

## Stochastic simulation for determining the design flood of cascade reservoir systems

Baohong Lu, Huanghe Gu, Ziyin Xie, Jiufu Liu, Lejun Ma and Weixian Lu

### ABSTRACT

Stochastic simulation is widely applied for estimating the design flood of various hydrosystems. The design flood at a reservoir site should consider the impact of upstream reservoirs, along with any development of hydropower. This paper investigates and applies a stochastic simulation approach for determining the design flood of a complex cascade of reservoirs in the Longtan watershed, southern China. The magnitude of the design flood when the impact of the upstream reservoirs is considered is less than that without considering them. In particular, the stochastic simulation model takes into account both systematic and historical flood records. As the reliability of the frequency analysis increases with more representative samples, it is desirable to incorporate historical flood records, if available, into the stochastic simulation model. This study shows that the design values from the stochastic simulation method with historical flood records are higher than those without historical flood records. The paper demonstrates the advantages of adopting a stochastic flow simulation approach to address design-flood-related issues for a complex cascade reservoir system.

**Key words** | cascade reservoir systems, design flood, historical flood records, multi-station auto-regression, stochastic simulation

**Baohong Lu** (corresponding author)

**Huanghe Gu**

**Lejun Ma**

College of Hydrology and Water Resources,

Hohai University,

Nanjing 210098,

China

E-mail: [Lubaohong@126.com](mailto:Lubaohong@126.com)

**Baohong Lu**

**Huanghe Gu**

**Ziyin Xie**

**Jiufu Liu**

**Lejun Ma**

**Weixian Lu**

State Key Laboratory of Hydrology-Water

Resources and Hydraulic Engineering,

Nanjing 210098,

China

**Ziyin Xie**

**Jiufu Liu**

**Weixian Lu**

Nanjing Hydraulic Research Institute,

Nanjing 210029,

China

### INTRODUCTION

The impacts of human activities on hydrologic environments increase as the human population grows and the economy develops. Many changes may occur in the water cycle of a watershed due to the human activities (Georgakakos & Smith 2001; Yu *et al.* 2006; Isik *et al.* 2008). Any modification in a river section affects not only upstream but also downstream conditions (Galay 1983; Simons & Senturk 1992). It is well known that construction of dams has important impacts on the downstream design floods. Determination of flood magnitude–frequency relationship is a fundamental engineering hydrology problem for design flood estimation (Bocchiola *et al.* 2003). If sufficiently long river flow records are available, the flow–frequency relation can be estimated directly. However, it is not uncommon that hydrosystem infrastructural project sites have limited or no information on flood flow records, and so the flow–frequency relation has to be estimated indirectly. The indirect way commonly applied in practice

for estimating design floods is by rainfall–runoff modeling, in which rainfall is treated as the statistical input and is converted to flow using a suitable deterministic catchment rainfall–runoff response model (Boorman *et al.* 1990).

In hydropower and hydroelectricity development, many systems in China involve an arrangement of several reservoirs in a cascade. The choice of design flood magnitude has a great impact not only on the investment and benefit of the hydropower project development, but also on the scale and safety of the infrastructures (Maidment 1993). Conventionally, methods to determine the design discharge for a system involving cascade reservoirs require searching for appropriate combinations of flood or rainfall frequency among the various watersheds that contribute inflows to the reservoirs in the system (Boughton & Droop 2003). In principle, the design flood for a cascade reservoir system should be the critical combination of flood–rainfall–frequency among the contributing watersheds that is

mostly likely to create adverse effects with respect to flood control objectives. However, this approach of determining design flood discharge is further complicated by different reservoir operation rules. Therefore, for a complex system involving a cascade of reservoirs the use of the conventional approach to search for a suitable design flood is impractical, if not impossible, especially when the condition of climate and land use changes is to be considered. Alternatively, stochastic simulation approaches allow simulating a large number of probabilistic plausible flood events through which all combinations of different contributing watershed areas and flood events are considered.

The stochastic simulation method can generate a large number of floods, considering flow inputs from different sub-areas (Chen 2000; Wang *et al.* 2007). With the advent of computing power, stochastic simulation has been increasingly applied in the determination of design flood of cascade reservoir systems (Kottegoda 1980; Ding & Deng 1988). In terms of flood hydrograph at selected dam sites, stochastic simulation can be used to synthesize multi-station annual maximum flood hydrograph and to generate sufficiently long periods of synthetic flood records. Through reservoir routing by incorporating operation rules of individual reservoir, the flood hydrograph at the downstream reach of the reservoir system can be obtained to derive the annual maximum discharge–frequency relation. Under the stochastic simulation framework, there is no need to be concerned with the issue of searching for the critical combination of rainfall–flood–frequency among different contributing catchments as done by the conventional approach.

In this paper a stochastic simulation model is described for the determination of the design flood of a complex cascade reservoir system in Longtan Watershed of Guangxi Province, China. The system is a typical cascade reservoirs system involving four large reservoirs. Furthermore, the paper describes a method to determine the design flood by incorporating the historical records into the stochastic simulation method.

