

# Joint modelling of flood peaks and volumes: A copula application for the Danube River

George Papaioannou<sup>1\*</sup>, Silvia Kohnová<sup>3</sup>, Tomáš Bacigál<sup>2</sup>, Ján Szolgay<sup>3</sup>, Kamila Hlavčová<sup>3</sup>, Athanasios Loukas<sup>1</sup>

<sup>1</sup> Laboratory of Hydrology and Aquatic Systems Analysis, Department of Civil Engineering, University of Thessaly, Pedion Areos, 38334 Volos, Greece.

<sup>2</sup> Slovak University of Technology in Bratislava, Faculty of Civil Engineering, Department of Mathematics and Constructive Geometry, Radlinského 11, 810 05 Bratislava, Slovakia.

<sup>3</sup> Slovak University of Technology in Bratislava, Faculty of Civil Engineering, Department of Land and Water Resources Management, Radlinského 11, 810 05 Bratislava, Slovakia.

\* Corresponding author. Tel.: +302421074153. E-mail: gpapaioa@uth.gr

**Abstract:** Flood frequency analysis is usually performed as a univariate analysis of flood peaks using a suitable theoretical probability distribution of the annual maximum flood peaks or peak over threshold values. However, other flood attributes, such as flood volume and duration, are necessary for the design of hydrotechnical projects, too. In this study, the suitability of various copula families for a bivariate analysis of peak discharges and flood volumes has been tested. Streamflow data from selected gauging stations along the whole Danube River have been used. Kendall's rank correlation coefficient ( $\tau$ ) quantifies the dependence between flood peak discharge and flood volume settings. The methodology is applied to two different data samples: 1) annual maximum flood (AMF) peaks combined with annual maximum flow volumes of fixed durations at 5, 10, 15, 20, 25, 30 and 60 days, respectively (which can be regarded as a regime analysis of the dependence between the extremes of both variables in a given year), and 2) annual maximum flood (AMF) peaks with corresponding flood volumes (which is a typical choice for engineering studies). The bivariate modelling of the extracted peak discharge - flood volume couples is achieved with the use of the Ali-Mikhail-Haq (AMH), Clayton, Frank, Joe, Gumbel, Hüsler-Reiss, Galambos, Tawn, Normal, Plackett and FGM copula families. Scatterplots of the observed and simulated peak discharge - flood volume pairs and goodness-of-fit tests have been used to assess the overall applicability of the copulas as well as observing any changes in suitable models along the Danube River. The results indicate that for the second data sampling method, almost all of the considered Archimedean class copula families perform better than the other copula families selected for this study, and that for the first method, only the upper-tail-flat copulas excel (except for the AMH copula due to its inability to model stronger relationships).

**Keywords:** Bivariate frequency analysis; Copulas; Dependence of flood peaks and volumes; Kendall's rank correlation coefficient; Danube River.

## INTRODUCTION

The design of flood protection structures where storage is involved requires an entire hydrograph or at least the flood volume/shape estimates related to the flood peaks. Therefore, the relationship between peaks of annual maximum floods (AMF) and their volumes is an interesting scientific research issue. From a broader perspective, however, the dependence between the annual extremes of both variables is also of interest when studied from the perspective of regimes.

In the past, identical marginal distributions for both of these random variables have often been used for modelling their dependence in hydrology (e.g., Goel et al., 1998; Yue et al., 2001). As flood peaks are the most commonly used data in hydrological frequency analysis, the majority of studies have analyzed flood peak frequency curves (Cunnane, 1988, 1989; Dawdy et al., 2012; Groupe de recherche en hydrologie statistique (GREHYS), 1996; Laio et al., 2011; Mediero and Kjeldsen, 2014) in contrast to flood volumes, for which fewer studies are available (e.g. Bačová-Mitková, 2011; Mediero et al., 2010). Under the assumption that flood peaks and volumes have the same type of marginal distributions, several authors have used bivariate distributions for a frequency analysis of these variables (e.g. Singh and Singh, 1991; Shiau et al., 2006; Yue et al., 2001). This requirement, however, is seldom ful-

filled in practice. In the event that the flood peaks and volumes do not have the same type of marginal probability distribution, the copula approach provides a flexible solution (Gaál et al., 2015; Giustarini et al., 2010; Szolgay et al., 2015; Zhang and Singh, 2006). Copulas allow for the combining of various marginal distributions into multivariate distributions (Favre et al., 2004; Genest and Favre, 2007; Nelsen, 2006; Salvadori and De Michele, 2004, 2010). Copulas have recently become a popular tool in hydrological analysis for modeling the relationship among hydrological characteristics. A methodology for using copulas in hydrology has been described e.g. by Dupuis (2007) and Genest and Favre (2007). Recently, a lot of studies have implemented the application of copulas in engineering practice (see, e.g., Aronica et al., 2012; Balistocchi and Baldassarre, 2011; Bezak et al., 2016; De Michele and Salvadori, 2003; De Michele et al., 2005; Gaál et al., 2010; Zhang and Singh, 2007). Favre et al., (2004) were among the first authors who instituted the use of a two-dimensional copula for describing the relationship between flood discharges and volumes. Zhang and Singh (2006) stressed that bivariate copula-based distributions of flood peaks vs. volumes and flood volumes vs. durations provide better results for showing agreement when plotting frequency estimates than traditional distributions. Bačová-Mitková and Halmová (2014) applied parametric families of Archimedean copulas for an analysis of the relationship between vol-

umes, peak discharges and durations at the Bratislava gauging station on the Danube River. Szolgay et al. (2012) used a joint analysis of maximum discharges and volumes through copulas for the estimation of design quantities. The analysis described a case study in the Vltava River Basin in the Czech Republic to estimate the design discharge for a return period of 10,000 years to assess the safety of the Orlik dam. The more recent study of Sraj et al. (2015) applied a bivariate copula analysis to the annual maximum discharges and volumes from a gauging station on the Sava River in Slovenia. The results from three families were compared, and the Gumbel-Hougaard copula was found to be the most appropriate for modeling peak discharges vs. volume and volume vs. duration.

