

A reliability-based selective index for regional flood frequency analysis methods

Gwo-Fong Lin* and Lu-Hsien Chen

Department of Civil Engineering, National Taiwan University, Taipei 10617, Taiwan

Abstract:

In this paper, a new index is proposed for the selection of the best regional frequency analysis method. First, based on the theory of reliability, the new selective index is developed. The variances of three regional T -year event estimators are then derived. The proposed methodology is applied to an actual watershed. For each regional method, the reliability of various T -year regional estimates is computed. Finally, the reliability-based selective index graph is constructed from which the best regional method can be determined. In addition, the selection result is compared with that based on the traditional index, root mean square error. The proposed new index is recommended as an alternative to the existing indices such as root mean square error, because the influence of uncertainty and the accuracy of estimates are considered. Copyright © 2003 John Wiley & Sons, Ltd.

KEY WORDS regional flood frequency analysis; reliability; selective index

INTRODUCTION

Regional flood frequency analysis is often needed to estimate T -year event magnitude at sites without observed data. Several regional flood frequency analysis methods have been proposed. How to choose the most suitable or the best regional method for an ungauged site is an important task. For the selection, usually one compares the regional and at-site estimates. For such a purpose, the most common measure of the performance of the regional estimates is the root mean square error (Wiltshire, 1986; Burn, 1988, 1990; Cunnane, 1988; Ribeiro-Correa *et al.*, 1995; GREHYS, 1996, 1997). The root mean square error (RMSE) of the T -year event estimate is:

$$\text{RMSE} = \sqrt{\frac{1}{N_s} \sum_{i=1}^{N_s} \left[\frac{\hat{q}_{R,T}^i - \hat{q}_{AS,T}^i}{\hat{q}_{AS,T}^i} \right]^2} \quad (1)$$

where N_s is the number of stations, $\hat{q}_{AS,T}^i$ and $\hat{q}_{R,T}^i$ are respectively the at-site and regional T -year event estimates of Station i . The regional method that yields the smallest value of root mean square error is the best method.

An important question facing the hydrologists in the frequency analysis, at-site or regional, is the natural uncertainty as well as the model and parameter uncertainty. For the selection of a regional method, the influence of uncertainty has been neglected in the past. Such a limitation has prompted a search for an improved approach to selecting the most suitable regional method. The objective of this paper is to find a new index that is based on reliability. A reliability-based selective index graph is constructed from which the best regional method can be determined. The advantages of the proposed methodology are demonstrated through an actual application.

* Correspondence to: Gwo-Fong Lin, Department of Civil Engineering, National Taiwan University, Taipei 10617, Taiwan.
E-mail: gflin@ntu.edu.tw

A RELIABILITY-BASED SELECTIVE INDEX

When performing a regional analysis, one expects that the regional T -year estimate be as close to the true T -year event as possible for an ungauged site. In addition, one desires that the regional estimate will not underestimate the true event. The probability that the regional estimate will not underestimate the true event is:

$$r = \text{Prob}[q_R \geq q_{AS}] \quad (2)$$

where q_R is the regional estimator and q_{AS} is the at-site estimator for a particular return period T -year (which is regarded as true relative to the regional estimate). Equation (2) can be further written as:

$$r = \text{Prob}[\text{SM} \geq 0] \quad (3)$$

where SM is the safety margin ($\text{SM} = q_R - q_{AS}$). The mean of SM is given by:

$$\mu_{\text{SM}} = E[q_R - q_{AS}] = \mu_{q_R} - \mu_{q_{AS}} \quad (4)$$

where μ_{q_R} is the mean of the regional estimator and $\mu_{q_{AS}}$ is the mean of the at-site estimator. Furthermore, the variance of SM can be obtained as:

$$\sigma_{\text{SM}}^2 = \sigma_{q_R}^2 + \sigma_{q_{AS}}^2 - 2\text{Cov}[q_R, q_{AS}] \quad (5)$$

where $\sigma_{q_R}^2$ is the variance of the regional estimator, $\sigma_{q_{AS}}^2$ is the variance of the at-site estimator, and $\text{Cov}[q_R, q_{AS}]$ is the covariance of the regional and at-site estimators. If the regional and at-site estimators are independent, then

