill similar to the climatology, while a forecast which is less (more) skillful than the climatology will result in negative (positive) skill score values. If BS_F denotes the forecast score, BS_{clim} is the scores of the climatological forecast of the same predictand. The BSS is defined as:

$$BSS = 1 - \frac{BS_F}{BS_{clim}} \quad (A.6)$$

Continuous ranked probability skill score (CRPSS). Similar to the BSS, the CRPSS is a skill score relative to the climatological forecast. The continuous ranked probability score (CRPS), generated using the cumulative distribution function (CDF) of the forecast and observed samples, is computed to compare how the distribution of an ensemble of forecasts to the observed value. It is sensitive to bias in the forecast values as well as variability (Hersbach, 2000). If n is the number of forecasts, x is the forecast variable, x_o is the observed variable, $F_i(x)$ is the i -th $i=1,2,3,\dots,n$, cumulative distribution function (CDF), and $H_i(x-x_o)$ is the Heaviside function, which has the value 0 when $(x-x_o)<0$ and 1 otherwise. The CRPS is defined as:

$$CRPS = \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{\infty} [F_i(x) - H_i(x - x_o)]^2 dx \quad (A.7)$$

The CRPSS is defined as:

$$CRPSS = 1 - \frac{CRPS_F}{CRPS_{clim}} \quad (A.8)$$

Cover rate (CR). The CR denotes the percentage of observations bracketed by the prediction interval. In this study, we use the 90% prediction interval based on the 5 and 95 percentiles. Therefore, a good calibration and predictive uncertainty is achieved when the CR is close to 90%.

d-factor. The d-factor represents the average width of the prediction interval and is defined as:

$$d - factor = \frac{\frac{1}{n} \sum_{t=1}^n (Q_{t,u} - Q_{t,l})}{\sigma_o} \quad (A.9)$$

where $Q_{t,u}$ and $Q_{t,l}$ are the upper and lower bounds of the 90% prediction interval, and σ_o represents the standard deviation of the observed streamflow. d-factor close to 1 is preferred.

References

- Ajami, N.K., Duan, Q., Sorooshian, S., 2007. An integrated hydrologic Bayesian multimodel combination framework: confronting input, parameter, and model structural uncertainty in hydrologic prediction. Water Resour. Res. 43, W01403.

- 862 Ashouri, H., Hsu, K.L., Sorooshian, S., Braitwaite, D.K., Knapp, K.R.,
863 Cecil, L.D., Nelson, B.R., Prat, O.P., 2015. PERSIANN-CDR: Daily
864 precipitation data record from multisatellite observations for
865 hydrological and climate studies. *Bull. Am. Meteorolog. Soc.* 96, 69–
866 83.
- 867 Bates, B.C., Campbell, E.P., 2001. A Markov Chain Monte Carlo scheme
868 for parameter estimation and inference in conceptual rainfall-runoff
869 modeling. *Water Resour. Res.* 37, 937–947.
- 870 Behrangi, A., Khakbaz, B., Jaw, T.C., AghaKouchak, A., Hus, K.,
871 Sorooshian, S., 2011. Hydrologic Evaluation of Satellite Precipitation
872 Products Over a mid-Size Basin. *J. Hydrol.* 397, 225–237.
- 873 Bitew, M.M., Gebremichael, M., 2011. Evaluation of satellite rainfall
874 products through hydrologic simulation in a fully distributed
875 hydrologic model. *Water Resour. Res.* 47.
- 876 Bourdin, D.R., Nipen, T.N., Stull, R.B., 2014. Reliable probabilistic
877 forecasts from an ensemble reservoir inflow forecasting system.
878 *Water Resour. Res.* 50, 3108-3130.
- 879 Box, G.E.P., Tiao, G.C., 1992. *Bayesian Inference in Statistical Analysis*.
880 John Wiley, New York.
- 881 Chen, Y., Yuan, W., Xia, J., Fisher, J.B., Dong, W., Zhang, X., Liang, S.,
882 Ye, A., Cai, W., Feng, J., 2015. Using Bayesian model averaging to