The main objective of this study is to investigate the suitability and evaluate the applicability of the selected copula models for the flood peak–volume relationship along the Danube River using streamflow data from seven gauging stations. The following eleven copula families were used for the analysis: Ali-Mikhail-Haq, Clayton, Frank, Joe, Gumbel, Hüsler-Reiss, Galambos, Tawn, Normal, Plackett and FGM copula families. In contrast to previous studies conducted in the Danube River area (e.g., Bačová-Mitková and Halmová, 2014; Gaál et al., 2015; Szolgay et al., 2015, 2016), the flood peak discharge and volume pairs are constructed by means of two essentially different approaches. In either case, flood peaks are represented by the AMF peak. The difference appears in the way the flood volumes are defined. In the first approach, the annual maxima of the volumes for different durations have been selected (which can be regarded as a regime analysis of the dependence of the extremes of both variables in a given year), while in the second approach, flood volumes corresponding to the annual maximum flood peaks have been used (which is a typical choice for engineering studies). A bivariate analysis has been applied to the extracted pairs of the flood peak discharge and flood volume data using the copulas to assess which copula family is the most suitable for reproducing the dependence

structure of the variables selected as well as observing changes in their applicability along the Danube River.

## STUDY AREA

The Danube River originates in the Black Forest (Schwarzwald) in Germany at an altitude of about 678 m a.s.l. at the confluence of the Brigach and Breg rivers in Donaueschingen. The Danube River flows through ten countries, and its delta is located in the Black Sea. The area of the entire watershed is 801,463 km<sup>2</sup> with a main channel length of 2830 km (ICPDR, 2009). The average discharge is 6.500 m<sup>3</sup> s<sup>-1</sup> before the delta. The climatic regime over the river basin is dominated by a clear seasonality, which is influenced in the northern parts by the Atlantic Ocean and in the eastern parts by the continental climate. The total annual precipitation in the catchment is between 3000 mm at the high altitudes and 400 mm in the lowlands (Sommerwerk et al., 2009). The hydrological regime of the Danube River is highly influenced by precipitation patterns and the orographic structure of its catchment (Pekárová et al., 2008, 2013).

The Danube Basin can be subdivided into three main parts: the Upper Danube region, between the springs and the Devín Gate (mean elevation = 133 m a.s.l., catchment area = 131,338 km<sup>2</sup>, mean annual discharge = 2,051 m<sup>3</sup> s<sup>-1</sup> at the Devín-Bratislava gauge), which is characterized by high precipitation in the west (up to 2500 mm of mean annual precipitation totals in the Alps), whereas the eastern regions have lower precipitation (700 mm of mean annual precipitation totals in the lowlands). The runoff regime corresponds to that of the Alpine tributaries with the maximum discharges in June and minimum discharges during the winter months. The Central Danube Region lies between the Devín Gate and the Iron Gate (60 m a.s.l., 444,894 km<sup>2</sup>, 5,585 m<sup>3</sup> s<sup>-1</sup> at the Turnu Severin/Orsova gauge). The runoff regime is characterized by two runoff peaks in June and April. The April runoff peak is mainly caused by

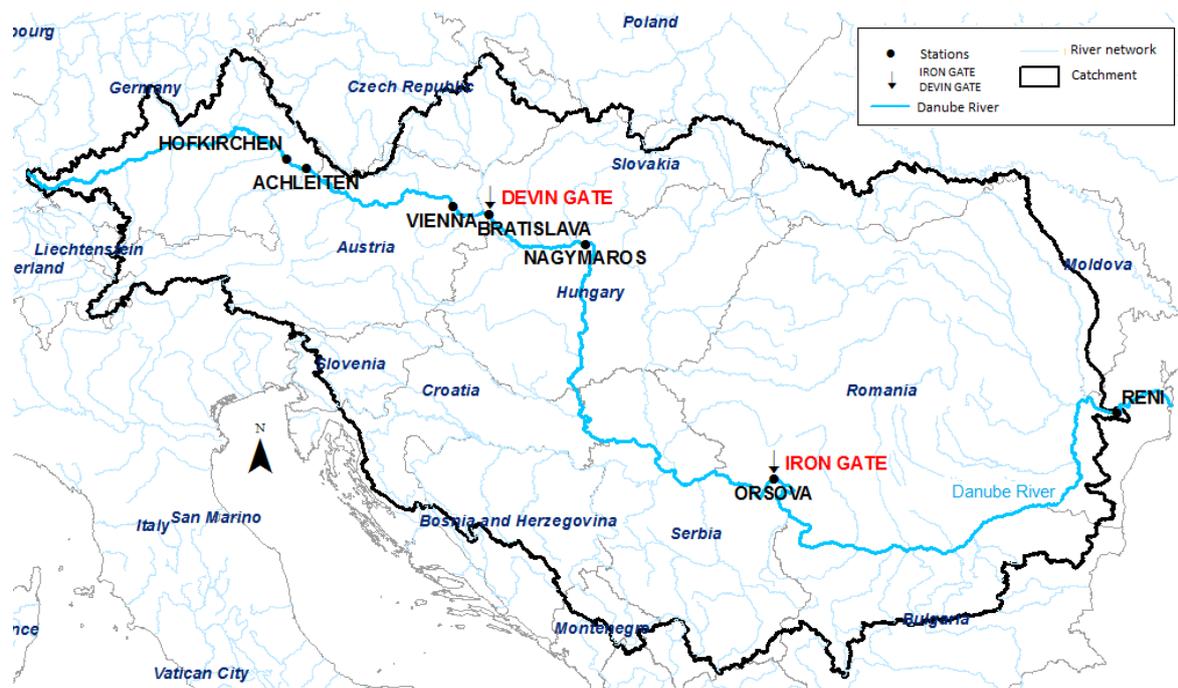


Fig. 1. The Danube River with the locations of the gauging stations analyzed.

