

## Joint modeling of flood peak discharges, volume and duration: a case study of the Danube River in Bratislava

Veronika Bačová Mitková\*, Dana Halmová

Institute of Hydrology SAS, Račianska 75, 831 02 Bratislava, Slovakia.

\* Corresponding author. Tel.: +421249268242. Fax: +421244259311. E-mail: mitkova@uh.savba.sk

**Abstract:** The study is focused on the analysis and statistical evaluation of the joint probability of the occurrence of hydrological variables such as peak discharge ( $Q$ ), volume ( $V$ ) and duration ( $t$ ). In our case study, we focus on the bivariate statistical analysis of these hydrological variables of the Danube River in Bratislava gauging station, during the period of 1876–2013. The study presents the methodology of the bivariate statistical analysis, choice of appropriate marginal distributions and appropriate copula functions in representing the joint distribution. Finally, the joint return periods and conditional return periods for some hydrological pairs ( $Q$ - $V$ ,  $V$ - $t$ ,  $Q$ - $t$ ) were calculated. The approach using copulas can reproduce a wide range of correlation (nonlinear) frequently observed in hydrology. Results of this study provide comprehensive information about flood where a devastating effect may be increased in the case where its three basic components (or at least two of them)  $Q$ ,  $V$  and  $t$  have the same significance.

**Keywords:** The Danube River; Joint distribution; Copula functions; Peak flow; Volume and duration of the wave; Multivariate frequency analysis.

### INTRODUCTION

The effect of the extreme hydrological event depends on many factors such as season, weather conditions, geographical location, relief, duration, volume, rainfall and so on. Many hydrologists analyzed mutual dependence between these components of the hydrological cycle to identify the flood-generating processes; see e.g. Maca and Torf (2009), Merz and Blöschl (2003) and Parajka et al. (2008). Understanding flood regime changes in Europe was addressed by Hall et al. (2013). The analysis and comparisons of the seasonality of hydrological characteristics in Slovakia and Austria or across the Alpine–Carpathian range can be found in the works of Parajka et al. (2009) or Parajka et al. (2010).

Therefore, for more comprehensive analysis of the flood, it is necessary to know and understand the relationship and dependence between the basic characteristics of the flood wave. In order to obtain this information, a joint probability density function and a joint cumulative distribution function of variables are needed. In a sense, every joint distribution function implicitly contains both: a description of the marginal behavior of individual variables and a description of their dependency structure. The joint probability distribution analysis is better use in some cases of the hydrological events, for example:

1. Interdependence and mutual correlation of the variables that characterized flood wave (discharge, volume, duration). Some flood events with a peak of return period of 100-years could be less damaging than floods with both peak and volume of return period of 10-years.

2. Mutual influences of extreme events that occur on the main river and its tributaries simultaneously. Some examples are provided in Favre et al. (2004), Le Clerc and Lang (2002), Prohaska et al. (1999).

In traditional method of bivariate flood frequency analysis, the parametric distribution functions of the hydrological variables such as bivariate normal, exponential, lognormal, gamma, Gumbell, etc. were used. Some existing studies which solves the bivariate flood frequency analysis considering specific families of the parametric distribution functions were by Ashkar and

Rousselle (1982), Faqir and Shazia (2007), Sackl and Bergmann (1987), Todorovic (1978), Yue et al. (1999). A drawback of this approach is that the same family of marginal distributions is assumed for all flood variables, while analyzed variables shows different margins.

The concept of copula functions relaxes the restriction of the traditional multivariate flood frequency analysis and connects marginal distribution of random variables with their joint distribution. The main advantage of constructing multivariate distributions via copulas is that univariate marginal distributions can be defined independently of their joint behaviour.

Therefore, copulas appear to be the most suitable tool for modeling dependencies between two or more variables. The word “copula” was first used in mathematical or statistical sense by Sklar (1959), and many of the basic results on copula can be traced to the early work of Heofinding (1940). However, a good introduction of the theory of copulas can be found in monographs Joe (1997) and Nelsen (1999). For example, De Michele et al. (2005) applied bivariate Archimedean copulas to simulate flood pairs of the peak and volume. Salvadori and De Michele (2004) studied the relation between univariate and bivariate return period in term of Archimedean copula. Zhang and Singh (2006) derived bivariate joint distributions of the flood peak and volume, and volume and duration by using copula method. Grimaldi and Serinaldi (2006) constructed a trivariate joint distribution of the flood event variables using the fully nested Archimedean copula functions. The dependence between discharge and volume using bivariate copulas was presented by Favre et al. (2004) or Genest and Favre (2007). Karmakar and Simonovic (2009) present the bivariate flood frequency analysis with the copula functions for evaluating joint distributions with mixed marginal distributions. There exist a huge number of parametric bivariate copulas, but the set of higher-dimensional copula is rather limited. Furthermore, there have been some attempts to construct multivariate extensions of the Archimedean bivariate copula, see e.g. Embrechts et al. (2001) and Savu and Trede (2006).

Flood risk assessment in Slovakia is defined and determined by the Sectorial Technical Norm of the Ministry of Environ-

ment OTN ŽP 3112-1:03 and Flood Risk directive 2007/60/EC. With the strategy of Europe 2020, it is one of the measures to meet the partial milestone of the RIS3 "Supporting research and innovation in environmental areas, including climate change adaptation". Thus, this is the reason the monitoring and evaluation of extreme hydrological phenomena in the form of flooding, flow increase, or dryness is very timely. We preclude that the results obtained by the bivariate hydrological analysis provide more reliable assessment of flood risk. The bivariate analyses treat flood events in a complex way, thus having a potential to be applied in the praxis of water management and the design of flood protective systems.

The aim of this study is to analyze and statistically evaluate the bivariate joint probability of the hydrological variables occurrence. The three main hydrological variables: peak discharge, volume and duration of the flood waves of the Danube River in Bratislava gauging station, during the period of 1876–2013 were analyzed. The study consists of two main parts. The first part deals with the estimation of the marginal distribution of hydrological variables, while the second part concerns itself with the estimation of the joint distribution of variables using copula function. The section "Methodology" is focused on the description, application and comparison of Archimedean copulas in bivariate statistical analysis. Finally, the joint return periods and conditional return periods of the hydrological pairs ( $Q-V$ ,  $V-t$ ,  $Q-t$ ) were calculated.

## METHODOLOGY

### Basic features of copulas

A copula function is a mathematical technique which offers a flexible way of describing nonlinear dependence among multivariate data in isolation from their marginal probability distributions, and serves as a powerful tool for modeling as well as simulating nonlinearly interrelated multivariate data. A bivariate copula can be represented as:

$$C:[0,1]^2 \rightarrow [0,1]. \quad (1)$$

It has to fulfill the following conditions: 1)  $C(u,1)=u$ ,  $C(1,v)=v$ ,  $C(u,0)=C(0,v)=0$ , and 2)  $C(u_1,u_2)+C(v_1,v_2)-C(u_1,v_2)-C(v_1,u_2) \geq 0$ , if  $u_1 \geq v_1$ ,  $u_2 \geq v_2$  and  $u_1, u_2, v_1, v_2 \in [0,1]$ . The first condition defines the copula limits. The second condition ensures that the probability corresponding to any rectangle in the unit square is non-negative (Karmakar and Simonovic, 2007). As mentioned above, we could also define copula as a joint distribution function of standard uniform random variables. The random variables may have different properties and thus need to be converted to variables having an interval of  $[0,1]$  by scaling the data.

### Sklar's theorem

Sklar's theorem (1959) is the foundation of the concept of the copula. Sklar showed that every  $n$ -dimensional distribution function  $H(x_1, \dots, x_d)$  could be written as:

$$H(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)). \quad (2)$$

For two dimensional (bivariate) distribution function  $H(x,y)$ , then we can write:

$$H(x, y) = C(F(x), F(y)), \quad (3)$$

where  $F(x) = u$  and  $F(y) = v$  are marginal distribution functions.

If  $F(x)$  and  $F(y)$  are continuous, then the copula function  $C$  is unique and has the following representation:

$$C(u, v) = H(F^{-1}(u), F^{-1}(v)), \quad 0 \leq u, v \leq 1, \quad (4)$$

where the  $F^{-1}(u)$  and  $F^{-1}(v)$  are inverse distribution functions of the marginal. Conversely, it can be proven that if  $C$  is a copula function and  $F(x)$  and  $F(y)$  are distribution functions, then the function  $H(x,y)$  defined by Eq. (1) is a *bivariate* distribution function with margins  $F(x)$  and  $F(y)$  (Nelsen, 2006).

