

A robust real-time flood forecasting method based on error estimation for reservoirs

Dandan Shen^{a,b,*}, Weimin Bao^a and Peng Ni^c

^a College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China

^b School of Water Conservancy & Environment Engineering, Zhejiang University of Water Resources and Electric Power, Zhejiang 310018, China

^c Shanghai Municipal Engineering Design Institute (Group) Co., Ltd, Shanghai 200092, China

*Corresponding author. E-mail: dandanshen00@163.com

ABSTRACT

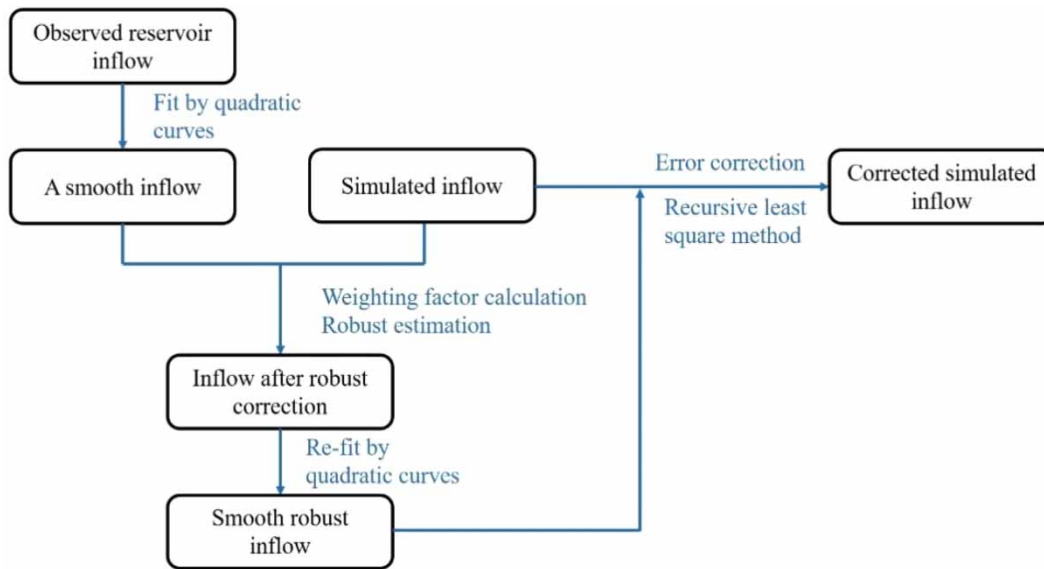
The observed discharge, an important input for flood forecasting systems, can significantly affect the accuracy of forecasting results. Since the reservoir inflow is not measured directly but calculated based on the observed reservoir stage and the reservoir outflow, it always contains gross errors causing some inflow to be outliers. In this study, a robust error estimation method for real-time flood forecasting has been developed for reservoirs' data processing. The method differs from the conventional flood forecasting method, in that it represses the gross errors by a robust loss function based on the real error distribution of measurements. Furthermore, a fluctuation coefficient has been proposed to quantify the degree of fluctuation of a jagged reservoir inflow. The performance of the method is evaluated by both synthetic data and real cases. By using floods generated synthetically by an ideal model in which the true values and errors are known, the method is shown to be efficient as theoretically expected. The results show that the method is efficient and universally applicable in different cases. And the degree of forecast improvement is positively related to the fluctuation coefficient of the floods. The more severe the fluctuation of the flood hydrograph, the more robust the method.

Key words: error distribution, flood forecasting, fluctuation coefficient, reservoir inflow, robust estimation

HIGHLIGHTS

- A robust error estimation method for real-time flood forecasting in reservoirs is developed.
- A fluctuation coefficient is proposed to quantify the degree of fluctuation of a jagged reservoir inflow.
- The method is demonstrated to be efficient by an ideal model and 10 reservoir basins of different characteristics.
- The performance of the proposed method is positively related to the fluctuation coefficient of the floods.

GRAPHICAL ABSTRACT



INTRODUCTION

In conventional real-time flood forecasting, parameter estimation and real-time flow updating are mostly carried out based on the residuals between measured discharge and discharge simulated by hydrological models (Bao & Zhao 2014). However, in modern forecasting systems, the discharge and precipitation are generally measured automatically by telemetering devices (Chao *et al.* 2008), which have the advantages of 24-h online monitoring, high precision, and labor-saving compared to the traditional measurement. Even so, the observed data inevitably contain some extreme errors and gross errors, which are irregular, unpredictable, and difficult to identify, especially in heavy rain. These errors have negative effects on the development of forecasting model, calibration of model parameters, and flow updating in real-time flood forecasting. Although there are many statistical tests in traditional least square estimation which can detect gross errors (Chan & Zhou 2009), the least square estimations are only unbiased, consistent, and valid when the residuals of observations follow the normal distribution. However, there are few observed data that strictly follow the normal distribution in real cases. There are always some extreme errors and gross errors in observed data due to some objective factors like instruments. The least square estimations are not valid for cases following a contaminated distribution (Chan & Zhou 2009). That is because the statistics used are calculated by the least square. Hence, it is almost impossible to separate the extreme errors and gross errors from observed data in the error estimation. And some gross errors are difficult to detect by the traditional methods in complex systems or multidimensional regression fitting problems, especially the gross errors in highly automated real-time flood forecasting systems.

In order to solve this problem, a new method known as the robust estimation method has been developed. It can detect and deal with gross errors better than the least square estimation method. However, it was not until the 1960s that substantial development was made to the robust estimation method when Tukey (1960) proposed the error distribution. After the development, the method can reduce the effects of gross errors with weighting factors based on the distribution of residuals. This robust method is different from the conventional recursive least square method, in which the residuals are transformed nonlinearly. Based on the robust theory, Zhou (1989) proposed the concept of equivalent weights to ameliorate the conventional least square algorithm. Then, Bao *et al.* (2003) introduced the robust estimation theory, which was well recognized in statistics, into hydrological forecasting. He pointed out that in applying robust estimation to hydrological forecasting, it is important to select a robustness function that is appropriate to the hydrological problem. Several widely used robust estimation methods, such as the Huber estimation method, the Tukey estimation method, and the IGG estimation method, were investigated and applied to the parameter estimation of the hydrological models (Bao *et al.* 2003). A recursive robust method for online estimation of the time-varying parameters of the auto-regressive updating model using a weighted least

square method with forgetting factors was developed (Chao *et al.* 2008). Moreover, based on the telemetry measured rainfall distribution characteristics and the robust estimation theory, a three-stepwise updating method was developed (Zhao *et al.* 2010). Furthermore, in the application of parameter estimation for the Muskingum Model (Shen *et al.* 2016), the robust least square algorithm was shown to be efficient and was able to handle the gross errors and extreme errors.

