The applicability of real-time flood forecasting correction techniques coupled with the Muskingum method

Ruixiang Yang, Baodeng Hou, Weihua Xiao, Chuan Liang, Xuelei Zhang, Baoqi Li and Haiying Yu

ABSTRACT

Improving flood forecasting performance is critical for flood management. Real-time flood forecasting correction techniques (e.g., proportional correction (PC) and Kalman filter (KF)) coupled with the Muskingum method improve the forecasting performance but have limitations (e.g., short lead times and inadequate performance, respectively). Here, particle filter (PF) and combination forecasting (CF) are coupled with the Muskingum method and then applied to 10 flood events along the Shaxi River, China. Two indexes (overall consistency and permissible range) are selected to compare the performances of PC, KF, PF and CF for 3 h lead time. The changes in overall consistency for different lead times (1–6 h) are used to evaluate the applicability of PC, KF, PF and CF. The main conclusions are as follows: (1) for 3 h lead time, the two indexes indicate that the PF performance is optimal, followed in order by KF and PC; CF performance is close to PF and better than KF. (2) The performance of PC decreases faster than that of KF and PF with increases in the lead time. PC and PF are applicable for short (1–2 h) and long lead times (3–6 h), respectively. CF is applicable for 1–6 h lead times; however, it has no advantage over PC and PF for short and long lead times, respectively, which may be due to insufficient training and increase in cumulative errors.

Key words | correction techniques, Muskingum method, Nash–Sutcliffe efficiency, permissible range, real-time flood forecasting

Ruixiang Yang
Baodeng Hou (corresponding author)
Weihua Xiao
Xuelei Zhang
Baoqi Li
State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin,
China Institute of Water Resources and Hydropower Research,
Beijing, China
E-mail: *houbaodeng@163.com*

Ruixiang Yang
Chuan Liang
Haiying Yu
State Key Laboratory of Hydraulics and Mountain River Engineering,
College of Water Resource and Hydropower,
Sichuan University,
Chengdu, China

Ruixiang Yang
Pearl River Comprehensive Technology and Network Information Centre of Pearl River Water Resources Commission,
Ministry of Water Resources,
Guangzhou, China

Haiying Yu
Engineering College,
Sichuan Normal University,
Chengdu, China

INTRODUCTION

Flood forecasting is a non-engineering measure for flood control development that resulted from the attempt to address the occurrence of flood disasters (Ryder 2009). Traditional forecasting (TF), based on a physically hydrological model using historical data, reflects only the general regularities of the forecasting area (Kan *et al.* 2017). TF errors may be generated and accumulated gradually for a variety of reasons, mainly including model structure errors, model parameter errors and the errors caused by water projects. These errors often increase over time and may eventually exceed the permissible range of precision specified by the standard or used in practice (Liu *et al.* 2015). With the development of automatic hydrological monitoring and information transmission technology, real-time flood forecasting is commonly achieved by proportional correction (PC) and Kalman filter (KF) to correct TF using real-time hydrological data (Calvo & Savi 2009; Liu *et al.* 2016). These two correction techniques have numerous advantages that improve the forecasting performance to some extent; for example, the PC approach boasts a low computational cost, and KF has an unbiased minimum variance (Ocio *et al.* 2017). However, some limitations, such as the short

doi: 10.2166/nh.2019.128

lead time of PC and the inadequate correction performance of KF, have been observed in practical applications, and the direct correction of TF values using PC may cause unexpected and unstable deviations (Perumal & Sahoo 2007).

Particle filter (PF), which has undergone considerable development in recent years, can be applied to any nonlinear and non-Gaussian system represented by a state-space model (Weerts & El Serafy 2006). Compared with the above correction techniques, PF updates only forecasting value weights and then maintains the actual forecasting values, which avoids the situations in which forecasting values exceed the physical range during the updating process (Peter & Luc 2009). Therefore, the researches on PF are valuable in hydrology. Since Moradkhani *et al.* (2005) applied PF to the evaluation of parameters and states in hydrological models, scholars have performed several studies on the same topic. Dechant & Moradkhani (2012) and Bi *et al.* (2015) used PF to estimate the discharge in two runoff models simultaneously (HyMod Model and the Sacramento Soil Moisture Accounting Model) and the soil moisture in a distributed variable infiltration capacity hydrological model, respectively. In the field of flood forecasting, Noh *et al.* (2014) and Xu *et al.* (2017) studied the combination of PF combined with distributed hydrological models to forecast short-term runoff and that of PF with a channel hydraulics model to construct a real-time probabilistic channel flood forecasting model, respectively.

Combination forecasting (CF) integrates the results of different models to obtain more stable and reliable results (Cloke & Pappenberger 2009). Many studies have reported that a single model can produce large forecasting uncertainty; accordingly, a multi-model approach has been shown to produce better results (Zsoter *et al.* 2016). Thus, CF has been increasingly utilized by forecasters; examples of this technique generally include the simple averaging method and the weighted averaging method (Ajami *et al.* 2006). In addition, Wu et al. (2015) applied a real-time correction CF to a case study of the Xiangjiaba station on the Jinsha River.

For updating schemes in the Muskingum model, Mazzoleni *et al.* (2018) compared different updating approaches (direct insertion, nudging, KF, Ensemble KF and Augmented Ensemble KF) in the Muskingum model using different assumptions on structure and error definition of the model. The result is that in case of lumped modelling, direct insertion performs best while nudging performs worst; KF is sensitive to model error estimation; and Ensemble KF and Augmented Ensemble KF are sensitive to ensemble spread definition and computationally expensive. However, PF and CF applied for updating the Muskingan method has not been reported yet.

It has been more than 30 years since the flood forecasting was carried out from the Yongan hydrological station to the Shaxian hydrological station along the Shaxi River; and abundant experience has been accumulated for reference. In the study area, direct insertion performs worse than PC and KF; however, there are problems to be solved for them. To improve the short lead time and the unstable discharge hydrograph of PC, as well as the inadequate correction performance of KF for practical forecasting, 10 typical flood events with the total time intervals of 1,155 h extending from the two stations are selected, and two real-time flood forecasting correction techniques (PF and CF) are introduced in this study.

First, the flood characteristics and error distribution of TF are analysed. Then, real-time flood forecasting models are constructed based on the three real-time flood forecasting correction techniques (PC, KF and PF) that are individually coupled with the Muskingum method, and CF is developed via a combination of these three correction techniques. The 6 h average flood travel time is used as the lead time for TF in practice. However, a loss in the lead time may occur during the real-time correction (Bao *et al.* 2014). Therefore, 3 h is set as the lead time for practical real-time forecasting. According to the basin characteristics, the correction performances of PC, KF, PF and CF for 3 h lead time are compared via two indexes: overall consistency and permissible range to discuss the advantages and disadvantages of these techniques in detail. In addition, the influence of different lead times (1–6 h) on the correction performances of these correction techniques is obtained by the change in overall consistency with the lead time. Finally, the applicability of each of these correction techniques individually coupled with the Muskingum method is discussed.