There exist a large number of the copula classes; thus commonly used copulas includes elliptical (normal, Gaussian and Student-t, ...), Archimedean (Clayton, Gumbel-Hougaard, Frank, Joe, and Ali-Mikhail-Haq, ...), extreme value (Hüsler-Reiss, Galambos, Tawn, and t-EV, ...) and other families (Plackett and Farlie-Gumbel-Morgenstern, ...).

Among existing types of copulas, the Archimedean one is the very popular class used in hydrological application (De Michele et al., 2005; Favre et al., 2004; Genest and Favre, 2007; Zhang and Singh, 2006). The Archimedean class of copulas is popular in empirical applications for flexibility and easy construction. This class of copulas includes a whole suite of closed-form copulas that covers a wide range of dependency structures, including comprehensive and non-comprehensive copulas, radial symmetry and asymmetry, and asymptotic tail dependence and independence. However, a bivariate asymmetrical Archimedean copula can be defined as:

$$C(u, v) = \varphi^{-1}(\varphi(u) + \varphi(v)), \quad (5)$$

where  $\varphi$  is the copula generator which is a continuous, strictly convex and monotonic decreasing function.

For each copula, the value of  $\theta$  (copula parameter) can be obtained by considering the mathematical relationship (Nelsen, 2006) between Kendall's coefficient of correlation  $\tau$  and the generating function  $\varphi(t)$ , which is given by:

$$\tau = 1 + 4 \int_0^1 [\varphi(t) / \varphi'(t)] dt, \quad (6)$$

where  $t = u$  or  $v$ .

Table 1 shows a well-known and most used bivariate one-parameter copula in hydrology. Some other distributions expressed as copulas were also listed, e.g., in Balakrishnan and Lai (2009).

### Dependence measures

Schweizer and Wolff (1981) showed that two standard non parametric correlation measures could be expressed solely in terms of the copula function. These are Kendall's correlation coefficient and Spearman's correlation coefficient. Kendall's coefficient of correlation ( $\tau$ ) is a well-known nonparametric measure of dependence or association in multivariate statistics. Kendall's  $\tau$  based on observation is estimated as:

$$\tau_n = \frac{1}{\binom{n}{2}} \sum_{i < j} \text{sign}[(x_i - x_j)(y_i - y_j)], \quad (7)$$

where  $\text{sign} = 1$  if  $[(x_i - x_j)(y_i - y_j)] > 0$ ,  $\text{sign} = -1$  if  $[(x_i - x_j)(y_i - y_j)] < 0$ ;  $i, j = 1, 2, \dots, n$ .

**Table 1.** Probability function, parameter space, generating function and relationship of non-parametric dependence measure with association parameter for the most frequently used Archimedean copulas in hydrology.

Copula	$C(u, v, \theta)$	parameter space $\theta$	Kendall's $\tau$	generator $\varphi(t)$
Clayton	$(u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$	$[-1, \infty) \setminus \{0\}$	$\frac{\theta}{\theta + 2}$	$\frac{1}{\theta}(t^{-\theta} - 1)$
Gumbel-Hougaard	$\exp\left\{-\left[(-\ln u)^\theta + (-\ln v)^\theta\right]^{1/\theta}\right\}$	$[1, \infty)$	$\frac{\theta - 1}{\theta}$	$(-\ln t)^\theta$
Frank	$-\frac{1}{\theta} \ln\left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{(e^{-\theta} - 1)}\right)$	$(-\infty, \infty) \setminus \{0\}$	$1 + \frac{4}{\theta}[D_1(\theta^*) - 1]$	$-\ln \frac{e^{-\theta t} - 1}{e^{-\theta} - 1}$

Debye function  $D_1 = \frac{1}{\theta} \int_0^\theta \frac{t}{e^t - 1} dt, \theta^* = -\log(\theta)$

Like Kendall's coefficient, another nonparametric measure of dependence is Spearman's  $\rho$  (Spearman, 1904). If two sets of data  $X_i$  and  $Y_i$  are converted to rankings  $x_i$  and  $y_i$ , and the differences  $d_i$  between the ranks of each observation of the two variables are calculated; thus if there are no tied ranks, i.e.  $\exists i, j (i \neq j \wedge (X_i = X_j \vee Y_i = Y_j))$ , then Spearman's  $\rho$  is given as (Myers and Well, 2003):

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}, \tag{8}$$

where  $d_i = (x_i - y_i)$  is the difference between the ranks of corresponding values  $X_i$  and  $Y_i$  and  $n$  is the number of values in each data set (same for both sets) (Karmakar and Simonovic, 2009).

Pearson correlation coefficient  $r$  is a measure of linear dependence; that is, if  $Y$  depends on  $X$  linearly, then the absolute value of  $r$  is equal to 1. We also know that linear correlation is invariant only under strictly increasing linear transformations. In addition, linear correlation is the correct dependence measure to use for multivariate normal distributions; but for other joint distributions, it can give very misleading impressions. Therefore, the Pearson correlation coefficient  $r$  can be expressed as:

$$R = \frac{\text{Cov}(x, y)}{\sqrt{\text{var}(X)} \sqrt{\text{var}(y)}}, \tag{9}$$

Pearson correlation coefficient depends not only on the copula but also on the marginal distributions. Thus, this measure is affected by nonlinear changes of scale. Moreover, it is not invariant to monotonic transformations to Kendall's and Spearman's measures (Favre et al., 2004).

Appropriateness of the choice of the distribution function was based on the "Gringorten plotting-position", which represents the empirical probability (Cunnane, 1978; Gringorten, 1963; Yue et al., 1999; Zhang and Singh 2007).

$$F(x, y) = P(X \leq x_i, Y \leq y_j) = \frac{\sum_{m=1}^i \sum_{l=1}^j n_{ml} - 0,44}{N + 0,12}, \tag{10}$$

where,  $N$  is the total number of the variables,  $j$  and  $i$  ascending ranks of  $X_i$  and  $Y_j$ ,  $n_{ml}$  is the number of occurrence of the combinations of  $X_i$  and  $Y_j$ .

**Return periods of the variables and conditional distribution**

In hydrological frequency analysis, the return period of the hydrological variable that occurs once in a year can be defined as:

$$T = \frac{1}{P(X \geq x)} = \frac{1}{(1 - F(x))}, \tag{11}$$

where,  $T$  is return period in years and  $F(x)$  is univariate cumulative distribution function.

In multivariate statistical analysis, we can determine the return period of the phenomenon in two ways. The first is a joint return period, while the second is a conditional return period.

*Joint return period* for two variables defined more authors (Salvadori, 2004; Salvadori and De Michele, 2006; Shiau, 2003) and it can be written in the form of:

$$T^{and}_{x,y} = \frac{1}{(1 - F(x) - F(y) + H(x, y))}, \tag{12}$$

or

$$T^{or}_{x,y} = \frac{1}{(1 - H(x, y))}. \tag{13}$$

Equation (12) represents the joint return period of  $X \geq x$  and  $Y \geq y$ . Equation (13) represents joint return period of  $X \geq x$  or  $Y \geq y$ . These relationships indicate, that different combinations of the numbers  $x$  and  $y$ , can take same return period.  $H(x, y)$  is the joint cumulative distribution function (can be expressed as copula function – Equation (3)).

*Conditional return period* for  $X$ , given  $Y \geq y$  may be expressed as (Shiau, 2003):

$$T_{(x|Y \geq y)} = \frac{1}{(1 - F(y))(1 - F(x) - F(y) + H(x, y))}, \tag{14}$$

where  $x$  and  $y$  are random variables and  $H(x, y)$  is the joint cumulative distribution function. Conditional cumulative distribution function of  $X$ , given  $Y \geq y$  may be expressed as:

$$F_{(x|Y \geq y)} = \frac{F(y) - H(x, y)}{1 - F(y)}, \tag{15}$$

where  $H(x, y)$  is the joint cumulative distribution function of the

random variables  $X$  and  $Y$ , and  $F(y)$  is cumulative distribution function of the variable  $Y$ . An equivalent formula for conditional return period of  $Y \geq y$ , given  $X \leq x$  can be obtained.