$$\sigma_{\text{SM}}^2 = \sigma_{q_R}^2 + \sigma_{q_{AS}}^2 \quad (6)$$

The assumption of a normal distribution for SM is quite satisfactory when $r \leq 0.99$ (Ang, 1973). If SM is assumed to be normally distributed, then the standardized random variable $U = (\text{SM} - \mu_{\text{SM}})/\sigma_{\text{SM}}$ is also normally distributed. Hence, Equation (3) becomes:

$$r = \text{Prob}[U \geq -(\mu_{\text{SM}}/\sigma_{\text{SM}})] = F_U(\mu_{\text{SM}}/\sigma_{\text{SM}}) \quad (7)$$

where U is the standardized normal random variable and $F_U(u)$ is the cumulative distribution function of the standardized random variable U .

Statistical estimates are often presented with an interval. The concept of interval estimates is adopted herein for the development of the new selective index. The new selective index ω_β is defined as the probability of the regional estimate lying in the interval of $(1 + \beta)q_{AS}$ and $(1 - \beta)q_{AS}$. In other words, ω_β is the probability of the regional estimate lying within βq_{AS} of the true event magnitude q_{AS} . Mathematically, the new selective index ω_β is written as:

$$\omega_\beta = r_{1-\beta} - r_{1+\beta} \quad (8)$$

where

$$r_{1\pm\beta} = \text{Prob}[q_R \geq (1 \pm \beta)q_{AS}] \quad (9)$$

From Equations (8) and (9), one can see that ω_β increases as the value of β increases. When performing the regional frequency analysis, one can construct the relationship of ω_β and β for each regional method. The graph containing the curves of ω_β versus β for various methods is called the 'selective index graph' herein. For a fixed value of β , the method with a larger value of ω_β is a better method according to the definition of the selective index ω_β (see Equation (8)). That is, the best method has the largest probability that the regional estimate lies in the interval of $(1 + \beta)q_{AS}$ and $(1 - \beta)q_{AS}$.

VARIANCE OF REGIONAL ESTIMATORS

As shown in the previous section, we need to know the means and variances of both the regional and at-site estimators. The means and variances of various at-site estimators are already available. In this section, we try to derive the variances of the regional estimators corresponding to three regional methods as follows.

The station-year method

For an ungauged site, the regional T -year event estimator, $\hat{q}_{R,T}$, is given by:

$$\hat{q}_{R,T} = \hat{x}_T \hat{\mu} \quad (10)$$

where $\hat{\mu}$ is the estimator of mean annual maximum flow for the ungauged site, and \hat{x}_T is called the standardized regional T -year event estimator. The standardized annual maximum flow for Year j at Station i is defined as:

$$x_{ij} = q_{ij}/\bar{q}_i \quad (11)$$

where \bar{q}_i is the mean annual maximum flow at Station i .

The generalized extreme value (GEV) distribution is adopted herein. The cumulative distribution function of the GEV distribution is:

$$F_X(x) = \exp \left[- \left(1 - \kappa \frac{x - \xi}{\alpha} \right)^{\frac{1}{\kappa}} \right] \quad (12)$$

where ξ is the location parameter, α is the scale parameter and κ is the shape parameter. Extreme value distributions Type 1, 2 and 3 correspond to $\kappa = 0$, $\kappa < 0$ and $\kappa > 0$, respectively. Equation (12) applies for $-\infty < x < \infty$, $\xi + \alpha/\kappa \leq x < \infty$ and $-\infty < x \leq \xi + \alpha/\kappa$ for extreme value distributions Type 1, 2 and 3, respectively. The at-site T -year event estimator, $\hat{q}_{AS,T}$, is expressed as:

$$\hat{q}_{AS,T} = \hat{\xi} + \frac{\hat{\alpha}}{\hat{\kappa}} \{1 - K_T^{\hat{\kappa}}\} \quad (13)$$

where

$$K_T = -\ln \left(1 - \frac{1}{T} \right) \quad (14)$$

The mean and approximate variance of $\hat{q}_{AS,T}$ can be obtained as (Lu and Stedinger, 1992):

