12th International Conference on Hydroscience & Engineering
Hydro-Science & Engineering for Environmental Resilience
November 6-10, 2016, Tainan, Taiwan.

Bivariate Discontinuous Flood Frequency Analysis Based on Archimedean Copula Functions

Yu Chen[1], *Pengzhi Lin*[2]
1. State Key Laboratory of Hydraulics and Mountain River Engineering; College of Hydraulic and Hydroelectric Engineering, Sichuan University
2. State Key Laboratory of Hydraulics and Mountain River Engineering, Sichuan University
Chengdu, Sichuan, China

ABSTRACT

Based on the flood data at the Danba station on the Dadu River in China, the flood characteristics (peak flow and flood volume) are extracted and analyzed. Considering the mutual correlation between peak flow and flood volume, copula-based flood frequency analysis, as an alternative to the commonly used univariate flood frequency analysis, is applied in the study to represent the joint distribution of the two correlated variables. Among different copula families, Archimedean copula family is chosen, wherein the Gumbel-Hougaard copula (GHC) is employed with parameter estimated by two-dimensional maximum likelihood (ML) method. On the basis of marginal distributions, the joint distribution of peak flow and flood volume can be deduced. The applicability of the proposed bivariate flood frequency analysis model is demonstrated through the case study, and the results indicate that the proposed model is reasonable, feasible and useful for describing the joint probabilistic behavior of bivariate flood events.

KEY WORDS: copula function; joint distribution; marginal distribution; Gumbel-Hougaard copula.

INTRODUCTION

The flood is a multi-attribute natural hazard and is characterized by mutually correlated flood properties peak flow, volume, and duration of flood hydrograph (Jozef, and Patrick, 2013). Therefore the most commonly used univariate frequency analysis can only provide limited assessment results, and copula is often regarded as its good alternative. Copula is applied in many fields, such as hydrology, decision-making, risk management, etc. A copula is very useful to implement efficient algorithms for simulating joint distributions in a more realistic way. In fact, copula is able to model the dependence structure independently of the margin distributions, and is a function that links univariate marginal distribution functions to construct a multivariate distribution function. It is then possible to build multidimensional distributions with different margins, the structure of dependence being mathematically formalized through the copula (Favre, Adlouni, Perreault, Thiémonge, and Bobée, 2004). The main advantage of copula function over classical bivariate frequency analyses is that the selection of marginal distributions and multivariate dependence modeling are two separate processes, giving additional flexibility to the practitioner (Favre, Adlouni, Perreault, Thiémonge, and Bobée, 2004).

A number of attempts have been made to perform bivariate and multivariate flood frequency analyses that take into consideration the dependence among flood variables but with restrictive assumptions (Goodarzi, Mirzaei, Shui, and Ziaei, 2011). In practice, extreme events such as flood peak and flood volume may be represented by the Gumbel distribution. Thus, it may be advantageous to directly use a bivariate extreme value distribution to analyze the joint behavior of two correlated Gumbel distributed random variables (Sheng, 2001). In this paper, a copula-based bivariate model for flood frequency analysis of peak flow and volume is developed. The objective of this paper is therefore two folds: 1) to investigate the possibility of applying the copula approach to model the joint distributions of flood flow and peak volume for a studied case; 2) to investigate the performance of the copula approach in flood frequency analysis by stochastic simulation.

BIVARIATE FREQUENCY ANALYSIS

The bivariate model for two random variables can be uniquely constructed based on chosen marginal distributions and the copula representing the dependence between variables independently (Dung, Merz, Bárdossy, and Apel, 2015).

Multivariate distribution construction using copulas was developed by Sklar. Every joint distribution can be written in a copula and its univariate marginal distributions (Li, Guo, Lu, and Guo, 2013). There are different families of copulas, and the Archimedean copula family is more desirable for hydrologic analyses because it can be easily constructed, and it can be applied whether the correlation among the hydrological variables is positive or negative. Gumbel-Hougarrd, Ali-Mikhail-Haq, Cook-Johnson, and Frank copulas were introduced as one-parameter Archimedean copulas (Nelsen 1997). Typically, Gumbel-Hougarrd copula (GHC) played an important role in hydrologic frequency analysis, and is found to be the most suitable dependence model for flood peak and volume because of its good property of upper tail-dependence and because of good performance in many applications (Li, Guo, Lu, and Guo, 2013). For this study the one-parameter GHC is chosen to be used. The GHC function is expressed as:

$$C(u,v) = \exp\{-[(-\ln u)^{\theta} + (-\ln v)^{\theta}]^{1/\theta}\}, \quad \theta \in [1,\infty) \tag{1}$$

where θ is the parameter describing the association between two random variables u and v.