Hitherto, the robust estimation method has been mainly used for the parameter estimation of hydrological models (Finsterle & Najita 1998; Cizek 2001; Bárdossy & Singh 2008; Tsanov *et al.* 2020; Krisnayanti *et al.* 2021) and flood risk management (Mens *et al.* 2011; Lamond & Penning-Rowsell 2014; Kerich 2020). It has rarely been applied to real-time flood forecasting. For reservoir inflow, it is usually not measured directly but calculated based on the observed water stage, the reservoir outflow, and the stage-storage curve (Chao *et al.* 2008). The errors in the water surface observation may greatly reduce the accuracy of real-time updating in flood forecasting.

In this study, in order to improve the accuracy of flood forecasting in reservoirs and preserve the stability of the forecasting system, a robust real-time flood forecasting method based on error distribution has been developed. This new method takes into account the contaminated distributions of outliers in the observed data and suppresses the outliers by a robust loss function. Furthermore, a fluctuation coefficient has been proposed, which can describe the fluctuation of reservoir inflow caused by gross errors. With the fluctuation coefficient, the relationship between the degree of fluctuation of a flood and the performance of the robust real-time flood forecasting method can be analyzed quantitatively. By applying the method to floods generated by an ideal model, the efficiency and robustness of the method has been assessed. Furthermore, by applying the method to 10 reservoir basins with different characteristics and conditions, the superiority and applicability of the method has been verified. Additionally, the results indicate that the performance of the proposed method is positively related to the fluctuation coefficient of floods.

METHODOLOGY

Reservoir inflow and fluctuation coefficient

For a reservoir in which the observed inflow is not measured directly, it is instead calculated based on the reservoir water balance according to the observed reservoir stage and the reservoir outflow. The reservoir inflow can be described as

$$\frac{\Delta V}{\Delta t} = \bar{I} - \bar{O} - Z \quad (1)$$

where ΔV is the change in the reservoir storage over the period Δt , \bar{I} is the average reservoir inflow over Δt , \bar{O} is the average reservoir outflow over Δt , Z is the water loss from the reservoir over Δt , which includes evaporation, leakage, and other losses. If the water loss is negligible, $Z = 0$.

In general, floods are often accompanied by strong winds, which lead to gross errors in the reservoir stage measurements. Since ΔV is calculated based on the reservoir storage-stage curve which is positive related, there is an amplification of error in the calculation from the reservoir stage measurement to the reservoir inflow. This amplification is positively correlated to the water surface area of the reservoir and negatively correlated to the time period. Therefore, the errors in the reservoir inflow usually display an unknown distribution with zero mean and a large variance.

In fact, the degree of reservoir inflow fluctuation varies from basin to basin. It can even vary from flood to flood. Floods during non-flooding seasons are relatively stable, because the positive and negative errors generally offset each other. However, during flooding seasons with strong winds, the errors in the reservoir stage observation are significantly larger. As the water surface area varies with the reservoir stage, the errors in the reservoir stage observation can cause fluctuations in the reservoir inflow, resulting in extreme errors and gross errors. Figure 1 shows a typical example of fluctuations in the reservoir inflow, which includes negative inflow.

In this study, a fluctuation coefficient that can quantitatively describe the irregular fluctuations of reservoir inflow has been developed. Although, due to observation error, the hydrograph of a reservoir inflow generally looks like a jagged wave, drawing a smooth curve through the mid-points between two adjacent points can generally reflect the actual reservoir inflow approximately. Therefore, on a piecewise basis, a quadratic curve is fitted to a jagged reservoir inflow so as to transform the reservoir inflow into a relatively smooth curve. Then, the dispersion degree of the measured inflow from the fitted curve is

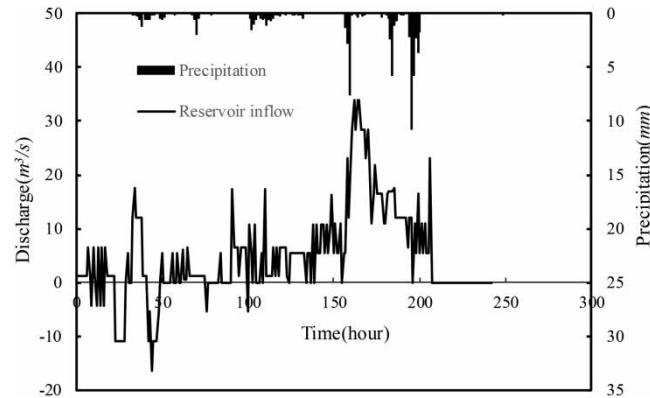


Figure 1 | Example of fluctuations in the reservoir inflow.

calculated, which describes the fluctuation of the reservoir inflow. The fluctuation coefficient is as follows:

$$\alpha = \sqrt{\frac{1}{m} \sum_{i=1}^m ((Q_o - Q_s)/Q_o)^2} \quad (2)$$

where α is the fluctuation coefficient, Q_o is the observed discharge, Q_s is the polynomial approximation of the inflow on the fitted curve, and m is the number of discharge observations.

Error distribution and robust estimation

In conventional estimation, the residuals are often assumed to follow a certain distribution (Tukey 1960). For example, in the least square algorithm, which is a typical parameter estimation method in hydrological forecasting, the residuals are often considered as random white noise. The least square algorithm then minimizes the sum of the squared residuals and gives all the residuals equal weightage. The residuals are always considered to follow the Gaussian distribution (Hampel 1973). However, in modern hydrological forecasting, observations are mostly automatically obtained by a telemetry system. The observed data, therefore, inevitably contain some extreme errors and gross errors. So, it is almost impossible for the residuals to follow a particular distribution closely. In fact, the residuals can follow a non-Gaussian distribution, which can be modeled as a mixture distribution of two fractions (Chan & Zhou 2009). The contaminated distribution often arises in real-time hydrological data, of which the probability density function can be expressed as follows (Tukey 1960):

$$J(\varepsilon) = [1 - \beta]h(\varepsilon) + \beta\eta(\varepsilon) \quad 0 \leq \beta < 1 \quad (3)$$

where $J(\varepsilon)$ is the actual distribution of the residual ε , α is the contamination rate, $h(\varepsilon)$ is the major distribution that a large portion of residuals follow, and $\eta(\varepsilon)$ is the interference distribution that outliers and extreme errors follow. For the major distribution $h(\varepsilon)$, it is often approximated to a normal distribution with zero mean and a small variance. For the interference distribution $\eta(\varepsilon)$, it is usually an unknown distribution with different and much bigger variance. In real-time flood forecasting, the α and $\eta(\varepsilon)$ are the unknowns. Therefore, the robustness estimation method first analyzes the actual distribution of residuals before searching for the optima. Furthermore, the search for the optima is carried out on the actual residual distribution and not on a hypothetical distribution. This is a fundamental difference between the robustness estimation method and the conventional estimation method (Chao *et al.* 2008).