$$E[\hat{q}_{AS,T}] = \xi + \frac{\alpha}{\kappa} [1 - \Gamma(1 + \kappa)] \quad (15)$$

$$\text{Var}[\hat{q}_{AS,T}] = \frac{a^2 \exp[a_0(p) + a_1(p) \exp(-\kappa) + a_2(p)\kappa^2 + a_3(p)\kappa^3]}{n} \quad (16)$$

where p is the cumulative probability; n is the sample size; $a_0(p)$, $a_1(p)$, $a_2(p)$ and $a_3(p)$ are coefficients depending on p . Equation (16) provides good estimates of $\text{Var}[\hat{q}_{AS,T}]$ for $n \geq 10$ and $|\kappa| \leq 0.3$. One can refer to Lu and Stedinger (1992) for a_0 , a_1 , a_2 and a_3 .

Assuming \hat{x}_T and $\hat{\mu}$ are independent, one can obtain the variance of the regional T -year event estimator, $\hat{q}_{R,T}$, as:

$$\text{Var}[\hat{q}_{R,T}] = E[\hat{x}_T]^2 \text{Var}[\hat{\mu}] + E[\hat{\mu}]^2 \text{Var}[\hat{x}_T] \quad (17)$$

where $\hat{\mu}$ can be obtained from a regression model which is a function of mean daily flow, basin area, basin slope, etc. The model is written as:

$$M = BC + \sigma^2 I \quad (18)$$

where

$$M = (\mu_1 \quad \mu_2 \quad \cdots \quad \mu_m)^T \quad (19)$$

$$B = \begin{bmatrix} 1 & b_{11} & b_{12} & \cdots & b_{1n} \\ 1 & b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & b_{m1} & b_{m2} & \cdots & b_{mn} \end{bmatrix} \quad (20)$$

$$C = (c_0 \quad c_1 \cdots c_n)^T \quad (21)$$

where μ_m is the mean annual maximum flow for Station m , b_{mn} is the n th basin characteristic for Station m , and C is a column vector containing the regression coefficients.

The variance of the estimated coefficient \hat{M} is given by:

$$\text{Var}[\hat{M}] = \sigma^2 B'(B'B)^{-1}B \quad (22)$$

where B' is the condition vector and σ^2 is the error variance of the regression model. Once $E[\hat{x}_T]$, $\text{Var}[\hat{x}_T]$ and $\text{Var}[\hat{\mu}]$ are obtained from Equations (15), (16) and (22), the variance of the regional estimator can be determined from Equation (17).

The index-flood method

The Gumbel distribution (extreme value distribution Type 1) is adopted herein. The at-site T -year event estimator can be written as:

$$\hat{q}_{AS,T} = \xi + \alpha y \quad (23)$$

where α and ξ are parameters estimated using the L -moment, and $y = -\ln[-\ln(1 - 1/T)]$ is the Gumbel reduced variate. In terms of sample mean \bar{x} and sample variance s , the estimators of α and ξ are written as:

$$\hat{\alpha} = 0.7797s \quad (24)$$

$$\hat{\xi} = \bar{x} - 0.5772\hat{\alpha} \quad (25)$$

The variance of the at-site T -year event estimator using L -moments is (Phien, 1987):

$$\text{Var}[\hat{q}_{AS,T}] = \frac{\alpha^2[(1.1128 - 0.9066/n) - (0.4574 - 1.1722/n)y + (0.8046 - 0.1855/n)y^2]}{n - 1} \quad (26)$$

where n is the number of years of record. As n increases, Equation (26) converges to:

$$\text{Var}[\hat{q}_{AS,T}] = \frac{\alpha^2[1.1128 - 0.4574y + 0.8046y^2]}{n} \quad (27)$$

The difference between Equations (26) and (27) is less than 5% for $n > 20$. The regional T -year event estimator can be written as:

$$\hat{q}_{R,T} = \hat{\xi}(1 + y\tilde{\sigma}) \quad (28)$$

$$\tilde{\sigma} = \frac{1}{N} \sum_{i=1}^M n_i \frac{\hat{\alpha}_i}{\hat{\xi}_i} \quad (29)$$

where n_i is the number of years of record for the i th station, N is the total years of record, and M is the number of stations for the region.

