

Rechecking the characteristic flood levels of a built reservoir after extending the hydrologic time series

Y. Zhou^{1,2}, J. Xu^{1,2}, D. Wang¹, X. Zhao³ & C. Yang¹

¹*Water Resources Department,*

Changjiang River Scientific Research Institute, China

²*Hubei Province Key Laboratory of River Basin Water Resources and Ecological Environment Science,*

Changjiang River Scientific Research Institute, China

³*Hubei Provincial Water Resources and Hydropower Planning Survey and Design Institute, China*

Abstract

Reservoirs are one of the most efficient infrastructures for integrated water resources development and management in a river basin. Rechecking characteristic flood levels plays a more and more important role in flood risk management, especially with extending hydrologic time series for a built reservoir. Hydrologic uncertainty is a key factor which impacts probability distribution of characteristic flood levels. Methodology of stochastic differential equation for flood routing is proposed in order to derive probability distribution of characteristic flood level for a built reservoir. The Three Gorges reservoir in the Changjiang River basin of China is selected for a case study. The uncertainty of flood is transformed into uncertainty of characteristic flood level so as to analyze probability distribution of characteristic flood level. One hundred and twenty-eight years of daily runoff data from 1882 to 2009 have been used to test the method. The results indicate that the proposed method can make an effective derivation of probability distribution of characteristic flood level by considering hydrologic uncertainty. Reliability of hydrologic time series and original checking of flood levels (180.4 m) increases with the increasing value of n . Reliabilities of the hydrologic time series and checking of flood levels are 93.19% and 99.17% under the sample size 120, respectively. The research results will provide



important theoretical and technical support for flood risk management of river basins.

Keywords: characteristic flood level, uncertainty, reliability, stochastic differential equation.

1 Introduction

Characteristic flood levels of reservoirs usually contain upper water levels for flood control, designing flood levels and checking flood levels. Hydrological uncertainty is one of the key factors which impacts probability distribution of characteristic flood levels. Rechecking characteristic flood levels has been studied by many relative scholars. Among them, Andrade [1], Li *et al.* [2], and Zhou and Guo [3] had confirmed it. Currently, designing flood frequency and flood risk are usually used for quantitative analysis of rechecking characteristic flood levels. However, risk analysis studies combined with reliability analysis for hydrologic time series and checking flood levels are still quite scarce. In this paper, a methodology of stochastic differential equation is proposed and developed to derive probability distribution of characteristic flood levels in order to recheck characteristic flood levels of a built reservoir after extending the hydrologic time series. The Three Gorges reservoir (TGR) in the Changjiang River basin of China is selected for a case study.

2 Methodology

To analyze the sampling property of reservoir water levels, a stochastic differential equation with a stochastic input term and a random initial condition must be established.

2.1 Sampling property of reservoir water level

The Wiener process with zero mean can be found from reservoir flood routing equation (1). The equation (1) is a typical Ito equation with a stochastic input term and a random initial condition as shown by Soong [4]:

$$\begin{cases} dH(t) = \frac{\mu_Q(t) - \mu_q(H, t)}{G(H)} dt + \frac{dB(t)}{G(H)} \\ H(t_0) = H_0 \end{cases} \quad (1)$$

where $H(t)$ is the reservoir water level hydrograph at time t , H_0 is the initial reservoir water level at time zero, $\mu_Q(t)$ is the inflow design flood hydrograph at time t and $\mu_q(H, t)$ is the outflow discharge hydrograph which is a function of H , $G(H)$ is equal to dV/dH and V is reservoir volume, $B(t)$ is that of Brownian motion.