## DESCRIPTIONS OF STUDY AREA

Longtan hydroelectric power station is located on the main stream of the Hongshui River in Tian'e, Guangxi Province,

south China. Figure 1 shows that the two main tributaries, namely South Panjiang River and North Panjiang River, which meet upstream of the power station site. The catchment area upstream of Longtan Station is  $9.85 \times 10^4 \text{ km}^2$ , the average annual precipitation is 1,200 mm, the average annual runoff is  $5.08 \times 10^{10} \text{ m}^3$ , the installed power-generating capacity reached 6,300 MW, and the total reservoir capacity is around  $2.73 \times 10^{10} \text{ m}^3$ . Along South Panjiang River and North Panjiang River, several hydroelectric power stations have been built and several others are under construction.

Under the influence of operations of upstream hydroelectric power stations at Tianshengqiao, Dongqing, Guangzhao, the natural flow characteristics at Longtan hydroelectric power station site have been greatly altered. Similarly, regulating flow at Guangzhao Reservoir also has altered the natural flow regime at Dongqing hydroelectric power station. The flow regulation in the reservoir group creates complexity in the determination of a design flood at the dam site for Longtan hydroelectric power station. By stochastic flood simulation a large amount of flood hydrographs characteristic of different contribution areas can be generated. This situation circumvents the difficulty of determining the critical combination of floods in a complex watershed–reservoir system.

As shown schematically in Figure 1, inflow to Longtan hydroelectric power station consists of three parts: (1) outflow from Tianshengqiao Reservoir located on the South Panjiang tributary; (2) outflow from Dongqing Reservoir which is the result of regulating the Guangzhao Reservoir outflow combined with runoff from the Gaoche tributary; and (3) runoff contributed from intermediate areas between Longtan and downstream of Dongqing and Tianshengqiao reservoirs. The flood contributed from the intermediate areas is obtained from back-calculation with the flow records at Dongqing, Tianshengqiao and Longtan stations because there is no streamflow gauging station in the intermediate areas.

## MODEL FRAMEWORK

The basic procedures of stochastic simulation are shown in Figure 2. A multi-station auto-regressive model is developed

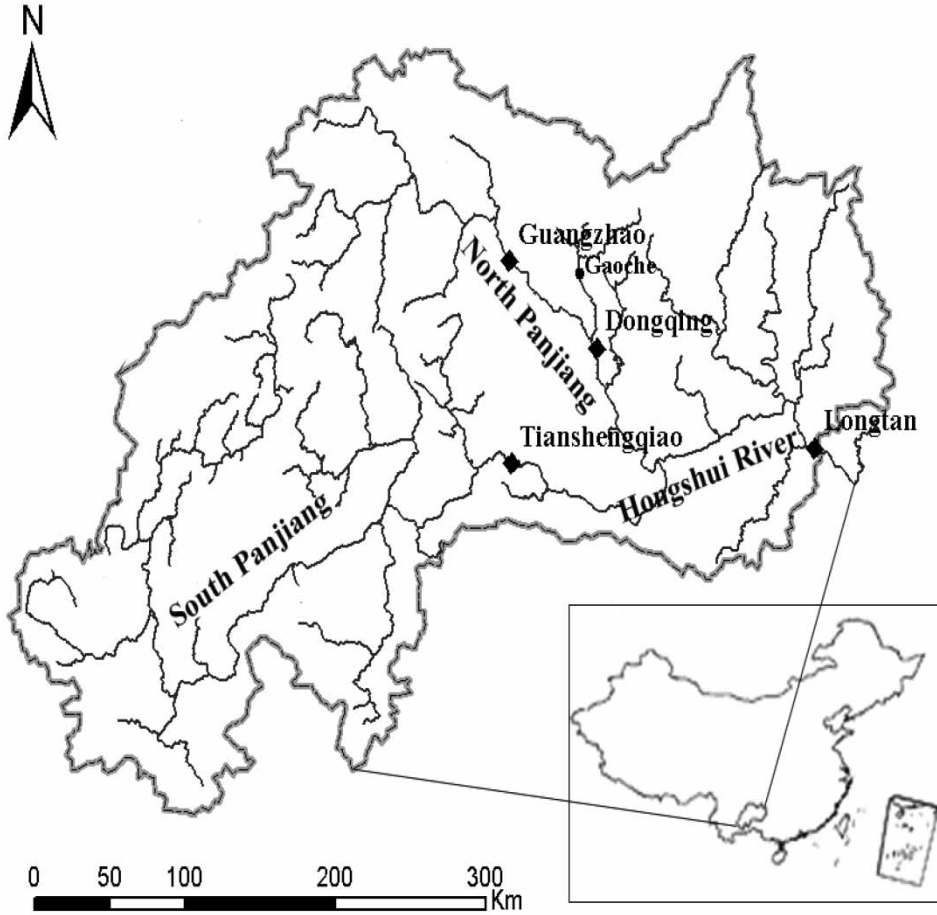


Figure 1 | Schematic illustration of the reservoir system in the Longtan watershed.

