Stochastic Environmental Resea (2004) 18: 61–66
DOI 10.1007/s00477-003-0124-x

ORIGINAL PAPER

S. Yue · C. Y. Wang

A comparison of two bivariate extreme value distributions

Abstract There are two distinct bivariate extreme value distributions constructed from Gumbel marginals, namely Gumbel mixed (GM) model and Gumbel logistic (GL) model. These two models have completely different structures and their dependence ranges are different. The product-moment correlation coefficient for the former is $\rho \in [0, 2/3]$ and the latter is $\rho \in [0, 1]$. It is natural to ask which one is more appropriate for representing the joint probabilistic behavior of two correlated Gumbel-distributed variables. This study compares these two models by numerical experiments. The comparison is based on that: (i) if the two distribution models are identical, then the joint probability and the joint return period computed by the GM model should be the same as those by the GL model; and (ii) if a selected distribution is the true distribution from which sample data are drawn, then the probabilities computed by the theoretical model should provide a good fit to empirical ones. Comparison results indicate that in the range of correlation coefficient $\rho \in [0, 2/3]$, both models provide identical joint probabilities and joint return periods, and both indicate a good fit to empirical probabilities; while for $\rho \in (2/3, 1)$, only the Gumbel logistic model can be used.

Keywords Gumbel distribution · Bivariate extreme value distribution · Gumbel mixed model · Gumbel logistic model · Correlation

S. Yue (✉)
US Environmental Protection Agency
Mid-Continent Ecology Division, 6201 Congdon Blvd.
Duluth, MN 55804, USA
E-mail: yue.sheng@epa.gov

C. Y. Wang
Water Resources & Hydropower Section
One-evaluation Dept, China Development Bank
Beijing 100073, China

1 Introduction

The hydrological extreme events such as floods and storms may appear to be multivariate events. To effectively fulfill hydraulic structure design and management, one needs to understand joint statistical properties of these multivariate hydrological events. Multivariate probability distributions can be implemented to represent joint statistical behavior of these events. The Gumbel distribution or extreme value type I (EVI) distribution is one of most frequently adopted distribution types for modeling hydrological extreme events such as floods and storms (Gumbel 1958; Todorovic 1978; Castillo 1988; Stedinger et al., 1993). The study of the bivariate extreme value distribution will be of interest to hydrological engineers for analyzing the joint probabilistic behavior of two correlated Gumbel distributed hydrological events.

In the statistical literature, a few bivariate extreme value distribution models have been developed and studied (see for examples, Gumbel, 1960a, b, 1961; Gumbel and Mustafi, 1967; Oliveria, 1975, 1982; Pickands, 1981; Buishand, 1984; Raynal-Villasenor and Salas, 1987; Tawn, 1988; Joe et al., 1992; and Coles and Tawn, 1991, 1994). These distribution models have mainly remained their theoretical development and have seldom been applied to resolve practical problems in multivariate frequency analyses in hydrological/environmental science. The recent studies of Yue et al. (1999, 2000) and Yue (2000a, b, 2001a, b) provided practical applications of two explicit bivariate extreme distributions, namely the Gumbel mixed (GM) model and the Gumbel logistic (GL) model proposed by Gumbel (1960a, b, 1961) to represent multivariate hydrological and meteorological events. These two models are constructed from Gumbel marginals and are only two differentiable bivariate extreme distribution models. As both models have the same marginal distributions, it is natural to ask which model is more appropriate for representing joint probabilistic properties of extreme

Table 1 Statistics of X and Y and parameters of the Gumbel distributions

ρ	X				Y			
	Statistics		Parameters		Statistics		Parameters	
	M	S	α	u	M	S	α	u
0.1	16.43	11.89	9.27	11.08	17.53	9.02	7.03	13.47
0.3	13.19	5.71	4.45	10.62	25.55	13.27	10.35	19.58
0.5	13.09	6.29	4.90	10.26	27.28	14.79	11.53	20.62
0.6	10.00	6.11	4.77	7.24	29.94	14.64	11.42	23.35
2/3	16.35	9.03	7.04	12.29	23.37	11.27	8.79	18.30
0.9	22.99	9.11	7.10	18.89	15.64	7.79	6.07	12.14

events. A referee of Yue (2001b) also suggested that the comparison of these two models should be made. A simple comparison between these two models was made there using only one practical sample.