**Table 1.** Gauging stations with the observation periods, catchment areas and elevation of the gauging stations used in the analysis.

Gauged Stations	Country	Length of the data series (Years)	Catchment area (km <sup>2</sup> )	Gauging station elevation (m a.s.l.)
Hofkirchen	Germany	1901–2007	47,496	299.17
Achleiten	Germany	1901–2007	76,653	287.27
Vienna	Austria	1828–2006	101,731	152
Bratislava	Slovakia	1876–2006	131,338	132.86
Nagyymaros	Hungary	1893–2007	183,534	99.37
Orsova	Romania	1904–2009	576,232	44.76
Reni	Ukraine	1931–1995 /1997–2002	805,700	0.2

the addition of waters from the snow melting in the Carpathians and from the early spring rains of the tributaries in the lowlands and low mountains of the area. The Lower Danube Region is located between the Iron Gate and the Danube’s embouchure into the Black Sea (1 m a.s.l., 230,768 km<sup>2</sup>, 6,499 m<sup>3</sup> s<sup>-1</sup> at the Ceatal Izmail gauge). The Danube Delta is considered to be region with temperate climate zones of a steppe character due to low precipitation totals (mean annual precipitation totals range from 600 mm to 300 mm) and the high summer temperatures over 35°C (ICPDR, 2009).

In this study the discharge data (annual maximum peak and daily discharges) of the following gauging stations: Hofkirchen (Germany), Achleiten (Germany), Vienna (Austria) from the Upper Danube Region, Bratislava (Slovakia), Nagymaros (Hungary) from the Central Danube Region and Orsova (Romania) and Reni (Ukraine) from the Lower Danube Region, were used, see Figure 1 and Table 1.

**METHODOLOGY**

This study applies a bivariate analysis of the AMF peak and volumes using various copula families. For all of the implementations of the bivariate analysis, the dependence of the two variables is assessed with Kendall’s tau correlation coefficient. A “blanket” goodness-of-fit (GOF) test, which is based on Kendall’s transformation with Cramér-von Mises’ measure of distance (Genest et al., 2009), has been used to evaluate and subsequently rank the best fitted copula families.

Initially, all the univariate data were tested for the presence of temporal dependence by the Mann-Kendall test (Mann, 1945) and the Ljung-Box test (Ljung and Box, 1978) due to prerequisite for a copula analysis that the data are independent and identically distributed (i.i.d.). Subsequently, the multivariate data of the flood peaks and volumes were tested for the significance of the (rank) correlation (Kendall, 1955) and for the asymmetry of their relationship (Genest et al., 2012) to justify the choice of modelling by (symmetrical) copulas.

In the next subchapter the applied copula families are described in detail, followed by the methods of input data selection and the annual maximum flood peaks and flow volume separation.

**Copulas**

A copula is a function that allows for the modelling of the dependence structure between stochastic variables. The main advantage is that the copula approach can split the problem of constructing multivariate probability distributions into a part containing the marginal one-dimensional distribution functions and a part containing the dependence structure. These two parts can be studied and estimated independently and then put together into a joint distribution function. The bivariate case can be written as:

$$F_{XY}(x,y) = C(F_X(x),F_Y(y)), \tag{1}$$

where  $F_X, F_Y$  are the respective marginal distribution functions of the random variables  $X, Y$ .  $F_{XY}$  represents the joint distribution function of the random vector  $(X,Y)$ , and copula  $C$  is a function  $C:[0,1]^2 \rightarrow [0,1]$  that satisfies the boundary conditions  $C(t,0) = C(0,t) = 0$  and  $C(t,1) = C(1,t) = t$  (uniform margins) for any  $t \in [0,1]$  and the so-called 2-increasing property. Thus a copula can be viewed as a standardized joint distribution function. Further details can be found in (Nelsen, 2006).

**Table 2.** List of the copulas applied in this study.

Archimedean copulas	Extreme-value copulas	Other copulas
Ali-Mikhail-Haq	Hüsler-Reiss	Normal
Clayton	Galambos	Plackett
Frank	Tawn	FGM
Joe		
Gumbel		

The current study applies ten commonly-used one-parametric families from several classes of copulas (Table 2, Appendix), namely, the Archimedean class (the Clayton, Frank, Gumbel-Hougaard, Joe and Ali-Mikhail-Haq (AMH) copulas), the extreme-value class (Gumbel-Hougaard, Galambos, Hüsler-Reiss), the elliptical class (normal or Gaussian family), and the unclassified Plackett and Farlie-Gumbel-Morgenstern (FGM) parametric family of copulas. Additionally, the three-parametric Tawn family (of the EV class) has also been used. The Archimedean copulas

$$C(u, v) = \varphi^{-1}(\varphi(u) + \varphi(v)), \forall u, v \in [0,1], \tag{2}$$

have the advantage of easy construction via a one-dimensional function  $\varphi: [0,1] \rightarrow [0, \infty)$  called a generator. The extreme-value (EV) copulas

$$C(u, v) = uv^{A(\log(u)/\log(uv))}, \forall u, v \in [0,1], \tag{3}$$

are uniquely defined through the so-called Pickands dependence function  $A: [0,1] \rightarrow [1/2,1]$  (see, e.g., Bacigál and Mesiar, 2012; Gudendorf and Segers, 2010; Tawn, 1988) and were theoretically derived to model the dependence between extremes of random variables, while the elliptical copulas are simply the copulas extracted from elliptically contoured distributions. All the Archimedean (except for the Frank) and EV copulas are non-symmetrical with respect to secondary diagonals; some accumulate more of a probability mass at points (0,0) or (1,1), which is denoted as a lower tail or upper tail dependence, respectively.

The parameters  $\theta$  of the copulas are estimated by maximizing the so-called pseudo-likelihood function:

$$L(\theta) = \sum_i \log[c_{\theta}(U_i, V_i)], \quad (4)$$

which, besides the copula density  $c_{\theta}$ , contains pseudo-observations  $U_i, V_i$  ( $i = 1, \dots, n, j = 1, 2$ ), which are a transformation of the  $n$  real observations of the respective random variable  $X, Y$ , through a corresponding empirical distribution function, also known as a plotting position.

