Interpretation of New Trends in Flood Frequency Analysis: A Case Study

A. Melih YANMAZ, Engin Gunindi, and Elcin KENTEL

Middle East Technical Univ., Civil Eng. Department, 06531, Ankara, Turkey,
email: myanz@metu.edu.tr

Abstract

Conventional flood frequency analyses are based on use of a suitable univariate probability density function for maximum annual discharges. This approach is not robust enough since it deals only with the statistical distribution of annual series of peak discharges. However, flood events are better characterized by the peak discharge, volume of runoff, and the duration of the flood jointly. That is why there is a growing trend in the formulation of multivariate flood frequency analyses. A bivariate flood frequency analysis is performed in which the annual maximum peak discharges and the corresponding surface runoff volumes of flood are handled as random variables. The results of both approaches, i.e. univariate and multivariate cases show differences. Therefore, choice of a method for such a study would influence the design and operation of hydraulic structures for high flows. In this paper, a case study has been performed for a basin in Turkey to account for the effect of the differences in the application of the aforementioned approaches. Various bivariate probability density functions have been applied and tested for goodness. Use of these analyses and suitability of these functions in characterizing the nature of flood events are discussed. The results of a bivariate analysis are also compared with the findings of a univariate analysis carried out using a number of probability density functions. Furthermore, an example is presented to illustrate the effect of the approach used in the flood frequency analysis on the dimensions of a flood detention dam located in the basin.

***Key Words**: Flood Frequency, Bivariate Analysis, Peak Discharge, Surface Runoff Volume*

Introduction

Design of hydraulic structures is based on the determination of design discharge, which corresponds to a certain return period, compatible to the local site characteristics. A flood frequency analysis needs to be carried out for this purpose. Conventional flood frequency analysis is performed using annual flood peak discharges to obtain extreme flood peaks. However, such applications are normally incapable of giving adequate information for the floods since the whole event would be modelled more correctly by the joint consideration of flood peaks, volumes, and durations (Yanmaz and Gunindi, 2006). That is why there is a growing research activity on multivariate flood frequency analysis. Number of previous studies, which dealt with multivariate flood frequency analysis, are limited. Sackl and Bergmann (1987) and Goel et al. (1998) used bivariate normal distribution. Application of trivariate gumbel distribution (Escalante-Sandoval and Raynal-Villaseñor, 1998), bivariate lognormal distribution (Yue, 2000; Yue, 2002), and bivariate gamma distribution (Yue, 2001; Yue et al., 2001) are some examples of sophisticated frequency analyses. This paper deals with application of some bivariate probability density functions to flood frequency analysis. To this end, five-parameter bivariate gamma, two-parameter bivariate lognormal, and two-parameter bivariate Gumbel probability density functions (PDFs) will be used.

Hydrologic events having skewed distributions, such as flood peak discharge and flood volume may follow gamma type probability distribution (Bobée and Ashkar, 1991; Stedinger et al., 1993). Gamma distribution has the advantage of having only positive values and it does not require logarithmic transformation. The joint probability density function (PDF) and cumulative density function (CDF) of the five-parameter bivariate gamma distribution with variates X and Y are as follows (Smith et al., 1982; Yue, 2001):

$$f(x,y)=\begin{cases}\frac{K_1}{K_2}\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}c_{jk}\left(\beta_x x\right)^j\left(\eta\beta_y y\right)^{j+k} & if\,\rho>0\\ f_x(x).f_y(y) & if\,\rho=0\end{cases} \tag{1}$$

$$F(x,y)=\begin{cases}J\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}d_{kj}H\left(\gamma_x+j,\frac{\beta_x x}{1-\eta}\right)H\left(\gamma_y+j+k,\frac{\beta_y y}{1-\eta}\right) & if\,\rho>0\\ F(x).F(y) & if\,\rho=0\end{cases} \tag{2}$$