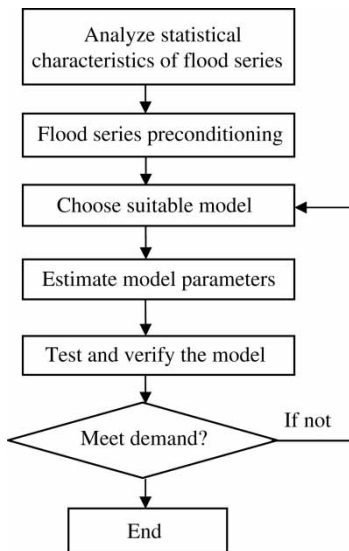


Figure 2 | Flow chart of stochastic simulation framework.

herein to simulate the inflows at all the stations (Tianshengqiao, Dongqing, Guangzhao and the intermediate areas) which can be expressed as (Yevjevich 1972)

$$Z(t) = \sum_{i=1}^p A_i Z(t-i) + B\epsilon(t) \tag{1}$$

where  $Z(t) = m \times 1$  column vector containing flood series at the  $m$ th station at time  $t$  during the flood period;  $A_i = m \times m$  matrix of the auto-regression (AR) parameters;  $B = m \times m$  lower triangular matrix of parameters;  $\epsilon(t) = m \times 1$  error column vector defined by normal variables with zero mean;  $p =$  the order of the AR model; and  $m =$  number of the stations ( $m = 4$  herein). The AR parameters matrix  $A_i$  is the function of auto-correlation coefficients and cross-correlation coefficients;  $B$  is the model's white noise

variance matrix. Explicitly, elements in Equation (1) are defined as

$$\mathbf{Z}(t) = \begin{bmatrix} Z_1(t) \\ Z_2(t) \\ \vdots \\ Z_m(t) \end{bmatrix} \quad (2)$$

$$\mathbf{A}_i = \begin{bmatrix} a_{1,1}^i & a_{1,2}^i & \cdots & a_{1,m}^i \\ a_{2,1}^i & a_{2,2}^i & \cdots & a_{2,m}^i \\ \vdots & \vdots & \cdots & \vdots \\ a_{m,1}^i & a_{m,2}^i & \cdots & a_{m,m}^i \end{bmatrix} \quad (3)$$

$$\mathbf{B} = \begin{bmatrix} b_{1,1} & 0 & \cdots & 0 \\ b_{2,1} & b_{2,2} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ b_{m,1} & b_{m,2} & \cdots & b_{m,m} \end{bmatrix} \quad (4)$$

$$\boldsymbol{\varepsilon}(t) = \begin{bmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \\ \vdots \\ \varepsilon_m(t) \end{bmatrix}. \quad (5)$$

The parameter matrices  $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_p$  and  $\mathbf{B}$  need to be estimated. By introducing matrices  $\mathbf{M}_i$  (Equation (6),  $i=1, 2, \dots, p$ ) with its elements relate to the auto-correlation and cross-correlation of the flood series of each station, the parameter matrix  $\mathbf{A}_i$  and  $\mathbf{B}$  could be solved by the Whittle recurrence method as (Whittle 1953)

$$[\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_p] = [\mathbf{M}_1, \mathbf{M}_2, \dots, \mathbf{M}_p] \begin{bmatrix} \mathbf{M}_0 & \mathbf{M}_1 & \cdots & \mathbf{M}_{p-1} \\ \mathbf{M}_1^T & \mathbf{M}_0 & \cdots & \mathbf{M}_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{M}_{p-1}^T & \mathbf{M}_{p-2}^T & \cdots & \mathbf{M}_0 \end{bmatrix}^{-1} \quad (6)$$

$$\mathbf{B}\mathbf{B}^T = \mathbf{M}_0 - \sum_{i=1}^p \mathbf{A}_i \mathbf{M}_i^T \quad (7)$$

where  $\mathbf{M}^T$  and  $\mathbf{M}^{-1}$  denote, respectively, the transpose

and inverse of matrix of  $\mathbf{M}$  defined by

$$\mathbf{M}_i = \begin{bmatrix} \rho_{1,1}^i & \rho_{1,2}^i & \cdots & \rho_{1,m}^i \\ \rho_{2,1}^i & \rho_{2,2}^i & \cdots & \rho_{2,m}^i \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{m,1}^i & \rho_{m,2}^i & \cdots & \rho_{m,m}^i \end{bmatrix} \quad (8)$$

where

$$\rho_{i,j}^0 = \frac{1}{T} \sum_{t=1}^T \left( \sum_{i,j=1}^m Z_i(t) Z_j(t) \right);$$

$$\rho_{i,j}^l = \frac{1}{T-1} \sum_{t=l+1}^T \left( \sum_{i,j=1}^m Z_i(t) Z_j(t-l) \right);$$

and  $T$  is the number of the time periods.