This study makes a detailed comparison of the suitability of these two models for representing joint properties of two correlated Gumbel-distributed random variables. As it is almost impossible to get a number of combinations of two correlated Gumbel-distributed random variables with expected correlation coefficients in the practical world, Monte Carlo simulation are employed to generate sample data with the Gumbel distribution. Simulation can provide us various devised combination scenarios of two random variables.

2 Comparison

The joint cumulative distribution functions of these two models are presented in Appendix A. Both the GM model and the GL model are constructed from the Gumbel marginal distributions. To mimic this situation, a number of Gumbel-distributed univariates X and Y are generated first. Then, a number of combinations of X and Y with certain dependence can be selected among these generated data.

2.1 Generation of Gumbel distributed random variables

A random variable $Z(=X, Y)$ is said to follow the Gumbel distribution (or EVI distribution) if Eq. (A6) (see Appendix A) can be used to represent the sample of Z. The inverse form of the distribution is given by

$$z = u - \alpha \ln(-\ln F_Z) \quad F_Z \in [0, 1]\,(z = x, y) \tag{1}$$

The random variable F_Z is uniformly distributed on the interval [0, 1]. It was generated with sample size of 100 by the approach of Lewis et al. (1969). Then, by using Eq. (1) with given location and scale parameters u and α, the corresponding value of Z was obtained. A number of samples of X and Y were generated by this approach, respectively. Then six pairs of two variables, X and Y, with Pearson's product-moment correlation coefficient (PPMCC) $\rho = 0.1,\ 0.3,\ 0.5,\ 0.6,\ 0.667(2/3),\ 0.9$ were selected among these generated sample data, respectively. The means and standard deviations of X and Y and parameters of the Gumbel distributions are presented in Table 1. The association parameters for the GM and the GL models were computed using Eqs. (A3) and (A4), respectively. They are given in Table 2.

2.2 Empirical and theoretical joint probabilities of X and Y

Empirical joint non-exceedance probabilities were computed using the approach proposed by Yue et al. (1999). A two-dimensional table is first constructed in which the variables X and Y are arranged in ascending order. The element (n_{ml}) in row m and column l of the table is the number of the concurrence of these two variables. The joint cumulative frequency (non-exceedance joint probability) of the combinations of x_i and y_j is then given as:

$$F(x, y) = \Pr(X \le x_i, Y \le y_j) = \frac{\sum_{m=1}^{i} \sum_{l=1}^{j} n_{ml} - 0.44}{N + 0.12} \tag{2}$$

where N is the total number of sample size ($N = 100$).

Theoretical joint probabilities of the real occurrence combinations of x_i and y_j for the GM and GL were estimated using Eqs. (A1) and (A2), respectively. Figures 1–5 displays the empirical and theoretical joint probabilities of X and Y with correlation coefficients

Table 2 Association parameter between X and Y

ρ	Association parameters between X and Y	
	Gumbel mixed model θ	Gumbel logistic model η
0.1	0.162	1.054
0.3	0.474	1.195
0.5	0.768	1.414
0.6	0.908	1.581
2/3	1.000	1.732
0.9	1.307	3.162

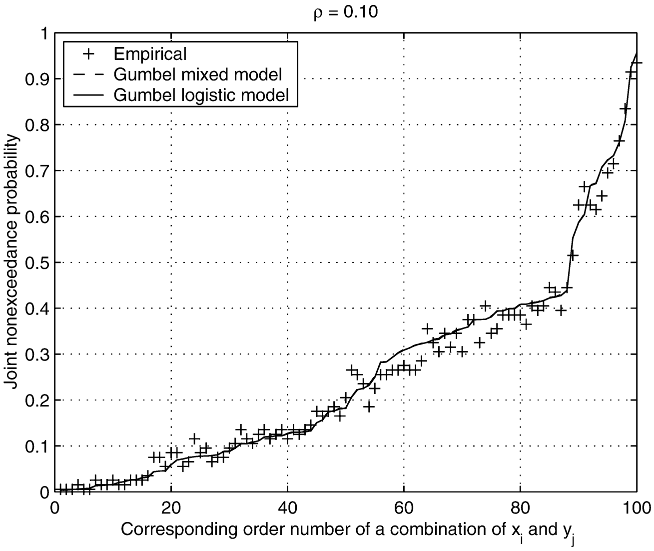

Fig. 1 Comparison of empirical and theoretical joint probabilities of X and Y with $\rho = 0.1$