MATERIALS AND METHODS

Study area and flood characteristics

The studied river reach extends from the upstream boundary at the Yongan hydrological station to the downstream

boundary at the Shaxian hydrological station, which is located in the middle and lower reaches of the Shaxi River, China (Figure 1). The average annual temperature and precipitation of the basin are 19.9 °C and 1,706.5 mm, respectively, and surface water is derived mainly from precipitation. The study area is 2,619 km^2, and the channel length is 78 km between the two stations. The section widths of the two stations are 186 and 242 m, respectively. The average flood travel time is 6 h, and there are five run-of-river hydropower stations with incomplete daily regulations. Ten typical flood events between the two stations, with observed time intervals of 1–5 h from 2002 to 2010, are analysed in this study. The first flood event occurred in 2002, the second and third flood events occurred in 2005, the fourth through sixth flood events occurred in 2006 and the remaining flood events occurred in 2010.

Based on the analysis of a total of 15 peaks from the 10 studied flood events between the two stations, 6 single-peak flood events, 3 double-peak flood events and 1 significant double-peak flood event with a hidden single-peak are observed. In addition, the relative size of a peak in the same double-peak flood event differs between the two stations. The peak times for five peaks observed at the Yongan hydrological station and almost all of the rise times occur later than the peak times and rise times observed at the Shaxian hydrological station. After the rainfall forecasting is implemented, the hydropower stations are pre-discharged before the flood to increase their storage capacity, which results in an early flood rise time observed at the Shaxian hydrological station. Additionally, the peak generated by the pre-discharge process occurs earlier than that of a natural flood, causing the peak time at the Shaxian hydrological station to occur earlier than that at the Yongan hydrological station. However, long-duration flood events cannot be completely controlled because of the incomplete daily regulation capacities at the five run-of-river hydropower stations. Consequently, the flood events observed at the Shaxian hydrological station still largely retain natural flood characteristics.

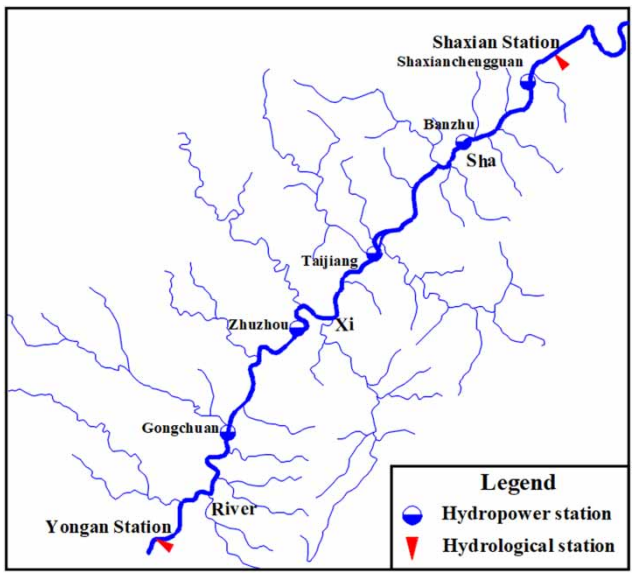

Figure 1 | Schematic diagram of the studied river reach from the Yongan hydrological station to the Shaxian hydrological station along the Shaxi River.

The flood forecasting error distribution during the calibration period (before 2002) using TF is analysed with the prior probability density function of PF. The mean error value is −156.20, and the standard deviation is 384.46. The error distribution is considered Gaussian, but it is rejected at the 0.05 significance level. Therefore, this distribution is considered to be noncompliant with the Gaussian distribution, which is similar to the results of Zhang *et al.* (2015). Considering the t location-scale distribution (representing the translation and expansion of the standard t distribution) as the error distribution, the probability density function is calculated by the following equation:

$$f(x) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sigma\sqrt{v\pi}\,\Gamma\left(\frac{v}{2}\right)}\left[1+\frac{\left(\frac{x-u}{\sigma}\right)^2}{v}\right]^{-\frac{v+1}{2}} \quad (1)$$

where u represents the location parameter; σ represents the scale parameter and v represents the degree of freedom. The maximum likelihood estimation results of the three parameters are $u = -130$, $\sigma = 287$ and $v = 4$, which are accepted at the 0.05 significance level. Therefore, the error distribution is considered to fit the t location-scale distribution. Figure 2 shows the error frequency histograms and the corresponding Gaussian and t location-scale fits.

Muskingum method and real-time correction techniques

A flood routing model using the Muskingum method is applied to TF in the study area (Zhao 1992). The basic

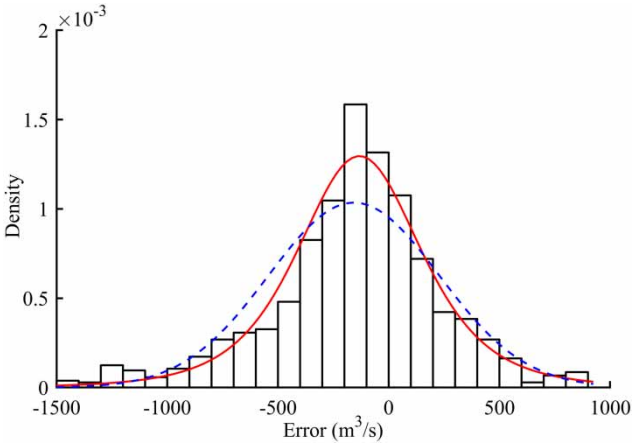

Figure 2 | Error distribution of TF in the calibration period. (The columns correspond to the distribution of error. The blue dotted line represents the Gaussian distribution. The red solid line represents the *t* location-scale distribution). Please refer to the online version of this paper to see this figure in colour: http://dx.doi.org/10.2166/nh.2019.128.

principle of the Muskingum method is to divide a river reach into n unit reaches. After $n-1$ iterations, the discharge of unit reach n is obtained. The Muskingum method is calculated as follows:

$$Q_{c,t+\Delta t}^{i+1} = C_0 Q_{c,t+\Delta t}^{i} + C_1 Q_{c,t}^{i} + C_2 Q_{c,t}^{i+1} \tag{2}$$

where i represents the unit reach i, $i = 1, \ldots, n-1$; t represents the time; Δt represents the time interval; c represents the forecasting value; Q represents the discharge (e.g., $Q_{c,t+\Delta t}^{i+1}$ represents the forecasting discharge at time $t+\Delta t$ for unit reach $i+1$); and C_1, C_2 and C_3 represent the model coefficients, which are estimated as follows:

$$C_0 = \frac{0.5\Delta t - kx}{0.5\Delta t + k - kx} \quad C_1 = \frac{0.5\Delta t + kx}{0.5\Delta t + k - kx} \quad C_2 = \frac{-0.5\Delta t + k - kx}{0.5\Delta t + k - kx} \tag{3}$$

where x represents the weighting factor of the discharge; and k represents the storage constant, reflecting the flood travel time for the storage volume of a river course in a steady flow state. The TF discharge is obtained by adding the local inflow, which is derived from the related rainfall forecasting and rainfall-runoff model in practice, to the discharge of unit reach n.