To establish the parameter matrices  $\mathbf{A}_i$  and  $\mathbf{B}$ , the order of the model  $p$  should be determined first. The common method to estimate  $p$  is the Bayesian Information Criterion (BIC) which can be expressed as (Schwarz 1978)

$$BIC(p) = n \ln [\text{Det}(\mathbf{S}_p)] + \ln(n) \left( p + \frac{1}{2} + \frac{1}{2m} \right) m^2 \quad (9)$$

where  $n$  denotes the length of the flood series;  $\text{Det}(\mathbf{S}_p)$  denotes the value of determinant of matrix  $\mathbf{S}_p = \mathbf{B}\mathbf{B}^T$ . If  $BIC(p_0) \leq BIC(p)$ , for all orders considered, and then  $p_0$  is the best order for the model.

To establish flow series at Longtan, streamflow records at gauging stations at Tianshengqiao, Guangzhao, Dongqing, together with the flow of the intermediate area, are used to set up a four-station auto-regressive model to simulate the daily flow process in the study area (i.e.  $m=4$ ). The annual main flood period is from May to September.

For reservoir operation and hydropower generation, flows at an hourly time step are needed. Hence, a disaggregating model is applied to decompose average daily discharge into an hourly flow hydrograph. Then, reservoir operation and the streamflow routing are performed. The decomposition coefficient is determined by using the actual observational record (Koutsoyiannis 1994;

Koutsoyiannis & Manetas 1996) as

$$K_i(t) = \frac{X_i(t)}{\sum_{i=1}^T X_i(t)} \quad (10)$$

where  $K_i(t)$  = the decomposition coefficient at hour  $t$  in the  $i$ th year;  $X_i(t)$  = discharge of hour  $t$  in the  $i$ th year; and  $T$  = the total number of hours in the  $i$ th year.

The term  $\varepsilon(t)$  is defined as independent normally distributed random variable. But the observational data do not follow the normal distribution, i.e. the skewness coefficients at the four stations are not zero. Therefore, a transformation to reduce this skewness closer to zero is needed (Raman & Sunilkumar 1995). In this study, the Wilson–Hilferty transformation is used (McGinnis & Sammons 1970). The procedure of the Wilson–Hilferty transformation (Salas et al. 1985) is as follows:

1. Log-transformation is applied to the original flows  $X_i(t)$  as:

$$w_i(t) = \log[X_i(t) + cX_i] \quad (11)$$

where  $c$  is a small positive increment to cater for zero flows to prevent numerical problem.

2. Standardize transformed data by:

$$y_i(t) = \frac{w_i(t) - w_i}{\sigma_i} \quad (12)$$

where  $y_i(t)$  have zero mean and unit variance, and  $w_i$  and  $\sigma_i$  are the mean and standard deviation of  $w_i(t)$ .

3. Use modified Wilson–Hilferty transformation (Kirby 1972):

$$Z(t) = \frac{6}{C_{sy_i(t)}} \left\{ \left[ \frac{C_{sy_i(t)}}{2} y_i'(t) + 1 \right]^{1/3} - 1 \right\} + \frac{C_{sy_i(t)}}{6} \quad (13)$$

where the  $C_{sy_i(t)}$  is the coefficient of skewness of  $y_i(t)$ . When  $C_{sy_i(t)} > 0$ ,

$$y_i'(t) = \max(y_i(t), -2/C_{sy_i(t)}) \quad (14)$$

while  $C_{sy_i(t)} < 0$ ,

$$y_i'(t) = \min(y_i(t), -2/C_{sy_i(t)}). \quad (15)$$

The procedure of determining a design flood consist of six steps: (1) L-moments are used to derive at-site statistics of the average daily discharges; (2) a modified Wilson–Hilferty transformation procedure is used for the normalization of daily discharge series; (3) generation of daily flow process by multi-station auto-regressive model; (4) daily discharge is decomposed into hourly discharges; (5) reservoir routing is performed by considering operation rules of each reservoir to obtain the flow series at Longtan hydro-power station; and (6) estimation of the design flood according to the generated flow series at Longtan.

## MODEL VERIFICATION

The purpose of model verification is to verify that the model is in compliance with the assumptions made in its development. Specifically, the multi-site AR model developed assumes that all residuals are spatially and temporally independent. Hence, a test of independence on the basis of the residuals' auto-correlation and cross-correlation will be made. Furthermore, a test of normality for residuals will be conducted.

### Test of temporal independence

To test temporal independence of residuals at a site, auto-correlation coefficients with lag time from 1 to 3 days are calculated. Taking the significance level  $\alpha=0.05$ , the residuals of time series at a site are independent when the auto-correlation coefficient  $\rho_{i,i}$  ( $i = 1, 2, \dots, m$ ) satisfies the following condition (Salas et al. 1985):

$$|\rho_p| < z_{\alpha/2} \sqrt{\frac{1}{n-p}} \quad (16)$$

where  $n$  = the number of the flood series of each year;  $p$  = time lag of the model;  $z_{\alpha/2}$  is the critical value associated with the standard normal variable with an exceedance

probability of  $\alpha/2$ . If annual  $\rho_{i,i}$  is in the range of  $(-\rho_p, \rho_p)$ , then temporal independence of the residuals at a site are accepted. The results are given in Table 1, which shows the passing rate, defined as the percentage of sample  $\rho_{i,i}$  lying within the range of  $(-\rho_p, \rho_p)$ , is higher than 80% among all record years considered. Hence, the model to a large extent meets independence condition.