$\rho = 0.10,\ 0.30,\ 0.50,\ 0.60$, and $0.667(2/3)$, respectively. In these diagrams, the dashed-line represents the theoretical joint probabilities computed by the GM model, which are arranged in ascending order. The corresponding joint probabilities calculated by the GL model are illustrated by the solid-line. The corresponding empirical joint probabilities are indicated by the plus sign. The x-axis is the corresponding order number of a combination of x_i and y_j. Figures 1–5 indicate that the computed theoretical joint probabilities by the GM model are almost the same as those computed by the GL model. No significant differences can be detected between the theoretical and empirical probabilities. To test the goodness of fit of the theoretical distribution to the empirical one, the Kolmogorov–Smirnov test (Kanji, 1993) was executed. All the above cases were accepted at the significance level of 0.05. Thus, both models may be appropriate for representing the joint distribution of two correlated Gumbel-distributed random variables whose PPMCC is: $0 \le \rho \le 2/3$.

Theoretically, the range of the PPMCC of X and Y of the GM model is: $0 \le \rho \le 2/3$ (Oliveria, 1975; 1982). When $\rho > 2/3$, the GM model is not suitable for representing the joint distribution of two correlated Gumbel-distributed variables. To illustrate this point, the 51×41 matrixes with $X = 0\ (1)\ 50$ and $Y = 0\ (1)\ 40$ were set up and the corresponding joint pdfs of X and Y with $\rho = 0.9$ were computed and

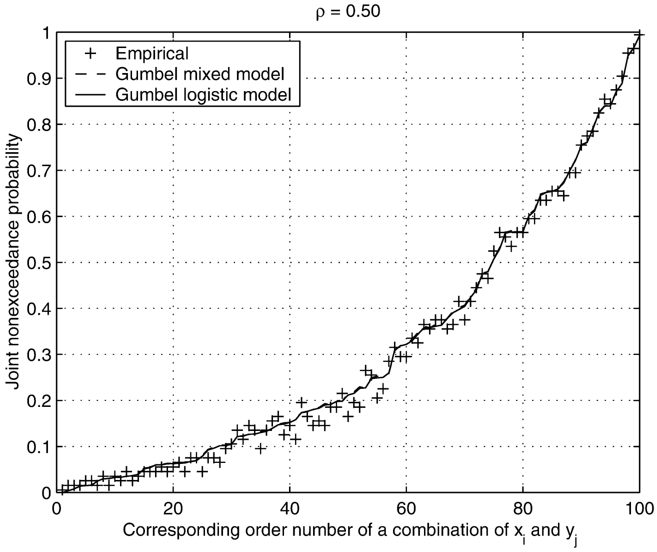

Fig. 3 Comparison of empirical and theoretical joint probabilities of X and Y with $\rho = 0.5$

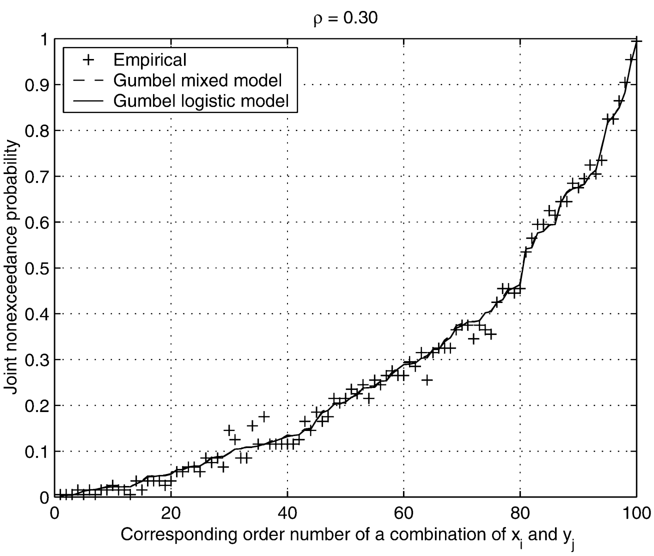

Fig. 2 Comparison of empirical and theoretical joint probabilities of X and Y with $\rho = 0.3$