The numerical solution of Euler is used to solve the stochastic differential equation (1) (Yan *et al.* [5]). The variance of $B(t)$ is equal to $\sigma_Q^2(t)t/G^2(H)$, when uncertainty of hydrological inflow is only considered as the main random factor of $B(t)$. According to the Wiener process, the reservoir water level $H(t)$ follows the normal distributed random variable with mean $(\mu_Q(t) - \mu_q(H, t)/G(H))t$ and variance $\sigma_Q^2(t)t/G^2(H)$, i.e.:

$$H(t) \sim \left((\mu_Q(t) - \mu_q(H, t)/G(H))t, \sigma_Q^2(t)t/G^2(H) \right) \quad (2)$$

2.2 Sampling error of characteristic flood level

Let CFL Y_j , with $j=1, 2, \dots, m$, be an independent and identically normal distributed process with mean μ and variance σ^2 , i.e.:

$$Y_1, Y_2, \dots, Y_m \sim N(\mu, \sigma^2) \quad (3)$$

Let's assume the mean μ and variance σ^2 are estimated by the Maximum Likelihood Estimation method (MLE), namely:

$$\hat{\mu} = \bar{Y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (4)$$

$$\hat{\sigma}^2 = S^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{Y})^2 \quad (5)$$

where y_1, y_2, \dots, y_n is a sample from (2), n is sample size, \bar{Y} is sample mean, S^2 is sample variance. With reference to a generic observation Y , not included in the estimation sample, the corresponding estimated CFL \hat{Z} is given by the sample equation based on the sample mean and standard deviation:

$$\hat{Z} = \frac{Y - \hat{\mu}}{\hat{\sigma}} = \frac{Y - \bar{Y}}{S} \quad (6)$$

The CFL value Z , based on the population mean and standard deviation of the underlying series is:

$$Z = \frac{Y - \mu}{\sigma} \quad (7)$$

Therefore, the sampling variability of the CFL can be characterized by investigating the distribution of the following random variable as a function of the estimation sample n :

$$D = Z - \hat{Z} = \frac{Y - \mu}{\sigma} - \frac{Y - \bar{Y}}{S} \quad (8)$$

The r.v. \hat{Z} is therefore the ratio of a normal r.v. to the squared root of a χ^2 r.v. and thus, after an appropriate rescaling, it is distributed as a Student's t . Indeed, it can be shown that:

$$Y - \bar{Y} \sim N\left(0, \frac{n+1}{n} \sigma^2\right) \quad (9)$$

$$\frac{Y - \bar{Y}}{S} \sqrt{\frac{n-1}{n+1}} \sim t(n-1) \quad (10)$$

From equation (10), it follows that $E\left[\frac{Y - \bar{Y}}{S}\right] = 0$ and

$$\text{Var}\left[\frac{Y - \bar{Y}}{S}\right] = \frac{n^2 - 1}{(n-1)(n-3)}.$$

Analytical derivation of the distribution of D is not an easy task, since D is the difference of two dependent r.v. with a standard normal and a student's t as marginal distribution, respectively. However, the first two moments of D can provide enough information to characterize the sampling variability of the CFL, since they allow to compute the bias and the Mean Squared Error (MSE) of estimation, as:

$$\text{Bias} = E[D] \quad (11)$$

$$\text{MSE} = E[D^2] = \text{Var}[D] + E[D]^2 \quad (12)$$

In practice it is preferable to use the Root Mean Squared Error (RMSE) of estimation which can be computed by taking the square root of the MSE:

$$\text{RMSE} = \sqrt{\text{MSE}} \quad (13)$$

The bias term $E[D]$ can be computed as:

$$E[D] = E\left[\frac{Y - \mu}{\sigma} - \frac{Y - \bar{Y}}{S}\right] = E\left[\frac{Y - \mu}{\sigma}\right] - E\left[\frac{Y - \bar{Y}}{S}\right] = 0 \quad (14)$$

since both expectations are zero. Thus, in the normal case, the CFL estimation given by equation (6) is unbiased. As a direct consequence, the MSE of estimation coincides with $\text{Var}[D]$.

On the basis of equation (12), the MSE can be written as:

$$\text{MSE} = \text{Var} \left[\frac{Y - \mu}{\sigma} - \frac{Y - \bar{Y}}{S} \right] = \text{Var} \left[\frac{Y - \mu}{\sigma} \right] + \text{Var} \left[\frac{Y - \bar{Y}}{S} \right] - 2 \text{Cov} \left[\frac{Y - \mu}{\sigma}, \frac{Y - \bar{Y}}{S} \right] \quad (15)$$

The first term in the above equation is obviously 1. The second term, as it has been shown previously, is equal to $\frac{n^2 - 1}{(n - 1)(n - 3)}$.