### Test of spatial independence

To test independence of residuals between different sites, sample cross-correlation of residuals among each pair of sites is calculated. Taking the significance level  $\alpha = 0.05$ , the residual of two sites are independent when the value of sample cross-correlation coefficient  $\rho_{i,j}$  ( $i, j = 1, 2, 3, 4; i \neq j$ ) satisfies the following condition (Salas et al. 1985):

$$\left| \sqrt{n-2} \frac{\rho_{i,j}^0}{\sqrt{1-\rho_{i,j}^2}} \right| < t_\alpha \quad (17)$$

where  $t_\alpha$  is the critical value obtained from the  $t$ -distribution table (the degree of freedom ( $\nu$ ) is equal to the sample capacity minus one, e.g.  $\nu = n - 1$ );  $\rho_{i,j}^0$  is the sample cross-correlation coefficient between stations  $i$  and  $j$  with zero time lag. Table 2 shows the passing rate of all years is 100% indicating that residuals are spatially independent.

### Test of residual normality

Calculating the coefficient of skewness ( $C_s$ ) of the residuals series and taking significance level of  $\alpha = 0.05$ , the residuals

**Table 1** | Independence test of residual auto-correlation

Station	Passing rate (%)		
	Time lag = 1	Time lag = 2	Time lag = 3
Dongqing	88.60	92.11	91.23
Guangzhao	92.11	90.35	88.60
Tianshengqiao	83.33	86.84	90.35
Intermediate area	88.60	89.47	83.33

**Table 2** | Independence test of residual auto-correlation (no correlation)

Station	Passing rate (%)		
	Guangzhao	Tianshengqiao	Intermediate
Dongqing	100	100	100
Guangzhao		100	100
Tianshengqiao			100

follow the normal distribution if the sample  $C_s$  satisfies the following criterion (Salas et al. 1985):

$$|C_s| < \sqrt{\frac{6}{n}} z_{\alpha/2} \quad (18)$$

The results of the normality test are shown in Table 3. The passing rate of the normality test is higher than 90%, except at Dongqing, which is slightly less than 80%. Therefore, the residuals of the multi-station autoregressive model of daily flows developed in the study area can be considered spatially and temporally independent following the normal distribution.

## STOCHASTIC SIMULATION WITHOUT CONSIDERING HISTORICAL FLOODS

The developed multi-site auto-regressive model is applied to synthesize daily discharge series at Panjingqiao, Dongqing, Tianshengqiao and the intermediate area for a period of 50,000 years. The synthesized flow series are routed through each of the three upstream reservoirs (respectively considering their operation rules) and the intermediate area to calculate the hourly flow series at Longtan hydroelectric power station site.

To assess the statistical features of design flood at Longtan (i.e., mean, coefficient of variation  $C_v$  and skewness coefficient  $C_s$ ), synthesized 50,000-yr hourly flow series at

**Table 3** | Normality test of residual

Station	Dongqing	Guangzhao	Tianshengqiao	Intermediate area
Passing rate (%)	79.82	95.61	98.25	93.86

the four sites are divided into one hundred 500-yr long flow sub-series from which the maximum hourly discharge in each flow sub-series is taken to be the design flood. Of particular interest in this study is to examine the effect of the three existing reservoirs on the statistical features of the design flood at the proposed dam site at Longtan. This examination can be done by routing each of the 100 hourly flow sub-series of 500 years length through the watershed in the natural condition (i.e., without reservoirs) and in the present condition (i.e., with the presence of three reservoirs). The statistical features of the design flood and the corresponding flood volumes of different durations are shown in Table 4. The original design values are obtained by analyzing only the flood records at the dam site without considering the influence of the upstream reservoirs, while flood records from four stations are used in the stochastic simulation.

As shown in Table 4, the mean flow volumes for different durations without considering the reservoirs are greater than those when the reservoirs are considered. This situation is because some water is stored in the upstream reservoirs during the periods of flooding. The values of coefficients of variation and skewness with and without considering the influence of upstream reservoirs are similar. They do not vary much with different durations implying that regulation capacity for the upstream reservoirs is not much. The results by stochastic simulation without considering the influence of the upstream reservoirs for different durations are close to the original design values, which do not consider the influence of the upstream reservoirs (only the flood records at dam site are used). The coefficients of variations are similar between the simulation and original

values, but the difference of coefficient of skewness are remarkable. This factor is the reason simulated flows are higher than original flow series at low frequency while less at high frequency as shown in Table 5.