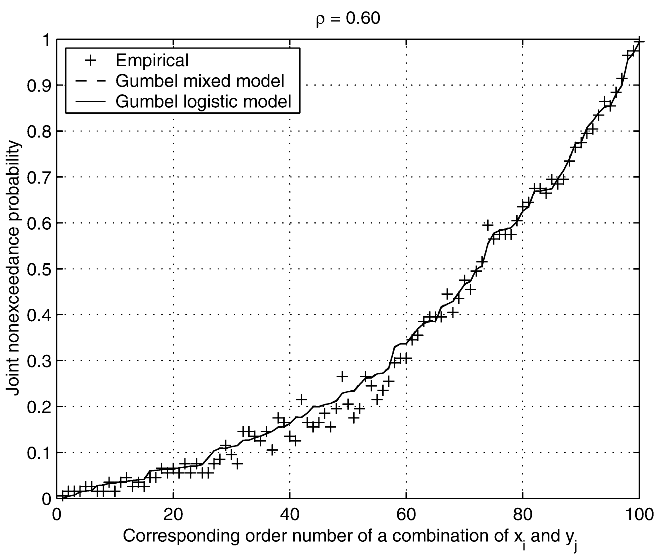

Fig. 4 Comparison of empirical and theoretical joint probabilities of X and Y with $\rho = 0.6$

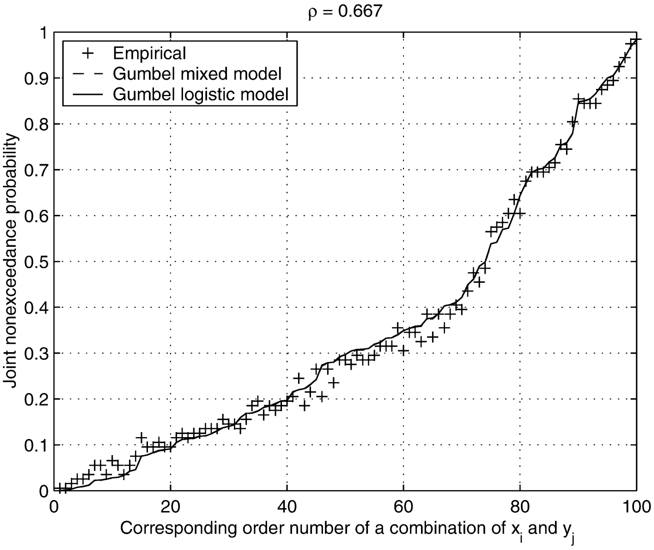

Fig. 5 Comparison of empirical and theoretical joint probabilities of X and Y with $\rho = 2/3$

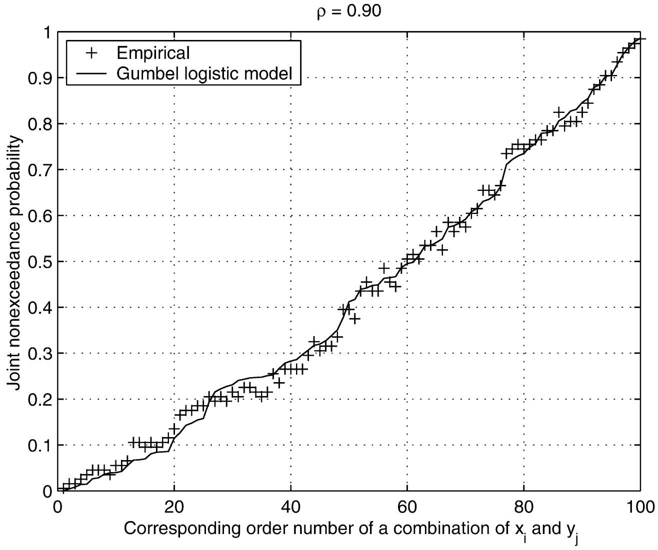

Fig. 7 Comparison of empirical and theoretical joint probabilities of X and Y by the GL with $\rho = 0.9$

were depicted in Fig. 6. It can be seen that in some domains, the joint probability density function (pdf) $f(x,y) < 0$. The minimum of the joint pdf is: $\min[f(x,y)] = -0.000086$. This violates the basic principle as a pdf, i.e., $f(x,y) \geq 0$.