Originating from the concept of robustness in statistics, the robust estimation method minimizes the influence of outliers on parameter estimation or error estimation as much as possible. With the remaining outliers, an appropriate estimation method is chosen so as to obtain an optimum parameter. There are two advantages in this approach. (1) When a hypothetical model is slightly different from an actual model, the performance of the estimation method is only slightly affected. There is stability in the estimation method. (2) When there are a number of outliers in the observed data, the estimated value is not severely affected. There is anti-interference in the estimation method.

The robustness estimation can be broadly classified into three types: (1) M-estimation (Chan & Zhou 2009), (2) L-estimation, and (3) R-estimation. The robust M-estimation is an extension of the classical maximum likelihood estimation. Due to the maturity of the theory and simple calculations, it is widely recognized and commonly used. Furthermore, it can be easily generalized to become a multidimensional parameter estimation method.

The least squares estimation method is susceptible to model errors and the influence of outliers. For the outliers that are far away from the sample observations, they have a greater influence on the residual sum of squares. On the other hand, if there is a function that can minimize the residual sum of squares (i.e. it is more robust than the quadratic function), then more valid and more robust estimates can be derived as compared to those by the conventional least square method. For this reason, the M-estimation has been used in this study. The maximum likelihood estimation criterion is expressed as follows:

$$\sum_{i=1}^n \ln f(X - Y_i) \rightarrow \max \tag{4}$$

where $Y = [Y_1, Y_2, \dots, Y_n]^T$ is the observation, X is the unknown vector, and $f(X - Y)$ is the distribution density function of the observation Y .

In order to make the equation general, a robust loss function $\rho(\cdot) = -\ln f(\cdot)$ was proposed in 1974 (Hampel 1974). This function has to provide high efficiency for the normal distribution portion in Equation (3), as well as for the outlier portion. Moreover, this function is symmetrical, convex, or non-decreasing on the positive axis. Thus, Equation (4) can be written in the following form:

$$\sum_{i=1}^n \rho(X - Y_i) \rightarrow \min \tag{5}$$

Equation (5) is the robust estimation criterion.

Furthermore, it is desirable that the extremal function's derivative $\varphi(\cdot) = \rho'(\cdot)$, namely the influence function, should be bounded and continuous. Boundedness ensures that the influence of a single observation on the estimation cannot be too large. Continuity ensures that rounding or quantization errors cannot have a major effect on the results. Thus, Equation (5) can be expressed as:

$$\sum_{i=1}^n \varphi(X - Y_i) = 0 \tag{6}$$

Furthermore, Equation (6) can also be written as:

$$\varphi(\varepsilon_i) = \frac{\partial}{\partial \varepsilon_i} \rho(\varepsilon_i) = \omega(\varepsilon_i) \cdot \varepsilon_i \tag{7}$$

where ε_i is the residual of the observation, and $\omega(\varepsilon_i) = \varphi(\varepsilon_i)/\varepsilon_i$ is the weighting factor. $\rho(\cdot)$, $\varphi(\cdot)$, and $\omega(\cdot)$ are collectively referred to as the robustness functions.

Many loss and influence functions with the desired properties have been proposed in the literature (Bao et al. 2003). In this study, a Huber estimator has been selected as follows:

$$\rho(\varepsilon_i) = \begin{cases} \varepsilon_i^2/2 & |\varepsilon_i| \leq k\sigma \\ k|\varepsilon_i| - k^2/2 & |\varepsilon_i| > k\sigma \end{cases} \tag{8}$$

$$\varphi(\varepsilon_i) = \begin{cases} \varepsilon_i & |\varepsilon_i| \leq k\sigma \\ k \cdot \text{sign}(\varepsilon_i) & |\varepsilon_i| > k\sigma \end{cases} \tag{9}$$

$$\omega(\varepsilon_i) = \begin{cases} 1 & |\varepsilon_i| \leq k\sigma \\ k\sigma/|\varepsilon_i| & |\varepsilon_i| > k\sigma \end{cases} \tag{10}$$

where k is a constant. According to the surveying error theory, $k = 1.5$ is a reasonable value. The efficiency of the Huber estimator with $k = 1.5$ is much higher than that of the conventional estimation methods for a residual probability density function

(Equation (3)) whose distribution is heavy-tailed non-Gaussian. Furthermore, its efficiency is the same as those of the conventional estimation methods for a probability density function whose distribution is Gaussian.

The effect of Equation (10) is to assign a smaller weight to a small portion of large residuals so that the impact of the outliers on the final estimates is less significant, while giving a unit weight to the bulk of small moderate residuals (Chao *et al.* 2008).

In addition, $\omega(\varepsilon_i)$ is dependent on σ , while σ is affected by $\omega(\varepsilon_i)$. So, $\omega(\varepsilon_i)$ has to be computed through iteration. In this study, the first iteration starts with $\omega(\varepsilon) = 1$, and σ is calculated by the following equation:

$$\sigma = \sqrt{\sum_{i=1}^t \omega_i \varepsilon_i^2 / (t - 1)} \tag{11}$$

Basic steps of the robust real-time flood forecasting method

In this study, by introducing the robust estimation to flood forecast, a robust real-time flood forecasting method has been developed in which the unwanted errors are assigned with weighting factors according to the error distribution theory. The calculation procedure of the robust real-time flood forecasting method is as follows:

1. Use a hydrological model to calculate the reservoir inflow, known as Q_{cal} . In this study, the Xinanjiang model, which was developed in 1973 by Zhao (Zhao 1992; Zhao *et al.* 1995) and improved by the researchers of Hydrological Forecast Research Institute in Hohai University (Bao *et al.* 2014; Si *et al.* 2015), has been used. The model consists of four components: (1) evapotranspiration, which consists of a three-layer soil moisture structure, (2) saturation excess runoff mechanism, (3) slope flow concentration, which consists of linear reservoirs, and (4) stream flow concentration, which uses the Muskingum method.
2. On a piecewise basis, fit the observed reservoir inflow by quadratic curves to generate a relatively smooth discharge curve. The number of sample points for each segment, known as smooth hour, is determined by the characteristics of each flood. The principle of segmentation is that the smooth hour should be long enough to eliminate the zigzag fluctuations of the reservoir inflow while keeping the overall contour of the flood process, especially the flood peak.
3. Calculate the residuals of the measurements and update the jagged inflow with the following equations:

$$\varepsilon(t) = Q_o(t) - Q_s(t) \tag{12}$$

$$Q_r(t) = \omega(t) \times Q_o(t) + (1 - \omega(t)) \times Q_s(t) \tag{13}$$

where Q_o is the observed discharge, Q_r is the discharge after the robust correction, and Q_s is the smooth discharge determined by Step 2. In Equations (12) and (13), $\omega(t)$ is calculated by Equations (10) and (11).