The variance of the regional T -year event estimator is expressed as:

$$\text{Var}[\hat{q}_{R,T}] = \left(1 + y \frac{\tilde{\alpha}}{\tilde{\xi}}\right)^2 \text{Var}[\hat{\xi}] + 2 \left(1 + y \frac{\tilde{\alpha}}{\tilde{\xi}}\right) (yE[\hat{\xi}])\text{Cov}\left(\hat{\xi}, \frac{\tilde{\alpha}}{\tilde{\xi}}\right) + \xi^2 y^2 \text{Var}\left(\frac{\tilde{\alpha}}{\tilde{\xi}}\right) \tag{30}$$

When $\hat{\xi}$ and $\tilde{\alpha}/\tilde{\xi}$ are independent, then Equation (30) becomes:

$$\text{Var}[\hat{q}_{R,T}] = \left(1 + y \frac{\tilde{\alpha}}{\tilde{\xi}}\right)^2 \text{Var}[\hat{\xi}] + \xi^2 y^2 \text{Var}\left(\frac{\tilde{\alpha}}{\tilde{\xi}}\right) \tag{31}$$

The variance of the at-site $\hat{\alpha}/\hat{\xi}$ can be estimated using the first-order approximation:

$$\text{Var}\left(\frac{\hat{\alpha}}{\hat{\xi}}\right) = \frac{1}{\xi^2} \text{Var}[\hat{\alpha}] + \frac{\alpha^2}{\xi^4} \text{Var}[\hat{\xi}] + 2 \frac{\alpha}{\xi^3} \text{Cov}[\hat{\alpha}, \hat{\xi}] \tag{32}$$

One can refer to Lu and Stedinger (1992) for $\text{Var}[\hat{\alpha}]$, $\text{Var}[\hat{\xi}]$ and $\text{Cov}[\hat{\alpha}, \hat{\xi}]$:

$$\text{Var}[\hat{\alpha}] = \frac{0.8046\alpha^2}{n} \tag{33}$$

$$\text{Var}[\hat{\xi}] = \frac{1.1128\alpha^2}{n} \tag{34}$$

$$\text{Cov}[\hat{\alpha}, \hat{\xi}] = \frac{0.2287\alpha^2}{n} \tag{35}$$

Substituting Equations (33)–(35) into Equation (32) gives $\text{Var}[\hat{\alpha}/\hat{\xi}]$ which is then substituted into Equation (29) to yield:

$$\text{Var}[\tilde{\sigma}] = \frac{1}{N^2} \sum_{i=1}^M \left[n_i \left(\frac{\hat{\alpha}_i}{\hat{\xi}_i}\right)^2 \left(0.8046 - 0.4574 \frac{\hat{\alpha}_i}{\hat{\xi}_i} + 1.1128 \left(\frac{\hat{\alpha}_i}{\hat{\xi}_i}\right)^2 \right) \right] \tag{36}$$

Finally, substituting Equations (34) and (36) into Equation (31), one can obtain the variance of the regional T -year event estimator, $\text{Var}[\hat{q}_{R,T}]$.

The regional-regression model

In this paper, the regional T -year event estimate is expressed as:

$$\hat{y}_T = \ln(\hat{q}_{R,T}) = \hat{\psi}_0 + \sum_{j=1}^n \hat{\psi}_j \ln B_{ij} \tag{37}$$

where $\hat{q}_{R,T}$ is the regional T -year event estimate, B_{ij} is the j th basin characteristic (e.g., basin area, lake index, watershed slope, longitude and latitude) at Station i , and $\hat{\psi}_0, \hat{\psi}_1, \dots, \hat{\psi}_n$ are regression coefficients. The regression coefficients can be obtained from the following model:

$$Y_T = X\Psi + \varepsilon \tag{38}$$

where

$$Y_T = (y_{T,1} \quad y_{T,2} \quad \cdots \quad y_{T,m})^T \quad (39)$$

$$X = \begin{bmatrix} 1 & \ln B_{11} & \cdots & \ln B_{1n} \\ \vdots & \vdots & & \vdots \\ 1 & \ln B_{m1} & \cdots & \ln B_{mn} \end{bmatrix} \quad (40)$$

$$\Psi = (\psi_0 \quad \psi_1 \quad \cdots \quad \psi_n)^T \quad (41)$$

$$\varepsilon = (\varepsilon_1 \quad \varepsilon_2 \quad \cdots \quad \varepsilon_m)^T \quad (42)$$

and $y_{T,m}$ is the T -year logarithmic flow at Station m . The estimator of the parameter Ψ is given by:

$$\hat{\Psi} = [X'X]^{-1}X'Y \quad (43)$$

If the estimate errors $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m$ are independent and each is with zero mean and variance σ_ε^2 , the variance of the parameter Ψ is:

$$\hat{\sigma}_\Psi^2 = \sigma_\varepsilon^2(X'X)^{-1} \quad (44)$$

Once the regression coefficients $\psi_0, \psi_1, \dots, \psi_n$ are obtained, from Equation (37) the regional T -year event estimator is given by:

$$\hat{q}_{R,T} = \exp[\hat{y}_T] \quad (45)$$

and its variance is:

$$\text{Var}[\hat{q}_{R,T}] = \exp[2\hat{y}_T] \text{Var}[\hat{y}_T] \quad (46)$$

where \hat{y}_T and $\text{Var}[\hat{y}_T]$ can be obtained from:

$$\hat{Y}_T = X\hat{\Psi} \quad (47)$$

$$\text{Var}[\hat{Y}_T] = \sigma_\varepsilon^2 X'(X'X)^{-1}X \quad (48)$$

Because σ_ε^2 is unknown usually, an unbiased estimate of σ_ε^2 is given by:

$$s^2 = \frac{\sum_{i=1}^{n_s} (y_{T,i} - \hat{y}_{T,i})^2}{n_s - k - 1} \quad (49)$$

where y_T and \hat{y}_T are the T -year observed and estimated logarithmic flows, respectively, n_s is the sample size and k is the number of independent variables, which is equal to the number of regression coefficients minus one.

APPLICATION AND DISCUSSION

In this paper, actual data of the Tanshui river basin in northern Taiwan are used. Locations of water-level stations are shown in Figure 1. The average slope of the Tanshui River is 1/45, and the area of the basin is 2726 km². There are 16 water-level stations in this basin.

First, at-site (i.e., single-site) frequency analysis for each station is performed. The 2-, 5-, 10-, 25-, 50-, 100- and 200-year floods (called at-site estimates) and their corresponding variances are then found. In dealing with the regional frequency analysis, each time one of the 16 stations is assumed to be ungauged. Based on the data of the remaining 15 stations, then the regional estimates (2-, 5-, 10-, 25-, 50-, 100- and 200-year