The empirical and theoretical joint probabilities of X and Y with $\rho = 0.9$ for the GL model are displayed in Fig. 7. There is no significant differences can be found between empirical and theoretical probabilities.

In practice, a quantile corresponding to a given return period is important for practitioners to do effective engineering planning and design. The joint return periods of these two models corresponding to the above PPMCCs and the marginal distributions are also computed using Eq. (A8). The joint return period contour lines are displayed in Figs. 8–13, which correspond to the given $\rho = 0.10$, 0.30, 0.50, 0.60, 0.667(2/3), and 0.90, respectively. In these diagrams, the solid lines represent the joint return periods computed by the GL model, while the dashed lines indicate the joint return periods by the GM model. When $\rho \leq 2/3$, there are no differences in the joint return periods computed by the two models, while for $\rho = 0.90$, the differences between the joint return periods by the two models are evident. The joint return periods by the GM model are smaller than by the GL model. This is due to that in some domains, the joint pdf of the GM model is negative and $F_{\mathrm{GM}}(x,y) < F_{\mathrm{GL}}(x,y)$. These results further demonstrate the observations in the preceding paragraphs, i.e., when $\rho \leq 2/3$, both models provide the same results, while $\rho > 2/3$, the GM model is invalid.

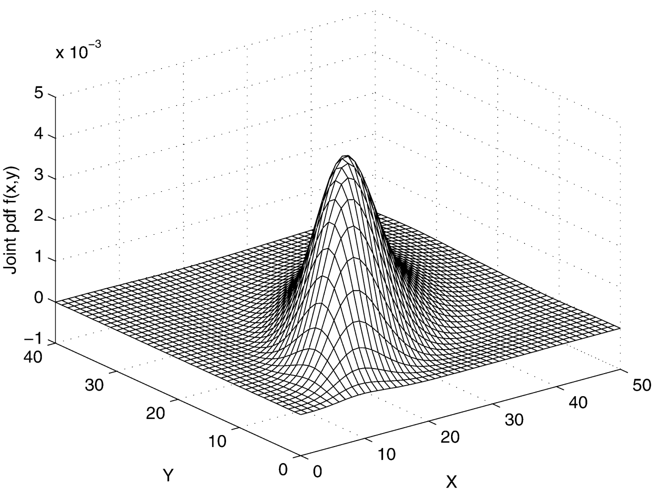

Fig. 6 Joint probability density function of X and Y by the GM model

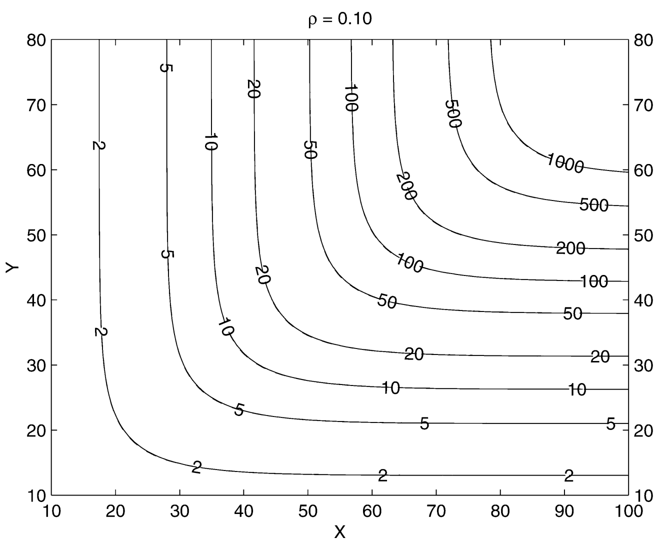

Fig. 8 Comparison of joint return periods by the two models with $\rho = 0.1$

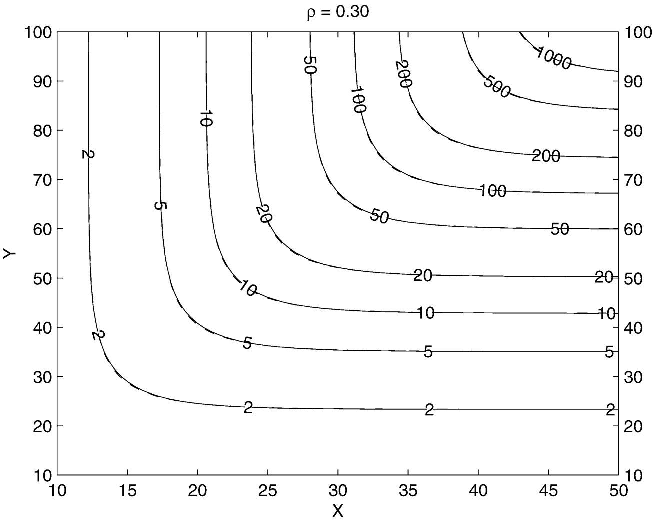

Fig. 9 Comparison of joint return periods by the two models with $\rho = 0.3$

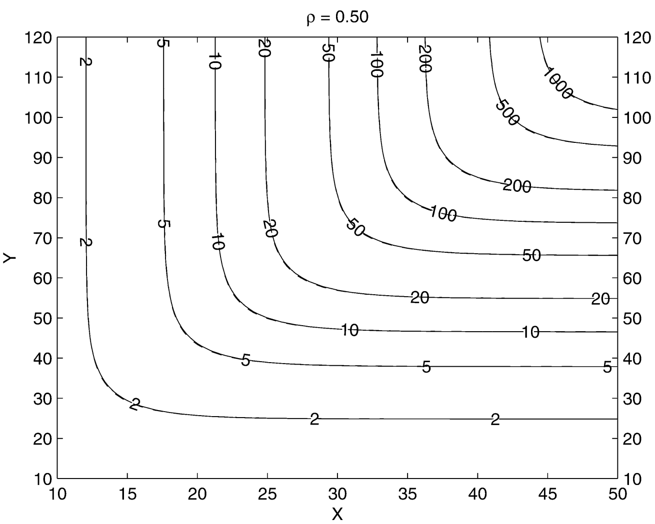

Fig. 10 Comparison of joint return periods by the two models with $\rho = 0.5$

3 Conclusion

This study presents the comparison of two bivariate extreme value distributions, termed as the Gumbel mixed (GM) model and the Gumbel logistic (GL) model. The applicability of these two models for representing the joint distribution of two correlated Gumbel distributed variables was examined by simulation experiments. Comparison results demonstrate that within the range of PPMCC: $0 \leq \rho \leq 2/3$, both models provide the same joint probabilities and joint return periods, and both may be useful for representing the joint statistical properties of the two random variables with Gumbel marginals. When $\rho > 2/3$, only the Gumbel logistic model can be applied to represent the joint distribution of the two Gumbel-distributed random variables.