in which $x, y \geq 0,\ \ 0<\eta<1,\ \ \gamma_y \geq \gamma_x,\ \ and\ \ 0 \leq \rho < \eta\sqrt{\gamma_x/\gamma_y}$; and

$$K_1=\left(\beta_x x\right)^{\gamma_x-1}\left(\beta_y y\right)^{\gamma_y-1}\exp\left(-\frac{\beta_x x+\beta_y y}{1-\eta}\right) \tag{3}$$

$$K_2=\left(1-\eta\right)^{\gamma_x-1}\ \Gamma\left(\gamma_x\right)\Gamma\left(\gamma_y-\gamma_x\right) \tag{4}$$

$$c_{kj}=\frac{\Gamma\left(\gamma_y-\gamma_x+k\right)}{\left(1-\eta\right)^{2j+k}\,\Gamma\left(\gamma_y+j+k\right)j!k!} \tag{5}$$

$$\eta=\rho\sqrt{\frac{\gamma_y}{\gamma_x}} \tag{6}$$

$$J=\frac{\left(1-\eta\right)^{\gamma_y}}{\Gamma\left(\gamma_x\right)\Gamma\left(\gamma_y-\gamma_x\right)} \tag{7}$$

$$d_{kj}=\frac{\eta^{j+k}\,\Gamma\left(\gamma_y-\gamma_x+k\right)}{\Gamma\left(\gamma_y+j+k\right)j!\,k!} \tag{8}$$

$$H\left(a,z\right)=\int_0^z t^{a-1}e^{-t}dt \tag{9}$$

$$\Gamma\left(z\right)=\int_0^\infty t^{z-1}e^{-t}dt \tag{10}$$

where γ and β are the scale and shape parameters of the marginal gamma distributions, respectively. In Eq. (6), η is defined as the association parameter between X and Y, and ρ is the product-moment correlation coefficient of X and Y.

Another suitable probability distribution, which can be applicable to extreme events, such as annual maximum discharges, is the Gumbel distribution. The PDF and CDF of the bivariate Gumbel distribution are (Yue and Rasmussen, 2002):

$$
\begin{aligned}
f(x,y) &= \frac{\partial^2 F(x,y)}{\partial x \partial y} \\
&= \frac{F(x,y)}{\alpha_x \alpha_y} \left\{ e^{-\left[m(x-u_x)\right]/\alpha_x} + e^{-\left[m(y-u_y)\right]/\alpha_y} \right\}^{(1-2m)/m} \left(\left\{ e^{-\left[m(x-u_x)\right]/\alpha_x} + e^{-\left[m(y-u_y)\right]/\alpha_y} \right\}^{1/m} + m - 1 \right) \\
&\quad \times \exp\left[-m\left(\frac{x-u_x}{\alpha_x} + \frac{y-u_y}{\alpha_y} \right) \right]
\end{aligned} \quad (11)
$$

In which *F (x, y)* is given by

$$
F(x,y) = \exp\left\{ -\left[\left(-\ln F_X(x)\right)^m + \left(-\ln F_Y(y)\right)^m \right]^{1/m} \right\} \qquad (m \geq 1) \quad (12)
$$

where $F_X(x)$ and $F_Y(y)$ are the marginal distributions of random variables *X* and *Y*, respectively, and *m* (m≥1) is the parameter describing the association between the two random variables *X* and *Y*.

Bivariate lognormal distribution is another choice that can be used to represent joint distributions of flood peaks and flood volumes. It has the advantage of having positive values and logarithmic transformation tends to reduce positive skewness (Chow et al., 1988). Equations used to calculate the CDF of bivariate lognormal distribution are given below (Yue, 2000). A positive random variable *X* is said to be lognormally distributed with the parameters mean (μ) and standard deviation (σ) if *Y=lnX* is normally distributed with mean μ and variance σ^2. If two correlated continuous random variables X_1 and X_2 are lognormally distributed with different parameters (mean and standard deviation) as follows:

$$
f(x_1) = \frac{1}{\sigma_{Y_1}\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\ln(x_1) - \mu_{Y_1}}{\sigma_{Y_1}} \right)^2 \right] \quad (X_1 > 0) \quad (13)
$$

$$
f(x_2) = \frac{1}{\sigma_{Y_2}\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\ln(x_2) - \mu_{Y_2}}{\sigma_{Y_2}} \right)^2 \right] \quad (X_2 > 0) \quad (14)
$$

where μ_Y and σ_Y are the mean and standard deviation of *Y*, respectively, then the joint probability distribution of these two variables can be represented by the bivariate lognormal distribution. The PDF of the bivariate lognormal distribution can be derived using the Jacobian of the transformation and is given by:

$$
f(x_1,x_2) = \frac{1}{2\pi x_1 x_2 \sigma_{Y_1}\sigma_{Y_2}\sqrt{1-\rho^2}} \exp\left[-\frac{q^2}{2} \right] \quad (-1 < \rho < +1) \quad (15)
$$

$$
q = \frac{1}{1-\rho^2} \exp\left[\left(\frac{\ln(x_1) - \mu_{Y_1}}{\sigma_{Y_1}} \right)^2 - 2\rho\left(\frac{\ln(x_1) - \mu_{Y_1}}{\sigma_{Y_1}} \right)\left(\frac{\ln(x_2) - \mu_{Y_2}}{\sigma_{Y_2}} \right) + \left(\frac{\ln(x_2) - \mu_{Y_2}}{\sigma_{Y_2}} \right)^2 \right] \quad (16)
$$

where μ_{Y_i} and σ_{Y_i} are the mean and standard deviation of Y_i (*i*=1,2), respectively. As the CDF of the bivariate distribution is not attainable analytically, it is calculated by numerically integrating the corresponding PDF. The following section is devoted to application of these functions.

Application

Use of the aforementioned bivariate distributions will be illustrated in a case study, which is selected as Fol Creek Basin in Eastern Black Sea region of Turkey. This basin is subject to frequent floods, which have resulted in loss of several lives and considerable damage to all types of facilities. Fol Creek is an elongated basin with an area of approximately 220 km^2. There is a recording gauging station close to the outlet of the basin. The main stream is situated in a V-shaped valley which has steep side slopes. The overall runoff coefficient in the basin is quite large due to the steep slopes, mainly clayey and rocky formations (Yanmaz and Coskun, 1995). Town of Vakfikebir, which is situated at the outlet of the basin, is adversely affected by the frequent floods since various structures and establishments are located within the floodplains. Several training and bank protection facilities have been proposed by Yanmaz and Bilen (2000) to increase the flood carrying capacity of the river and hence to inhibit the inundation of the adjacent land. In addition to these facilities, implementation of a flood detention dam is assumed to be an effective measure to cope with floods. To this end, local topographic and geologic conditions were investigated and construction of a flood detention dam at a suitable axis, which is approximately 2.5 km upstream of the Black Sea shoreline, is proposed (Yanmaz and Gunindi, 2006). Yanmaz and Gunindi (2006) carried out a flood frequency analysis to observe the effect of frequency analysis on the design of the upstream flood detention dam. The annual series of peak discharges, which were recorded by the stream-gauging station (Q_p), are presented in Table 1 together with the corresponding direct runoff volume (V).

Table 1. Annual Maximum Flood Events of the Fol Creek.