PC is a real-time correction technology; this technique is used for practical forecasting in the study area. The main assumption is that the ratio of the observed value to the forecasting value is relatively stable over a short period of time. The ratio at the current time is equal to that at a lead time before the current time. Multiplying the forecasting value by the ratio achieves the correction value, and the correction value for unit reach n is as follows:

$$Q_{u,t+t_0} = Q_{c,t+t_0} \times \frac{Q_{o,t}}{Q_{u,t}} \tag{4}$$

where u represents the correction value; o represents the observed value and t_0 represents the lead time.

KF is a modern analytical technology for dynamic systems that mainly includes multivariable controls, optimal controls, and estimation and adaptive controls. To describe the linear dynamic process description, a state equation and a measurement equation are constructed with KF. The steps are as follows (Wu *et al.* 2013): (1) select the initial value; (2) use TF to calculate the forecasting value; (3) estimate the forecasting error covariance; (4) calculate the gain matrix; (5) estimate the filtering error covariance; (6) calculate the filtered value and (7) repeat steps (2)–(6) until the steps have been operated for all moments. Steps (3)–(5) are expressed in the following equations:

$$P_{t+t_0/t} = \Phi_{t+t_0/t} P_{t/t} \Phi_{t+t_0/t}^{T} + \Gamma_{t+t_0} O_{t+t_0} \Gamma_{t+t_0}^{T} \tag{5}$$

$$K_{t+t_0} = P_{t+t_0/t} H_{t+t_0}^{T} (H_{t+t_0} P_{t+t_0/t} H_{t+t_0}^{T} + R_{t+t_0})^{-1} \tag{6}$$

$$P_{t+t_0/t+t_0} = (I - K_{t+t_0} H_{t+t_0}) P_{t+t_0/t} \tag{7}$$

where P represents the error covariance matrix; Φ represents the state transition matrix; Γ represents the state noise distribution matrix; O represents the state noise covariance matrix; K represents the gain matrix; H represents the observation matrix; R represents the observation noise covariance matrix; and I represents the identity matrix. The correction value for unit reach n is as follows:

$$Q_{u,t+t_0} = Q_{c,t+t_0} + K_{t+t_0}(Q_{o,t} - H_{t+t_0} Q_{u,t}) \tag{8}$$

PF is an algorithm that uses the Monte Carlo algorithm to implement Bayesian estimation theory, which can be applied to any type of state-space model (Han & Li 2008). The main idea is to select a set of weighted and random

sample particles from the state space to achieve an approximation of the state probability density distribution and then replace the integral operation with the sample mean to obtain the minimum variance estimate of the state. The steps are as follows (van Leeuwen 2009): (1) particle initialization. This process considers the number of particles, which is set to N, and the weight is equal to $1/N$. The perturbation term is added for this step. (2) Particle weight updating. This process involves assuming that N particles can approximate the importance probability density function of the state values that is commonly expressed in practice with a prior probability density function. After obtaining the observed values, the weight of each particle is recalculated according to the difference between the particle and the observed value, where particles with small differences are assigned large weights. Then, the particle weights are normalized. (3) Resampling (i.e., copying the particles according to the weights). This process is commonly performed when the number of effective particles is less than the set threshold. The greater the weight is, the larger the number of copies. After resampling, the weight of each particle is equally set to $1/N$. (4) Calculate the filtered value. The state transfer equation of the system is used to forecast the state of each particle to obtain a forecasting value. Based on the forecasting value, the weight of the particle is updated using the same methods in step (2). (5) Steps (2)–(4) should be repeated until the steps have been performed for all moments. The correction value for unit reach n is as follows:

$$Q_{u,t+t_0} = \sum_{j=1}^{N} w^j Q^j_{u,t+t_0} = \sum_{j=1}^{N} w^j (Q^j_{c,t+t_0} + Q^j_{u,t} - Q_{o,t}) \tag{9}$$

where w represents the normalized weight of each particle and j represents a particle.

The multiple linear regression is used to construct weighted average CF. The time series are divided into training and validation periods (Qu *et al.* 2017). The weighted coefficients of each correction technique involved in CF are obtained by the least square method during the training period, and the CF value is obtained with the weighted average equation during the verification period (Chen *et al.* 2015). The correction value for unit reach n is as follows:

$$Q_{u,t} = \sum_{m=1}^{M} a^m Q^m_{u,t} \tag{10}$$

where a^m represents the weight of correction technique m and M represents the number of correction techniques employed.

Performance indexes

In this paper, the flood forecasting performance is evaluated with two indexes: the overall consistency and the permissible range. The overall consistency reflects the degree of consistency between the forecasting value and the observed value and is evaluated based on the Nash–Sutcliffe efficiency (NSE) and the precision grade (Fares *et al.* 2014). The NSE is calculated by the following equation:

$$\mathrm{NSE} = 1 - \frac{\sum_{t=1}^{T} (Q_{u,t} - Q_{o,t})^2}{\sum_{t=1}^{T} (Q_{o,t} - \bar{Q}_o)^2} \tag{11}$$

where T represents the total time interval and $\bar{Q}_o$ represents the average observed value. The precision grade is divided into four categories based on the corresponding NSE values: A ($\mathrm{NSE} \geq 0.90$), B ($0.90 > \mathrm{NSE} \geq 0.70$), C ($0.70 > \mathrm{NSE} \geq 0.50$) and D ($0.50 > \mathrm{NSE}$). The permissible range is the range of allowable error based on the application requirements for the forecasting results and the technique levels, reflecting the degree of similarity between the forecasting and observed values. The permissible range is divided into three indexes: a peak discharge permissible error of 15% with respect to the observed value, a peak time permissible error of 3 h and a process discharge permissible error of 10% with respect to the observed value expressed as the process pass rate.

Experimental setup

The observed data of 1–5 h time intervals are interpolated into a 1 h interval; and the time interval between the two next corrections is also 1 h. The flood forecasting begins with a 24 h warm-up period; that is, the forecasting begins to forecast without real-time correction for 24 h before the flood rise start time. The purpose of this step is to ensure

that the precipitation affecting the flood process can be considered, thereby making the channel discharge obtained by the Muskingum method more reliable.