- 883 estimate terrestrial evapotranspiration in China. *J. Hydrol.* 528, 537-
884 549.
- 885 DeChant, C.M., Moradkhani, H., 2014. Hydrologic Prediction and
886 Uncertainty Quantification. In: Eslamian, S. (Ed.), *Handbook of*
887 *Engineering Hydrology: Modeling, Climate Change, and Variability*.
888 CRC Press, pp. 387–414.
- 889 Devineni, N., Sankarasubramanian, A., Ghosh, S., 2008. Multimodel
890 ensembles of streamflow forecasts: role of predictor state in
891 developing optimal combinations. *Water Resour. Res.* 44, W09404.
- 892 Dinku, T., Ruiz, F., Connor, S.J., Ceccato, P., 2010. Validation and
893 intercomparison of satellite rainfall estimates over Colombia. *J. Appl.*
894 *Meteorol. Climatol.* 49, 1004–1014.
895 <http://doi.org/10.1175/2009JAMC2260.1>.
- 896 Duan, Q., Ajami, N.K., Gao, X., Sorooshian, S., 2007. Multi-model
897 ensemble hydrologic prediction using Bayesian model averaging.
898 *Adv. Water Resour.* 30, 1371-1386.
- 899 Evin, G., Kavetski, D., Thyer, M., Kuczera, G., 2013. Pitfalls and
900 improvements in the joint inference of heteroscedasticity and
901 autocorrelation in hydrological model calibration. *Water Resour. Res.*
902 49, 4518–4524.

- 903 Evin, G., Thyer, M., Kavetski, D., McInerney, D., Kuczera, G., 2014.
904 Comparison of joint versus postprocessor approaches for hydrological
905 uncertainty estimation accounting for error autocorrelation and
906 heteroscedasticity. *Water Resour. Res.* 50, 2350–2375.
- 907 Falck, A.S., Maggioni, V., Tomasella, J., Vila, D.A., Diniz, F.L., 2015.
908 Propagation of satellite precipitation uncertainties through a
909 distributed hydrologic model: A case study in the Tocantins–Araguaia
910 basin in Brazil. *J. Hydrol.* 527, 943–957.
- 911 Gelman, A.G., Rubin, D.B., 1992. Inference from iterative simulation
912 using multiple sequences. *Stat. Sci.* 7, 457–472.
- 913 Hamill, T.M., 2001. Interpretation of rank histograms for verifying
914 ensemble forecasts. *Mon. Weather Rev.* 129, 550–560.
- 915 He, S., Guo, S., Liu, Z., Yin, J., Chen, K., Wu, X., 2018. Uncertainty
916 analysis of hydrological multi-model ensembles based on CBP-BMA
917 method. *Hydrol. Res.* <http://doi.org/10.2166/nh.2018.160>.
- 918 Hersbach H., 2000. Decomposition of the continuous ranked probability
919 score for ensemble prediction systems. *Weather Forecast.* 15, 559–
920 570.
- 921 Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C. T., 1999. Bayesian
922 model averaging: a tutorial. *Stat. Sci.* 14, 382–401.

- 923 Hong, Y., Hsu, K.L., Moradkhani, H., Sorooshian, S., 2006. Uncertainty
924 quantification of satellite precipitation estimation and Monte Carlo
925 assessment of the error propagation into hydrologic response. *Water*
926 *Resour. Res.* 42.
- 927 Hossain, F., Anagnostou, E.N., 2006. Assessment of a multi-dimensional
928 satellite rainfall error model for ensemble generation of satellite
929 rainfall data. *Geosci. Remote Sens. Lett.* 3, 419–423.
- 930 Hou, Y., Kakar, R.K., Neeck, S., Azarbarzin, A.A., Kummerow, C.D.,
931 Kojima, M., Oki, R., Nakamura, K., Iguchi, T., 2014. The global
932 precipitation measurement mission. *Bull. Am. Meteorol. Soc.* 95,
933 701–722.
- 934 Huffman, G.J., Adler, R.F., Bolvin, D.T., Gu, G.J., Nelkin, E.J., Bowman,
935 K.P., Hong, Y., Stocker, E.F., Wolff, D.B., 2007. The TRMM multi-
936 satellite precipitation analysis (TMPA): quasi-global, multi-year,
937 combined-sensor precipitation estimates at fine scales. *J.*
938 *Hydrometeorol.* 8, 38–55.
- 939 Jiang, S., Ren, L., Hong, Y., Yang, X., Ma, M., Zhang, Y., Yuan, F., 2014.
940 Improvement of multi-satellite real-time precipitation products for
941 ensemble streamflow simulation in a middle latitude basin in South
942 China. *Water Resour. Manag.* 28, 2259–2278.