A “blanket” test (Genest et al., 2009) provides the goodness-of-fit test with the Cramér-von Mises’ measure of distance:

$$S_n = \sum_i \{C_{\theta}(U_i, V_i) - C_n(U_i, V_i)\}^2, \quad (5)$$

between the parametric copula  $C_{\theta}$  and the empirical copula, which is defined as

$$C_n(u, v) = \sum_i 1(U_i < u) 1(V_i < v) / n. \quad (6)$$

Given the validity of the null hypothesis that  $C_{\theta}$  fits well, the probability distribution of  $S_n$  is unknown and needs to be calculated from bootstrap simulations.

The correlation is quantified by Kendall’s rank correlation coefficient  $\tau \in [-1, 1]$ , which is a measure of concordance that is also able to detect nonlinear dependence (to the contrary of the standard Pearson correlation coefficient). Recall that there is a direct relation between Kendall’s  $\tau$  and a copula,

$$\tau = \iint_{[0,1]^2} C(u, v) dC(u, v). \quad (7)$$

The analysis is performed with the ‘acopula’ package (Bacigál, 2013) under GNU R (R Core Team, 2014). Supplementary details about the copulas used in this study can be found in the Appendix.

## The annual maximum flood volumes of selected durations

In the first approach the selecting of the annual maximum flow volumes of fixed durations around the peak wave were done separately, instead of defining the volume of each flood (thus conducting a process-oriented analysis of the hydrograph and defining the beginning and end of each runoff event). Here, this simplified algorithm for volume selection was adopted to emulate the volume of a wave, since this paper is more technically oriented towards studying the suitability of theoretical models for such a type of analysis.

Both variables, AMF and food volumes of the fixed duration, represent the annual maxima, but they are not necessarily linked to the same hydrological event. Such an approach for the construction of a data set is more preferred among statisticians, since the selected bivariate sample satisfies the conditions for a rigorous adoption of the extreme-value class copulas. Moreover, in our case, it allows for studying the regime of extreme values within each year and defining the possibility that in a year with a high flood, there will also be a flow volume of a predefined duration. For notational simplicity, we will subsequently refer to these flow volumes as flood volumes.

The annual maximum flood volumes of the fixed duration dataset used in this study was prepared in the frame of the UNESCO project “Flood regime of rivers in the Danube River Basin” (Pekárová et al., 2008). The annual flow volumes were separated using daily discharge data and fixed around the flood peaks at seven different durations: 5, 10, 15, 20, 25, 30 and 60 days, respectively, using a method described in Bačová-Mitková (2002). According to this methodology for all flood peaks in each year, using a moving window, the flood volumes of certain duration were calculated and largest annual flood

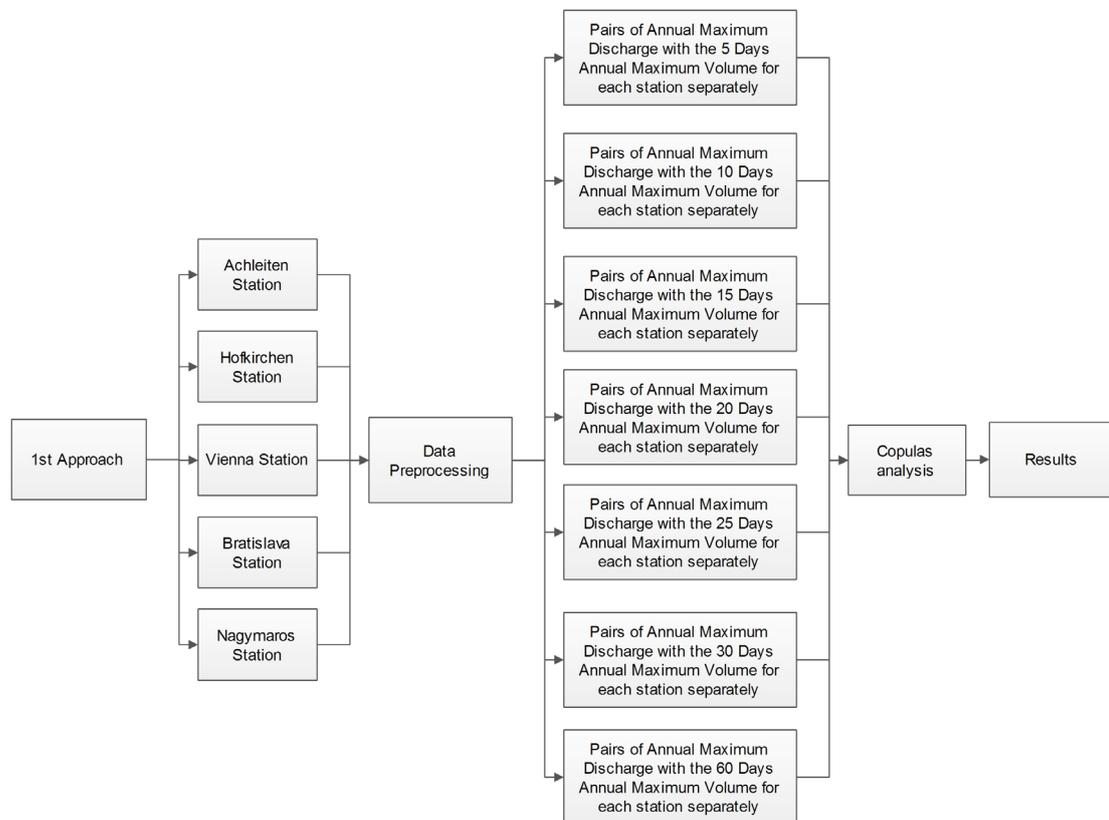
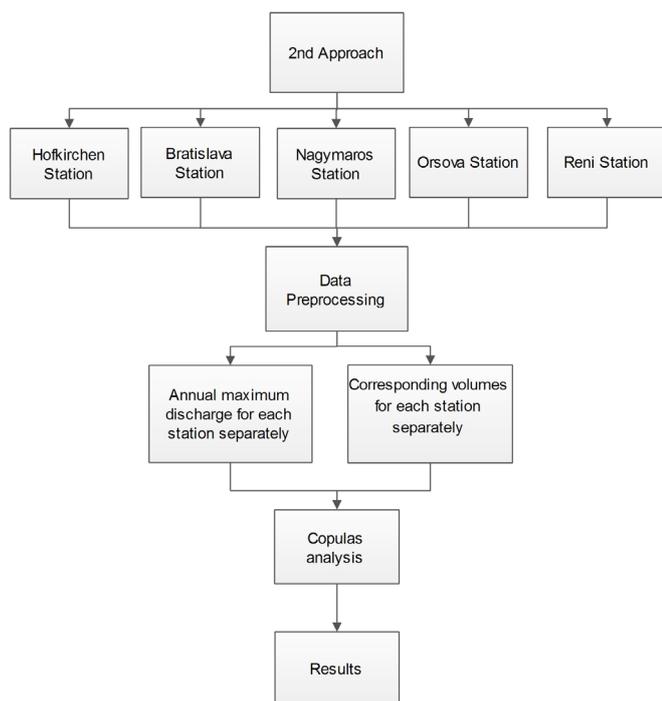