Based on the forecasting scheme of the Hydrologic and Water Resources Survey Bureau, Sanming Branch, the initial parameters of the Muskingum method for TF are set as constants for the total reach ($K = 6$), the weighted factor of the total reach ($x = 0.4$), the routing time ($\Delta t = 1$ h) and the number of unit reaches ($n = 6$). Then, the unit reach parameters are $K_l = \Delta t = 1$ h and $x_l = -0.1$, which are determined according to the observations during the calibration period (before 2002). Ten typical flood events with a total time interval of 1,155 h are used for TF in this study. The same observations are then used to measure the performance of the correction methods.

PC does not have initial parameters to set. The initial parameters of KF include the discharge Q and matrixes P, O and R. In general, Q is set as the observed value, while the elements of P are set as large values, and O and R are determined according to the observations during the calibration period (before 2002). The initial parameters of PF include the number of particles and the resampling threshold. Theoretically, the greater the number of particles is, the better the correction performance; however, the calculation time increases considerably with greater numbers of particles (Snyder *et al.* 2008). Moreover, high resampling frequencies not only increase the computational costs but also reduce the particle diversity (Zuo 2013). Accordingly, one flood event is selected here to analyse the computing time of PF. The number of particles is set to 100, 1,000 and 10,000, and the required computation times are 0.9, 2.8 and 19.2 s, respectively. However, the change in the NSE is negligible, and the largest number of effective particles is greater than two-thirds of the total. Therefore, the number of particles is set to 1,000, and the resampling threshold is two-third of the number of particles.

RESULTS AND DISCUSSION

Correction performance for 3 h lead time

For 3 h lead time, TF results are corrected by the real-time correction techniques of PC, KF and PF. The performances of these techniques, in terms of the overall consistency and permissible range indexes, are shown in Table 1.

For the NSE index of the overall consistency, the NSE of PC is larger than that of TF (except for the second flood event), and the number of flood events with the highest NSE is 2. KF improves the NSE for all flood events to varying degrees, and the number of flood events with the highest NSE is 2 (1 parallel). The NSE of PF is also improved for all flood events, and the number of flood events with the highest NSE is 7 (1 parallel). For the precision grade of the overall consistency, the precision grade of 1 flood event is A and those of 5 flood events are grade B using TF; the precision grades of 4 flood events are A using PC, KF and PF; and the precision grades of 4, 5 and 6 flood events are B using PC, KF and PF, respectively. Compared with TF, these three correction techniques improve the performance significantly, and the use of PF improves the precision grades of all flood events to at least a grade of B. In summary, for the overall consistency index, the performance of PF obviously exceeds those of the other two correction techniques, followed by KF, while PC performs the worst.

For the peak discharge index of the permissible range, the qualified numbers of flood events using TF, PC, KF and PF are 12, 10, 13 and 14, respectively; the numbers of peaks with the absolute error percentages below 5% are 7, 4, 9 and 7, respectively; the numbers of peaks with percentages between 5 and 10% are 2, 3, 4 and 6, respectively. Therefore, the performances of KF and PF are similar in terms of the peak discharge index; KF performs slightly better, whereas PC performs even worse than TF. For the peak time index of the permissible range, the qualified numbers of flood events using TF, PC, KF and PF are 8, 12, 13 and 14, respectively, and the numbers of flood events with the minimum absolute error values are 4 (2 parallel), 5 (3 parallel), 6 (3 parallel) and 9 (6 parallel), respectively. Therefore, in terms of the peak time index, the performance of PF is optimal; none of the remaining correction techniques shows obvious advantages, but they perform better than TF. For the process discharge index of the permissible range, the numbers of flood events with qualified rates in the 75–90% range are 1, 2, 3 and 3 for TF, PC, KF and PF, respectively, and the numbers of flood events with qualified rates in the 60–75% range are 2, 6, 4 and 3, respectively, while the numbers of flood events with the highest qualified rates are

Table 1 | Correction performance of the correction techniques for 3 h lead time

	Overall consistency				Permissible range											
	NSE (precision grade)				Peak discharge error (%)				Peak time error (h)				Process pass rate (%)			
No.	TF	PC	KF	PF	TF	PC	KF	PF	TF	PC	KF	PF	TF	PC	KF	PF
1	0.88 (B)	0.97 (A)	0.95 (A)	0.96 (A)	−5.4	0.1	−3.5	−2.5	0	−2	−1	3	73	85	81	84
2	0.82 (B)	0.77 (B)	0.89 (B)	0.88 (B)	−8.9 11.8	54.0 7.4	−1.3 9.2	−6.9 9.6	2 2	−5 −1	0 0	2 1	53	69	74	56
3	0.73 (B)	0.77 (B)	0.85 (B)	0.89 (B)	−4.5 −4.4	12.6 6.1	2.6 −0.9	7.7 4.9	3 5	−3 3	−2 4	−2 3	44	60	60	70
4	0.68 (C)	0.77 (B)	0.72 (B)	0.79 (B)	−26.4	0.2	−4.7	−13.6	1	3	−2	3	57	54	57	54
5	0.97 (A)	0.97 (A)	0.98 (A)	0.98 (A)	1.5	4.5	3.1	3.6	4	3	3	3	82	81	86	82
6	0.83 (B)	0.95 (A)	0.92 (A)	0.93 (A)	−0.8	5.7	0.2	0.2	−1	−4	−3	−3	32	65	49	64
7	0.66 (C)	0.77 (B)	0.76 (B)	0.82 (B)	3.2	22.5	5.4	3.3	6	2	2	2	44	70	64	55
8	0.40 (D)	0.47 (D)	0.65 (C)	0.70 (B)	15.9	52.7	18.3	26.4	5	2	3	2	12	49	21	49
9	0.85 (B)	0.90 (A)	0.92 (A)	0.95 (A)	−11.3 −10.6	2.6 13.9	−4.6 −2.5	−0.6 −0.5	2 7	2 5	2 5	1 5	45	63	65	73
10	0.55 (C)	0.62 (C)	0.77 (B)	0.87 (B)	−0.6 18.3 2.6	40.6 24.6 13.8	26.6 −7.6 6.5	9.4 −9.1 6.2	8 7 0	−1 3 −1	−1 3 −1	0 2 0	65	69	76	84

1 (1 parallel), 4 (1 parallel), 3 (1 parallel) and 4 (1 parallel), respectively. Therefore, for the process discharge index, the performances of PC and PF are better than that of KF. In conclusion, for the permissible range index, the order of performance is the same as that for the overall consistency.