Let $u = F_X(x)$ and $v = F_Y(y)$ be marginal cumulative probability distributions of peak flow and volume, then the cumulative distribution function (CDF) of the bivariate GHC can be expressed as (Zhang and Singh 2007):

$$C_\theta^2(u,v) = C_\theta^2(F_X(x), F_Y(y)) = \varphi^{-1}[\varphi(u) + \varphi(v)] = \exp\{-[(-\ln u)^\theta + (-\ln v)^\theta]^{1/\theta}\}, \quad (\theta \geq 1) \tag{2}$$

and its probability density function (PDF) is presented as:

$$C_\theta^2(u,v) = \frac{\partial C_\theta^2}{\partial u \partial v} = \frac{(-\ln u \ln v)^{\theta-1}}{uv} \exp(-w^{1/\theta})[w^{2/\theta-2} + (\theta-1)w^{1/\theta-2}], (\theta \geq 1) \tag{3}$$

where $w = (-\ln u)^\theta + (-\ln v)^\theta$; θ is the unknown parameter to be estimated.

In order to build up the copula-based bivariate statistical model for flood frequency analysis for a particular case study comprising two studied random variables, two fundamental and steps need to be taken: parameter estimation and goodness-of-fit testing. The former is to estimate parameter θ assuming that the fitting distribution belongs to a known distribution class. The selection of a candidate distribution class (marginal, copula, bivariate) for the fitting practically depends on studied variables. The latter is to test the validity of that assumption (Dung, Merz, Bárdossy, and Apel, 2015).

Different method exist for estimating the copula parameters, such as direct method based on common rank correlation measures like Kendall's τ and Spearman's ρ, maximum likelihood (ML) (Dung, Merz, Bárdossy, and Apel, 2015). For this study, the ML method is used to estimate the parameter of GHC. The ML estimation method does not require any prior assumptions regarding marginal distributions of the dependent variables. The procedure consists of transforming the marginal variables into uniformly distributed vectors using its empirical distribution function (Jozef, and Patrick, 2013). The likelihood function of the copula function can be expressed as:

$$\ln C_\theta^2(u,v;\theta) = -\ln uv + (\theta+1)\ln(\ln u \ln v) - w^{1/\theta} + \ln[w^{2/\theta-2} + (\theta-1)w^{1/\theta-2}] \tag{4}$$

where $w = (-\ln u)^\theta + (-\ln v)^\theta$; C_θ^2 denotes the density function; θ is the parameter of the GHC (Kong, Huang, Fan, and Li, 2015).
The parameter θ of GHC can be obtained by:

$$\theta = \arg\max \sum_{i=1}^{n} \ln C_\theta^2(u,v;\theta) \tag{5}$$

Akaike information criterion (AIC) is computed from the ML value of the fitted distribution. The goodness-of-fit evaluation was performed by means of Kolmogorov-Smirnov (K-S) tests (Dung, Merz, Bárdossy, and Apel, 2015).

AIC is a measure of the relative quality of statistical models for a given set of data, and deals with the trade-off between the goodness of fit of the model and the complexity of the model. It can be expressed as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left[F_{emp}(x_{i1}, x_{i2}, \text{L}, x_{im}) - C(y_{i1}, y_{i2}, \text{L}, y_{im})\right]^2$$
$$AIC = n\ln(MSE) + 2k \tag{6}$$

where $F_{emp}(x_{i1}, x_{i2}, \text{L}, x_{im})$ and $C(y_{i1}, y_{i2}, \text{L}, y_{im})$ are empirical frequency and theoretical frequency; m is the function's dimension; k is the number of model parameters. The smaller the ACI value, the better the fitting result of copula function.

The K-S test is a nonparametric probability distribution free test, and quantifies the largest vertical difference between the specified and empirical distributions. Given n increasing ordered data points, the K-S test statistic is defined as:

$$T = \sup_x \left|F^*(x) - F_n(x)\right| \tag{7}$$

where $F^*(x)$ stands for the specified distribution; $F_n(x)$ stands for the empirical distribution; and 'sup' stands for supermum (Kong, Huang, Fan, and Li, 2015).