The covariance term in equation (17) can be written as:

$$\text{Cov} \left[\frac{Y - \mu}{\sigma}, \frac{Y - \bar{Y}}{S} \right] = E \left[\frac{Y - \mu}{\sigma} \cdot \frac{Y - \bar{Y}}{S} \right] = \text{Cov} \left[\frac{Y}{\sigma}, \frac{Y}{S} \right] - \text{Cov} \left[\frac{Y}{\sigma}, \frac{\bar{Y}}{S} \right] \quad (16)$$

The second term is obviously zero, since the observation Y is not included in the estimation sample, and therefore it is uncorrelated with the sample mean \bar{Y} or the sample standard deviation S .

By means of conditional expectation concepts, the first covariance term in equation (16) can be rewritten as:

$$\begin{aligned} \text{Cov} \left[\frac{Y}{\sigma}, \frac{Y}{S} \right] &= \frac{1}{\sigma} \text{Cov} \left[Y, \frac{Y}{S} \right] = \frac{1}{\sigma} E \left[\text{Cov} \left(Y, \frac{Y}{S} \right) \middle| \frac{1}{S} \right] + \text{Cov} \left[E \left[Y \middle| \frac{1}{S} \right], \frac{1}{S} \right], \\ E \left[\frac{Y}{S} \middle| \frac{1}{S} \right] &= \frac{1}{\sigma} E \left[\frac{1}{S} \text{Var}[Y] \right] = \sigma \cdot E \left[\frac{1}{S} \right] = \sigma \cdot E \left[\left(S^2 \right)^{\frac{1}{2}} \right] = \frac{\Gamma \left(\frac{n-2}{2} \right)}{\Gamma \left(\frac{n-1}{2} \right)} \cdot \sqrt{\frac{n}{2}} \quad (17) \end{aligned}$$

The latter expectation can be computed by reminding that S^2 is distributed according to a rescaled χ^2 distribution and therefore it follows Cancelleri and Bonaccorso [6]:

$$E \left[\frac{1}{S} \right] = E \left[\left(S^2 \right)^{\frac{1}{2}} \right] = \frac{1}{\sigma} \frac{\Gamma \left(\frac{n-2}{2} \right)}{\Gamma \left(\frac{n-1}{2} \right)} \cdot \sqrt{\frac{n}{2}} \quad (18)$$

Finally, combining equations (15), (17) and (18), it follows that:

$$\text{MSE} = 1 + \frac{n^2 - 1}{(n - 1)(n - 3)} - \frac{\Gamma \left(\frac{n-2}{2} \right)}{\Gamma \left(\frac{n-1}{2} \right)} \cdot \sqrt{2n} \quad (19)$$

thus



$$RMSE = \left[1 + \frac{n^2 - 1}{(n-1)(n-3)} - \frac{\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \cdot \sqrt{2n} \right]^{\frac{1}{2}} \quad (20)$$

It can be inferred from equations (19) and (20) that, in the normal case, the MSE and RMSE of estimation of CFL does not depend on the parameters μ and σ^2 of the underlying variable, but only on the sample size n ($n \geq 4$). Moreover, the values of MSE and RMSE decrease with the value of n and are equal to maximum with $n=4$.

2.3 Rechecking indicators of characteristic flood level

Two rechecking indicators are developed to verify reliability of hydrologic time series and CFL for built reservoir after extending hydrologic time series.

The first rechecking indicator is reliability of hydrologic time series. The first reliability α is defined as:

$$\alpha = \frac{RMSE(n=4) - RMSE(n)}{RMSE(n=4)} \times 100\% \quad (21)$$

It can be also inferred from equation (21) that, in the normal case, the value of α does not depend on the parameters μ and σ^2 of the underlying variable, but only on the sample size n ($n \geq 4$).

The second rechecking indicator is reliability of the i th CFL. The second reliability β is defined as:

$$\beta = \Phi(z_i) \times 100\% \quad (22)$$

where the value of z_i is standardized by equation (6), $\Phi(\cdot)$ is standardization normal distribution function.

3 Results and discussion

This section briefly introduces the study area, after which simulation results for TGR are presented and discussed.