The discharge processes and performances of TF, PC, KF and PF for the 10 flood events are shown in Figure 3. Table 1 and Figure 3 show that the discharge hydrographs of TF and KF are more stable than those of PC and PF, which exhibit sudden discharge fluctuations. The best performance is observed for the first and fifth flood events, which is consistent with the typical discharge process. The error in the first peak of the second flood event using PC is excessively large. The observed values at time intervals 60–70 suddenly plateau after an abrupt increase; as a result, PC forecasting values are only one-half of the observed values, with an error of 54% after proportional upscaling. The performances for the third and tenth flood events using PC are similar to those of the second flood event. The duration of the fourth flood event is relatively short, and the local inflow peak occurs much earlier than the observed peak. Therefore, the forecasting values from the three correction techniques all oscillate near the peak, which creates a large forecasting error in the peak time. The eighth flood event resembles the fourth flood event to an extent. The difference between them is that the local inflow peak in the eighth flood event occurs later than the observed peak, and the fluctuation is also inconsistent, leading to the worst performances for all indexes among the tenth flood events both before and after correction. Due to the small local inflow during the sixth flood event, the forecasting values obtained using TF are all smaller than the observed values during the rising flood stage; additionally, as a consequence of the lag effect during the lead time, none of the three correction techniques match the fluctuation trend of observed values. The local inflow peak time of the seventh flood event does not correspond to the observed peak time; hence, the three correction techniques oscillate significantly near the observed values during the rising stage. In addition, the performance of the ninth flood event is similar to that of the seventh flood event.

Based on the flood events 1–8 used for training, the weights of PC, KF and PF are 0.1467, 0.0811 and 0.7655, respectively, which are verified with the evaluation results of the two indexes in the above section. The CF performance for the ninth to tenth flood events is shown in Table 2.

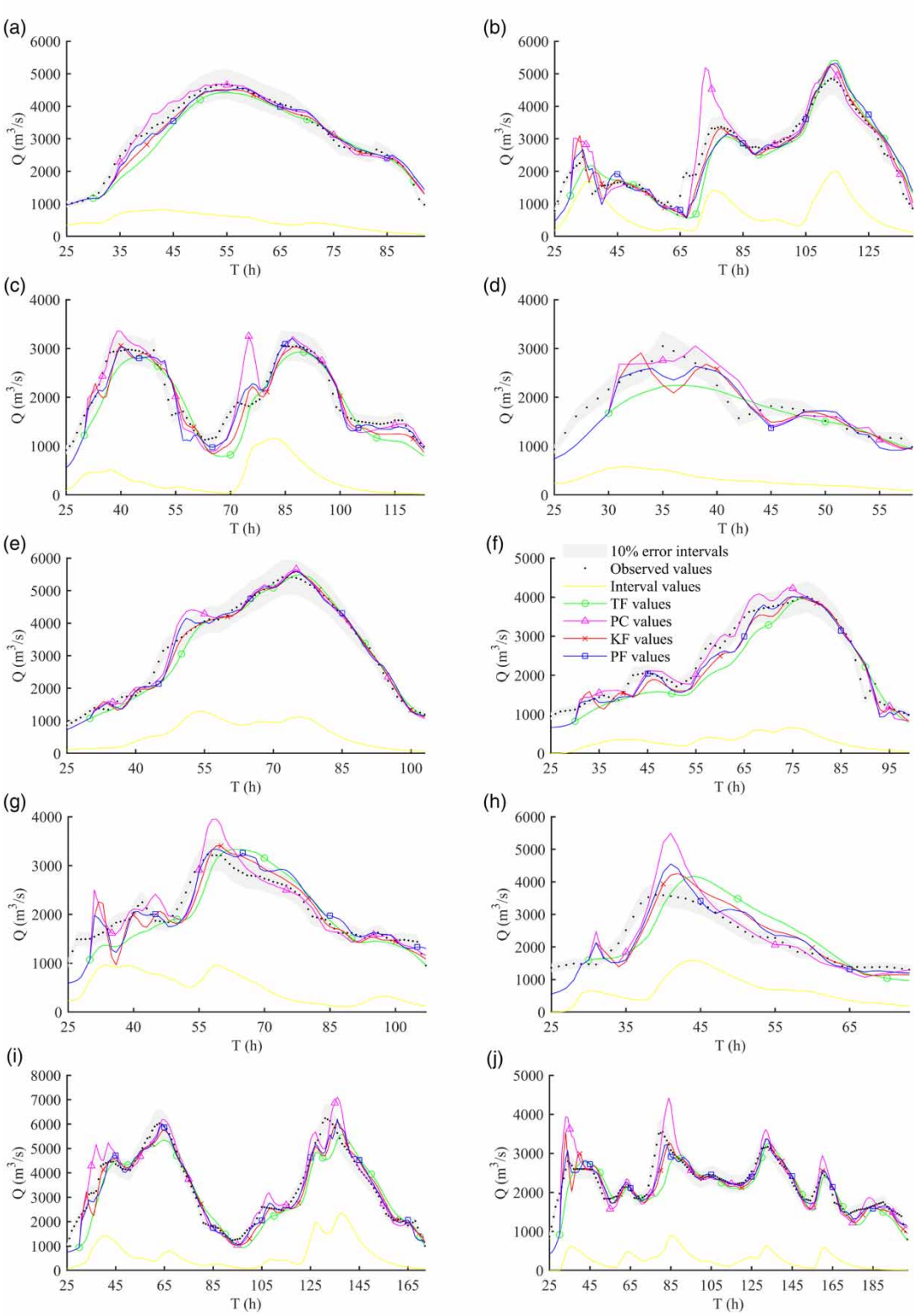

Figure 3 | Discharge processes for 3 h lead time.

Table 2 | Performance of CF for 3 h lead time

	Overall consistency		Permissible range					
	NSE (precision grade)		Peak discharge error (%)		Peak time error (h)		Process pass rate (%)	
No.	CF	Order	CF	Order	CF	Order	CF	Order
9	0.95 (A)	1* (1*)	−1.2 0.8	2 (1) 2 (1)	1 5	1* (1*) 1* (1*)	74	1 (2)
10	0.87 (B)	1* (1*)	12.7 −5.7 6.4	3 (2) 1 (3) 3 (2)	0 2 −1	1* (1*) 1* (1*) 3* (1*)	84	1* (1*)

Note: *Parallel; the number in parentheses: the order of PF.

Compared with PF, which boasts the best correction performance, the performance of the weighted average CF is not comprehensively better, and the differences among most indexes are not large. For the NSE index of the overall consistency, all 2 flood events are the top 1 (parallel). For the peak discharge index of the permissible range, the poor performance is influenced mainly by large peak discharge errors of PC. For the peak time index of the permissible range, the orders of 4 peaks are the top 1 (parallel), and the order of the remaining peak is the last (parallel); however, the absolute error is only 1 h. Both flood events account for more than 70% of the process discharge index of the permissible range. In total, although there is no direct comparison between CF and KF in Table 2, the performance of CF is clearly better than KF.