**CASE STUDY**

**Hydrological characteristics of the Danube River**

To illustrate the bivariate hydrological analysis, 138 years (1876–2013) of daily discharge data of the Danube River at Bratislava in Slovakia were used.

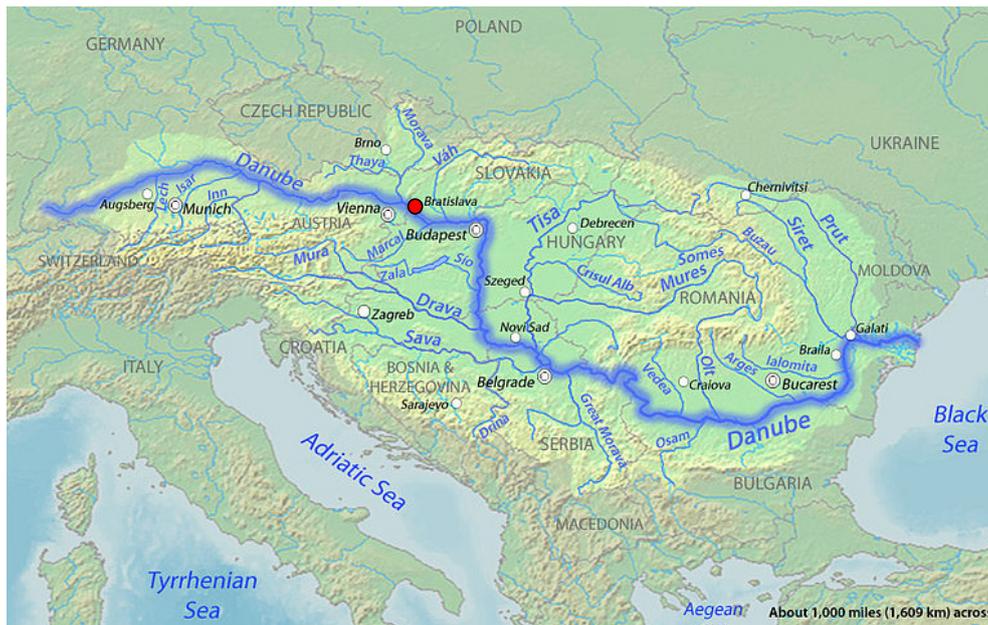
The Danube River is the second greatest river in Europe. The basin covers an area of 817 000 km<sup>2</sup>. The river originates from the Black Forest in Germany at the confluence of the Briga and the Breg streams. The Danube then flows southeast for 2,872 km (1,785 mi), passing through four Central European capitals before emptying into the Black Sea via the Danube Delta in Romania and Ukraine (Figure 1). The Slovak part of the Danube River is situated from river-km 1708.2 to river-km 1880.2. The length of around 7.5 km of the river creates a natural border to Austria, 22.5 km is in Slovakia and the rest of 142 km is the state border to Hungary. Between the Vienna and the Danube lowland, the Danube flows in concentrated channel with relatively high bed slope. After leaving the Small Carpathians, it keeps the slope and flows over its alluvial cone

through a complicated network of branches and meanders downstream to the town of Medvedov.

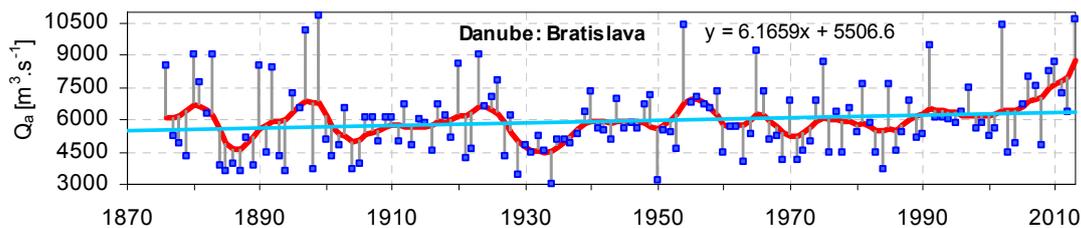
The different physical features of the river basin affect the amount of water runoff in its three sections. In the upper Danube, the runoff corresponds to that of the Alpine tributaries, where the maximum occurs in June when melting of snow and ice in the Alps is very intensive. Runoff drops to its lowest point during the winter months. In the middle basin, the phases last up to four months, with two runoff peaks in April and June. The April peak is local. It is caused by the addition of waters from the melting snow in the plains and from the early spring rains of the lowland and the low mountains of the area.

**Relationship between hydrological characteristics**

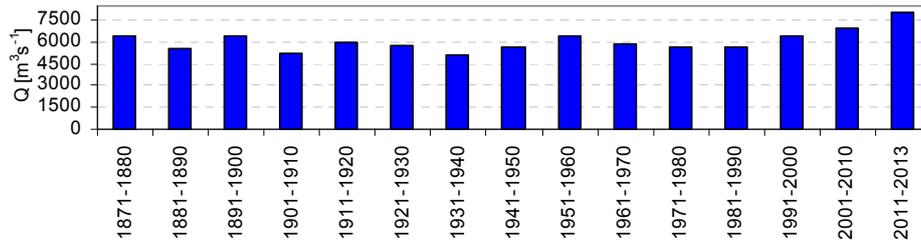
The number of 138 flood events during the period of 1876–2013 on the Danube River at Bratislava in Slovakia was analyzed. The events were based on the annual maximum discharge series method. The trend of the annual flood peaks of the selected flood events during the period of 1876–2013 is shown in Figure 2 and the decadal averages of annual peak discharges is illustrated in Figure 3. The annual peak discharges of the selected floods show a slight upward trend (linear line) since 1876 and ranging from 3000 m<sup>3</sup>s<sup>-1</sup> to 10 870 m<sup>3</sup>s<sup>-1</sup>.



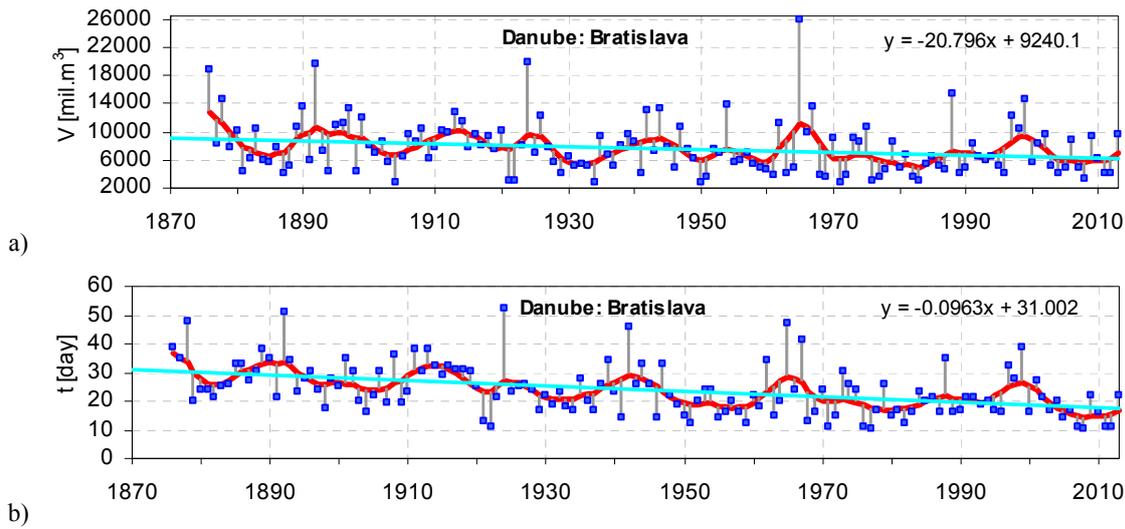
**Fig. 1.** Map of the Danube River basin (red point indicates Bratislava location).