Without considering historical flood records, the frequency relations of flood characteristics (i.e., peak discharge and the corresponding flow volumes of different durations) under the existing condition are given in Table 5. The frequency relationships are determined through the Pearson Type 3 distribution. As can be seen from Table 5, the relative differences between the simulated and the original flood quantiles are within the range of 20%, except for the 10-yr 3-day flow volume which is -22%. But the frequency interval from 0.02 to 5% is mainly concerned in the design stage for the project. Therefore, the quantile values of simulated flow are larger in the range of low frequency whereas they are smaller in the interval of high frequency. Referring to the original design values, application of the stochastic simulation method, considering the influence of the upstream reservoirs, may result in the designed projects being more conservative for flood control.

## STOCHASTIC SIMULATION CONSIDERING HISTORICAL FLOODS

When historical floods are considered, the challenge is how to convert maximum flood volume for a known time period or maximum flood peak in the historical record to the maximum mean daily flood. According to the existing literature investigating historical floods in the study area, the four

**Table 4** | Comparison between simulated and original design flood characteristics

Flow variables <sup>a</sup>	Simulation values Without reservoirs			With reservoirs			Original values		
	Mean	C <sub>v</sub> <sup>b</sup>	C <sub>s</sub> <sup>b</sup>	Mean	C <sub>v</sub> <sup>b</sup>	C <sub>s</sub> <sup>b</sup>	Mean	C <sub>v</sub> <sup>b</sup>	C <sub>s</sub> <sup>b</sup>
Q <sub>p</sub>	10,390	0.33	1.16	9,833	0.34	1.28	11,100	0.28	0.70
V <sub>3-d</sub>	21.1	0.33	1.11	20.1	0.33	1.14	23.6	0.27	0.68
V <sub>7-d</sub>	43.1	0.32	1.04	41.8	0.33	1.09	46.2	0.27	0.68
V <sub>15-d</sub>	76.4	0.31	0.97	75.8	0.31	0.98	82.8	0.27	0.68
V <sub>30-d</sub>	123.1	0.26	0.85	122.5	0.27	0.80	134.0	0.27	0.68

<sup>a</sup>Q<sub>p</sub> = Peak discharge (m<sup>3</sup>/s); V<sub>n-d</sub> = n-day flow volume (10<sup>8</sup> m<sup>3</sup>).

<sup>b</sup>C<sub>v</sub> = Coefficient of variation; C<sub>s</sub> = Coefficient of skewness.

**Table 5** | Frequency relations of flood characteristics (with reservoirs)

Frequency (%)	$Q_m$			$V_{3d}$			$V_{7d}$			$V_{15d}$		
	O	S	R	O	S	R	O	S	R	O	S	R
0.01	27,500	31,942	16.2	56.8	61.8	8.8	111.0	126.0	13.5	199.0	214.6	7.8
0.02	26,400	30,275	14.7	54.7	58.8	7.5	107.0	119.9	12.1	192.0	204.8	6.7
0.05	25,000	28,052	12.2	51.8	54.7	5.6	101.0	111.7	10.6	182.0	191.6	5.3
0.1	23,800	26,354	10.7	49.5	51.6	4.2	96.9	105.5	8.9	174.0	181.5	4.3
0.2	22,700	24,637	8.5	47.2	48.4	2.5	92.3	99.1	7.4	166.0	171.2	3.1
0.5	21,800	22,333	2.4	45.4	44.1	-2.9	88.9	90.5	1.8	159.0	157.3	-1.1
1	21,100	20,559	-2.6	44.0	40.8	-7.3	86.1	83.9	-2.6	154.0	146.4	-4.9
2	19,900	18,478	-7.2	41.5	37.4	-9.9	81.2	77.0	-5.2	146.0	135.2	-7.4
5	18,600	16,278	-12.5	38.9	32.7	-15.9	76.1	67.6	-11.2	136.0	119.7	-12.0
10	17,800	14,324	-19.5	37.3	29.0	-22.3	72.9	60.0	-17.7	131.0	107.2	-18.2

Frequency = Exceedance probability, O = Original flows, S = Simulated flows, R = Relative difference, %.

sites (namely, Tianshengqiao, Dongqing, Guangzhao and the intermediate area) have instantaneous flood peak data. Except for Tianshengqiao, none of the other sites has a flood volume record with durations of 3, 7 and 15 days. The flood volumes with different durations of other sites are estimated by the relationship between instantaneous flood peak and flood volumes.

To incorporate historical floods at a gauging site, the following equations are used to calculate the sample statistical moments of floods (Liang *et al.* 2006):

$$\bar{x} = \frac{1}{N} \left( \sum_{j=1}^a x_j + \frac{N-a}{n} \sum_{i=1}^n x_i \right) \quad (19)$$

$$C_v = \frac{1}{\bar{x}} \sqrt{\frac{1}{N-1} \left[ \sum_{j=1}^a (x_j - \bar{x})^2 + \frac{N-a}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right]} \quad (20)$$

$$C_s = \frac{N \left[ \sum_{j=1}^a (x_j - \bar{x})^2 + ((N-a)/n) \sum_{i=1}^n (x_i - \bar{x})^3 \right]}{(N-1)(N-2)\bar{x}^3 C_v^3} \quad (21)$$

where  $C_v$  is the coefficient of variation;  $C_s$  is the coefficient of skewness;  $N$  is the length of record extended to cover the historical events (in years);  $a$  is the number of the historical floods;  $n$  is systematic record length (in years).