Fig. 2. Flowchart of the first approach for the sample selection.



**Fig. 3.** Flowchart of the second approach for the sample selection: flood volumes corresponding to flood peaks.

volume values for the selected durations were then taken for the analysis. The mean daily streamflow data used in this approach are from the following Danube River stations: 1) Hofkirchen, 2) Achleiten, 3) Vienna, 4) Bratislava and 5) Nagymaros. Unfortunately, the annual maximum flood volume data of selected durations for the Orsova and Reni stations were not available within the frame of the mentioned UNESCO project.

Next, the applicability of different copula families from the Archimedean class and extreme-value class as well as three other families (Normal, Plackett and FGM) for all of the above data combinations was investigated. Figure 2 presents the flowchart of the first approach.

### The flood volumes corresponding to the annual maximum flood peaks

The second approach refers to a bivariate analysis using the AMF peaks and their corresponding flood volumes for each station (Figure 3). In other words, this is an event-based analysis, since each year at the given site is represented by a single flood event with the largest peak discharge value (and its corresponding flood volume). This is rather an intuitive approach more preferred among engineering hydrologists in practical applications, and it enables the definition of conditional probabilities for design purposes.

The mean daily streamflow data used for estimating the flood volume separation in this approach are taken from the following Danube River stations: 1) Hofkirchen, 2) Bratislava, 3) Nagymaros, 4) Orsova and 5) Reni. A local-minimum method developed by Willems (2009) was used for the separation of the base flow, and for a better estimation of the beginning and end of the flood waves. Eleven different copula families (Table 2, Appendix) for all of the various data combinations were investigated for their applicability.

## RESULTS

In this section the results indicating the suitability of the various copula families for the bivariate analysis of the AMF discharges and flood volumes are presented. First, the preliminary data testing for the presence of temporal dependence, and the significance of (rank) correlation and asymmetry are discussed. Then, four different rank categories, which were created according to the GOF test, are described. The detailed results of the bivariate analyses of the AMF peaks and volumes are presented separately, according to the approach of volume separation, which was described in the previous section.

The preliminary tests revealed only a few violations of the copula analysis prerequisites, namely, the original flood peak data at the Vienna station seemed to contain a trend. No peaks at any station were indicated for autocorrelation. As for the volumes, within the both approaches no volume data shows a temporal dependence. Within the second approach a downward trend was detected at the Orsova and Reni stations due to decreasing volatility in recent decades.

The correlation between the AMF peaks and all the flood volumes was quantified with Kendall's Tau and results are summarized within the corresponding subsections.

According to the GOF test statistic, which serves as a measure of the distance to the empirical distribution (the lower the better), four different rank categories were created: a) The first to third ranks, and b) the *total rank*, which refers to the frequency with which a copula family scored with one of the previous ranks. For instance, at the Hofkirchen station (see Table 4, column 3), copulas from the Frank family scored three times with the lowest the GOF test statistic (1<sup>st</sup> rank) among all the copulas estimated for each fixed volume class, then two times with the second lowest value of the GOF test statistic (i.e. the 2<sup>nd</sup> rank), and just once with the third lowest value (the 3<sup>rd</sup> rank) – a total of scoring 6 times in the copula competition, which is summarized in Table 5. The total rank then represents the number of the scores across all the fixed volume classes and all the stations. We adopted this ranking approach to be able to select the best models among all those applied, as more than one model was always suitable. The analysis was performed with a focus on the stations spread along the river, different flood volumes (annual maximum volumes of fixed durations and flood volumes corresponding to flood peaks), and diverse (yet commonly implemented) copula models. In the following paragraphs, a detailed analysis of the results of both approaches will be presented.

### First approach: the annual maximum flood volumes of selected durations

One of the findings in this approach is that the (always significant) correlation between the flood variables (i.e., the discharge and volume) decreases as the flood's duration increases (Table 3). A notable decrease in the correlation can be observed in the Austrian reach, which would require further analysis, e.g., with respect to the lateral inflow in this reach. Overall the values (except for this reach) are more or less constant for all the durations. The results concerning Kendall's Tau correlation analysis for all the stations is that although the upper tail dependence (reading from the properties of the best fitting copula) disappears when the floods' durations increase, the lower tail dependence remains unchanged. The tests for asymmetry rejected the exchangeability of peaks and volumes in two out of 35 cases; thus the copula families considered are taken as appropriate in this respect.