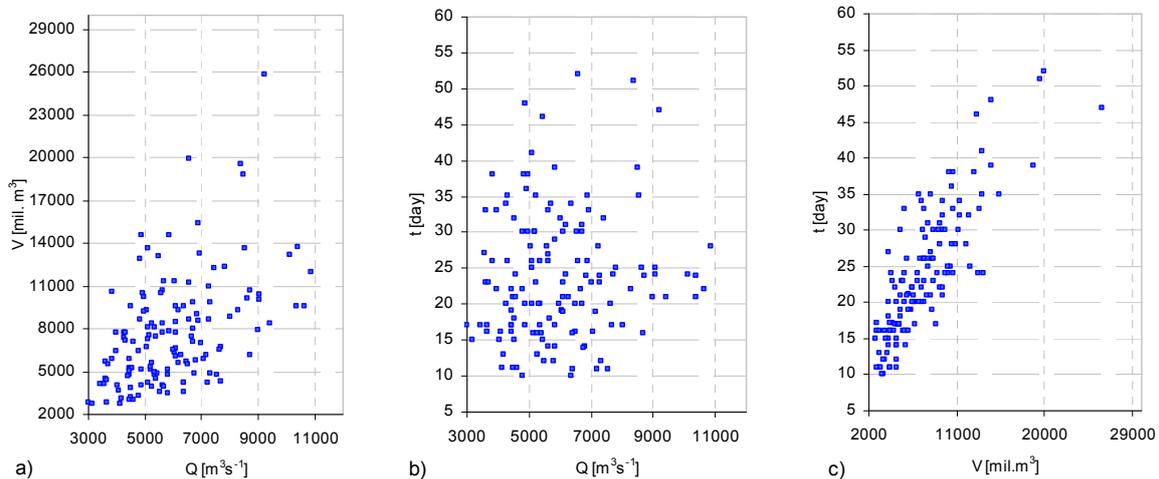
**Fig. 2.** Temporal variability of annual peak discharges (square points), their averages through a 7-year moving window, and a linear trend line (period of 1876–2013, Danube River: Bratislava).



**Fig. 3.** Decadal averages of the annual peak discharges of selected flood events during the period of 1876–2013 (Danube River: Bratislava).



**Fig. 4.** Top panel a) volumes and down panel b) durations corresponding to the annual peaks discharges during the period of 1876–2013 (Danube River: Bratislava) shown in Fig. 2.



**Fig. 5.** Relationships between analyzed pairs of the selected flood events during the period of 1876–2013 (Danube River: Bratislava): a)  $Q$ - $V$ , b)  $Q$ - $t$  and c)  $V$ - $t$ .

From the flood hydrograph of the each wave, the corresponding volume and duration values were determined. Theoretically, the determination of the flood duration involves the identification of the time of occurrence of the start and the end of flood runoff. Generally, the start and the end of a flood event is determined by: (1) fixing some threshold discharge and considering the upper part of the hydrograph as a flood event, and the corresponding duration value as the flood duration (Ashkar and

Rousselle, 1982; Grimaldi and Serinaldi, 2006; Todorovic and Zelenhasic, 1970), and (2) by finding the time boundaries marked by a rise in discharge from the base flow denoting the start of a flood runoff, and a return to base flow, denoting the end of the flood runoff (Linsley et al., 1975; Yue et al., 1999).

In our study, the duration of the each flood wave is determined as the difference between time  $t_0$  (start - increasing of the discharge from level  $\bar{Q}_a$  - long-term average annual discharge)

and time  $t_n$  (end - decreasing of the discharge on level  $\bar{Q}_a$ ). Corresponding volume  $V$  for the particular flood wave was calculated for derived flood wave duration.

**Table 2.** Basic statistic characteristics of the analyzed hydrological variables during the period of 1876–2013 (Danube River: Bratislava).

	$Q$	$V$	$t$
Mean	5972.14	7670.05	23.69
Standard Deviation	1690.77	3826.03	8.83
Coefficient of skewness	0.79	1.55	0.86
Minimum	3000	2722	10
Maximum	10870	25879	52
Coefficient of variation	0.28	0.50	0.37
Count	138	138	138

**Table 3.** Correlation matrix of the analyzed pairs during the period of 1876–2013 (Danube River: Bratislava).

Correlation matrix	$Q-V$	$Q-t$	$V-t$
Spearman $\rho$	0.48	0.12	0.82
Kendall $\tau$	0.34	0.07	0.65
Pearson	0.49	0.13	0.83

The corresponding volume and duration values of selected flood waves are illustrated in Figure 4 a) and b). Figure 4 reveals the downward trends in the corresponding volume and duration of the selected flood waves (linear line). Volume values range is from 2 722 to 25 789 mil.m<sup>3</sup> and duration values range is from 10 to 52 days. Some basic statistical characteristics of these variables are presented in Table 2. The correlations between analyzed pairs  $Q-V$ ,  $Q-t$ ,  $V-t$  are shown in Figure 5a–c. Correlation coefficients of pairs are presented in Table 3. Analyzed pairs  $Q-V$  and  $V-t$  show an intermediate dependence or higher dependence. Analyzed pairs  $Q-t$  show very low dependence.

## RESULTS

### Univariate analysis

The first step of bivariate analysis is to identify marginal distributions for three main hydrological variables. The random variables may have different properties and thus need to be converted to variables having interval of [0,1] by scaling the data. Knowing the marginal distribution, we are able to separate marginal behavior and dependence structure. The dependence structure is fully described by the joint distribution of uniform variables obtained from marginal distribution.

In order to determine univariate parametrical distribution functions (marginal distributions), standard MLM (maximum likelihood method) method was used.

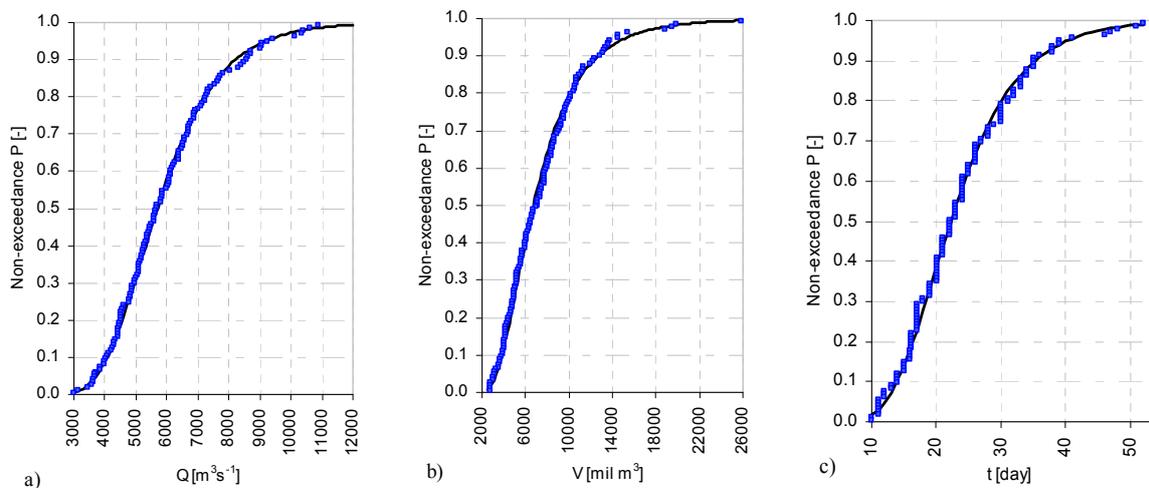
The two-parameter Gumbel distribution was used to describe the peak discharges  $Q$  and duration  $t$ . This distribution is one of the most frequently used and well suited for modeling extreme hydrologic events in hydrology. The three-parameter LogGamma distribution was used for volume  $V$  modeling. The Log-Gamma distribution is a very important model in statistical hydrology. It is a flexible three-parameter family capable of taking many different shapes and has been widely used in many countries for flood series modeling.

According to the goodness of fit tests (Kolmogorov-Smirnov and  $\chi^2$ ), we cannot reject the hypothesis that both selected distributions fit well to the observed data at a 5% significance level. Parameters of these marginal distributions are illustrated in Table 4.

Therefore, the empiric distribution evaluated with Cunnane (1988) formula have been fitted with selected parametrical cumulative distribution functions (Fig. 6 a)–c)).

**Table 4.** Parameters of univariate cumulative distribution functions and  $P_{\text{value}}$  of the Kolmogorov-Smirnov test ( $K-S$ ) and Chi square test ( $\chi^2$ ) (Danube River: Bratislava).