The 50,000-yr daily flow series at the four stations are generated in the same way as described previously. Following the operation rules of each reservoir and the parameters of stream routing, the statistics of the design flood characteristics at Longtan are assessed according to the simulated inflows. A comparison of simulated design flood characteristics with and without considering historical floods are presented in Tables 6 and 7, respectively.

**Table 6** | Statistical features of design flood characteristics

Item	Without historical floods			With historical floods			Original design		
	Mean	$C_v$	$C_s$	Mean	$C_v$	$C_s$	Mean	$C_v$	$C_s$
$Q_p$	9,833	0.34	1.28	10,124	0.35	1.24	11,500	0.34	1.19
$V_{3-d}$	20.08	0.33	1.14	20.96	0.34	1.15	24.9	0.30	1.05
$V_{7-d}$	41.81	0.33	1.09	43.69	0.34	1.11	50.6	0.35	1.05
$V_{15-d}$	75.78	0.31	0.98	78.72	0.32	1	91.5	0.35	1.05
$V_{30-d}$	122.49	0.27	0.8	126.21	0.28	0.85	150	0.35	1.05



**Table 7** | Comparison of frequency relations of flood characteristics at Longtan under different conditions.

F (%)	$Q_p$			$V_{3-d}$			$V_{7-d}$			$V_{15-d}$		
	Y	N	O	Y	N	O	Y	N	O	Y	N	O
0.01	32,860	31,942	36,500	66.0	61.8	70.3	135.0	126.0	158.0	227.3	214.6	285.0
0.02	31,158	30,275	34,600	62.7	58.8	67.0	128.4	119.9	150.0	216.8	204.8	272.0
0.05	28,888	28,052	32,200	58.3	54.7	62.6	119.5	111.7	140.0	202.6	191.6	253.0
0.1	27,151	26,354	30,300	54.9	51.6	59.3	112.7	105.5	132.0	191.8	181.5	239.0
0.2	25,395	24,637	28,400	51.5	48.4	55.9	105.8	99.1	124.0	180.7	171.2	224.0
0.5	23,036	22,333	26,900	46.9	44.1	53.3	96.5	90.5	118.0	165.8	157.3	213.0
1	21,217	20,559	25,800	43.3	40.8	51.3	89.2	83.9	113.0	154.1	146.4	205.0
2	19,358	18,478	23,800	39.6	37.4	47.7	81.8	77.0	105.0	142.2	135.2	189.0
5	16,817	16,278	21,800	34.6	32.7	44.1	71.6	67.6	96.0	125.6	119.7	174.0
10	14,804	14,324	20,500	30.6	29.0	41.8	63.4	60.0	90.8	112.2	107.2	164.0

F = Exceedance probability, Y = Simulated flows with reservoirs, N = Simulated flows without reservoirs, O = Original flows.

As can be seen from Table 6, the values of design flood characteristics from considering historical floods are higher than those without considering them. The stochastic simulation results are slightly smaller than the original values because the stochastic simulation considers the influence of the storage effects of the upstream reservoirs. The values of  $C_v$  and  $C_s$  are similar among stochastic simulation with/without considering historical floods and original design.

Table 7 compares the frequency relations of design flood characteristics at Longtan under the various conditions. The simulation results are different from the original design because only the systematic flood records at Longtan Station were used. The simulated flood quantiles considering the historical records are higher than those without considering the historical records. This situation indicates that the consideration of historical records is important for determining the design flood properties. Because literature on historical floods is available at Gaoche and Tianshengqiao, historical floods in stochastic generation of daily streamflow are incorporated only at the two locations. No consideration of historical floods is given at Guangzhao because of lack of such information.

## SUMMARY AND CONCLUSIONS

The development of the system of reservoirs upstream of Longtan hydroelectric power station site has altered the

flow regime at the site due to man-made regulation of river flows. This situation makes the determination of design flood at the proposed dam site difficult. The conventional trial-and-error approach attempting to find a suitable combination of flow occurrence frequencies in defining the design flood is not practical. In this complex reservoir system, the use of stochastic flow generation model, along with the reservoir routing, is a practical alternative to determine design flood properties at a site.

In this study, a multi-site auto-regressive model is developed that takes into account the time and space correlations of daily streamflows at four locations in the study area. In particular, the method of Whittle is used for its advantage that matrix inversion does not increase with the increase in model order. This factor not only reduces the computing time, but also improves the precision of the model. The stochastic model is used to simulate 50,000-year long hourly flow time series at the three reservoir sites during the flood season between May 20th and September 10th. The simulated flow series are routed through the three reservoirs by taking into account their respective operation rules.