Water Year	Q_p (m^3/s)	V (10^6 m^3)
1972	35.30	3.92
1976	31.00	1.50
1979	14.80	2.39
1980	73.50	3.19
1983	177.00	8.02
1984	41.15	3.82
1985	35.40	2.21
1986	45.00	2.23
1987	72.50	2.66
1988	52.80	3.27
1989	55.20	4.17
1993	41.79	2.10
1995	102.09	3.90
1996	101.96	2.54
1999	71.60	1.79
2000	196.72	3.21
2001	31.66	1.78
2002	74.08	1.48
2003	65.52	5.66
2004	71.12	4.10

Bivariate flood frequency analyses are carried out using the aforementioned three different joint probability distributions for the flood peak discharges and direct runoff volumes. To examine the goodness of fit of each one of these distributions to the flood peak discharges and volumes, the

Kolmogorov-Smirnov (KS) test (Johnson and Kotz, 1972) is applied for two significance levels of 0.05 and 0.10. The critical values of the KS test are 0.29 and 0.26 for significance levels of 0.05 and 0.10, respectively. The KS test statistics for bivariate gamma, Gumbel, and lognormal distributions are given in Table 2. As can be seen from Table 2, the KS statistics for both flood peak discharge and flood volume are smaller than the critical values for all three distributions. Thus, the null hypothesis that the underlying distribution is equivalent to the specified distribution is accepted. Therefore, the marginal and joint distributions of flood peak discharge and flood volume can be taken as either bivariate gamma, Gumbel or lognormal.

Table 2. Kolmogorov-Smirnov Statistics.

Distribution	Q_p	V
Gamma	0.236	0.247
Gumbel	0.190	0.123
Lognormal	0.146	0.105

To obtain a relation between Q_p and V, successive values of CDF are considered and the corresponding values of these variables are found. As a result, sets of equal return period curves correlating Q_p-V pairs are obtained as shown in Figures 1 through 3 in which T_r is the return period. Although these curves extend asymptotically along the axes, they should be limited by proper upper and lower bounds since very large values of these variables have no physical significance. To this end, the criterion suggested by Hable (2001) is used in which the aforementioned curves are bounded by lines passing through the origin having slopes of min $\{V_i/Q_{pi}$, i=1,2,.., N$\}$ and max $\{V_i/Q_{pi}$, i=1,2,.., N$\}$, where N=20 is the total number of sample data (Figures 1 through 3). Since the available record length is limited by 20 years, forecasts for very big return periods may be subject to high uncertainty. Recommendation of a suitable distribution function may be based on relative comparison of their results obtained under various return periods. Bivariate Gumbel and lognormal distributions resulted in very close values, whereas the lower bound for V obtained from bivariate gamma distribution was found to be greater than that of the remaining distributions for T_r=5 years. For T_r=25 years, the results of bivariate lognormal and gamma distributions were close to each other. On the other hand, three distributions yielded almost relatively close results for T_r=50 years. In case of T_r=100 years, bivariate gamma distribution gave results, which were almost the average of the other two distributions. Prior to a sensitivity analysis to observe the effects of such differences on the design, it may be helpful to compare the previous findings with the results of univariate analyses.

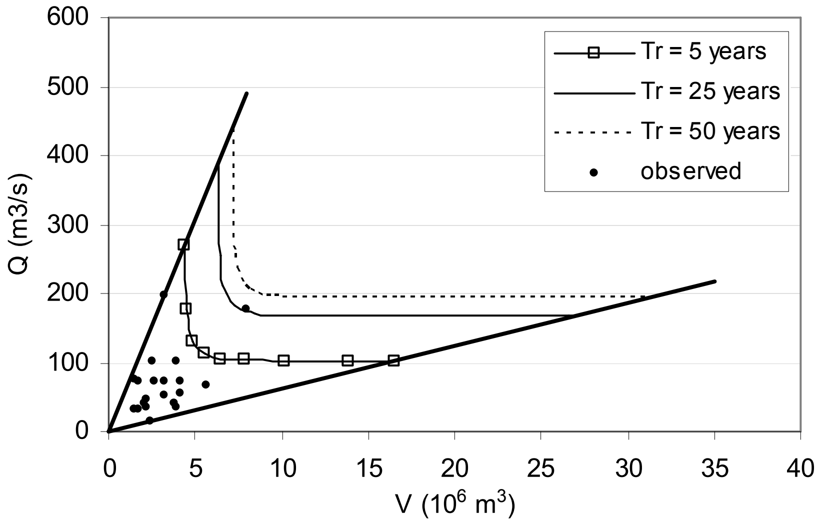

Figure 1. Correlation between peak discharges and volumes for bivariate gamma distribution.