Variables (cdf)	Location	Scale	Shape	$P_{\text{value}} K-S$	$P_{\text{value}} \chi^2$
$Q$ (Gumbel)	5195.67	1341.26		0.97	0.99
$t$ (Gumbel)	19.65	6.94		0.95	0.81
$V$ (LogGamma (3P))	273.8	0.0318	741.32	0.84	0.76



**Fig. 6.** Comparison between Cunnane empirical distribution and a) theoretical Gumbel cumulative distribution function for  $Q$ , b) theoretical LogGamma cumulative distribution function for  $V$ , and c) theoretical Gumbel cumulative distribution function for  $t$  (point-empirical data, line-theoretical distribution).

**Bivariate analysis**

Copula functions were used as mathematical tool for determining a joint cumulative distribution of two dependent variables. The copulas considered in the study are included into the class of Archimedean copulas. The Clayton, Gumbel-Hougaard and Frank copulas were selected for this case study. The copula parameters (Table 5) were estimated using a mathematical relationship between the Kendall's coefficient of rank correlation and the generating function.

**Table 5.** Copula parameters.

Pairs	Clayton	Gumbel-Hougaard	Frank
<i>Q-V</i>	<b>1.01</b>	1.51	3.34
<i>Q-t</i>	<b>0.114</b>	1.057	0.486
<i>V-t</i>	3.333	<b>2.67</b>	8.63

**Table 6.** Information criterions of the model.

Pairs		Clayton	Gumbel-Hougaard	Frank
<i>Q-V</i>	SIC	<b>15.63</b>	24.76	26.47
	AIC	<b>21.4</b>	30.35	32.23
	HQIC	<b>19.11</b>	28.24	29.94
<i>Q-t</i>	SIC	<b>-9.68</b>	-9.45	-8.53
	AIC	<b>-3.92</b>	-3.69	-2.77
	HQIC	<b>-6.21</b>	-5.98	-5.08
<i>V-t</i>	SIC	124.68	<b>107.9</b>	140.34
	AIC	130.44	<b>113.72</b>	146.11
	HQIC	128.15	<b>111.43</b>	143.82

For the goodness-of-fit-test, three criterions: SIC - Schwarz information criterion, AIC - Akaike information criterion, and HQIC- Hannan-Quinn information criterion were used. These information criteria are based on the methods MSE (mean squared error); and the basic principle of the criterions is to minimize the mean rate variations between the filter output values and input values (Table 6). Clayton copula shows the

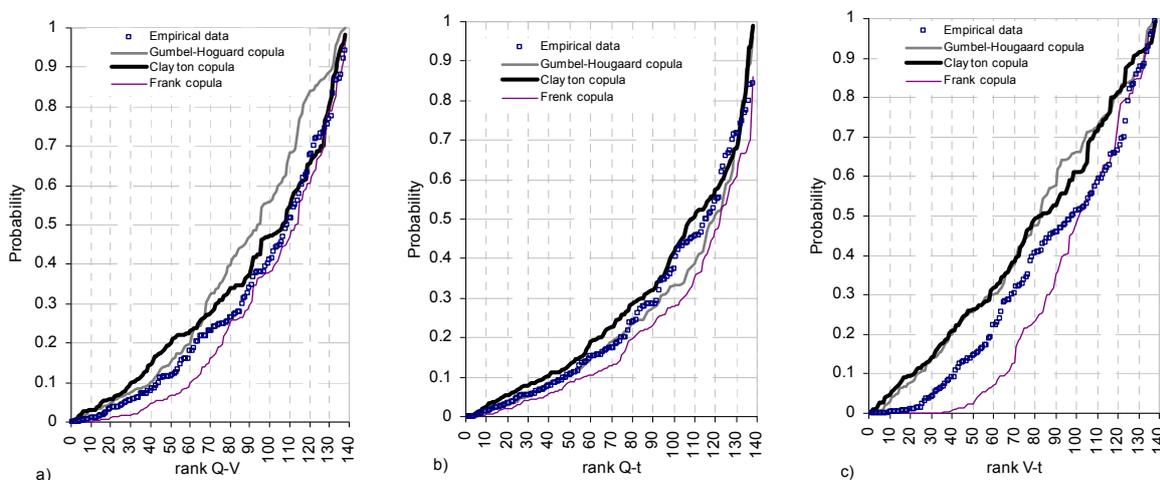
lowest values of calculated criterions for pairs *Q-V* and *Q-t*, and Gumbel-Hougaard copula shows the lowest values of calculated criterions for pairs *V-t*. Testing of the goodness of fit of the copulas is still discussed as a topic in literature. Kojadinovic and Yan (2011) compared the empirical copula with a parametric estimate of the copula derived under the null hypothesis. They extended the multiplier approach to multivariate multi-parameter copulas and study the finite-sample performance of the resulting goodness of fit test. For example, Chowdhary et al. (2011), Karmakar and Simonovic (2009), Shiau et al. (2010) used some of these criterions for testing the goodness of fit of the copulas. Chen and Fan (2005) introduced the so called "pseudo-likelihood ratio test", inspired by a semi-parametric adaptation of the criterion AIC.

Secondly, the Gringorten plotting-position formula (Eq. 10) was used for the assessment of copula function fitting (Karmakar and Simonovic, 2007; Yue et al., 1999; Zhang and Singh, 2007). Comparison of empirical joint distribution for observed pairs *Q-V*, *Q-t*, *V-t* and the corresponding probability values derived by copulas are presented in Fig. 7. Based on the information criteria and visual comparison, the next copulas were selected as the most suitable to illustrate the joint distribution of the variables: *Q-V* – Clayton copula, *Q-t* – Clayton copula, *V-t* – Gumbel-Hougaard copula.

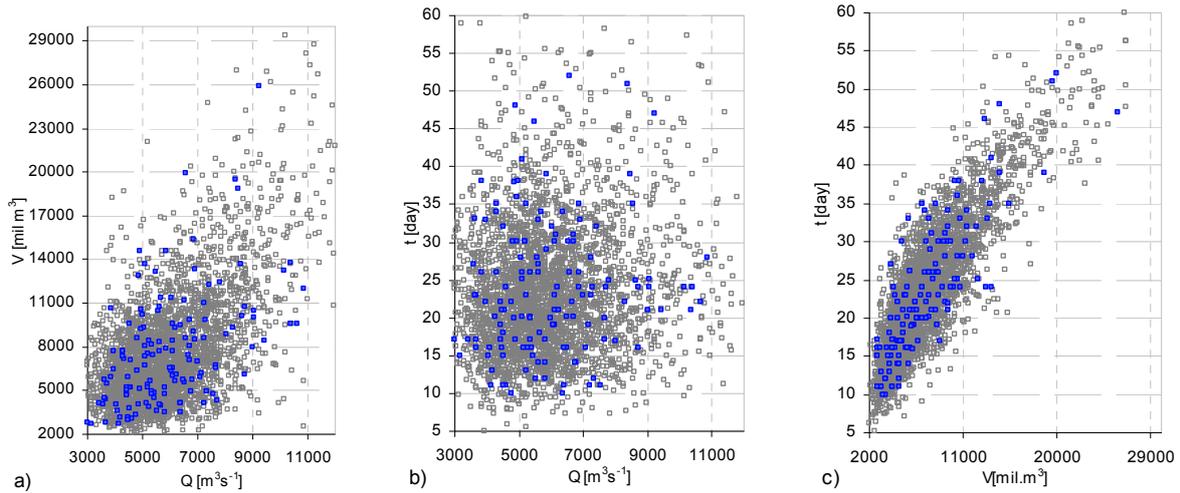
Then, the simulations of 3000 pairs using the selected copulas were done (Fig. 8). The main purpose of simulations using copulas is to describe and to illustrate the interdependences of two hydrological variables. Correlation coefficients of the simulated pairs from appropriate copulas are summarized in Table 7. Also, Figure 9a-c illustrates contour maps of analyzed pairs with same joint cumulative distribution.