**Table 3.** The Kendall's Tau correlation coefficient values for both approaches and all the stations analysed.

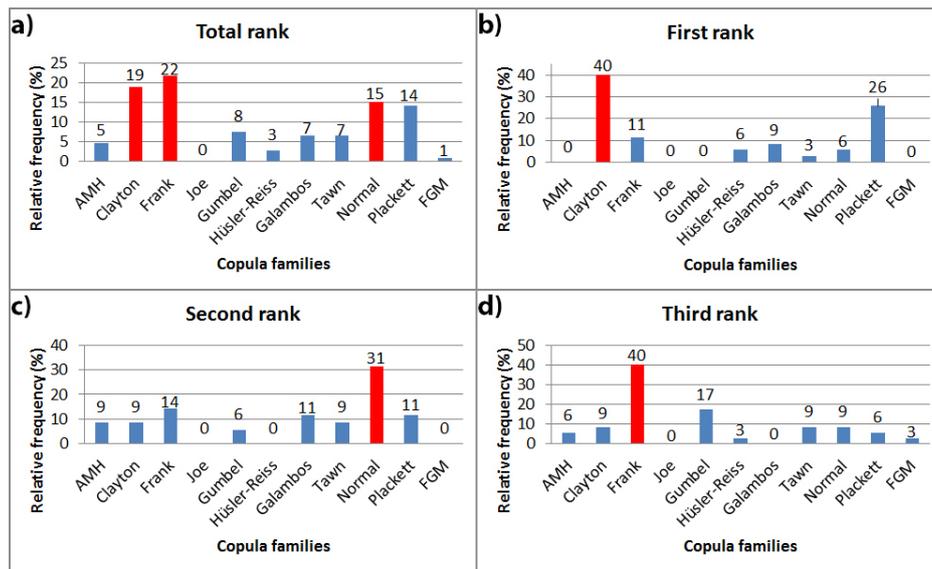
Volumes	Hofkirchen	Achleiten	Vienna	Bratislava	Nagyvaros	Orsova	Reni
5D	0.85	0.75	0.75	0.85	0.88	–	–
10D	0.79	0.68	0.68	0.78	0.82	–	–
15D	0.67	0.54	0.53	0.64	0.67	–	–
20D	0.64	0.49	0.49	0.61	0.63	–	–
25D	0.61	0.46	0.46	0.59	0.6	–	–
30D	0.6	0.43	0.44	0.57	0.58	–	–
60D	0.52	0.36	0.36	0.46	0.5	–	–
Corresponding volumes to AMF	0.26	–	–	0.36	0.35	0.41	0.27

**Table 4.** The rank values of the copula families according to the GOF test for both approaches.

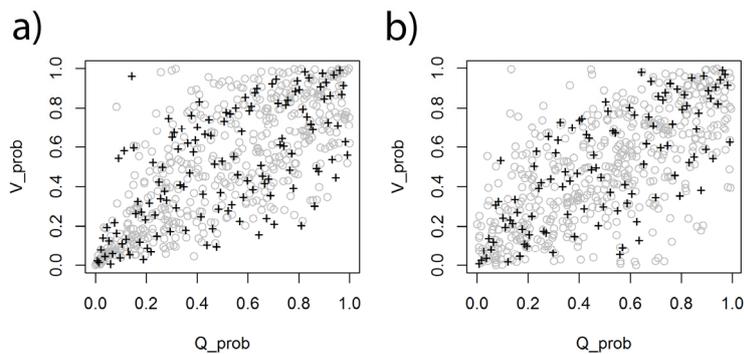
Volume	Ranks	Hofkirchen Copula family	Achleiten Copula family	Vienna Copula family	Bratislava Copula family	Nagyvaros Copula family
5D	1 <sup>st</sup>	Frank	Hüsler-Reiss	Hüsler-Reiss	Galambos	Plackett
	2 <sup>nd</sup>	Galambos	Galambos	Galambos	Gumbel	Normal
	3 <sup>rd</sup>	Gumbel	Tawn	Gumbel	Tawn	Gumbel
10D	1 <sup>st</sup>	Galambos	Plackett	Galambos	Tawn	Frank
	2 <sup>nd</sup>	Tawn	Clayton	Gumbel	Galambos	Tawn
	3 <sup>rd</sup>	Gumbel	Frank	Hüsler-Reiss	Gumbel	Plackett
15D	1 <sup>st</sup>	Frank	Clayton	Clayton	Plackett	Normal
	2 <sup>nd</sup>	Normal	Plackett	Normal	Tawn	Frank
	3 <sup>rd</sup>	Clayton	Frank	Frank	Gumbel	Tawn
20D	1 <sup>st</sup>	Plackett	Clayton	Clayton	Plackett	Plackett
	2 <sup>nd</sup>	Normal	Normal	Plackett	Normal	Normal
	3 <sup>rd</sup>	Frank	Frank	Frank	Clayton	Frank
25D	1 <sup>st</sup>	Clayton	Clayton	Clayton	Normal	Plackett
	2 <sup>nd</sup>	Frank	Frank	Plackett	Clayton	Frank
	3 <sup>rd</sup>	Normal	AMH	Frank	Frank	Normal
30D	1 <sup>st</sup>	Plackett	Clayton	Clayton	Clayton	Plackett
	2 <sup>nd</sup>	Frank	Normal	AMH	Normal	Clayton
	3 <sup>rd</sup>	Clayton	AMH	Frank	Frank	Frank
60D	1 <sup>st</sup>	Frank	Clayton	Clayton	Clayton	Clayton
	2 <sup>nd</sup>	Plackett	AMH	AMH	Normal	Normal
	3 <sup>rd</sup>	Normal	Plackett	FGM	Frank	Frank
Corresp. volumes to AMF	1 <sup>st</sup>	AMH	–	–	Frank	AMH
	2 <sup>nd</sup>	Frank	–	–	Normal	Frank
	3 <sup>rd</sup>	Clayton	–	–	AMH	Clayton

**Table 5.** The frequency (total) and relative frequency (total %) of the overall rankings according to the GOF test for the data samples created on the basis of the first approach.