- 943 Jiang, S., Ren, L., Hong, Y., Yong, B., Yang, X., Yuan, F., Ma, M., 2012.
944 Comprehensive evaluation of multi-satellite precipitation products
945 with a dense rain gauge network and optimally merging their
946 simulated hydrological flows using the Bayesian model averaging
947 method. J. Hydrol. 452–453, 213–225.
948 <http://doi.org/10.1016/j.jhydrol.2012.05.055>.
- 949 Jiang, S., Ren, L., Xu, C.Y., Liu, S., Yuan, F., Yang, X., 2018. Quantifying
950 multi-source uncertainties in multi-model predictions using the
951 Bayesian model averaging scheme. Hydrol. Res. 49, 954–970.
- 952 Joyce, R.J., Janowiak, J.E., Arkin, P.A., Xie, P., 2004. CMORPH: a
953 method that produces global precipitation estimates from passive
954 microwave and infrared data at high spatial and temporal resolution.
955 J. Hydrometeorol. 5, 487–503.
- 956 Kavetski, D., 2018. Parameter Estimation and Predictive Uncertainty
957 Quantification in Hydrological Modelling, in: Duan, Q.,
958 Pappenberger, F., Thielen, J., Wood, A., Cloke, H., Schaake, J. (Eds.),
959 Handbook of Hydrometeorological Ensemble Forecasting. Springer,
960 Berlin, Heidelberg, pp. 1–42.
- 961 Kidd, C.K., Kniveton, D.R., Todd, M.C., Bellerby, T.J., 2003. Satellite
962 rainfall estimation using combined passive microwave and infrared
963 algorithms. J. Hydrometeorol. 4, 1088–1104.

- 964 Kidd, C., Huffman, G., 2011. Global Precipitation Measurement. Meteorol.
965 Appl. 18, 334–353.
- 966 Kim, J., Mohanty, B.P., Shin, Y., 2015. Effective soil moisture estimate
967 and its uncertainty using multimodel simulation based on Bayesian
968 Model Averaging. J. Geophys. Res. – Atmos. 120, 8023–8042.
- 969 Krause, P., Boyle, D.P., Bäse, F., 2005. Comparison of different efficiency
970 criteria for hydrological model assessment. Adv. Geosci. 5, 89–97.
- 971 Kubota, T., Shige, S., Hashizume, H., Aonashi, K., Takahashi, N., Seto, S.,
972 Hirose, M., Takayabu, Y.N., Nakagawa, K., Iwanami, K., Ushio, T.,
973 Kachi, M., Okamoto, K., 2007. Global Precipitation Map using
974 Satelliteborne Microwave Radiometers by the GSMaP Project:
975 Production and Validation. IEEE Trans. Geosci. Remote Sens. 45,
976 2259–2275.
- 977 Kucera, P.A., Ebert, E.E., Turk, F.J., Levizzani, V., Kirschbaum, D.,
978 Tapiador, F.J., Loew, A., Borsche, M., 2013. Precipitation from Space:
979 Advancing Earth System Science. Bull. Am. Meteorol. Soc. 94, 365–
980 375.
- 981 Kuczera, G., Renard, B., Thyer, M., Kavetski, D., 2010. There are no
982 hydrological monsters, just models and observations with large
983 uncertainties! Hydrol. Sci. J. 55, 980–991.
984 <http://doi.org/10.1080/02626667.2010.504677>.