4. Re-fit the discharge after the robust correction, Q_r in Step 3, to generate a smooth robust discharge, known as Q_{rs} .
5. Based on the differences between Q_{rs} and the discharge Q_{cal} calculated by the hydrological model, use the recursive least square method to update the real-time flood.

Evaluation criteria

While the mean square deviation is commonly used to calculate the deviation of calculated discharge relative to measured discharge, this evaluation criterion works on the basis of equal weights to all the measurements, including the outliers. Actually, in evaluating the performance of a method, the participation of the outliers should be limited or completely prohibited. Therefore, it is necessary to select a more appropriate criterion to evaluate the performance of the method developed in this study. As such, a weighted mean square deviation, namely the robust mean square error, has been selected to evaluate the rationality and the validity of the robust real-time flood forecasting method. It has also been used in the further analysis of the relationship between the fluctuation coefficient and the robust effect. The robust mean square error reflects the deviation of calculated discharge relative to the effective measurements rather than all the measurements. Mathematically, it can be expressed as

$$V = \sqrt{\frac{1}{m - 1} \sum_{i=1}^m \omega_i (Q_{rs,i} - Q_{c,i})^2} \tag{14}$$

where ω_i is the weighting factor, $Q_{rs,i}$ is the smooth discharge after robust correction, $Q_{c,i}$ is the calculated discharge after real-time flood updating, and m is the number of discharge measurements.

Furthermore, the residual reduction rate E_v has been used to evaluate the performance of the developed method, which can be described as:

$$E_v = (V_{ex} - V_r) / V_{ex} \tag{15}$$

where V_{ex} is the robust mean square error without robust updating and V_r is the robust mean square error after robust updating.

RESULTS AND DISCUSSION

In order to investigate the validity and effectiveness of the developed method, it has been applied to floods generated by a theoretical model and in 10 selected reservoirs.

Ideal model

In this study, an ideal model has been used to synthetically generate floods. The reason for using the ideal model is that in an ideal model, the true values and the error structures are known and they can be varied to test different hypotheses. Ten floods without outliers from 1990 to 1997 at the Shaxian station in the Minjiang River in China have been used as the true values of discharge. Next, the following two error distributions have been added to the observed discharge data to generate discharge series with outliers that follow a non-Gaussian distribution:

$$\delta_i \sim N(0, \text{var}) \quad i = 1, 2, \dots, n \tag{16}$$

$$e_i = \begin{cases} (r - 0.5) \cdot \bar{Q} \cdot b & , i = \text{int}(i/T)T \\ 0 & , i \neq \text{int}(i/T)T \end{cases} \tag{17}$$

where δ_i is the random Gaussian error of low intensity and high frequency, e_i is the outlier of high intensity and low frequency that follows a non-Gaussian distribution, r is a random number of (0,1), \bar{Q} is the average of observed discharge, b is a constant that controls the maximum of e_i , and T is the frequency of the outliers. By adjusting b and T , the outliers of different

Table 1 | Fluctuation coefficient and robust mean square error for the floods with different error magnitudes in ideal model

Error magnitude (%)	α	V_{ex}	V_r	E_v (%)
10	0.034	7.29	4.03	44.69
20	0.070	14.46	7.82	45.89
30	0.110	22.27	11.56	48.07
40	0.133	28.65	14.79	48.38
50	0.179	35.11	17.65	49.72
60	0.217	41.32	19.62	52.51
70	0.255	48.90	24.07	50.78
80	0.272	53.42	24.07	54.94
90	0.328	69.24	24.70	64.32
100	0.386	72.89	32.39	55.57
110	0.414	101.31	37.96	62.53
120	0.415	106.25	33.93	68.07
130	0.467	114.69	32.30	71.84
140	0.512	113.27	35.39	68.76
150	0.495	117.64	31.72	73.04
160	0.529	120.73	38.10	68.45
170	0.557	112.51	33.10	70.58

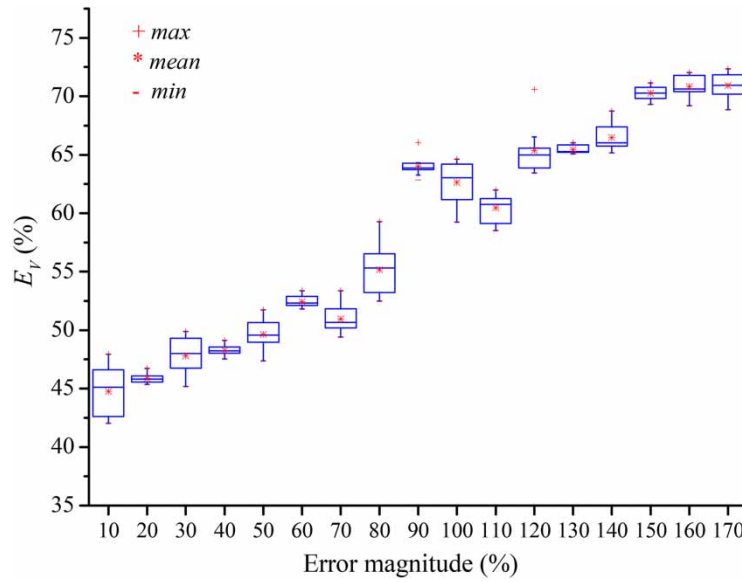


Figure 2 | Residual reduction rate of floods with different error magnitudes generated synthetically in the ideal model.

magnitudes and frequencies can be generated. The T and b can be randomly selected within a reasonable range according to the characteristic of floods. In this study, $T = 8$, and $b = 0.2 m$ ($m = 1, 2, \dots, 17$).

Table 1 shows the average fluctuation coefficient and the performance of the developed method for the 10 floods with different error magnitudes generated synthetically in the ideal model.

Based on the results in Table 1, two primary conclusions can be drawn:

1. Since the fluctuation coefficient increases with increasing error magnitudes, the fluctuation coefficient can quantitatively describe the degree of fluctuation of floods.
2. For the floods with different error magnitudes generated synthetically in the ideal model, the robust mean square error is reduced significantly after robust updating. This result shows that the robust real-time flood forecasting method is efficient and steady. The method can, therefore, be applied to real-life reservoirs.