3.1 Study area

The Three Gorges reservoir (TGR) is a vitally important and backbone project in the development and harnessing of the Changjiang River in China. The location map and characteristics of the TGR is shown in Figure 1. The TGR is the largest water conservancy project ever undertaken in the world. The main functions of

TGR are flood control, power generation, water supply as well as navigation, etc. The characteristic parameter values of TGR are given in Table 1. The current flood control operating rule of TGR is referenced by Zhou and Guo [3].



Figure 1: The location map and characteristics of the TGR.

Table 1: List of characteristic parameter values of TGR.

Flood limited water level/m	Upper water level for flood control/m	Designing flood level/m	Checking flood level/m
145.0	166.9	175.0	180.4

3.2 Analysis on rechecking of characteristic flood level

One hundred and twenty-eight years of daily runoff data in TGR from 1882 to 2009 have been used to test the method. As analysis on the rechecking of the original checking flood level (180.4 m) in TGR for example, the checking flood level is obtained from the maximum flood level by flood routing operation according to flood control operating rules. Sample sizes of flood data in four schemes are 30, 60, 90 and 120, respectively. The boxplot of maximum flood levels under four kinds of sample sizes are shown in Figure 2. Figure 2 shows that the mean and quartile of maximum flood levels increase with the increasing value of n , however, the upper of maximum flood levels is less than or equal to 180.4 m and the lower of maximum flood levels is larger than or equal to 175 m.

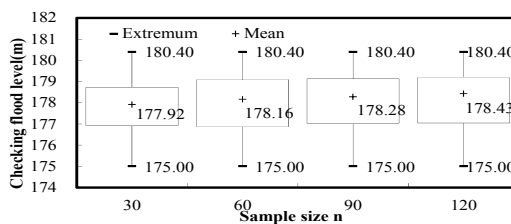


Figure 2: Boxplot for checking flood levels under four kinds of sample sizes.

Density probability curve of normal distribution for checking flood levels under four kinds of sample sizes is shown in Figure 3. Figure 3 shows that the variance of maximum flood levels increases with the increasing value of n . Besides, the sharp curves become shorter and fatter with the increasing value of n .

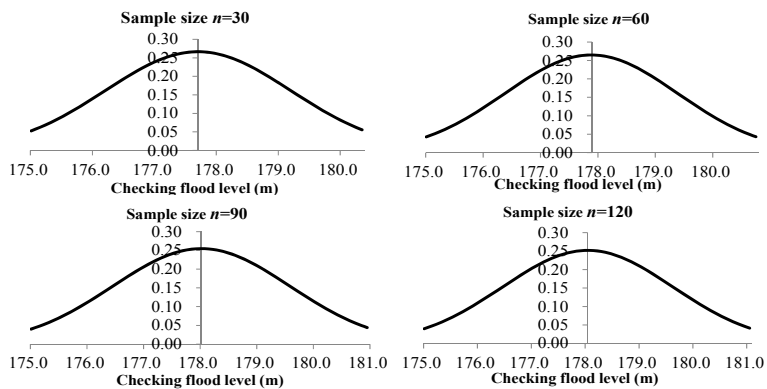


Figure 3: Probability density curve of normal distribution for checking flood level under four kinds of sample sizes.

The statistic result about reliabilities of hydrologic time series and original checking flood level (180.4 m) under four kinds of sample sizes is shown in Table 2. Table 2 shows that (a) reliability α of hydrologic time series increases with the increasing value of n and information of flood sample, (b) reliability β of original checking flood level (180.4 m) also increases with the increasing value of n and (c) both reliabilities α and β monotonically increase with the increasing value of n , reliability α is equal to 85.46% and 93.19% for sample sizes with 30 and 120, respectively as well as reliability β is equal to 96.77% and 99.17% for sample sizes with 30 and 120, respectively, both reliabilities α and β will approximate 100% with sample size $n \rightarrow +\infty$ when the maximum flood level is less than or equal to 180.4 m in TGR.

Table 2: Results of dual reliability under four kinds of sample sizes.