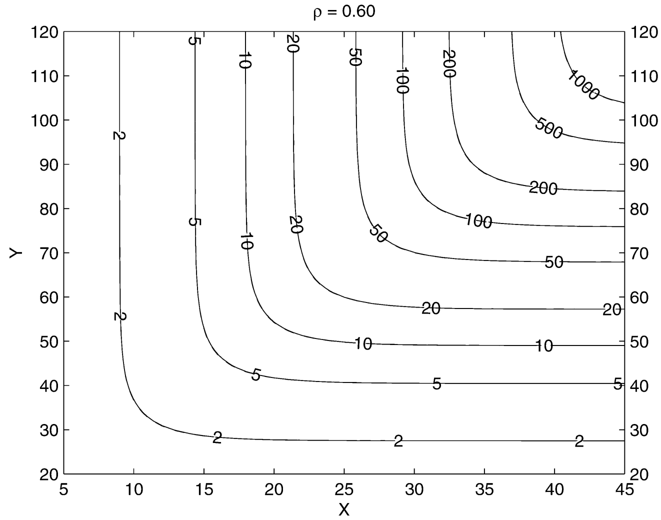

Fig. 11 Comparison of joint return periods by the two models with $\rho = 0.6$

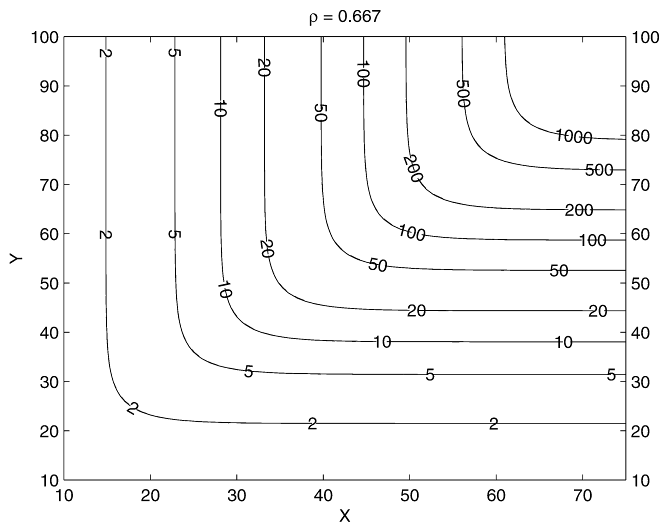

Fig. 12 Comparison of joint return periods by the two models with $\rho = 0.667$

Appendix A

The cumulative distribution functions (cdfs) of the Gumbel mixed (GM) model and the Gumbel logistic (GL) model are respectively

$$F_{\mathrm{GM}}(x,y) = F_X(x) \cdot F_Y(y) \times \exp\left\{-\theta \cdot \left[\frac{1}{\ln F_X(x)} + \frac{1}{\ln F_Y(y)}\right]^{-1}\right\} \quad (0 \leq \theta \leq 1) \tag{A1}$$

$$F_{\mathrm{GL}}(x,y) = \exp\left\{-[(-\ln F_X(x))^{\eta} + (-\ln F_Y(y))^{\eta}]^{\frac{1}{\eta}}\right\} \quad (1 \leq \eta) \tag{A2}$$

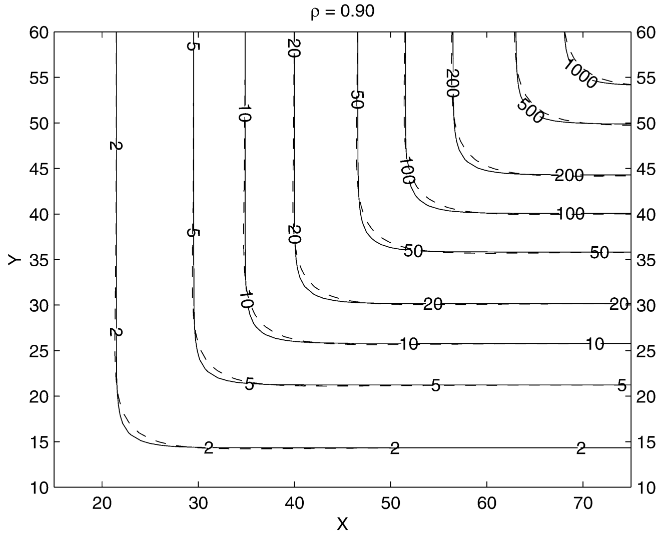

Fig. 13 Comparison of joint return periods by the two models with $\rho = 0.9$

Their joint probability density functions can be derived by differentiating the corresponding cdfs. In the two models, θ and η are the association parameters of the GM model and the GL model, respectively, which describe the dependence between two random variables and are presented by (see Gumbel 1960a, b, 1961; Gumbel and Mustafi 1967; Johnson and Kotz 1972; Oliveria 1975, 1982)

$$\theta = 2 \cdot \left[1 - \cos\left(\pi \cdot \sqrt{\frac{\rho}{6}}\right)\right] \quad (0 \leq \rho \leq 2/3) \tag{A3}$$

$$\eta = \frac{1}{\sqrt{(1-\rho)}} \quad (0 \leq \rho \leq 1) \tag{A4}$$

where ρ is the Pearson's product moment correlation coefficient (PPMCC):

$$\rho = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \tag{A5}$$

in which $(\mu_X,\ \sigma_X)$ and $(\mu_Y,\ \sigma_Y)$ are the population mean and standard deviation of X and Y, respectively. In practice, the population mean and standard deviation are often replaced by the sample mean (M) and standard deviation (S), respectively. $F_X(x)$ and $F_Y(y)$ are the marginal distribution functions of X and Y, respectively. They are given by

$$F_Z(z) = \exp\left[-\exp\left(-\frac{z - u_Z}{\alpha_Z}\right)\right] \quad (z = x, y) \tag{A6}$$

where u_Z and $\alpha_Z(Z = X, Y)$ are the location and scale parameters of Z. When the PPMCC $\rho = 2/3$, the association parameter θ of the Gumbel mixed model reaches its upper limitation and is equal to 1. When the PPMCC $\rho = 0$, both models become

$$F(x,y) = F_X(x)F_Y(y) \tag{A7}$$

The joint return period of X and Y associated with the event that either x or y or both is exceeded ($X > x, Y > y$, or $X > x$ and $Y > y$) can be given by

$$T(x,y) = \frac{1}{1 - F_B(x,y)} (F_B(x,y) = F_{GM}(x,y), F_{GL}(x,y)) \tag{A8}$$