Model verification, including the tests for temporal and spatial independence of residuals and for normality, proved that the multi-site AR model is in compliance with the assumptions made in the model development. The values of simulated flow quantile considering the influence of upstream reservoirs are higher in the range of high frequency, and lower in the general frequency range, than

those without considering the influence of upstream reservoirs. The presence of reservoirs upstream increases the flow at Longtan during dry years and reduces the flow in high-flow years due to the storage effect of the reservoirs. The relative differences between the simulated design flood and the original design are less than 20%.

In the stochastic simulation for determining the design flood properties at Longtan hydroelectric power station site, flow series from systematic records were employed to simulate flows. In addition, historical floods extracted from literature were also used to modify the parameters of stochastic model for flow simulation. The design flood value considering the historical floods is larger than that without considering the historical floods. The model appears to be capable of effective utilization of the information in the extended flood data for the catchment.

## ACKNOWLEDGEMENTS

The study is sponsored in part by the Ministry of Water Resources, Public Welfare Fund 200801003, the Science and Technology Progress of Hydro-China Cooperation Group, and NSFC 50379008 and 50979023, as well as Key Program of National Natural Science Foundation of China (NSFC 40830639) and the National Basic Research Program of China (Grant No: 2010CB951101).

## REFERENCES

- Bocchiola, D., De Michele, C. & Rosso, R. 2003 [Review of recent advances in index flood estimation](#). *J. Hydrol. Earth Syst. Sci.* **7** (3), 283–296.
- Boorman, D. B., Acreman, M. C. & Packman, J. C. 1990 A review of design flood estimation using the FSR rainfall-runoff method. *Institute of Hydrology Wallingford Report No. 111*.
- Boughton, W. & Droop, O. 2003 [Continuous simulation for design flood estimation – a review](#). *Environ. Model. Softw.* **18**, 309–318.
- Chen, Y. F. 2000 Analysis of effects of uncertainty of simulating model and parameters estimated on precision of design flood control storage. *J. Hohai University* **28**, 32–35 (in Chinese with English abstract).
- Ding, J. & Deng, Y. Z. 1988 *Stochastic Hydrology*. Chengdu University of Science and Technology Press, Chengdu, China (in Chinese).
- Galay, V. J. 1983 [Causes of river bed degradation](#). *Water Resour. Res.* **19** (5), 1057–1090.
- Georgakakos, K. P. & Smith, D. E. 2001 [Soil moisture tendencies into the next century for the conterminous United States](#). *J. Geophys. Res.* **106** (D21), 27367–27382.
- Isik, S., Dogan, E. & Kalin, L. 2008 [Effects of anthropogenic activities on the Lower Sakarya River](#). *Catena* **75**, 172–181.
- Kirby, W. 1972 [Computer-oriented Wilson-Ailferty transformation that preserves the first three moments the lower bound of the Pearson Type 3 distribution](#). *Water Resour. Res.* **8** (5), 1251–1254.
- Kottegoda, N. T. 1980 *Stochastic Water Resource Technology*. Wiley, New York.
- Koutsoyiannis, D. 1994 [A stochastic disaggregation method for design storm and flood synthesis](#). *J. Hydrol.* **156**, 193–225.
- Koutsoyiannis, D. & Manetas, A. 1996 [A simple disaggregation accurate adjusting procedures](#). *Water Resour. Res.* **32** (7), 2105–2117.
- Liang, Z. M., Zhong, P. A. & Hua, J. P. 2006 *Hydrological Computation*. China Water Power Press, China (in Chinese).
- Maidment, D. R. 1993 *Handbook of Hydrology*. McGraw-Hill, New York.
- McGinnis, D. F. & Sammons, W. H. 1970 Discussion of 'Daily streamflow simulation'. *ASCE J. Hydraul. Div.* **96** (HY5), 1201–1206.
- Raman, H. & Sunilkumar, N. 1995 [Multivariate modelling of water resources time series using artificial neural networks](#). *Hydrol. Sci.* **40** (2), 145–163.
- Salas, J. D., Tabios, Q. P. & Bartolini, P. 1985 Approaches to multivariate modeling of water resources time series. *Water Resour. Bull.* **21**, 683–708.
- Schwarz, G. 1978 [Estimating the dimension of a model](#). *Ann. Stat.* **6**, 461–464.
- Simons, D. B. & Senturk, F. 1992 *Sediment Transport Technology*. Water Resources Publications, Littleton, CO.
- Wang, W. S., Jin, J. L. & Li, Y. Q. 2007 Advances in stochastic simulation of hydrology. *J. Adv. Water Sci.* **18**, 768–775 (in Chinese with English abstract).
- Whittle, P. 1953 The analysis of multiple stationary time series. *J. R. Stat. Soc.* **15**, 125–139.
- Yevjevich, V. M. 1972 *Stochastic Processes in Hydrology*. Water Resources Publications, Fort Collins, CO.
- Yu, Z., Pollard, D. & Cheng, L. 2006 [On continental-scale hydrologic simulations with a coupled hydrologic model](#). *J. Hydrol.* **331**, 110–124.

First received 7 January 2010; accepted in revised form 21 April 2010. Available online December 2011