Sample size n	Reliability α (%)	Reliability β (%)	Maximum flood level (m)
30	85.46	96.77	180.4
60	90.16	98.36	180.4
90	92.08	98.90	180.4
120	93.19	99.17	180.4



Monte Carlo simulation is used to validate the equations (21) and (22) which the RMSE and α for estimation of hydrological time series do not depend on the parameters μ and σ^2 of the underlying variable, but only on the sample size n . The steps of Monte Carlo simulation are designed as follows: (a) 50,000 of checking flood level Z with sample sizes 30, 60, 90 as well as 120 subjected to standardization normal distribution are generated by Monte Carlo simulation, respectively, as well as another estimated checking flood level \hat{Z} is given by the sample equation based on flood routing operation, (b) the RMSEs based on observation (flood routing operation) and theoretical RMSE are computed by equations (14) as well as (21), respectively and (c) the reliabilities α based on observation and theoretical RMSE are computed by equation (22). Comparison for reliabilities α of hydrologic time series based on observation and theoretical RMSE under four kinds of sample sizes is shown in Figure 4.

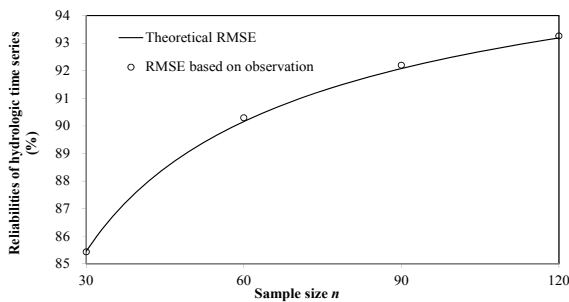


Figure 4: Comparison for reliabilities of hydrologic time series based on observation (flood routing operation) and theoretical RMSE (equation 22) under four kinds of sample sizes.

Figure 4 shows that the reliability α of hydrologic time series based on observation is in good coincidence with the reliability α of hydrologic time series based on theoretical RMSE. Therefore, the result of Figure 4 demonstrates that the RMSE and reliability α for estimation of hydrological time series do not depend on the parameters μ and σ^2 of the underlying variable, but only on the sample size n .

4 Conclusion

A stochastic differential equation for flood routing is proposed and developed in order to derive the probability distribution of characteristic flood levels for a built reservoir. The Three Gorges reservoir in the Changjiang River basin of China is selected as a case study. The following conclusions are drawn:

- (a) Both reliabilities α of hydrologic time series and β of original checking flood level (180.4 m) increase with the increasing value of n and information

of flood sample. Reliabilities α and β will approximate 100% with sample size $n \rightarrow +\infty$ when the maximum flood level is less than or equal to 180.4 m in TGR.

- (b) The RMSE and reliability α for estimation of hydrological time series do not depend on the parameters μ and σ^2 of the underlying variable, but only on the sample size n .

Acknowledgements

This study is financially supported by the National Natural Science Foundation of China (51179012) and Open Foundation of State Key Laboratory of Water Resources and Hydropower Engineering Science in Wuhan University (2014SWG02) and the International Cooperation in Science and Technology Special Project of China (2014DFA71910).

References

- [1] Andrade, M. G., A stochastic approach to the flood control problem. *Applied Mathematical Modelling*, **25**(6), pp. 499–511, 2001.
- [2] Li, X., Guo, S., Liu, P. & Chen, G., Dynamic control of flood limited water level for reservoir operation by considering inflow uncertainty. *Journal of Hydrology*, **391**(1–2), pp. 124–132, 2010.
- [3] Zhou Y. & Guo, S., Risk analysis for flood control operation of seasonal flood-limited water level incorporating inflow forecasting error. *Hydrological Sciences Journal*, **59**(5), pp. 1006–1019, 2014.
- [4] Soong, T.T., *Random Differential Equations in Science and Engineering*. McGraw-Hill: New York, pp. 25–31, 1973.
- [5] Yan, B., Guo, S. & Chen, L., Estimation of reservoir flood control operation risks with considering inflow forecasting errors. *Stochastic Environmental Research & Risk Assessment*, **28**(2), pp. 359–368, 2014.
- [6] Cancelliere, A. & Bonaccorso, B., Uncertainty analysis of the standardized precipitation index in the presence of trend. *Hydrology Days*, pp. 14–26, 2009.