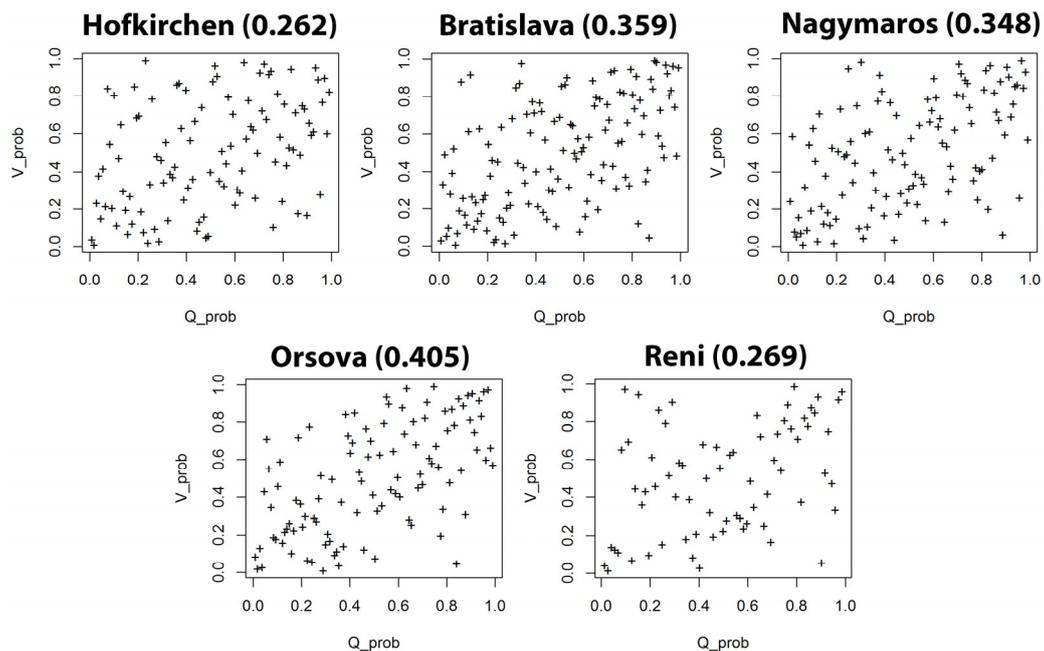
Copula families	Hofkirchen	Achleiten	Vienna	Bratislava	Nagyvaros	Total	Total (%)
AMH	0	3	2	0	0	5	5
Clayton	3	6	5	4	2	20	19
Frank	6	4	4	3	6	23	22
Joe	0	0	0	0	0	0	0
Gumbel	2	0	2	3	1	8	8
Hüsler-Reiss	0	1	2	0	0	3	3
Galambos	2	1	2	2	0	7	7
Tawn	1	1	0	3	2	7	7
Normal	4	2	1	4	5	16	15
Plackett	3	3	2	2	5	15	14
FGM	0	0	1	0	0	1	1



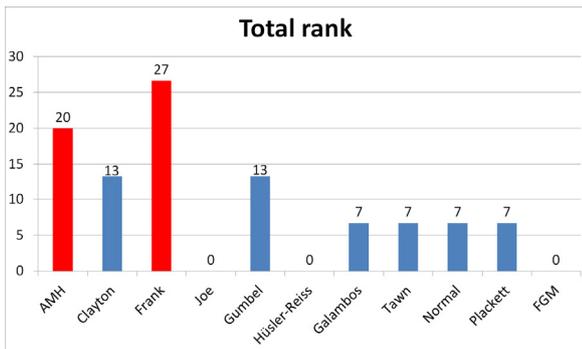
**Fig. 4.** The relative frequency of the ranks (in percentages) of different copula families as a measure of their applicability to all the stations (first approach).



**Fig. 5.** Scatter plots of 500 pair simulations according to the fitted copulas (grey color) and measured data (black color) for the data sample representing 60 days' fixed volume duration. a) Bratislava Station, the Clayton fit, GOF p-value 0.042, Kendall's Tau correlation coefficient 0.464 b) Hofkirchen Station, the Frank fit, GOF p-value 0.061, Kendall's Tau correlation coefficient 0.521.



**Fig. 6.** Scatter plots of the pseudo-observation for all the stations (the AMF peak vs. the corresponding flood volume) with the Kendall's Tau correlation coefficient (in parentheses).



**Fig. 7.** Ranking of the relative frequency (in percentages) of the applicability of different copula families to all the stations for the data samples created based on the second approach.

The copula family rankings are presented in Table 4, where, in the category of the first rank, the Clayton copula prevails with 14 counts,  $N = 35$  (a 40% relative frequency) (Figure 4b, Table 4). At the Achleiten and Vienna stations, the Clayton family has five counts with  $N = 7$  (a 70% relative frequency). In the second rank, the Normal family, which dominates in the ranking, has 11 counts with  $N = 35$  (a 31% relative frequency) (Figure 4c, Table 4), and at the stations of Bratislava and Nagymaros, the same family has three counts with  $N = 7$  (a 43% relative frequency). Moreover, in the third rank, the leading copula is that of the Frank family, having 14 counts with  $N = 35$  (a 40% relative frequency) (Figure 4.d, Table 4), and at the Vienna station, the same family has four counts with  $N = 7$  (a 57% relative frequency). Finally, the GOF test indicates that for the total ranking, the Frank copula performs better than the others with 23 counts,  $N = 105$  (a 22% relative frequency) (Figure 4a, Table 5). The Clayton copula is in the second place of the total ranking with 20 counts (a 19% relative frequency), and the Normal copula is in the third place with 16 counts (a 15% relative frequency) (Figure 4a, Table 5). In Figure 5 two scatter plot examples of the simulated/observed data can be seen.