**Table 7.** Correlation matrix of the 3000 simulated pairs *Q-V*, *Q-t*, *V-t* (Danube River: Bratislava).

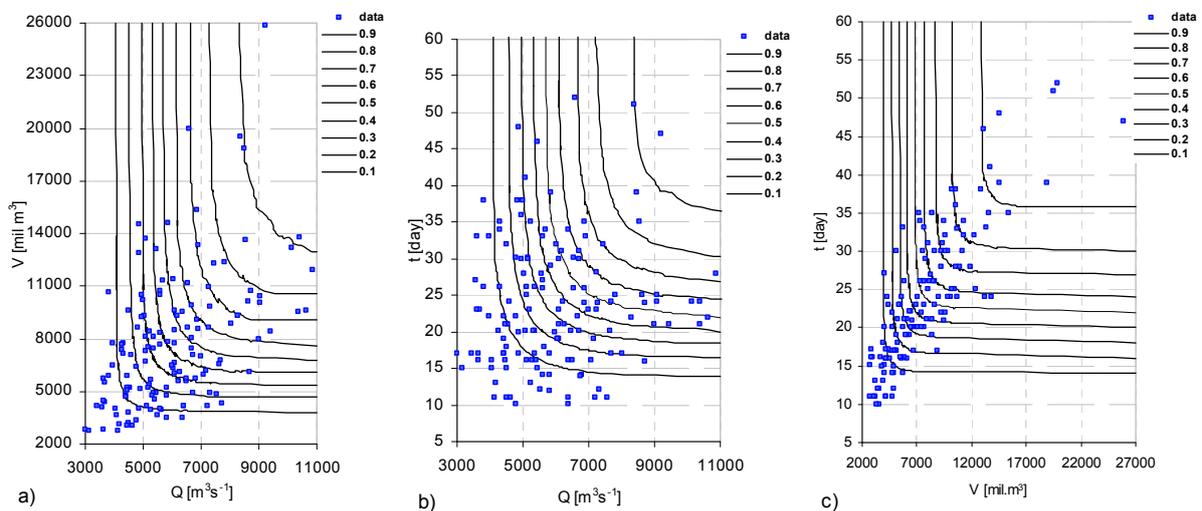
Correlation matrix	<i>Q-V</i> (Clayton)	<i>Q-t</i> (Clayton)	<i>V-t</i> (Gumbel-Hougaard)
Spearman $\rho$	0.477	0.101	0.82
Kendall $\tau$	0.342	0.071	0.621
Pearson	0.472	0.102	0.819



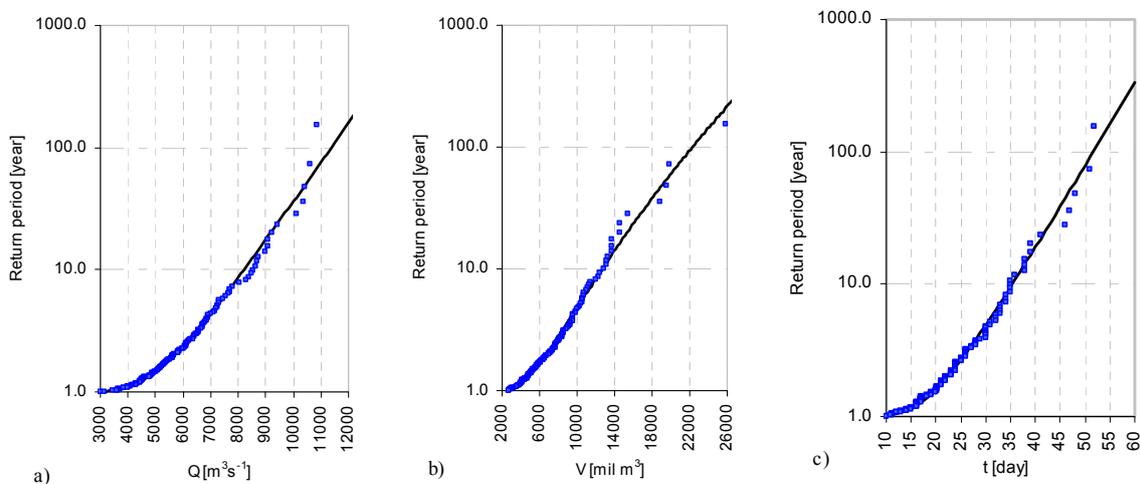
**Fig. 7.** Comparison of the empirical joint distribution for observed pairs and selected copulas (Danube River: Bratislava): a) *Q-V*, b) *Q-t* and c) *V-t*.



**Fig. 8.** Scatter plot of 3000 pair simulations according copula and measured data during the period of 1876–2013 (Danube River: Bratislava): a)  $Q$ - $V$ : Clayton copula, b)  $Q$ - $t$ : Clayton copula, c)  $V$ - $t$ : Gumbel-Hougaard copula.



**Fig. 9.** Contour map of the pairs with same joint cumulative distribution using selected copula functions (Danube River: Bratislava): a) Clayton copula, b) Clayton copula and c) Gumbel-Hougaard copula (point- empirical data, line- same joint distribution).



**Fig. 10.** Univariate return period of the hydrological variables during the period of 1876–2013 (Danube River: Bratislava): a) Gumbel distribution for peak discharge  $Q$ , b) LogGamma distribution for duration  $t$  and c) Gumbel distribution for volume  $V$  (point-empirical data, line- theoretical distribution).

**Table 8.** Joint and conditional return periods for pairs  $Q$ - $V$  (Danube River: Bratislava).

$T$ [year]	$F_x$	$F_y$	$Q$ [ $\text{m}^3\text{s}^{-1}$ ]	$V$ [mil. $\text{m}^3$ ]	$C_c$	$T_{xy}^{and}$ [year]	$T_{xy}^{or}$ [year]	$T_{Q/V}$ [year]
2	0.5	0.5	5687	6720.7	0.334	3	2	6
5	0.8	0.8	7207	10120.2	0.667	15	3	8
10	0.9	0.9	8214	12680.1	0.818	55	6	12
20	0.95	0.95	9179	15358.9	0.905	209	11	22
<b>50</b>	<b>0.98</b>	<b>0.98</b>	<b>10429</b>	<b>19162.3</b>	<b>0.961</b>	<b>1268</b>	<b>26</b>	<b>52</b>
100	0.99	0.99	11366	22276.7	0.980	5020	51	102

$C_C$  - Clayton copula,  $C_G$  - Gumbel-Hougaard copula

**Table 9.** Joint and conditional return periods for pairs  $Q$ - $t$  (Danube River: Bratislava).

$T$ [year]	$F_x$	$F_y$	$Q$ [ $\text{m}^3\text{s}^{-1}$ ]	$t$ [day]	$C_c$	$T_{xy}^{and}$ [years]	$T_{xy}^{or}$ [year]	$T_{Q/t}$ [year]
2	0.5	0.5	5687.26	22	0.263	4	1	4
5	0.8	0.8	7207.49	30	0.644	23	3	6
10	0.9	0.9	8214.01	35	0.811	91	5	11
20	0.95	0.95	9179.49	40	0.903	361	10	21
50	0.98	0.98	10429.20	47	0.960	2250	25	51
100	0.99	0.99	11365.69	52	0.980	8988	50	101

$C_C$  - Clayton copula,  $C_G$  - Gumbel-Hougaard copula

**Table 10.** Joint and conditional return periods for pairs  $V$ - $t$  (Danube River: Bratislava).

$T$ [year]	$F_x$	$F_y$	$V$ [mil. $\text{m}^3$ ]	$t$ [day]	$C_G$	$T_{xy}^{and}$ [years]	$T_{xy}^{or}$ [year]	$T_{V/t}$ [year]
2	0.5	0.5	6720.72	22	0.4070	2	2	11
5	0.8	0.8	10120.19	30	0.7487	7	4	19
10	0.9	0.9	12680.01	35	0.8723	14	8	36
20	0.95	0.95	15358.87	40	0.9356	28	16	70
50	0.98	0.98	19162.31	47	0.9741	71	39	171
100	0.99	0.99	22276.68	52	0.9870	142	77	339

**Table 11.** Joint and conditional return periods for the maximum hydrological variables during period of 1876–2013 on the Danube River in Bratislava

$T_x$	$T_y$	$F_x$	$F_y$	$Q$ [ $\text{m}^3\text{s}^{-1}$ ]	$V$ [mil. $\text{m}^3$ ]	$C_c$	$T_{xy}^{and}$ [years]	$T_{xy}^{or}$ [years]	$T_{Q/V}$ [years]
69	210	0.985	0.995	10870	25879	0.981	7272	52	216
				$Q$ [ $\text{m}^3\text{s}^{-1}$ ]	$t$ [day]	$C_c$			
69	105	0.986	0.990	10870	52	0.976	6514	42	107
				$V$ [mil. $\text{m}^3$ ]	$t$ [day]	$C_G$			
210	105	0.995	0.990	25879	52	0.989	236	99	189

$C_C$  - Clayton copula,  $C_G$  - Gumbel-Hougaard copula

### Joint and conditional return periods

Traditionally, a univariate return period is used to describe the hydrological extreme. The univariate return periods for hydrological variables estimated from selected distributions for this case study are illustrated in Fig. 10a–c. This approach is useful when only one extreme random variable is a significant design criterion. It means that, if the correlation between random variables is insignificant in the design criterion, the return periods for the extreme events can be defined rather than using only one random variable.