- 985 Li, X., Zhang, Q., Xu, C., 2012. Suitability of the TRMM satellite rainfalls
986 in driving a distributed hydrological model for water balance
987 computations in Xinjiang catchment, Poyang lake basin. *J. Hydrol.*
988 426, 28–38.
- 989 Li, Z., Yang, D., Gao, B., Jiao, Y., Hong, Y., Xu, T., 2015. Multiscale
990 hydrologic applications of the latest satellite precipitation products in
991 the Yangtze River basin using a distributed hydrologic model. *J.*
992 *Hydrometeorol.* 16, 407–426. [http://doi.org/10.1175/jhm-d-14-](http://doi.org/10.1175/jhm-d-14-0105.1)
993 [0105.1](http://doi.org/10.1175/jhm-d-14-0105.1).
- 994 Liang, X., Lettenmaier, D.P., Wood, E.F., Burges, S.J., 1994. A simple
995 hydrologically based model of land surface water and energy fluxes
996 for general circulation models. *J. Geophys. Res. – Atmos.* 99, 14415–
997 14428.
- 998 Liang, X., Lettenmaier, D.P., Wood, E.F., 1996. One-dimensional
999 statistical dynamic representation of subgrid spatial variability of
1000 precipitation in the two-layer variable infiltration capacity model. *J.*
1001 *Geophys. Res. – Atmos.* 101, 21403–21422.
- 1002 Liang, Z., Wang, D., Guo, Y., Zhang, Y., Dai, R., 2013. Application of
1003 Bayesian model averaging approach to multimodel ensemble
1004 hydrologic forecasting. *J. Hydrol. Eng.* 18, 1426-1436.

- 1005 Laio, F., Tamea, S., 2007. Verification tools for probabilistic forecasts of
1006 continuous hydrological variables. *Hydrol. Earth Syst. Sci.* 11, 1267–
1007 1277.
- 1008 Ma, Y., Hong, Y., Chen, Y., Yang, Y., Tang, G., Yao, Y., Long, D., Li, C.,
1009 Han, Z., Liu, R., 2018a. Performance of optimally merged
1010 multisatellite precipitation products using the dynamic Bayesian
1011 model averaging scheme over the Tibetan Plateau. *J. Geophys. Res. –*
1012 *Atmos.* 123, 814-834.
- 1013 Ma, Y., Yang, Y., Han, Z., Tang, G., Maguire, L., Chu, Z., Hong, Y., 2018b.
1014 Comprehensive evaluation of ensemble multi-satellite precipitation
1015 dataset using the dynamic Bayesian model averaging scheme over the
1016 Tibetan Plateau. *J. Hydrol.* 556, 634-644.
- 1017 Madadgar, S., Moradkhani, H., 2014. Improved Bayesian multimodeling:
1018 Integration of copulas and Bayesian model averaging. *Water Resour.*
1019 *Res.* 50, 9586-9603.
- 1020 Maggioni, V., Massari, C., 2018. On the performance of satellite
1021 precipitation products in riverine flood modeling: A review. *J. Hydrol.*
1022 558, 214-224.
- 1023 Maggioni, V., Vergara, H.J., Anagnostou, E.N., Gourley, J.J., Hong, Y.,
1024 Stampoulis, D., 2013. Investigating the applicability of error
1025 correction ensembles of satellite rainfall products in river flow

- 1026 simulations. J. Hydrometeorol. <http://doi.org/10.1175/JHM-D-12->
 1027 [074.1](http://doi.org/10.1175/JHM-D-12-074.1).
- 1028 Mantovan, P., Todini, E., 2006. Hydrological forecasting uncertainty
 1029 assessment: Incoherence of the GLUE methodology. J. Hydrol. 330,
 1030 368–381.
- 1031 McMillan, H., Jackson, B., Clark, M., Kavetski, D., Woods, R., 2011.
 1032 Rainfall uncertainty in hydrological modelling: An evaluation of
 1033 multiplicative error models. J. Hydrol. 400, 83-94.
- 1034 Mega, T., Ushio, T., Kubota, T., Kachi, M., Aonashi, K., Shige, S., 2014.
 1035 Gauge adjusted global satellite mapping of precipitation
 1036 (GSMaP_Gauge). General Assembly and Scientific Symposium
 1037 (URSI GASS), 2014 XXXIth URSI, Beijing.
- 1038 Mei, Y., Nikolopoulos, E.I., Anagnostou, E.N., Borga, M., 2016.
 1039 Evaluating satellite precipitation error propagation in runoff
 1040 simulations of mountainous basins. J. Hydrometeorol. 17, 1407-1423.
- 1041 Nikolopoulos, E.I., Anagnostou, E.N., Borga, M., 2013. Using high-
 1042 resolution satellite rainfall products to simulate a major flash flood
 1043 event in northern Italy. J. Hydrometeorol. 14, 171-185.
- 1044 Raftery, A., Gneiting, T., Balabdaoui, F., Polakowski, M., 2005. Using
 1045 bayesian model averaging to calibrate forecast ensembles. Mon.
 1046 Weather. Rev. 133, 1155–1175.