CASE STUDY

In this study, daily streamflow records from 1960 to 2008 at the Danba Hydrometric Station on the Dadu River Basin in China are analyzed, and the data is from the reference (Zhang, Lu, Yan, Mu, Chen, 2015). The Gumbel-Hougaard copula is selected as the most appropriate for the pair of peak flow and volume (Q-V).

The Kendall correlation coefficient (τ) and Spearman correlation coefficient (ρ) between the peak flow and volume are estimated using Eq. 8, and are 0.502 and 0.546 respectively. Thus, we can employ the bivariate extreme distribution to model the joint distribution of the peak flow and volume.

$$\tau = 4\int_{[0,1]^2} C(u,v)dC(u,v) - 1$$
$$\rho = 12\int_{[0,1]^2} C(u,v)dC(u,v) - 3 \tag{8}$$

The P-III distribution has been recommended by MWR (2006) to model the occurrences of flood in China. The probability density function (PDF) of the P-III distribution is defined as

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)}(x-\delta)^{\alpha-1}e^{-\beta(x-\delta)}, \quad \alpha > 0, \beta > 0, \delta \leq x < \infty \tag{9}$$

where $\Gamma(\alpha)$ is gamma function; α , β and δ are shape, scale and location parameters of the P-III distribution, respectively (Li, Guo, Lu, and Guo, 2013).

Te parameters of the P-III marginal distributions for peak flow and volume are estimated by using the L-Moments approach respectively as:

$$\begin{aligned} &\text{peak flow:} \quad \overline{x} = 3060.70, \quad C_v = 0.22, \quad C_s = 0.49, \\ &\text{flood volume:} \quad \overline{x} = 7900.00, \quad C_v = 0.64, \quad C_s = 1.45. \end{aligned} \tag{10}$$

GHC is used to establish the joint distributions of peak flow and volume, and the copula parameter θ is estimated by the ML method as: $\theta = 1.557$.

To test the goodness-of-fit of the GHC distribution, the K-S test is executed. The critical K-S value is D = 0.106. The goodness-of-fit tests indicate a good agreement between observed and theoretical probabilities for both marginal and joint distributions. Therefore, it can be concluded that all these two characteristics of flood events can be represented by the GHC distribution. The comparisons of theoretical frequency and empirical frequency for two marginal distributions of peak flow and volume are shown in Fig.1.

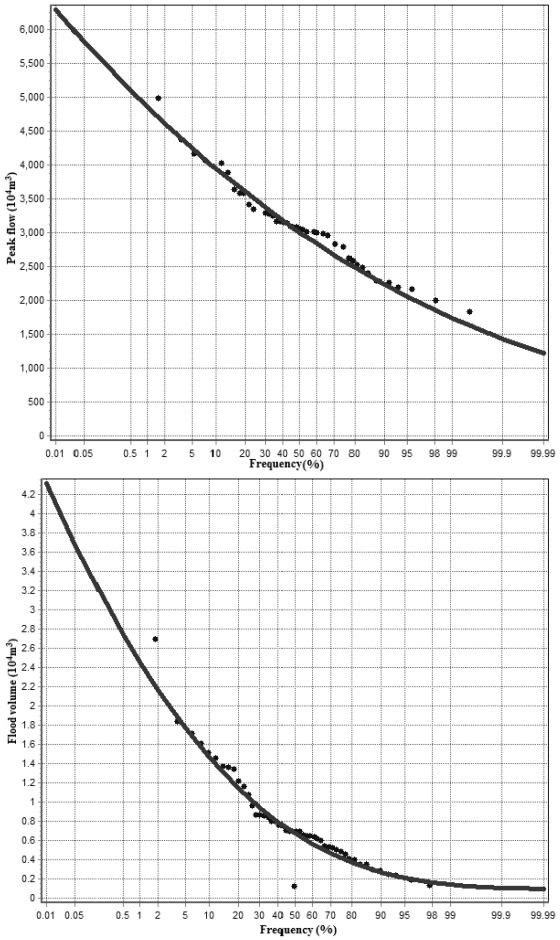

Fig. 1 Comparison of empirical frequency and theoretical frequency for peak flow and flood volume

The empirical and theoretical joint probabilities are plotted in Fig. 2, in which the solid-line represents the theoretical joint frequencies of peak flow and volume, which are arranged in ascending order, and the corresponding empirical joint frequencies are expressed by the dot. The x-axis is the corresponding order number of a combination of u_i and v_i. It can be seen that the theoretical frequencies fit the empirical ones well. It is therefore concluded that the model is suitable for representing the joint distribution of peak flow and volume.