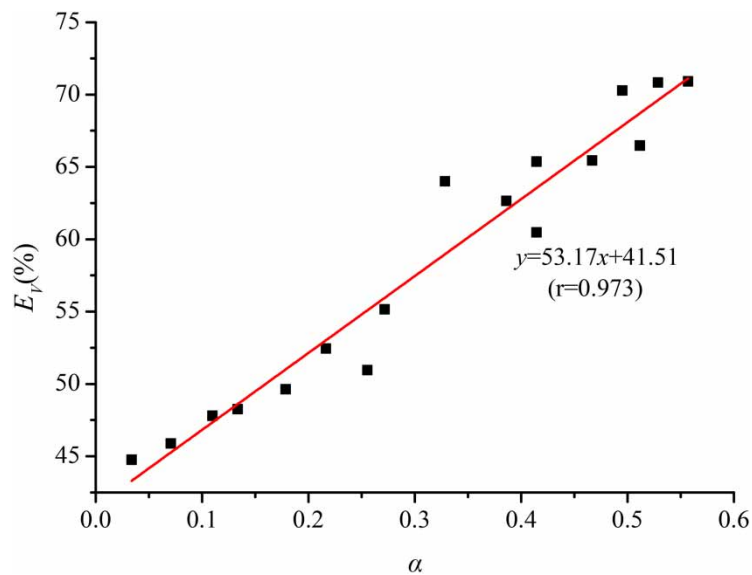


Figure 3 | Relationship between the average residual reduction rate and the average fluctuation coefficient of floods with different error magnitudes generated synthetically in the ideal model.

Table 2 | Characteristics of 10 selected reservoirs

Reservoir code	Basin name	Lat. N.	Long. E.	Climatic characteristics	Mean annual precipitation (mm)	Basin area (km ²)	Storage capacity (×10 ⁸ m ³)	Average water area (×10 ⁵ m ²)
1	Longjingshang	23°41'02.3"	116°03'38.1"	Subtropical monsoon humid climate	1,723	285	1.19	30.5
2	Qingshitan	25°30'50.6"	110°11'27.3"	Subtropical monsoon climate	1,800	474	6	150.6
3	Dongxi	27°47'04"	118°04'35"	Mid-subtropical monsoon humid climate	1,000	554	1.02	31.7
4	Dakai	23°24'38.3"	109°42'46.9"	South subtropical monsoon climate	1,600	427	4	144.9
5	Duihekou	30°31'32.7"	119°52'43.0"	Subtropical monsoon climate	1,100–1,150	148.7	1.16	36.3
6	Lishimen	29°03'41.9"	120°44'00.8"	Subtropical monsoon climate	1,895.4	296	1.99	34.7
7	Lushui	29°40'57.7"	113°56'40.9"	Subtropical maritime monsoon climate	1,550	3,400	7.06	351.2
8	Naban	22°08'11"	108°00'08"	South subtropical monsoon climate	1,715	490	8.32	148.6
9	Dongzhen	25°29'11.3"	118°56'28.0"	South subtropical maritime monsoon climate	1,200	321	4.35	108.8
10	Nanjiang	29°07'14.9"	120°26'45.0"	Subtropical monsoon humid climate	1,200–2,200	210	1.17	37.7

Figure 2 shows the residual reduction rates of the 10 floods with different error magnitudes generated synthetically in the ideal model. Figure 3 shows the relationship between the average residual reduction rate and the average fluctuation coefficient of the 10 floods. The results show that there is a linear relationship between the residual reduction rate and the fluctuation coefficient, and they are positively related.

Real cases

As a way to further investigate the efficiency and robustness of the developed method, 10 reservoirs of different characteristics have been selected. The selected reservoirs are distributed over different river systems in humid regions. Furthermore, the area, climatic conditions, and storage capacity of these reservoirs are very different from each other. The characteristics of the selected reservoirs are shown in Table 2, and the locations of the study areas are shown in Figure 4.

Table 3 shows the results of the simulations by the robust real-time flood forecasting method for the floods in the Qingshitan Reservoir. The results of the simulations for the floods in the other reservoirs are similar. The results show that the robust mean square errors of all the floods are significantly reduced after robust updating. From Table 3, it can also be seen that the simulation error of the runoff (e_{sx} and e_{kc}) is very small. This is an indication that the simulation results satisfy the water balance equation. Hence, the simulations by the robust real-time flood forecasting method are accurate.

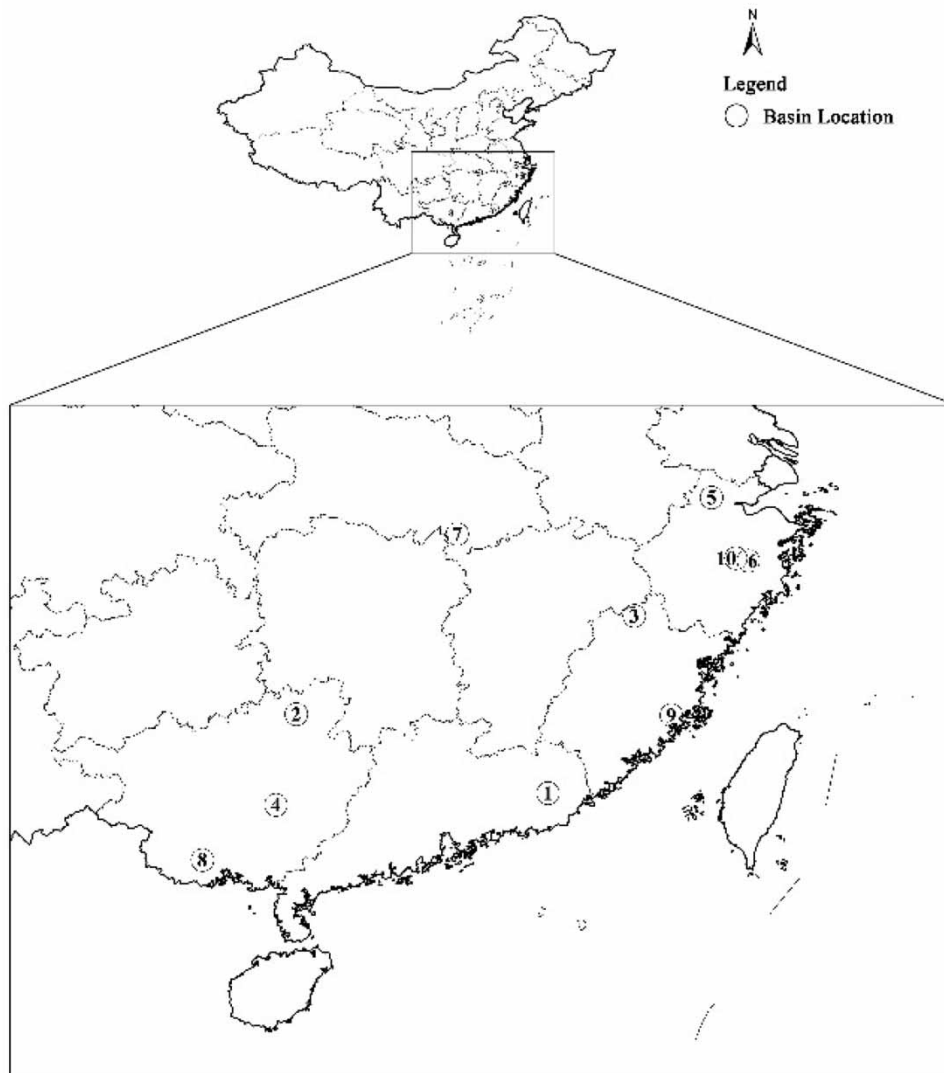


Figure 4 | Locations of 10 selected reservoirs.