- 1047 Rings, J., Vrugt, J.A., Schoups, G., Huisman, J.A., Vereecken, H., 2012.
1048 Bayesian model averaging using particle filtering and Gaussian
1049 mixture modeling: Theory, concepts, and simulation experiments.
1050 Water Resour. Res. 48.
- 1051 Roy, T., Serrat-Capdevila, A., Gupta, H., Valdes, J., 2017. A platform for
1052 probabilistic Multimodel and Multiproduct Streamflow Forecasting.
1053 Water Resour. Res. 53, 376-399.
- 1054 Schoups, G., Vrugt, J.A., 2010. A formal likelihood function for parameter
1055 and predictive inference of hydrologic models with correlated,
1056 heteroscedastic, and non-Gaussian errors. Water Resour. Res. 46,
1057 W10531.
- 1058 Sikorska, A.E., Seibert, J., 2016. Value of different precipitation data for
1059 flood prediction in an alpine catchment: a Bayesian approach. J.
1060 Hydrol. <http://doi.org/10.1016/j.jhydrol.2016.06.031>.
- 1061 Sorooshian, S., Dracup, J.A., 1980. Stochastic parameter estimation
1062 procedures for hydrologic rainfall-runoff models: Correlated and
1063 heteroscedastic error cases. Water Resour. Res. 16, 430-442.
- 1064 Sorooshian, S., Hsu, K.L., Gao, X.G., Gupta, H.V., Imam, B., Braithwaite,
1065 D., 2000. Evaluation of PERSIANN system satellite-based estimates
1066 of tropical rainfall. Bull. Am. Meteorol. Soc. 81, 2035-2046.

- 1067 Stedinger, J.R., Vogel, R.M., Lee, S.U., Batchelder, R., 2008. Appraisal of
1068 the generalized likelihood uncertainty estimation (GLUE) method.
1069 Water Resour. Res. 44. <https://doi.org/10.1029/2008WR006822>.
- 1070 Strauch, M., Bernhofer, C., Koide, S., Volk, M., Lorz, C., Makeschin, F.,
1071 2012. Using precipitation data ensemble for uncertainty analysis in
1072 SWAT streamflow simulation. J. Hydrol. 414, 413-424.
- 1073 Sun, R., Yuan, H., Liu, X., Jiang, X., 2016. Evaluation of the latest
1074 satellite-gauge precipitation products and their hydrologic
1075 applications over the Huaihe River basin. J. Hydrol. 536, 302-319.
- 1076 Sun, R., Yuan, H., Liu, X., 2017. Effect of heteroscedasticity treatment in
1077 residual error models on model calibration and prediction uncertainty
1078 estimation. J. Hydrol. 554, 680-692.
- 1079 Thyer, M., Renard, B., Kavetski, D., Kuczera, G., Franks, S.W., Srikanthan
1080 S., 2009. Critical evaluation of parameter consistency and predictive
1081 uncertainty in hydrological modeling: A case study using Bayesian
1082 total error analysis. Water Resour. Res. 45.
1083 <https://doi.org/10.1029/2008WR006825>.
- 1084 Tian, Y.D., Christa, D.P., John, B.E., 2010. Real-time error reduction for
1085 satellite-based precipitation estimates. J. Hydrometeorol. 11, 1275-
1086 1285.