Acknowledgement The anonymous reviewers' comments improved the quality of the paper.

References

Buishand TA (1984) Bivariate extreme-value data and the station-year method. J Hydrol 69: 77–95

Castillo E (1988) Extreme Value Theory in Engineering. Academic Press Inc., San Diego

Coles SG, Tawn JA (1991) Modelling extreme multivariate events. J R Stat Soc, Ser B 53(2): 377–392

Coles SG, Tawn JA (1994) Statistical methods for multivariate extremes: An application to structural design. Appl Stat 43(1): 1–48

Gumbel EJ (1958) Statistics of Extremes. Columbia University Press, New York

Gumbel EJ (1960a) Multivariate extremal distributions. Bull Inst Internat De Statistique, 37(2): 471–475

Gumbel EJ (1960b) Distributions del valeurs extremes en plusieurs dimensions. Publications de L'Institute de Statistique, Paris 9: 171–173

Gumbel EJ (1961) Bivariate logistic distributions. J. Amer. Stat. Assoc. 56: 335–349

Gumbel EJ, Mustafi CK (1967) Some analytical properties of bivariate extreme distributions. J Am Stat Assoc 62: 569–588

Joe H, Smith RL, Weissman I (1992) Bivariate threshold models for extremes. J R Stat Soc Ser B 54(1): 171–183

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

Kanji GK (1993) 100 Statistical tests, SAGE Publications, USA

Lewis PA, Goodman AS, Miller JM (1969) A pseudorandom number generator for the System/360. IBM Syst J 8(2): 136–146

Oliveria JTD (1975). Bivariate extremes: extensions. Bull Inter Stat Inst 46(2): 241–251

Oliveria JTD (1982) Bivariate Extremes: Models and Statistical Decision, Tech. Report No. 14, Center for Stochastic Processes, Dept. of Statistics, University of North Carolina, Chapel Hill, North Carolina, USA

Raynal-Villasenor JA, Salas JD (1987) Multivariate extreme value distributions in hydrological analyses. In: Rodda JC, Matalas NC eds. Water for the Future: Hydrology in Perspective, Proceedings of the Rome Symposium, IAHS Publ No. 164, pp. 111–119

Pickands J (1981) Multivariate extreme value distributions. Proc. 43rd sess. Int Statist Inst 49: 859–878.

Stedinger JR, Vogel RM, Foufoula-Georgiou E (1993) Chap. 18: Frequency analysis of extreme events. In: Maidment DR (Ed.) Handbook of Hydrology, McGraw-Hill Inc, USA

Tawn JA (1988) Bivariate extreme value theory: models and estimation. Biometrika 75(3): 397–415

Todorovic P (1978) Stochastic models of floods. Water Resour Res 14(2): 345–356

Yue S (2000a) The Gumbel mixed model applied to storm frequency analysis. Water Resour Management 14: 377–389

Yue S (2000b) The Gumbel logistic model for representing a multivariate storm event. Adv Water Res 24: 179–185

Yue S (2001a) A statistical measure of severity of El Nino events. Stochastic Environmental Res Risk Anal (SERRA) 15(2): 153–172

Yue S (2001b) A bivariate extreme value distribution applied to flood frequency analysis. Nordic Hydrology 32(1): 49–64

Yue S, Ouarda TBMJ, Bobée B, Legendre P, Bruneau P (2000b) Corrigendum to "The Gumbel mixed model for flood frequency analysis". J Hydrol 226(1–2): 88–100; J Hydrol 228(3–4): 283

Yue S, Ouarda TBMJ, Bobée B, Legendre P, Bruneau P (1999) The Gumbel mixed model for flood frequency analysis. J Hydrol 226(1–2): 88–100