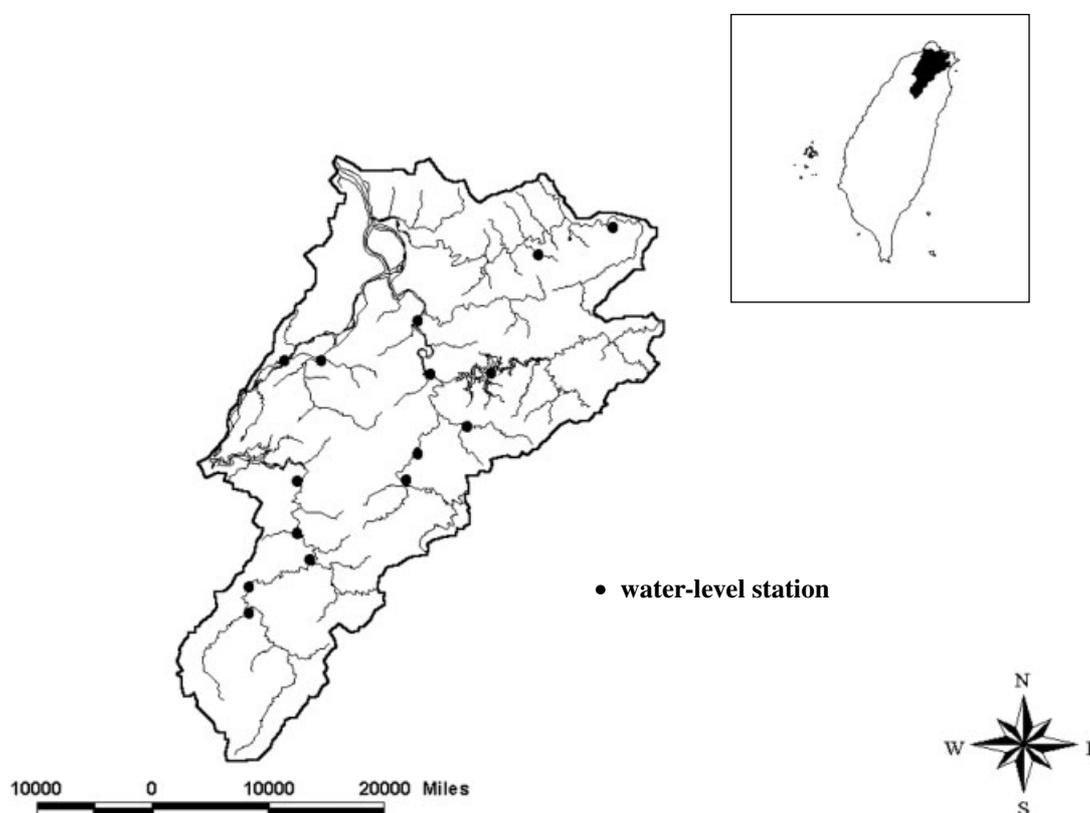


Figure 1. Locations of water-level stations in the Tanshui river basin

floods) and their corresponding variances for the ungauged station are determined. In total, we performed regional analysis 16 times for each regional method. Regarding the regression of the basin characteristics, we choose basin area, lake index, basin slope, longitude and latitude as the basin characteristics. However, we find that lake index, basin slope, longitude and latitude can be neglected according to the statistical t -test.

Once the means and variances of the at-site and regional estimators are found, the reliability of regional estimates can be determined from Equation (7). Table I shows the *average* reliability of regional estimates for the three regional methods based on only the observed data. We use the term 'average' because we perform the regional analysis 16 times. Hence, a value of reliability corresponding to a T -year event in Table I is the average of 16 values. Among three methods the index-flood method yields the highest values of reliability for various T -year estimates as shown in Table I. That is, the index-flood method has the lowest risk of underestimating a T -year event.

To go further into the assessment of the regional estimates, the technique of Monte Carlo simulation is applied to generate synthetic streamflows. We assume these models are correct and simulate from them using the parameters estimated. The average reliability of regional estimates for the three regional methods based on simulated data is summarized in Table II. The result in Table II is similar to that in Table I. Figure 2 compares the reliability of station-year estimates based on the observed and simulated data, respectively. In a like manner, the comparisons for the index-flood estimates and regional-regression estimates are shown in Figures 3 and 4. It is found that the results based on the observed data are in good agreement with those based on the simulated data. The comparison indicates that the observed data closely follow the model for the statistics we calculated.

Table I. The reliability of regional estimates based on the observed data

Return period (year)	Reliability		
	Station-year method	Index-flood method	Regional-regression model
2	0.452	0.524	0.411
5	0.447	0.531	0.368
10	0.440	0.532	0.429
25	0.428	0.533	0.433
50	0.419	0.533	0.451
100	0.413	0.534	0.469
200	0.409	0.534	0.484
Mean	0.430	0.532	0.435

Table II. The reliability of regional estimates based on the simulated data

Return period (year)	Reliability		
	Station-year method	Index-flood method	Regional-regression model
2	0.422	0.504	0.390
5	0.437	0.521	0.348
10	0.430	0.509	0.379
25	0.408	0.553	0.424
50	0.444	0.548	0.461
100	0.430	0.547	0.486
200	0.426	0.554	0.501
Mean	0.428	0.534	0.427