- 1087 Tong, K., Su, F., Yang, D., Hao, Z., 2014. Evaluation of satellite
1088 precipitation retrievals and their potential utilities in hydrologic
1089 modeling over the Tibetan Plateau. *J. Hydrol.* 519, 423–437.
- 1090 Ushio, T., Kubota, T., Shige, S., Okamoto, K., Aonashi, K., Inoue, T.,
1091 Takahashi, N., Iguchi, T., Kachi, M., Oki, R., Morimoto, T.,
1092 Kawasaki, Z., 2009. A Kalman filter approach to the Global Satellite
1093 Mapping of Precipitation (GSMaP) from combined passive
1094 microwave and infrared radiometric data. *J. Meteorol. Soc. Jpn.* 87A,
1095 137–151.
- 1096 Vergara, H., Hong, Y., Gourley, J.J., Anagnostou, E.N., Maggioni, V.,
1097 Stampoulis, D., Kirstetter, P.E., 2014. Effects of resolution of
1098 satellite-based rainfall estimates on hydrologic modeling skill at
1099 different scales. *J. Hydrometeorol.* 15, 593–613.
1100 <http://doi.org/10.1175/JHM-D-12-0113.1>.
- 1101 Vrugt, J.A., Robinson, B., 2007. Treatment of uncertainty using ensemble
1102 methods: Comparison of sequential data assimilation and Bayesian
1103 model averaging. *Water Resour. Res.* 43, W01411.
- 1104 Vrugt, J.A., ter Braak, C.J.F., Clark, M.P., Hyman, J.M., Robinson, B.A.,
1105 2008a. Treatment of input uncertainty in hydrologic modeling: doing
1106 hydrology backward with Markov chain Monte Carlo simulation.
1107 *Water Resour. Res.* 44. <https://doi.org/10.1029/2007WR006720>.

- 1108 Vrugt, J.A., Diks, C.G., Clark, M.P., 2008b. Ensemble Bayesian model
1109 averaging using Markov chain Monte Carlo sampling. *Environ. Fluid.*
1110 *Mech.* 8, 579-595.
- 1111 Vrugt, J.A., Ter Braak, C.J.F., Diks, C.G.H., Robinson, B.A., Hyman, J.M.,
1112 Higdon, D., 2009. Accelerating Markov chain Monte Carlo
1113 simulation by differential evolution with self-adaptive randomized
1114 subspace sampling. *Int. J. Nonlin. Sci. Num.* 10, 273–290.
- 1115 Wang, S., Liu, S., Mo, X., Peng, B., Qiu, J., Li, M., Liu, C., Wang, Z.,
1116 Bauer-Gottwein, P., 2015. Evaluation of remotely sensed
1117 precipitation and its performance for streamflow simulations in basins
1118 of the southeast Tibetan Plateau. *J. Hydrometeorol.* 16, 2577-2594.
- 1119 Wang, Z., Zhong, R., Lai, C., Chen, J., 2017. Evaluation of the GPM
1120 IMERG satellite-based precipitation products and the hydrological
1121 utility. *Atmos. Res.* 196, 151-163.
- 1122 Xie, P., Joyce, R., Wu, S., 2013. A 15-year high-resolution gauge–satellite
1123 merged analysis of precipitation. In: 27th Conf. on Hydrology, Austin,
1124 TX, Amer. Meteor. Soc. 3B.4
1125 <[https://ams.confex.com/ams/93Annual/webprogram/Paper220506.h](https://ams.confex.com/ams/93Annual/webprogram/Paper220506.html)
1126 [tml](https://ams.confex.com/ams/93Annual/webprogram/Paper220506.html)>.
- 1127 Yen, H., Wang, X., Fontane, D.G., Harmel, R.D., Arabi, M., 2014. A
1128 framework for propagation of uncertainty contributed by

parameterization, input data, model structure, and calibration/validation data in watershed modeling. *Environ. Modell. Softw.* 54, 211-221.

Yu, M.Y., Chen, X., Li, L.H., Bao, A.M., Paix, M.J., 2011. Streamflow simulation by SWAT using different precipitation sources in large arid basins with scarce rain gauges. *Water. Resour. Manag.* 25, 2669–2681.

Yu, J., Shen, Y., Pan, Y., Xiong, A., 2015. Comparative assessment between the daily merged precipitation dataset over China and the world's popular counterparts. *Acta Meteorol. Sin.* 73, 394–410 (in Chinese).

Zubieta, R., Getirana, A., Espinoza, J. C., Lavado, W., 2015. Impacts of satellite-based precipitation datasets on rainfall–runoff modeling of the Western Amazon basin of Peru and Ecuador. *J. Hydrol.* 528, 599–612.