Correction performance for different lead times

For different lead times (1–6 h), the overall consistency index values of the three correction techniques and CF are shown in Table 3. However, comparing these three techniques individually is difficult because there are 10 flood events and 6 lead times. Therefore, the following three indexes are generalized: the slope of the average decrease in NSE, the number

Table 3 | NSE (precision grade) of the correction techniques for different lead times

LT (h) /	WT /	CT TF	1 0.88 (B)	2 0.82 (B)	3 0.73 (B)	4 0.68 (C)	5 0.97 (A)	6 0.83 (B)	7 0.66 (C)	8 0.40 (D)	9 0.85 (B)	10 0.55 (B)
1	0.7561	PC	0.99 (A)	0.97 (A)	0.97 (A)	0.95 (A)	0.99 (A)	0.99 (A)	0.97 (A)	0.94 (A)	0.99 (A)	0.96 (A)
	0.0245	KF	0.94 (A)	0.90 (A)	0.85 (B)	0.84 (B)	0.98 (A)	0.91 (A)	0.82 (B)	0.67 (C)	0.93 (A)	0.78 (B)
	0.2093	PF	0.99 (A)	0.94 (A)	0.98 (A)	0.96 (A)	0.99 (A)	0.98 (A)	0.89 (B)	0.81 (B)	0.95 (A)	0.90 (A)
	/	CF	/	/	/	/	/	/	/	/	0.99 (A)	0.97 (A)
2	0.5535	PC	0.98 (A)	0.90 (A)	0.89 (B)	0.85 (B)	0.98 (A)	0.96 (A)	0.91 (A)	0.77 (B)	0.95 (A)	0.84 (B)
	0.0526	KF	0.93 (A)	0.90 (A)	0.84 (B)	0.81 (B)	0.99 (A)	0.90 (A)	0.80 (B)	0.64 (C)	0.92 (A)	0.76 (B)
	0.3844	PF	0.96 (A)	0.90 (A)	0.92 (A)	0.86 (B)	0.99 (A)	0.93 (A)	0.89 (B)	0.70 (B)	0.96 (A)	0.84 (B)
	/	CF	/	/	/	/	/	/	/	/	0.97 (A)	0.90 (A)
3	0.1467	PC	0.97 (A)	0.77 (B)	0.77 (B)	0.77 (B)	0.97 (A)	0.95 (A)	0.77 (B)	0.47 (D)	0.90 (A)	0.62 (C)
	0.0811	KF	0.95 (A)	0.91 (A)	0.85 (B)	0.72 (B)	0.98 (A)	0.92 (A)	0.76 (B)	0.65 (C)	0.92 (A)	0.77 (B)
	0.7655	PF	0.96 (A)	0.88 (B)	0.89 (B)	0.79 (B)	0.98 (A)	0.93 (A)	0.82 (B)	0.70 (B)	0.95 (A)	0.87 (B)
	/	CF	/	/	/	/	/	/	/	/	0.95 (A)	0.87 (B)
4	0.0232	PC	0.94 (A)	0.58 (B)	0.62 (B)	0.70 (B)	0.96 (A)	0.93 (A)	0.65 (B)	0.13 (D)	0.84 (A)	0.35 (C)
	0.1267	KF	0.93 (A)	0.87 (B)	0.80 (B)	0.70 (B)	0.97 (A)	0.89 (B)	0.67 (C)	0.42 (D)	0.91 (A)	0.68 (B)
	0.8452	PF	0.95 (A)	0.90 (A)	0.82 (B)	0.73 (B)	0.97 (A)	0.94 (A)	0.74 (B)	0.53 (C)	0.93 (A)	0.82 (B)
	/	CF	/	/	/	/	/	/	/	/	0.93 (A)	0.82 (B)
5	0.0075	PC	0.92 (A)	0.32 (D)	0.46 (D)	0.59 (D)	0.94 (A)	0.91 (A)	0.44 (D)	−0.43 (D)	0.78 (B)	−0.03 (D)
	0.1354	KF	0.93 (A)	0.84 (B)	0.79 (B)	0.81 (B)	0.97 (A)	0.89 (B)	0.61 (C)	0.39 (D)	0.90 (A)	0.60 (C)
	0.8525	PF	0.94 (A)	0.86 (B)	0.84 (B)	0.76 (B)	0.96 (A)	0.92 (A)	0.67 (C)	0.38 (D)	0.91 (A)	0.77 (B)
	/	CF	/	/	/	/	/	/	/	/	0.91 (A)	0.77 (B)
6	0.0001	PC	0.89 (B)	0.02 (D)	0.31 (D)	0.56 (C)	0.93 (A)	0.89 (B)	0.32 (D)	−0.90 (D)	0.70 (B)	−0.42 (D)
	0.1517	KF	0.92 (A)	0.80 (B)	0.79 (B)	0.91 (A)	0.97 (A)	0.87 (B)	0.49 (D)	0.18 (D)	0.89 (B)	0.51 (C)
	0.8419	PF	0.90 (A)	0.85 (B)	0.82 (B)	0.76 (B)	0.95 (A)	0.90 (A)	0.65 (C)	0.30 (D)	0.88 (B)	0.56 (C)
	/	CF	/	/	/	/	/	/	/	/	0.89 (B)	0.59 (C)

Note: LT: lead time; WT: weight; CT: correction technique; /: not applicable.

of the largest NSE values and the number of grade A precision categorizations in each lead time.

When the lead time increases from 1 to 6 h, the NSE values of all three correction techniques decrease. For PC, the slope of the average decrease in NSE is −0.131, the number of largest NSE values decreases from 8 to 0 and the number of grade A precision categorizations decreases from 10 to 1; for KF, the respective categories are −0.026, from 0 to 4 and from 5 to 3; for PF, the respective categories are −0.036, from 4 (2 parallel) to 6 and from 8 to 3. In general, for the overall consistency index, the decline in the performance of PC is much faster than that for KF and PF. For 1 and 2 h lead times, the order of correction performance is PC, PF and KF, but for 3–6 h lead times, the order is PF, KF and PC.

When the lead time increases from 1 to 6 h, the NSE values of CF decrease; and the weights of PC decrease from 0.7561 to 0.0001, while those of KF and PF increase from 0.0245 and 0.2093 to 0.1517 and 0.8419, respectively. For 1 and 2 h lead times, the NSE values of CF are higher than those of PC, KF and PF; and the weights of PC are the largest and far greater than those of KF and PF. For 3–6 h lead times, the NSE values of CF are also the top 1 (but parallel); and the weights of PF are the largest and far greater than those of PC and KF. The weights also reflect that PC performs significantly better for 1–2 h lead time, while PF performs significantly better for 3–6 h lead time.

For NSE values of correction techniques lower than that of TF, when the proportion reaches 40% (i.e., the number reaches 4 for PC, KF and PF and 1 for CF), the correction technique is considered to be inapplicable during the corresponding lead time. For 3 h lead time, the number of flood events is 1 using PC, while for 4 h lead time, the number is 7; therefore, PC is applicable for only 1–3 h lead times. Using KF, the number for 5 h lead time is 2, while for 6 h lead time, the number is 4; therefore, KF is applicable in only 1–5 h lead times. Using PF, the number is only 3 for 6 h lead time; therefore, PF is applicable for all lead times (1–6 h). Using CF, the number is 0 for all lead times; therefore, CF is applicable for all lead times (1–6 h).