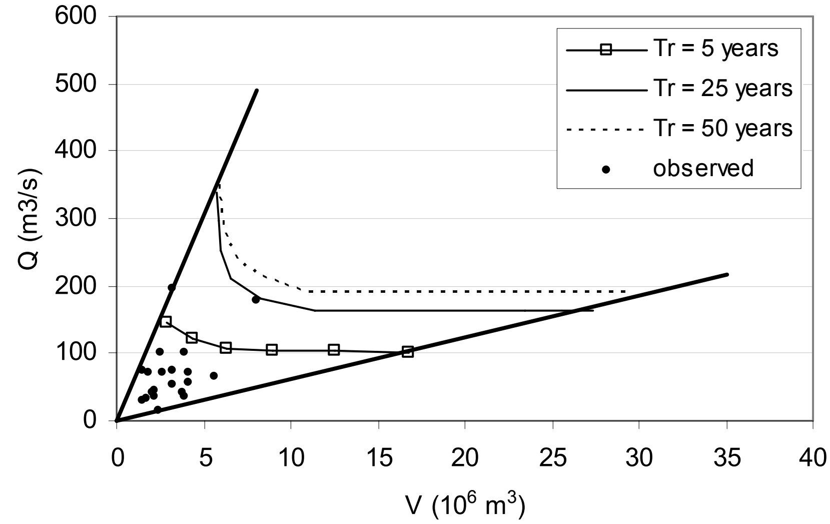

Figure 2. Correlation between peak discharges and volumes for bivariate Gumbel distribution.

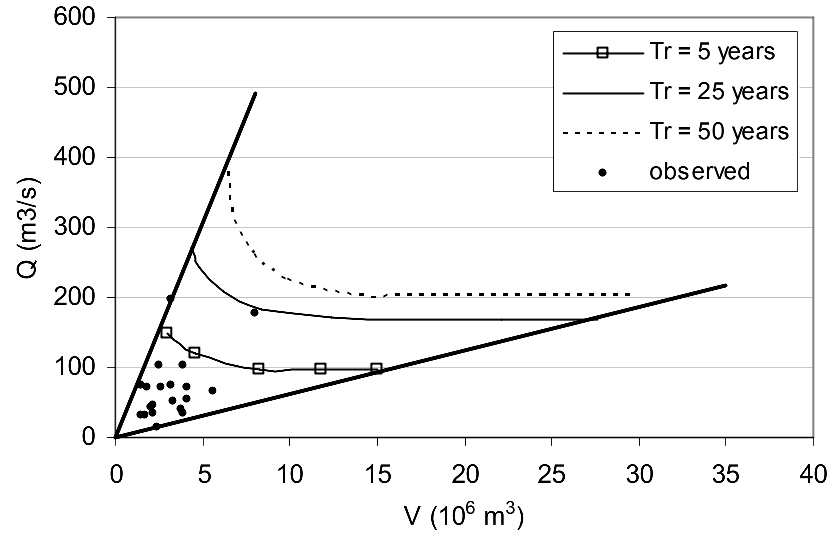

Figure 3. Correlation between peak discharges and volumes for bivariate lognormal distribution.