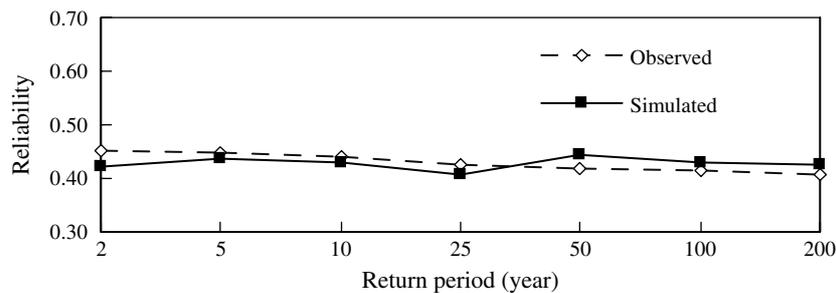


Figure 2. The reliability of station-year estimates based on the observed and simulated data, respectively

Finally, we proceed to establish the relationship of ω_β and β for each regional method. The variation of selective index ω_β with β (from 0.1 to 1.0) for various regional methods is presented in Table III and Figure 5 (called the reliability-based selective index graph). We can see that the selective index ω_β increases as the value of β increases for any method as expected. This is in agreement with the theoretical analysis. As indicated in the aforementioned theoretical analysis, the method with a larger value of ω_β is a better method when the value of β is fixed. Hence, according to Table III and Figure 5, the regional-regression model is the best one. To compare the selection result with that based on the traditional index, root mean square error,

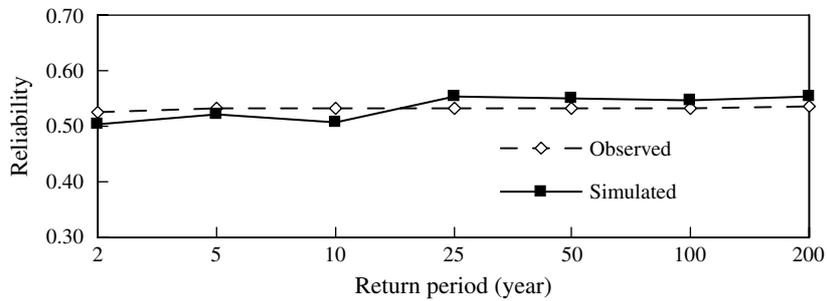


Figure 3. The reliability of index-flood estimates based on the observed and simulated data, respectively

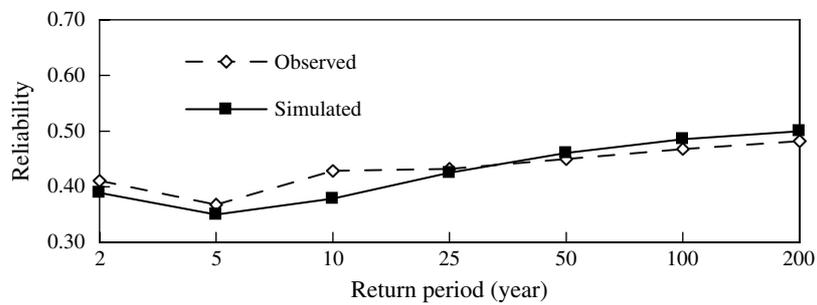


Figure 4. The reliability of regional-regression estimates based on the observed and simulated data, respectively

Table III. Values of the reliability-based selective index ω_β for various regional methods

β	ω_β		
	Station-year method	Index-flood method	Regional-regression model
0.1	0.12	0.15	0.18
0.2	0.24	0.29	0.36
0.3	0.34	0.41	0.52
0.4	0.43	0.51	0.65
0.5	0.50	0.59	0.77
0.6	0.57	0.65	0.85
0.7	0.62	0.70	0.90
0.8	0.67	0.73	0.93
0.9	0.71	0.77	0.95
1.0	0.73	0.79	0.96

we compute the root mean square error of various T -year event estimates for the three regional methods. As shown in Table IV, the station-year method is the best method if we use RMSE as the selective index.