Table 4 shows the average results of the method for the floods in the 10 selected reservoirs. It can be seen that the mean robust mean square errors of all the floods in different reservoirs are reduced after robust updating, and the average residual reduction rate is 20.97%. For a single flood, the reduction rate can reach 60% (e.g. Flood 850603 of the Lushui Reservoir). Moreover, the observations with significant robust effects are mostly close to the flood peaks, and for the floods with rapid confluence and precipitous fluctuation, the residual reduction rate is larger. This is because close to the flood peak, the fluctuation of the water surface is greater. Since the fluctuation phenomenon near the flood peak is more severe, the observed discharge data around the flood peak are more likely to contain outliers, especially for floods with rapid confluence and precipitous fluctuation. Based on the above analysis, the robust real-time flood forecasting method is rational and effective.

Figure 5 shows the relationship between the residual reduction rate and the fluctuation coefficient of the floods in the Qingshitan Reservoir. Figure 6 shows the statistical results for the floods in 10 selected reservoirs. From Figure 5, it can be seen that the larger the fluctuation coefficient, the higher the residual reduction rate. In other words, the robust effect is more

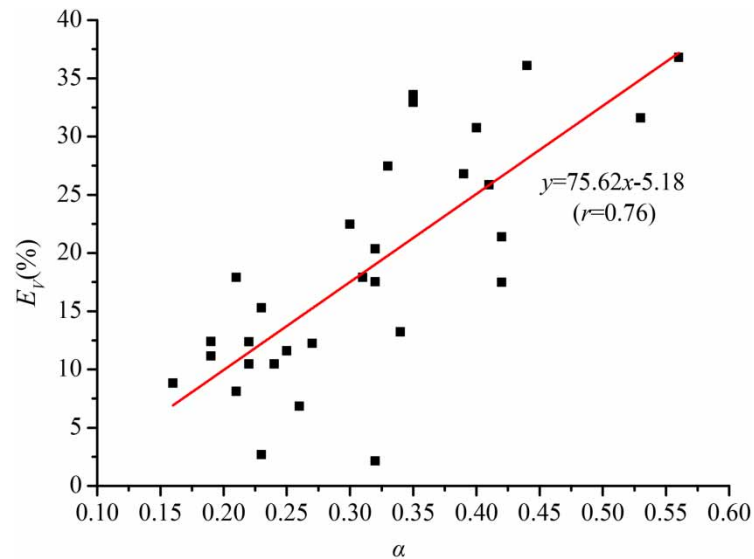
Table 3 | Results of the simulations by the robust real-time flood forecasting method for the floods in the Qingshitan Reservoir

Flood code	e_{sx} (%)	e_{kc} (%)	α	V_{ex} (m ³ /s)	V_r (m ³ /s)	E_v (%)
870629	0.26	-1.32	0.56	82.628	52.216	36.81
880622	3.56	3.51	0.23	40.250	39.173	2.68
890629	-3.36	-3.39	0.19	41.238	36.123	12.40
900529	4.61	4.10	0.35	37.200	24.942	32.95
900606	-0.62	-0.52	0.27	43.204	37.917	12.24
910706	1.55	1.40	0.44	25.995	16.612	36.10
920502	5.87	6.11	0.26	37.962	35.364	6.84
920702	0.88	0.72	0.32	44.972	37.085	17.54
930510	-0.22	-0.20	0.16	33.503	30.545	8.83
930611	8.73	8.05	0.42	55.941	43.969	21.40
930704	0.85	0.76	0.19	37.151	33.007	11.15
940521	4.06	4.26	0.22	38.696	34.644	10.47
940611	-3.09	-3.24	0.25	75.675	66.892	11.61
950524	-0.16	-0.23	0.32	70.316	68.806	2.15
950606	2.15	2.14	0.34	32.732	28.401	13.23
960623	-4.81	-4.85	0.22	58.725	51.456	12.38
960714	-7.73	-7.80	0.32	96.200	76.612	20.36
970512	4.74	4.06	0.33	29.356	21.298	27.45
970608	-1.53	-1.66	0.41	47.516	35.235	25.85
970703	-3.84	-3.69	0.24	39.941	35.765	10.46
980510	5.17	4.82	0.42	48.145	39.726	17.49
980520	-1.56	-1.45	0.23	81.555	69.082	15.29
980619	0.56	0.31	0.21	86.964	71.398	17.90
980723	5.49	5.27	0.21	54.351	49.937	8.12
000507	0.47	0.15	0.30	23.316	18.076	22.47
000523	0.22	-0.13	0.39	32.190	23.566	26.79
000607	-1.05	-1.04	0.35	33.351	22.144	33.60
000619	-3.99	-3.86	0.40	48.302	33.450	30.75
000815	3.60	2.69	0.53	20.390	13.944	31.61
Average	2.92	2.82	0.31	48.199	39.565	17.91

$$e = (R_0 - R_c)/R_0, \text{ where } R_0 \text{ is the measured runoff; } R_c \text{ is the calculated runoff.}$$

Table 4 | Average results for the floods in 10 selected reservoirs

Reservoir code	Basin name	α	V_{ex} (m ³ /s)	V_r (m ³ /s)	E_v (%)
1	Longjinshang	0.10	85.66	70.61	17.56
2	Qingshitian	0.31	48.20	39.57	17.91
3	Dongxi	0.13	21.35	16.33	23.50
4	Dakai	0.26	15.94	11.94	25.10
5	Duihekou	0.54	8.62	6.26	27.46
6	Lishimen	0.24	19.52	15.21	22.12
7	Lushui	0.37	112.91	73.84	34.61
8	Naban	0.44	72.00	61.31	14.85
9	Dongzhen	0.11	48.51	40.19	17.16
10	Nanjiang	0.12	19.89	18.02	9.39
	Average	0.26	45.26	35.33	20.97

**Figure 5** | Relationship between the residual reduction rate and the fluctuation coefficient of floods in the Qingshitian Reservoir.

significant. This phenomenon is also observed in the other nine reservoirs. Furthermore, from Figure 6, it can be seen that for different reservoirs, the robust performance of the method is positively correlated to the average fluctuation coefficient of floods in the reservoir. This is an indication that the method is robust.