In a previous study carried out by Yanmaz and Bilen (2000) a univariate flood frequency analysis for the annual maximum flows of the Fol Creek basin has been carried out using a computer program developed by Hosking, which was available at http://lib.stat.cmu.edu/general/lmoment. The program determines the magnitude of discharges corresponding to specified return periods for various probability distributions using the theory of L-moments for parameter estimation. It is applicable to either regional or at-site frequency analysis. Since there is only one stream-gauging station in the Fol Creek basin, the “at-site” option of the program was run for 12 different probability distributions (gamma (G), Generalized Extreme Value (GEV), Generalized Normal (GN), Extreme Value Type 1 (EV1), Normal (N), Pearson Type III (PE3), Generalized Pareto (GPA), Generalized Logistic (GL), Kappa (KAP), Wakeby (WAK), 2-Parameter Log-normal (LN2), and log-Pearson Type III (LPT3)). The average flood peak values of the univariate flood frequency analysis are compared with the approximate peak discharge ranges obtained from the bivariate analysis in Table 3. As can be seen from this table, bivariate analysis yielded a wide range of peak discharges within the upper and lower bounds for the return periods considered. Another result of this analysis is that the average peak discharges obtained from the univariate analysis are smaller than the lower limit of the bivariate cases. Therefore, the design performed using univariate analysis would lead to underestimates compared to the design obtained by using bivariate analysis.

Table 3. Comparison of Q_p values

T_r (years)	Univariate	Gamma	Gumbel	Lognormal
5	92	103-269	102-145	95-149
25	145	168-392	164-340	169-272
50	166	194-440	190-350	203-379
100	188	220-486	215-371	238-479

The effect of this analysis on the dam height can be investigated using the preliminary dimensions of the spillway having a height and length of 25 m and 40 m, respectively. The dam crest elevation, H_c, is computed by adding a 2 m of freeboard to the maximum lake elevation that can occur under various discharges listed in Table 3. To this end, the lower and upper values of the bounds of peak discharges defined for the aforementioned bivariate functions are used. Computations are performed relative to the univariate case for T_r=100 years, for which the peak discharge is 188 m^3/s. The results of this analysis are shown in Figure 4. As can be observed from this figure, percent change in dam crest elevation is almost linearly proportional to the percent change of discharges relative to the univariate case. The results clearly indicate that bivariate probability density functions result in more conservative values compared to the average discharges obtained from the univariate analysis. Similar discussions would also apply to an existing structure. The safety margin or the freeboard of the structure decreases for bivariate functions. These results would also imply the necessity of verification of the capacity of existing spillways, which were designed using conventional approaches, with respect to the aforementioned bivariate functions.

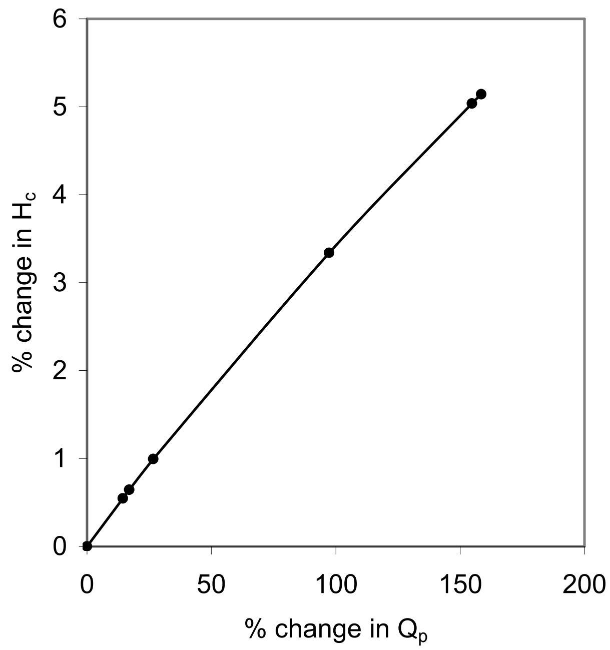

Figure 4. Variation of percent change in H_c with respect to percent change in Q_p.