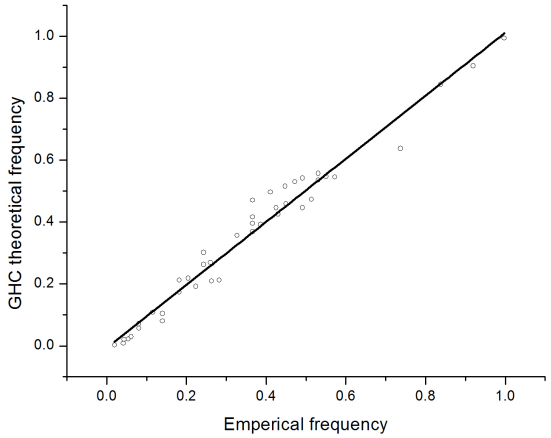

Fig. 2 Fit of empirical frequency and GHC theoretical frequency between peak flow and volume

It can be seen from Fig.2 that the empirical frequency points and theoretical frequency points are distributed in the vicinity of the 45 degree line, which shows that the fitting of two variables of empirical frequency and theoretical frequency is well, and indicates the joint distribution function is reasonable.

CONCLUSIONS

This paper proposes a bivariate flood frequency analysis model based on the GHC. Application of the GHC method involves: constructing marginal distributions of peak flow and volume; establishing the joint distributions of peak flow and volume; estimating the parameters of marginal distribution and joint distribution by using L-matrix method and ML method respectively; and performing the goodness-of-fit statistic test for both marginal and joint distributions.

One advantage of applying such technique in flood frequency analysis is the separation of the margins and dependence structure which simplifies the analysis greatly. The flood data from 1958 to 2008 from Danba station, Dadu River Basin, China, are used to demonstrate the usefulness of the copula technique. The analyzing results demonstrate that the proposed model is useful for representing the joint distributions of peak flow and volume. The proposed method also provides additional information of two correlated flood characteristics, which cannot be obtained by univariate flood frequency analysis.

ACKNOWLEDGEMENTS

This research was substantially supported by the National Basic Research Program of China Grant No. 2013CB036401 and National Natural Science Foundation of China (Grant No. 41501554), Sichuan

University, China (Grant No. 2015SCU11045) and the Ensemble Estimation of Flood Risk in a Changing Climate (EFRaCC) project funded by the British Council as part of its Global Innovation Initiative.

REFERENCES

Dung, N.V., Merz, B., Bárdossy, A., and Apel, H. (2015). Handling uncertainty in bivariate quantile estimation – An application to flood hazard analysis in the Mekong Delta. Journal of Hydrology, 527,704-717.

Favre, A.C., Adlouni, S.E, Perreault, L., Thiémonge, N., and Bobée, B. (2004). Multivariate hydrological frequency analysis using copulas. Water Resources Research, 40, W01101.

Goodarzi, E., Mirzaei, M., Shui, L.T., and Ziaei, M. (2011). Evaluation dam overtopping risk based on univariate and bivariate flood frequency analysis. Hydrol. Earth Syst. Sci. Discuss. 8, 9757-9796.

Jozef, V.D., and Patrick, W. (2013). Probabilistic flood risk assessment over large geographical regions. Water Resources Research, 49,3330-3344.

Kong, X.M., Huang, G.H., Fan, Y.R., and Li, Y.P. (2015). Maximum entropy- Gumbel-Hougaard copula method for simulation of monthly streamflow in Xiangxi river, China. Stoch Environ Res Risk Assess, 29,833-846.

MWR (The Ministry of Water Resources of People's Republic of China). (2006). Regulation for calculating design flood of water resources and hydropower projects. Beijing: China Water Power Press.

Nelsen, R.B. (1997). Dependence and order in families of Archimedean copulas. J Multivariate Anal, 60,111-122.

Sheng, Y. (2001). A Bivariate Extreme Value Distribution Applied to Flood Frequency Analysis. Nordic Hydrology , 32,49-64.

Sraj, M., Bezak, N., and Bivariate, M.B. (2015). Flood Frequency Analysis with Historical Information Based on Copula. Hydrol. Process. 29, 225-238.

Zhang, D.D., Lu, F., Yan, D.H., Mu, W.B., Chen, X.J. (2015). Research on multi-dimensional joint distribution of flood characteristics based on Archimedean copula. China Rural Water and Hydropower, 1, 68-74.

Zhang, L. and Singh, V.P. (2007). Gumbel-Hougaard copula for trivariate rainfall frequency analysis. J. Hydro. Eng. 12,409-419.