Comparing Figures 4 and 5, three mentionable differences can be seen. First, the average residual reduction rates of the 10 floods with different error magnitudes generated by the ideal model are between 40 and 75%, while the residual reduction rates of the floods in the Qingshitian Reservoir are all smaller than 40%, illustrating that the performance of the method in an ideal model is overall better than in real cases. Second, the slope of the trend line in Figure 4 (i.e. 53.17) is smaller than that in Figure 5 (i.e. 75.62), suggesting that the method is more robust in the ideal model than in real cases. Third, the correlation coefficient of the trend line in Figure 4 (i.e. 0.973) is much larger than that in Figure 5 (i.e. 0.76). This is an indication that the linear correlation between the residual reduction rate and the fluctuation coefficient for the floods generated by the ideal model is more significant than that for the floods in the real case. Underlying the differences is that the true values, the errors, and the error structure are known in the ideal model, so the gross errors are easier to detect and locate as compared to the floods in real cases. Hence, the method performs better in the ideal model than that in real cases.

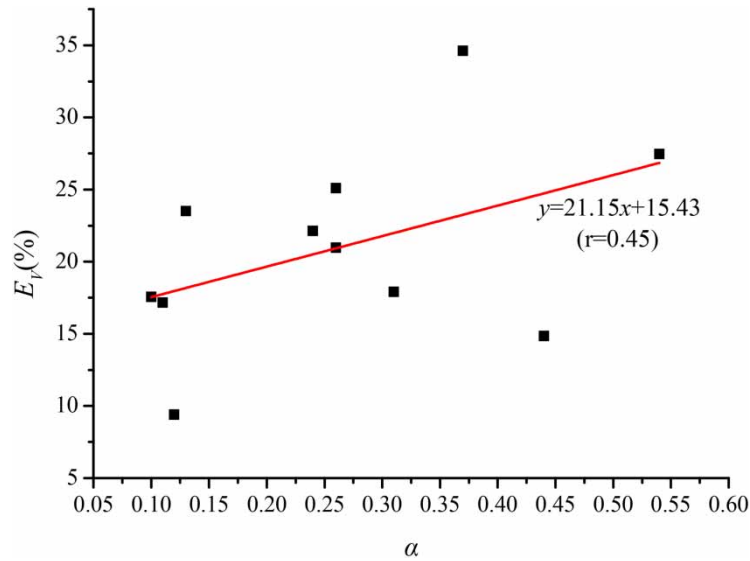


Figure 6 | Relationship between the average residual reduction rate and the average fluctuation coefficient of floods in 10 selected reservoirs.

In addition, comparing Figures 5 and 6, it can be seen that the slope of the trend line in Figure 5 is steeper than that in Figure 6. In fact, this observation applies to the floods in the other nine reservoirs. This is an indication that the relationship between the fluctuation coefficient and the robust effect is more significant in a single reservoir than in a group of 10 reservoirs. This is because for the group of 10 reservoirs, the relationship between the residual reduction rate and the fluctuation coefficient is an average. Hence, the relationship is less significant.

Furthermore, the slope of the trend line between the mean residual reduction rate, E_v , and the mean fluctuation coefficient, α , of all 10 selected reservoirs have been calculated. Figure 7 shows the relationship between the slope of E_v and α and the average fluctuation coefficient of 10 selected reservoirs. The results show that the slope of E_v and α is inversely proportional to the average fluctuation coefficient of reservoirs. That is because, for reservoirs with a large average fluctuation coefficient, the number and magnitude of the outliers in the observed flood data is large. As the residual reduction rate of different floods is generally large, the robustness of the method remains steady for different floods. On the other hand, for reservoirs with

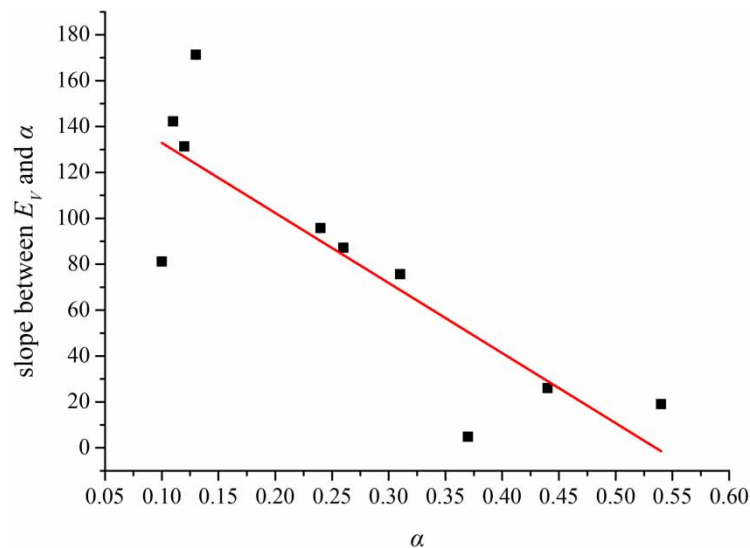


Figure 7 | Relationship between the average fluctuation coefficient and the slope of E_v and α for floods in 10 selected reservoirs.

small average fluctuation coefficient, the frequency and magnitude of the outliers is small. So, not all of the observed flood data contain outliers. The residual reduction rate for the observed flood data with outliers is large, while the residual reduction rate for the observed flood data with little or no outliers is small. Therefore, the robustness of the method on different floods changes greatly, resulting in a large slope of E_v and α , as shown in Figure 7.

Many methods and techniques have been developed to improve the accuracy of real-time flood forecasting in recent years. Most of the methods are focused on real-time correction, which can be classified into terminal error correction and process error correction. For example, the dynamic system response curve (DSRC) method, a process error correction method, has been developed to improve flood forecasting. There have been many achievements and applications of the DSRC method in the hydrological model and real-time flood forecasting in recent years (Si *et al.* 2013, 2015; Liang *et al.* 2021). Most of these process error correction methods rely on the accuracy of observed data. So, when they are applied in reservoirs, the accuracy is limited. However, the method proposed in our study is a terminal error correction method that can detect and correct the gross errors in data directly based on the actual error distribution of the observations. So, it is robust and applicable enough for any system with random error and outliers.