Fig 1. The upper Huaihe River basin of the Bengbu hydrological station.

Fig 2. The calibrated parameter ranges of the VIC model for the four precipitation input within the initial parameter range (y-axis domain).

Fig 3. The 90% uncertainty intervals associated with estimated parameters (shown in blue) and 90% prediction intervals (shown in orange)

calibrated with (a) TMPA, (b) GSMaP Gauge, (c) CMORPH BLD and (d) CMORPH CMA in the calibration period.

Fig 4. The BMA mean flow and mean flow of the posterior streamflow distribution calibrated with the four precipitation products in the (a) calibration period and (b) validation period.

Fig 5. BMA weights for the three satellite-gauge precipitation products.

Fig 6. The 90% BMA probabilistic prediction interval in the (a) calibration period and (b) validation period.

Fig 7. PIT histograms for the four precipitation products and BMA in the calibration period (left column) and validation period (right column).

Fig 8. QQ plots for the four precipitation products and BMA in the (a) calibration period and (b) validation period.

Fig 9. Brier skill scores for the four precipitation products and BMA in (a) calibration period and (b) validation period. Thresholds are set using the observed streamflow values when the corresponding quantiles of observed streamflow CDF are 20%, 40%, 60% and 80%.

1. Three precipitation products are assessed using the Bayesian uncertainty analysis.

2. Precipitation uncertainty and model errors are modeled jointly with
a new error model.

3. Using multi-satellite precipitation ensemble with BMA improves
predictive performance.

Table 1. Annual basin average precipitation for each year during 2003–
2012.

		2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
TMPA	Mean(m m)	1345	938	1212	937	1058	1059	977	949	790	821
GSMaP Gauge	Mean(m m)	1283	871	1126	882	1005	951	890	933	752	758
CMORH BLD	Mean(m m)	1259	892	1181	896	1031	972	926	955	771	794
CMORH CMA	Mean(m m)	1338	901	1141	909	1049	956	912	995	806	774

Table 2. Comparison of maximum-likelihood simulated streamflow calibrated with the four precipitation products.

	Calibration		Validation	
	NSCE	RE (%)	NSCE	RE (%)
TMPA	0.623	-24.1	0.397	-26.5
GSMaP Gauge	0.681	-23.4	0.414	-19.0
CMORPH BLD	0.685	-12.0	0.436	4.7
CMORPH CMA	0.802	-22.3	0.510	-27.6

Table 3. Comparison of mean flows of the posterior streamflow distribution calibrated with the four precipitation products and the BMA mean.

	Calibration		Validation	
	NSCE	RE (%)	NSCE	RE (%)
TMPA	0.635	-22.1	0.394	-24.4
GSMaP Gauge	0.64	-22.7	0.366	-17.8
CMORPH BLD	0.741	-6.7	0.483	11.4
CMORPH CMA	0.781	-22.9	0.503	-28.0
BMA	0.859	0.0	0.639	-10.2

1189 Table 4. BMA weights of all members selected from the streamflow
 1190 uncertainty intervals corresponding to the three satellite-gauge
 1191 precipitation products

	5%	50%	95%
TMPA	0.0043	0.0063	0.238
GSMaP Gauge	0.006	0.009	0.061
CMORPH BLD	0.00027	0.00036	0.675

1192

Table 5. Statistical diagnostics of the 90% streamflow prediction intervals calibrated with the four precipitation products and the BMA 90% streamflow prediction interval

	π_{reliab}	CRPSS	CR (%)	d-factor
Calibration period				
TMPA	0.095	0.50	97.0	2.07
GSMaP Gauge	0.129	0.46	97.6	2.48
CMORPH BLD	0.227	0.47	97.8	2.90
CMORPH	0.138	0.62	95.0	1.37
CMA				
BMA	0.126	0.63	91.0	1.29
Validation period				
TMPA	0.084	0.47	97.7	2.53
GSMaP Gauge	0.164	0.44	99.3	3.35
CMORPH BLD	0.273	0.41	99.4	3.75
CMORPH	0.124	0.52	93.2	1.58
CMA				
BMA	0.113	0.57	85.5	1.21