Conclusions

Applicability of bivariate probability density functions to flood frequency analysis is investigated. To this end, bivariate forms of gamma, Gumbel, and lognormal distributions are applied to a flood frequency analysis in a case study. Tests of goodness are carried out for these functions using the Kolmogorov-Smirnov test under 5 and 10% significance levels and it was observed that all three distributions fit to the sample data. Based on this verification, the flood peak discharges and flood volumes are determined using these functions for various return periods. Equal return period curves relating these variables are then generated for the aforementioned distributions. Meaningful data are obtained for pairs of flood peak discharges and flood volumes in certain bounds for physical significance. Solution to bivariate Gumbel distribution is easier than the remaining functions since its analytical form exists, whereas numerical integration is required for gamma and lognormal distributions. The results of bivariate analyses are also compared with the findings of a univariate analysis, which was performed using 12 different probability distributions. The average peak discharge values of the univariate analysis are found to be smaller than the lower bounds of all bivariate probability density functions tested in this study. Therefore, it may be concluded that use of bivariate flood frequency analysis can provide more conservative results in the design of a dam.

References

Bobée, B., and Ashkar, F. (1991). The Gamma Family and Derived Distributions Applied in Hydrology, Water Resources Publications, Littleton, CO.

Chow, V. T., Maidment, D. R., and Mays, L. W. (1988). Applied Hydrology. Singapore, McGraw Hill.

Escalante-Sandoval, C.A., J. A. Raynal Villaseñor. (1998). *Multivariate estimation of floods: the trivariate gumbel distribution*. J. Statist. Comput. Simul., OPA N.V., India, 61, 313-340.

Goel, N. K., Seth, S. M., and Chandra, S. (1998). *Multivariate modeling of flood flows*. Journal of Hydraulic Engineering, 124(2), 146-155.

Hable, O. (2001). Multidimensional probabilistic design concept for the estimation of the overtopping probability of dams. Ph.D. thesis, Technical University of Graz.

Johnson, N. L. and Kotz, S. (1972). Distribution in Statistics: Continuous Multivariate Distributions. John Wiley and Sons, New York.

Sackl, B. H., Bergmann, (1987). A bivariate flood model and its application in Hydrologic Frequency Modeling. V. P. Singh (ed.), D. Reidel Pub. Company, The Netherlands, 571-582.

Smith, O. E., S.I. Adelfang J.D., and Tubbs, (1982). A Bivariate gamma probability distribution with application to gust modeling, NASA Technical Memorandum NASA TM-82483, Alabama.

Stedinger, JR, Vogel, R.M., and Foufoula-Georgiou, E. (1993). Frequency analysis of extreme events, Handbook of Hydrology.

Yanmaz, A. M., and Coskun, F. (1995). *Hydrological aspects of bridge design: case study*, Journal of Irrigation and Drainage Engineering, 121(6), 1-8.

Yanmaz, A. M. and Bilen, S. (2000). *Selection of optimum riverbank stabilization facility*, Proc. 4th Int. Conf. on Hydroscience and Engineering, Seoul, Korea.

Yanmaz, A. M., and Gunindi, E. (2006). *Probabilistic Assessment of Overtopping Reliability of a Dam*. Proceedings of International Conference on Water Observation and Information System for Decision Support, Ohrid, Macedonia, 2006.

Yue, S. (2000). *The bivariate lognormal distribution to model a multivariate flood episode*. Hydrological Processes. John Wiley and Sons Ltd., 14, 2575-2588.

Yue, S. (2001). *A bivariate gamma distribution for use in multivariate flood frequency analysis*. Hydrological Processes. John Wiley and Sons Ltd., 15, 1033-1045.

Yue, S., Ouarda, T.B.M.J., and Bobée, B. (2001). *A review of bivariate gamma distributions for hydrological application*, Journal of Hydrology, Elsevier Science Ltd., 246, 1-18.

Yue, S. and Rasmussen, P. (2002). *Bivariate frequency analysis: discussion of some useful concepts in hydrological application*. Hydrological Processes, John Wiley and Sons Ltd., 16, 2881-2898.

Yue, S. (2002). *The bivariate lognormal distribution for describing joint statistical properties of a multivariate storm event*. Environmetrics. John Wiley and Sons Ltd., 13, 811-819.