CONCLUSIONS

Due to the errors in the water level measurements and the inverse calculation to obtain the reservoir inflow, the reservoir inflow data inevitably contain some outliers and gross errors, resulting in a jagged flood hydrograph. In this study, by introducing the robustness functions into real-time flood updating to limit the influence of gross errors on the forecasting results, a new robust real-time forecasting method has been developed. Furthermore, the fluctuation coefficient has been proposed, which can describe the irregular fluctuation quantitatively. The performance of the developed robust real-time flood forecasting method has been assessed by applying it to the floods generated by the ideal model and to floods in 10 selected reservoirs. The application results for both synthetic and real floods show that the robust real-time flood forecasting method is efficient and universally applicable. Even for flood data with outliers, the flood forecasting using the developed method is stable and accurate. Moreover, the robustness of the method is proportional to the fluctuation coefficient. In other words, the more severe the fluctuation of the flood hydrograph, the more robust is the method. And the robust real-time flood forecasting method can be a flexible method for different systems with disparate error structures through the selection of diverse robustness functions. In addition, the method is merely applied in humid regions and only utilized in conjunction with the Xinanjiang model in this study. The relationship between the fluctuation coefficient and the robust effect of the method in other regions of different characteristics as well as in other hydrological models, such as in arid areas, remains to be further verified.

ACKNOWLEDGEMENTS

This study is supported by the Postgraduate Research & Practice Innovation Program of Jiangsu Province (Grant No. SJKY19_0466), the Fundamental Research Funds for the Central Universities (Grant No. 2019B72514), the National Key R&D Program of China (Grant No. 2016YFC0402703), the fundamental research funds for central public welfare research institutes (Grant No. HKY-JBYW-2017-12), Open Fund of Key Laboratory for Technology in Rural Water Management of Zhejiang Province (Grant No. ZJWEU-RWM-20200202B), and the National Natural Science Foundation of China (Grant Nos. 51709077, 41371048, 51479062, and 51709076).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the hydrological stations in the study basins. Restrictions apply to the availability of these data, which were used under license for this study. Data are available from the corresponding author with the permission of the hydrological stations in the study basins.

CONFLICT OF INTEREST STATEMENT

All authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

REFERENCES

- Bao, W. & Zhao, L. 2014 Application of linearized calibration method for vertically mixed runoff model parameters. *Journal of Hydrologic Engineering* **19** (8). doi:10.1061/(asce)he.1943-5584.0000984.
- Bao, W., Haixiang, J. I., Qimei, H. U., Simin, Q. U. & Zhao, C. 2003 Robust estimation theory and its application to hydrology. *Advances in Water Science* **14** (4), 428–432.
- 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. doi:10.1061/(asce)he.1943-5584.0000848.
- Bárdossy, A. & Singh, S. K. 2008 Robust estimation of hydrological model parameters. *Hydrology & Earth System Sciences* **12** (6), 1273–1283.
- Chan, S. C. & Zhou, Y. 2009 On the performance analysis of the least mean M-estimate and normalized least mean M-estimate algorithms with Gaussian inputs and additive Gaussian and contaminated Gaussian noises. *Journal of Signal Processing Systems* **60** (1), 81–103. doi:10.1007/s11265-009-0405-9.
- Chao, Z., Hua-sheng, H., Wei-min, B. & Luo-ping, Z. 2008 Robust recursive estimation of auto-regressive updating model parameters for real-time flood forecasting. *Journal of Hydrology* **349** (3–4), 376–382. doi:10.1016/j.jhydrol.2007.11.011.
- Cizek, P. 2001 Robust estimation in nonlinear regression models. *CERGE-EI Working Papers* **142** (5 Suppl 1), S-749.
- Finstlerle, S. & Najita, J. 1998 Robust estimation of hydrogeologic model parameters. *Water Resources Research* **34** (11), 2939–2947.
- Hampel, F. R. 1973 Robust estimation: a condensed partial survey. *Zeitschrift Für Wahrscheinlichkeitstheorie Und Verwandte Gebiete* **27** (2), 87–104.
- Hampel, F. 1974 The influence curve and its role in robust estimation. *Publications of the American Statistical Association* **69** (346), 383–393.
- Kerich, E. C. 2020 Households drinking water sources and treatment methods options in a regional irrigation scheme. *Journal of Human, Earth, and Future* **1** (1), 10–19.
- Krisnayanti, D. S., Bunganaen, W., Frans, J., Seran, Y. A. & Legono, D. 2021 Curve number estimation for ungauged watershed in semi-arid region. *Civil Engineering Journal* **7** (6), 1070–1083.
- Lamond, J. & Penning-Rowsell, E. 2014 The robustness of flood insurance regimes given changing risk resulting from climate change. *Climate Risk Management* **2**, 1–10.
- Liang, Z., Huang, Y., Singh, V. P., Hu, Y. & Wang, J. 2021 Multi-source error correction for flood forecasting based on dynamic system response curve method. *Journal of Hydrology* **594**, 125908.
- Mens, M. J. P., Klijn, F., de Bruijn, K. M. & van Beek, E. 2011 The meaning of system robustness for flood risk management. *Environmental Science & Policy* **14** (8), 1121–1131. doi:10.1016/j.envsci.2011.08.003.
- Shen, D. D., Bao, W. M., Liu, K. X., Gong, T. T., Zhang, Q. & Chen, W. D. 2016 Research on parameter robust estimation for Muskingum model. *China Rural Water & Hydropower* **7**, 72–78.
- Si, W., Bao, W., Wang, H. & Qu, S. 2013 The research of rainfall error correction based on system response curve. *Applied Mechanics and Materials* **368–370** (1), 335–339. doi:10.4028/www.scientific.net/AMM.368-370.335.
- Si, W., Bao, W. & Gupta, H. V. 2015 Updating real-time flood forecasts via the dynamic system response curve method. *Water Resources Research* **51** (7), 5128–5144. doi:10.1002/2015wr017234.
- Tsanov, E., Ribarova, I., Dimova, G., Ninov, P. & Makropoulos, C. 2020 Water stress mitigation in the Vit River Basin based on WEAP and MatLab simulation. *Civil Engineering Journal* **6** (11), 2058–2071.
- Tukey, J. W. 1960 A survey of sampling from contaminated distribution. In: *Contributions to Probability & Statistics* (I. Olkin, ed.). Stanford University Press, Stanford, pp. 448–485.
- Zhao, R. 1992 The Xinanjiang model applied in China. *Journal of Hydrology* **135** (1–4), 371–381.
- Zhao, R. J., Liu, X. R. & Singh, V. P. 1995 The Xinanjiang model. *Computer Models of Watershed Hydrology* **135** (1), 371–381.
- Zhao, C., Hong, H. S. & Zhu, M. L. 2010 A three-stepwise robust statistical method for outlying rainfall observation. *Journal of the Graduate School of the Chinese Academy of Sciences* **27** (1), 17–26.
- Zhou, J. 1989 Classical theory of errors and robust estimation. *Acta Geodaetica Et Cartographica Sinica* **18** (2), 115–120.

First received 24 November 2021; accepted in revised form 8 March 2022. Available online 23 March 2022