SUMMARY AND CONCLUSIONS

In this paper, a new selective index is proposed for the selection of the best regional flood frequency analysis method. The new selective index is based on the theory of reliability. Actual application of the proposed methodology is performed. Three regional frequency analysis methods, namely the station-year,

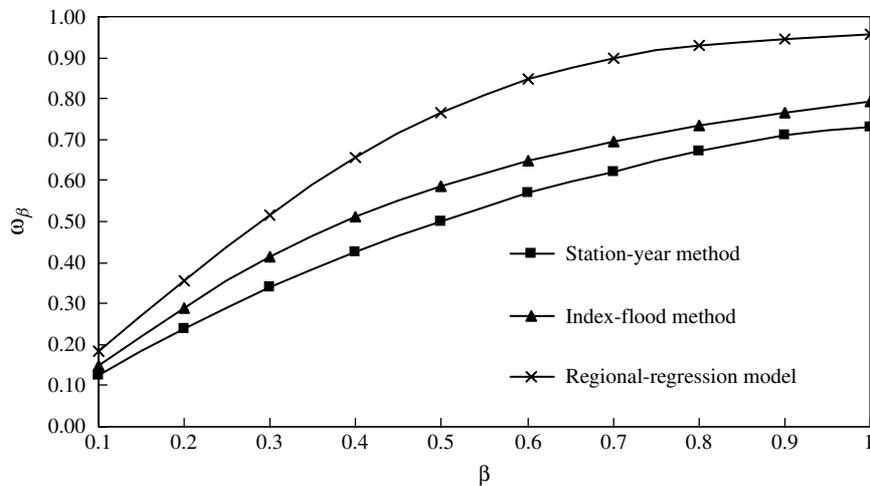


Figure 5. The variation of the reliability-based selective index ω_β with β for various regional methods

Table IV. The RMSE for various regional methods

Return period (year)	RMSE		
	Station-year method	Index-flood method	Regional-regression model
2	0.225	0.285	0.289
5	0.225	0.221	0.277
10	0.205	0.191	0.182
25	0.225	0.200	0.188
50	0.225	0.266	0.204
100	0.225	0.278	0.252
200	0.225	0.288	0.334
Mean	0.222	0.247	0.246

index-flood and regional-regression models, are considered in the application. The selection result based on the new index is different from that based on the traditional index (root mean square error). The proposed new index has advantages over existing indices, because it considers the influence of variance of estimates.

REFERENCES

- Ang AH-S. 1973. Structural risk analysis and reliability-based design. *Journal of the Structural Engineering Division, ASCE* **99**(ST9): 1891–1910.
- Burn DH. 1988. Delineation of groups for regional flood frequency analysis. *Journal of Hydrology* **104**: 345–361.
- Burn DH. 1990. An appraisal of the 'region of influence' approach to flood frequency analysis. *Hydrological Science Journal* **35**(2): 149–165.
- Cunnane C. 1988. Methods and merits of regional flood frequency analysis. *Journal of Hydrology* **100**: 269–290.
- GREHYS (Groupe de recherche en hydrologie statistique). 1996. Presentation and review of some methods for regional flood frequency analysis. *Journal of Hydrology* **186**: 63–84.
- GREHYS (Groupe de recherche en hydrologie statistique). 1997. Intercomparison of regional flood frequency procedures Canadian rivers. *Journal of Hydrology* **186**: 85–103.
- Lu LH, Stedinger JR. 1992. Variance of two- and three-parameter GEV/PWM quantile estimators: formulae, confidence intervals, and a comparison. *Journal of Hydrology* **138**: 247–267.

- Phien HN. 1987. A method of parameter estimator for the extreme value type I distribution. *Journal of Hydrology* **90**: 251–268.
- Ribeiro-Correa J, Cavadas G, Rousselle J. 1995. Identification of hydrological neighborhoods using canonical correlation analysis. *Journal of Hydrology* **173**: 71–89.
- Wiltshire SE. 1986. Identification of homogeneous regions for flood frequency analysis. *Journal of Hydrology* **84**: 287–302.