However, in many cases for hydrologic engineering, a flood may be considered as dangerous if its discharge is large either its volume or duration is also large. Therefore, in this present

case study, few particular joint and conditional return periods are evaluated via the Clayton and Gumbel-Hougaard copulas. The joint return periods for pairs  $Q$ - $V$ ,  $Q$ - $t$ ,  $V$ - $t$  were calculated according to the equations (12) and (13). Conditional return periods were calculated according to equation (14). The joint and conditional return periods for some specific hydrological variables, which can occur simultaneously with univariate return periods (and univariate probability of exceedence) of 2 (50%), 5 (20%), 10 (10%), 20 (2%), 50 (5%) and 100 (1%) years were investigated from Tables 8–10. The result shows the following inequality:

$$T^{or}_{(x,y)} \leq \min [T_x, T_y] \leq \max [T_x, T_y] \leq T^{and}_{(x,y)}$$

For example, for  $Q = 10\,429\text{ m}^3\text{s}^{-1}$  and  $V = 19\,162.3\text{ mil. m}^3$  is  $T^{or}_{Q,V} = 26$  years and  $T^{and}_{Q,V} = 1\,262$  years (bold in Table 8). It seems that these two magnitudes are also the lower limits for a 50-year flood defined by  $T^{or}$  and the upper limits for a 50-year flood defined by  $T^{and}$ . The same inequality may also be applied for other analyzed hydrological pairs. Next, the joint distribution functions and appropriate return periods (joint and conditional) for maximum values of discharge, volume and duration measured and derived on flood waves during the period of 1876–2013 are presented in Table 11.

## CONCLUSIONS AND DISCUSSION

In this present case study, the bivariate flood frequency analysis of the peak discharges, volumes and durations via copula function was done. The time period of 1876–2013 daily discharges data from the Danube River at Bratislava were analyzed. The first maximum annual discharges ( $Q$ ) were selected and the flood waves on these maximum annual discharges were determined. Then, the time duration ( $t$ ) and volume ( $V$ ) for each selected wave was computed.

Marginal distribution functions of the data were selected first. Observed and derived hydrological data  $Q$ ,  $V$ ,  $t$  were fitted by the two-parameter Gumbel (discharge, duration) and three-parameter LogGamma (volume) distributions. The Kolmogorov-Smirnov and  $\chi^2$  tests were used to test the goodness-of-fit of the observed data and the data obtained by theoretical distribution functions. Based on these tests and graphical comparison between measurements and derived data, selected parametrical distribution functions were used as the marginal distribution functions for bivariate frequency analysis.

Three most popular Archimedean copulas in hydrology (Clayton, Gumbel-Hougaard and Frank) were employed to construct the joint distribution of hydrological pairs of  $Q$ - $V$ ,  $Q$ - $t$  and  $V$ - $t$ . The joint CDFs (Cumulative Distribution Functions) evaluated from copula method were compared with their empirical probabilities (Eq. 10) and AIC, MSE and HQIC information criteria were used to test the goodness of fit of the model. The result shows that Clayton copula function has best fit for pairs  $Q$ - $V$  and  $Q$ - $t$  and Gumbel copula has the best fit for pair  $V$ - $t$  in our case study. Clayton copula is a suitable model for simulation pairs of  $Q$ - $V$  reported in Chowdhary et al. (2011). In Karmakar and Simonovic (2007), the Gumbel-Hougaard copula has the best fit for joint distribution of the pairs  $Q$ - $V$  and  $V$ - $t$ . Genest and Favre (2007) tested some extreme-value copulas and they suggested that these copulas (including Gumbel-Hougaard copula) are good tools for bivariate modeling of the hydrological extreme pairs of  $Q$ - $V$ . Also, Requena et al. (2012) suggests that Frank copula function has the best fit for pairs  $Q$ - $V$  in dams. However, we can assume that the choice of a particular copula might depend on the nature of the particular data set. This would be useful for further studying for more cases of selected data set and other copula functions.

This case study also presents basic relations to define a joint return period and conditional period for hydrological bivariate analysis. As mentioned above, the flood is an extreme hydrological event, where a devastating effect may be increased in the case where its three basic components (or at least two of them)  $Q$ ,  $V$  and  $t$  have the same significance. Therefore, it is appropriate to describe the extreme hydrological events jointly using two random variables. Return periods in a bivariate flood frequency analysis can be defined in two ways. The first one defines joint return periods as: the return periods using one random variable equaling or exceeding a certain magnitude and/or using another random variable equaling or exceeding another

certain magnitude. The second one is conditional return periods for one random variable, given that another random variable equals or exceeds a specific magnitude. The results of these return periods for specific values of  $Q$ ,  $V$  and  $t$  (univariate return period of 2, 5, 10, 20, 50 and 100 years) are summarized in Tables 8–10.

The applications of the copula functions in hydrology for extreme frequency analysis are still growing. In recent years, a large number of applications, besides the already mentioned in present case study have been done (see: Reddy and Ganguli (2012), Serinaldi and Kilsby (2013), Xiao et al (2013), Yan et al. (2013)). The results obtained from the multidimensional analysis of the variables, which characterize the hydrological waves (flow, volume, time) can contribute to more reliable assessment of flood risks. Hence, they give an overview of the flood event as a whole and might be practically used in water management and in the design of flood protective systems.

*Acknowledgements.* This work was supported by the Slovak Research and Development Agency under the contract No. APVV-0496-10 and results from the project implementation of the “Centre of excellence for integrated flood protection of land” (ITMS 26240120004) supported by the Research & Development Operational Programme funded by the ERDF.

## REFERENCES

- Ashkar, F., Rousselle, J., 1982. A multivariate statistical analysis of flood magnitude, duration, and volume. In: Statistical Analysis of Rainfall and Runoff. Water Resource Publication, Fort Collins, CO, pp. 651–669.
- Balakrishnan, N., Lai, C.D., 2009. Continuous Bivariate Distributions. Distributions Expressed as Copulas. Springer Science + Business Media, Philadelphia, pp. 67–103, doi: 10.1007/b101765\_3.
- Chen, X., Fan, Y., 2005. Pseudo-likelihood ratio tests for semi-parametric multivariate copula model selection. *Can. J. Stat.*, 33, 3, 389–414.
- Chowdhary, H., Escobar, L.A., Singh, V.P., 2011. Identification of suitable copulas for bivariate frequency analysis of flood peak and flood volume data. *Hydrol. Res.*, 42, 2–3, 193–215.
- Cunnane, C., 1978. Methods and merits of regional flood frequency analysis. *J. Hydrol.*, 100, 269–290.
- De Michele, C., Salvadori, G., Canossi, M., Petaccia, A., Rosso, R., 2005. Bivariate statistical approach to check adequacy of dam spillway. *J. Hydrol. Eng.*, 10, 1, 50–57.
- Embrechts, P., Lindskog, F., McNeil, A.J., 2001. Modeling dependence with copulas and applications to risk management. Department of mathematic. Working paper CH–8092, ETH, Zürich, 41 p.
- Faqir, M., Shazia, A., 2007. Modeling annual maximum peak flows at various dams and barrages in Pakistan. *J. Hydrol. Hydromech.*, 55, 1, 43–53.
- Favre, A.C., El Adlouni, S., Perreaut, L., Thiémonge, N., Bobée, B., 2004. Multivariate hydrological frequency analysis using copulas. *Water Resour. Res.*, 40, 1, W01101.
- Genest, C., Favre, A.C., 2007. Everything you always wanted to know about copula modeling but were afraid to ask. *J. Hydrol. Eng.*, 12, 347–368.
- Grimaldi, S., Serinaldi, F., 2006. Asymmetric Copula in Multivariate Flood Frequency Analysis. *Adv. Water Resour.*, 29, 1155–1167.
- Gringorten, I.I., 1963. A plotting rule for extreme probability paper. *J. Geophys. Res.*, 68, 3, 813–814.
- Hall, J., Arheimer, B., Borga, M., Brazdil, R., Claps, P., Kiss, A.,