DISCUSSION

For 3 h lead time, for the peak discharge index, the peak discharges of 6, 15 (6 are the same as TF), 8 (6 are the same as TF) and 9 (5 are the same as TF) flood events obtained by TF, PC, KF and PF, respectively, are larger than the observed values. The reason is that the TF forecasting values are small and the forecasting errors are large when the observed values rise sharply with rising flood waters. However, the TF forecasting errors decrease approaching the flood peak because the rate of increase in the observed values will slow down. The correction values by PC are not involved in the subsequent calculations. When the forecasting error of TF is large at time t and small at time $t+t_0$, or vice versa, a large PC correction error is observed, especially for the peak. The similarity between KF and PF is that the correction value at time $t+t_0$ is involved in subsequent calculations. In addition, their correction performance is less affected by TF forecasting error, and the changes in their correction values are relatively stable over time for a flood event. Therefore, the positive and negative peak discharge errors of KF are essentially the same as those of TF. The changes in the correction values by PF are larger than that by KF because PF uses a large number of particles to approximate the true value. For the peak time index, the peak times by TF, PC, KF and PF of 12, 8, 7 and 11 h, respectively, occur later, indicating that the forecasted peak time is unfavourably delayed. Nevertheless, the change trend is not obvious and needs further study. For the process discharge index, PC has the largest number of events with the highest qualified rates. The correction performances of KF and PF are affected mainly by the gain matrix and the importance probability density function, respectively. The responses to sudden discharge fluctuations are slower when using KF and PF than when using PC, which leads to low qualified rates for process discharge forecasting. The large peak discharge error for CF is affected mainly by the large error in the peak discharge for PC. PC performs better in the process discharge index of the permissible range, and thus, the weight obtained by the least square method when considering this index is relatively large. However, the poor correction performance of PC in the peak discharge index combined with the large weight leads to the above results.

For different lead times (1–6 h), the assumption of the same ratio between the times t and $t+t_0$ in PC is difficult to satisfy with an increase in the lead time; as a result, a decline is observed in the correction performance in

conjunction with poor results. Both KF and PF are applicable for long lead times. The difference between KF and PF lies in their different realization methods. The KF gain matrix, which can be regarded simply as a 'coefficient', is computed by an error covariance matrix; this 'coefficient' is multiplied by the error at time t and added to the forecasting value at time $t + t_0$ to obtain the correction value at time $t + t_0$. For PF, the particle weight corresponding to the forecasting value is calculated according to the error at time t. The error probability is obtained from the importance probability density function (i.e., the TF error distribution in the section 'Study area and flood characteristics'). In addition, the error probability is multiplied by the initial weight of $1/N$ and normalized to obtain the updated weight; and the correction value at time $t + t_0$ is obtained by multiplying the forecasting value at time $t + t_0$ by the updated weight and summing the results. Therefore, the correction performances of KF and PF are less affected by the lead times than that of PC.

One of the reasons why CF does not perform best for 3–6 h lead time is probably that the number of flood events used for training is too small. In addition, the cumulative errors of the eight flood events used for training by different correction techniques vary regularly with the increase in the lead time. When the lead time increases from 1 to 6 h, the cumulative errors of PC decrease from 7,610 to $-78{,}036\ m^3/s$, and those of KF decrease from 60,247 to $41{,}386\ m^3/s$, while those of PF increase from $-18{,}416$ to $37{,}492\ m^3/s$; as a result, the weighted average cumulative error increase from 3,376 to $37{,}834\ m^3/s$, which may be one of the reasons why CF does not perform well enough for 3–6 h lead time.

CONCLUSIONS

Real-time flood forecasting correction techniques play a considerable role in fully utilizing real-time observed hydrologic data, improving forecasting performance and creating strategies to optimize disaster prevention and better make decisions. However, there are some limitations when using PC and KF coupled with the Muskingum method; for example, PC suffers from a short lead time, and KF displays an inadequate correction performance. In this study, PF is introduced into the field of real-time flood forecasting and coupled with the Muskingum method. PC, KF and PF are applied to the real-time forecasting of 10 flood events from the Yongan hydrological station to the Shaxian hydrological station along the Shaxi River in China to compare their performances and evaluate their applicability. In addition, the performance of the weighted average CF method is studied. The main conclusions are as follows:

1. For 3 h lead time, the performance of PF in the overall consistency index, including the NSE value and the precision grade, is optimal, followed by those of KF and PC. For the permissible range index, the peak discharge performances using KF and PF are similar and those using PF are slightly superior; however, PC performs the worst among these techniques. The peak time performance using PF is better than that using the other two techniques, which exhibit similar performances. The process discharge performance of PC is slightly better than that of PF and much better than that of KF. Furthermore, the weighted average CF including the three correction techniques is constructed. The CF weights decrease in the order of PF, PC and KF, and its performance is similar to PF. Generally, the comprehensive performances of all the considered techniques (from adequate to poor) are in the order of PF, CF, KF and PC.
2. When the lead time increases from 1 to 6 h, the decline in the performance of PC is much faster than those of KF and PF, indicating that the influence of the lead time on PC is much larger than those on the other two techniques. The performance of PC is optimal for 1 and 2 h lead times, while the performance of PF is optimal for 3–6 h lead times. For 1 and 2 h lead times, the NSE values of CF are higher than those of PC, KF and PF; and the weights of PC are the largest and far greater than those of KF and PF. For 3–6 h lead times, the NSE values of CF are also the top 1 (but parallel); PC, KF, PF and CF are applicable for 1–3, 1–5, 1–6 and 1–6 h lead times, respectively. For short lead times (1 and 2 h), PC is applicable; its performance is the best except for CF, and the computation requires minimal time for accurate forecasting. For long lead times (3–6 h), PF is applicable because of its highest performance. In contrast, CF, with its long computation time, should not be

selected for short and long lead times. PC and KF lead times are increased by 3 and 1 h, respectively, by using PF. Furthermore, PF coupled with the Muskingum method can provide high-precision forecasting results for longer lead times to support flood control strategies.

The importance of flood peak discharge and peak time forecasting is self-evident, but their performance is not satisfactory. Therefore, improving the performance of real-time correction techniques for peak discharge and peak time forecasting should be the focus of future studies. Comparisons among extended KF, Ensemble KF and PF applied to nonlinear systems should be performed, and the application of PF in channels with longer flood travel times should also be further studied. In addition, the impacts of the methods on the uncertainty estimates and the uncertainties arising from the quality and amount of observed data should be considered. A greater number of flood events are needed to study the performance of CF and multimethod CF (i.e., different indexes should be evaluated by a combination of results from different forecasting methods), which represent two methods of forecasting using different techniques.

ACKNOWLEDGEMENTS

This research was funded by the National Key Research and Development Program of China (grant no. 2017YFC0404701) and the National Natural Science Foundation of China (grant nos 51779271 and 51509267). The authors thank the Hydrologic and Water Resources Survey Bureau, Sanming Branch for providing the hydrologic database.

REFERENCES

Ajami, N. K., Duan, Q., Gao, X. & Sorooshian, S. 2006 Multimodel combination techniques for analysis of hydrological simulations: application to distributed model intercomparison project results. *Journal of Hydrometeorology* **7** (4), 755–768.

Bao, W., Si, W. & Qu, S. 2014 Flow updating in real-time flood forecasting based on runoff correction by a dynamic system response curve. *Journal of Hydrologic Engineering* **19** (4), 747–756.

Bi, H., Ma, J. & Wang, F. 2015 An improved particle filter algorithm based on ensemble Kalman filter and Markov chain Monte Carlo method. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **8** (2), 447–459.

Calvo, B. & Savi, F. 2009 Real-time flood forecasting of the Tiber river in Rome. *Natural Hazards* **50** (3), 461–477.

Chen, L., Zhang, Y., Zhou, J., Singh, V. P., Guo, S. & Zhang, J. 2015 Real-time error correction method combined with combination flood forecasting technique for improving the accuracy of flood forecasting. *Journal of Hydrology* **521**, 157–169.

Cloke, H. & Pappenberger, F. 2009 Ensemble flood forecasting: a review. *Journal of Hydrology* **375** (3–4), 613–626.

DeChant, C. M. & Moradkhani, H. 2012 Examining the effectiveness and robustness of sequential data assimilation methods for quantification of uncertainty in hydrologic forecasting. *Water Resources Research* **48** (4), W4518.

Fares, A., Awal, R., Michaud, J., Chu, P. S. & Fares, S. 2014 Rainfall-runoff modeling in a flashy tropical watershed using the distributed HL-RDHM model. *Journal of Hydrology* **519** (Supplement), 3436–3447.

Han, X. & Li, X. 2008 An evaluation of the nonlinear/non-Gaussian filters for the sequential data assimilation. *Remote Sensing of Environment* **112** (4), 1434–1449.

Kan, G., He, X., Ding, L., Li, J., Liang, K. & Hong, Y. 2017 Study on applicability of conceptual hydrological models for flood forecasting in humid, semi-humid semi-arid and arid basins in China. *Water* **9** (10), 719.

Liu, J., Wang, J., Pan, S., Tang, K., Li, C. & Han, D. 2015 A real-time flood forecasting system with dual updating of the NWP rainfall and the river flow. *Natural Hazards* **77** (2), 1161–1182.

Liu, Z., Guo, S., Zhang, H., Liu, D. & Yang, G. 2016 Comparative study of three updating procedures for real-time flood forecasting. *Water Resources Management* **30** (7), 2111–2126.

Mazzoleni, M., Noh, S. J., Lee, H., Liu, Y., Seo, D., Amaranto, A., Alfonso, L. & Solomatine, D. P. 2018 Real-time assimilation of streamflow observations into a hydrological routing model: effects of model structures and updating methods. *Hydrological Sciences Journal* **63** (3), 386–407.

Moradkhani, H., Hsu, K., Gupta, H. & Sorooshian, S. 2005 Uncertainty assessment of hydrologic model states and parameters: sequential data assimilation using the particle filter. *Water Resources Research* **41** (5), W5012.

Noh, S. J., Rakovec, O., Weerts, A. H. & Tachikawa, Y. 2014 On noise specification in data assimilation schemes for improved flood forecasting using distributed hydrological models. *Journal of Hydrology* **519** (Supplement), 2707–2721.

Ocio, D., Le Vine, N., Westerberg, I., Pappenberger, F. & Buytaert, W. 2017 The role of rating curve uncertainty in real-time flood forecasting. *Water Resources Research* **53** (5), 4197–4213.

Perumal, M. & Sahoo, B. 2007 Limitations of real-time models for forecasting river flooding from monsoon rainfall. *Natural Hazards* **42** (2), 415–422.

Peter, S. & Luc, F. 2009 Assessing parameter, precipitation, and predictive uncertainty in a distributed hydrological model using sequential data assimilation with the particle filter. *Journal of Hydrology* **376** (3–4), 428–442.

Qu, B., Zhang, X., Pappenberger, F., Zhang, T. & Fang, Y. 2017 Multi-model grand ensemble hydrologic forecasting in the Fu river basin using Bayesian model averaging. *Water* **9** (2), 74.

Ryder, P. 2009 Flood forecasting and warning. *Meteorological Applications* **16** (1), 1–2.

Snyder, C., Bengtsson, T., Bickel, P. & Anderson, J. 2008 Obstacles to high-dimensional particle filtering. *Monthly Weather Review* **136** (12), 4629–4640.

van Leeuwen, P. J. 2009 Particle filtering in geophysical systems. *Monthly Weather Review* **137** (12), 4089–4114.

Weerts, A. H. & El Serafy, G. Y. H. 2006 Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models. *Water Resources Research* **42** (9), 123–154.

Wu, X., Xiang, X., Wang, C., Chen, X., Xu, C. & Yu, Z. 2013 Coupled hydraulic and Kalman filter model for real-time correction of flood forecast in the three gorges interzone of Yangtze river, China. *Journal of Hydrologic Engineering* **18** (11), 1416–1425.

Wu, J., Zhou, J., Chen, L. & Ye, L. 2015 Coupling forecast methods of multiple rainfall-runoff models for improving the precision of hydrological forecasting. *Water Resources Management* **29** (14), 5091–5108.

Xu, X., Zhang, X., Lai, R., Lai, R. & Zhang, Y. 2017 A real-time probabilistic channel flood-forecasting model based on the Bayesian particle filter approach. *Environmental Modelling & Software* **88**, 151–167.

Zhang, J., Chen, L., Singh, V. P., Cao, H. & Wang, D. 2015 Determination of the distribution of flood forecasting error. *Natural Hazards* **75** (2), 1389–1402.

Zhao, R. J. 1992 The Xinanjiang model applied in China. *Journal of Hydrology* **135** (1–4), 371–381.

Zsoter, E., Pappenberger, F., Smith, P., Emerton, R. E., Dutra, E., Wetterhall, F., Richardson, D., Bogner, K. & Balsamo, G. 2016 Building a multimodel flood prediction system with the TIGGE archive. *Journal of Hydrometeorology* **17** (11), 2923–2940.

Zuo, J. 2013 Dynamic resampling for alleviating sample impoverishment of particle filter. *IET Radar, Sonar & Navigation* **7** (9), 968–977.

First received 25 August 2018; accepted in revised form 20 October 2019. Available online 9 December 2019