- Kjeldsen, T.R., Kriaučiūnienė, J., Kundzewicz, Z.W., Lang, M., Llasat, M.C., Macdonald, N., McIntyre, N., Mediero, L., Merz, B., Merz, R., Molnar, P., Montanari, A., Neuhold, C., Parajka, J., Perdigão, R.A.P., Plavcova, L., Rogger, M., Salinas, J.L., Sauquet, E., Schär, C., Szolgay, J., Viglione, A., Blöschl, G., 2013. Understanding flood regime changes in Europe: a state of the art assessment. *Hydrol. Earth Syst. Sci. Discuss.*, 10, 15525–15624.
- Hoefding, W., 1940. Scale of invariant correlation theory. *Schr. Math. Angew. Math., Univ. Berlin, Berlin*, 5, 3, 181–233. (In German.)
- Joe, H., 1997. *Multivariate Models and Dependence Concepts*. Chapman and Hall, New York, 399 p.
- Karmakar, S., Simonovic, S.P., 2007. Flood frequency analysis using copula with mixed distributions. Project report No. 055. The University of Western Ontario, Department of Civil and Environmental Engineering. London, Ontario, Canada 58 p.
- Karmakar, S., Simonovic, S.P., 2009. Bivariate flood frequency analysis. Part 1: Determination of marginals by parametric and nonparametric techniques. *J. Flood Risk Manag.*, 1, 190–200.
- Kojadinovic, I., Yan, J., 2011. A Goodness-of-Fit test for multivariate multiparameter copulas based on multiplier central limit theorems. *Statistics and Computing*, 21, 1, 1–26.
- Le Clerc, S., Lang, M., 2002. Flood frequency analysis downstream confluences. Comparison between bivariate densities and experimental data. In: *International Conference on Flood Estimation*, Berne, pp. 295–304.
- Linsley, R.K., Kohler, M.A., Paulhus, J.L.H., 1975. *Applied Hydrology*. McGraw-Hill, New York.
- Maca, P., Torf, P., 2009. The influence of temporal rainfall distribution in the flood runoff modelling. *Soil and Water Res.*, 4 (Special Issue 2), 102–110.
- Merz, R., Blöschl, G., 2003. A process typology of regional floods. *Water Resour. Res.*, 39, 12, 1340–1360.
- Myers, J.L., Well, A.D., 2003. *Research Design and Statistical Analysis*. 2nd ed. Lawrence Erlbaum, Hillsdale, NJ, 508 p.
- Nelsen, R.B., 1999. *An Introduction to Copulas*. Lecture Notes in Statistics. Springer, New York.
- Nelsen, R.B., 2006. *An Introduction to Copula*. 2nd ed. Springer, New York.
- Parajka, J., Merz, R., Szolgay, J., Blöschl, G., Kohnová, S., Hlavčová, K., 2008. A comparison of precipitation and runoff seasonality in Slovakia and Austria. *Meteorological Journal*, 4, 9–14.
- Parajka, J., Kohnová, S., Merz, R., Szolgay, J., Hlavčová, K., Blöschl, G., 2009. Comparative analysis of the seasonality of hydrological characteristics in Slovakia and Austria. *Hydrological Sciences Journal*, 54, 3, 456–473.
- Parajka, J., Kohnová, S., Bálint, G., Barbuc, M., Borga, M., Claps, P., Cheval, S., Dumitrescu, A., Gaume, E., Hlavčová, K., Merz, R., Pfaundler, M., Stancalie, G., Szolgay, J., Blöschl, G., 2010. Seasonal characteristics of flood regimes across the Alpine-Carpathian range. *Journal of Hydrology*, 394, 1–2, 78–89.
- Prohaska, S., Isailović, D., Srna, P., Marčetić, I., 1999. Coincidence of flood flow of the Danube River and its tributaries. A Hydrological monograph of the Danube River Basin, Belgrade University, Belgrade.
- Reddy, M.J., Ganguli, P., 2012. Bivariate flood frequency analysis of upper Godavari River flows using Archimedean Copulas. *Water Resources Management*, 26, 14, 3995–4018.
- Requena, A., Mediero, L., Garrote, L., 2012. A Monte Carlo procedure based on the use of copulas for risk assessment of dam overtopping. In: *Proc. 3<sup>rd</sup> STAHY International Workshop on Statist. Methods for Hydrol. and Water Resour. Manag.*, Tunis, pp. 1–11.
- Sackl, B., Bergmann, H., 1987. *A Bivariate Flood Model and its Application*. Hydrologic Frequency Modeling. Riedel, Dordrecht, The Netherlands, pp. 571–582.
- Salvadori, G., 2004. Bivariate return periods via 2-Copulas. *Stat. Methodol.*, 1, 129–144.
- Salvadori, G., De Michele, C., 2004. Frequency analysis via copulas: theoretical aspects and applications to hydrological events. *Water Resour. Res.*, 40, W12511.
- Salvadori, G., De Michele, C., 2006. Statistical characterization of temporal structure of storms. *Adv. Water Resour.*, 29, 6, 827–842.
- Savu, C., Trede, M., 2006. *Hierarchical Archimedean Copulas*. Institute of Econometrics, University of Muenster, Muenster, Germany.
- Schweizer, B., Wolff, E.F., 1981. On nonparametric measures of dependence for random variables. *Ann. Stat.*, 9, 879–885.
- Serinaldi, F., Kilsby, C.G., 2013. The intrinsic dependence structure of peak, volume, duration, and average intensity of hyetographs and hydrographs. *Water Resour. Res.*, 49, 6, 3423–3442.
- Shiau, J.T., 2003. Return period of bivariate distributed extreme hydrological events. *Stoch. Envir. Res. Risk Assess.*, 17, 42–57.
- Shiau, J.T., Wang, H.Y., Tsai, C.T., 2010. Copula-based depth-duration-frequency analysis of typhoons in Taiwan. *Hydrol. Res.*, 41, 5, 414–422.
- Sklar, A., 1959. Fonction de répartition à n dimensions et leurs marges. [Distribution functions, dimensions and margins]. Publications of the Institute of Statistics, University of Paris, Paris, pp. 229–231. (In French.)
- Spearman, C., 1904. The proof and measurement of association between two things. *American Journal of Psychology*, 15, 72–101.
- Todorovic, P., 1978. Stochastic models of floods. *Water Resour. Res.*, 14, 2, 345–356.
- Todorovic, P., Zelenhasic, E., 1970. A stochastic model for flood analysis. *Water Resour. Res.*, 6, 6, 1641–1648.
- Xiao, M., Zhang, Q., Chen, Y., Chen, X., 2013. Hydrological drought frequency analysis of east river basin based on trivariate copulas function. *Journal of Natural Disasters*, 22, 2, 99–108.
- Yan, B., Guo, S., Yu, W., 2013. Coincidence risk of flood hydrographs between Yangtze River and Qing River Shuili Fadian Xuebao. *Journal of Hydroelectric Engineering*, 32, 1, 50–53.
- Yue, S., Ouarda, T.B.M.J., Bobée, B., Legendre, P., Bruneau, P., 1999. The Gumbel mixed model for flood frequency analysis. *J. Hydrol.*, 226, 1–2, 88–100.
- Zhang, L., Singh, V.P., 2006. Bivariate flood frequency analysis using the copula method. *J. Hydrol. Eng. ASCE*, 11, 2, 150–164.
- Zhang, L., Singh, V.P., 2007. Trivariate flood frequency analysis using the Gumbel-Hougaard copula. *J. Hydrol. Eng.*, 12, Special Issue on Copulas in Hydrology, 431–439.

Received 12 March 2014

Accepted 14 May 2014

Note: Colour version of Figures can be found in the web version of this article.