**Second approach: the flood volumes corresponding to annual maximum flood peaks**

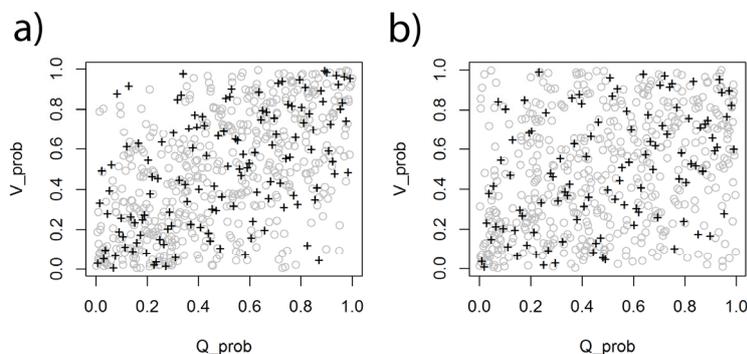
Concerning the second approach, according to Kendall’s Tau correlation coefficient (which is again always significant), the correlation between the AMF peak and the corresponding flood volume increases along the Danube River (downstream), except for the Reni station (Figure 6, Table 3). No relation

appears to be asymmetric with respect to the main diagonal (as suggested by the asymmetry test). Another finding from this analysis is that there is an increase in the upper tail dependence at the river stations located in the lower part of the river (see Figure 6), which is supported by the GOF test preference of extreme-value copulas at the Orsova and Reni stations (while at the Hofkirchen, Bratislava and Nagymaros stations, no copula with an upper tail dependence modelling capability was preferred), see Table 4. Furthermore, the analysis of the GOF test indicates that for the total rank, the Frank copula with four counts performs better than the others,  $N = 15$  (a 27% relative frequency) while the AMH is in the second place with three counts,  $N = 15$  (a 20% relative frequency) (Figure 7, Table 4). In Figure 8 two scatter plot examples of the simulated/observed data can be seen. In comparison with the first approach, the correlation of the AMF peak and the corresponding flood volumes is notably lower (see the p-values in Figure 8). Moreover, a result that follows from the analysis is that in the low correlation samples (the Hofkirchen, Bratislava and Nagymaros stations) the AMH copula family is mostly preferred (the Reni station is an exception due to the upper tail behaviour) (Figure 6, Table 3, Table 4).

**DISCUSSION AND CONCLUSIONS**

In this study we provided a bivariate analysis of the stream-flow data (the AMF peak and flood volumes) for modeling extreme flood events with the use of various copula families. First, the annual maximum flood volumes of the fixed durations were separated from the flood waves, which were not necessarily linked to the same hydrological event. This approach leads to the construction of the data set allowing us to study the regime of extreme values within each year and to investigate the chance that in a year with a high flood, there would also be a flood with a high volume. Since this paper is more technically oriented towards studying the suitability of theoretical models for such a type of analysis, this simplified volume selection was adopted instead of defining the volume of each flood (thus conducting a process-oriented analysis of the hydrograph and defining the beginning and end of each runoff event).

In the second approach to flood volume separation, the AMF peaks and their corresponding flood volumes for each station were sampled. This approach is mostly preferred in engineering studies in practical applications, since it enables the definition of conditional probabilities for design purposes. The positive and rather significant dependence of the AMF and annual



**Fig. 8.** Scatter plots of 500 pair simulations according to the fitted copula (grey color) and measured data (black color), for the data samples representing annual flood peaks and the corresponding flood volumes. a) Bratislava Station, the Frank fit, GOF p-value 0.045, the Kendall’s Tau correlation coefficient 0.359 b) Hofkirchen Station, the AMH fit, GOF p-value 0.04, Kendall’s Tau correlation coefficient 0.262.

maximum flood volumes of a fixed duration indicates that a detailed comparative analysis of such peaks and volumes along the river and between distinct periods bears a potential for new information on the regime of extremes and its changes.

As for the technical part, the main outcome of the analysis is that for all of the applications, the most favored copula family is the Frank one, and the least preferred is the Joe, one which failed to give any good fitting (Table 5, Figures 4 and 7). The AMH family is mostly favored in the low correlation samples.

In the first approach, the families with the best fit for all of the applications are the: 1) Frank, 2) Clayton and 3) Normal. Furthermore, the most favored copula families, taking into account their ranks, are: first rank – the Clayton, second rank – the Normal, and third rank – the Frank. The correlation analysis showed that there is: a) a certain decline in the correlation between the AMF peak and the flood volumes of fixed durations when the flood duration increases, and b) the persistence of a lower tail dependence with an increasing flood volume. When focusing the analysis on the stations, the Clayton family is the most applicable for the stations at Achleiten and Vienna. The second approach to the flood peak - volume sample construction shows that the Frank and AMH are the best fitted families. The correlation of the AMF peak – corresponding flood volume pairs rises as the stations downstream of the Danube River are passed.

The result that the Frank copula is the most favored family agrees with the study of Reddy and Ganguli (2012), which concerning the GOF p-value, underlined the Frank copula as the most suitable in terms of the best fit for the flood peak-volume pairs. Moreover, in the studies of Chowdhary et al. (2011) and Bačová-Mitková and Halmová (2014), the Clayton copula family was reported as the most suitable choice for simulation of the flood peak-volume pairs. The results of the unacceptable performance (in fitting) of the Joe copula family support the results of Szolgay et al. (2015). Finally, a valuable outcome from the results of this study, which agree with the results of Favre et al. (2004) and Szolgay et al. (2015, 2016), could be that a further investigation of the choice of the “best” copula families for the flood peak-volume pairs is essential.

For a more comprehensive and complete analysis of flood characteristics (peaks, volumes and durations), the following steps could be implemented in the near future: 1) a seasonal analysis splitting the year into several logical periods followed by applying the same methods of constructing the volumes and copula-fitting approach as in this study, 2) assessing the impacts of hydrotechnical projects on the study area (i.e., an examination of the correlation patterns between discharges – volumes due to dam construction), and 3) the use of partial duration series for a peak-over-threshold analysis using bivariate or trivariate copulas.

*Acknowledgement.* This research was funded by the “COST Action ES0901, European procedures for flood frequency estimation (FloodFreq)” research program, within the Short-Term Scientific Mission collaboration of the University of Thessaly, Department of Civil Engineering, and the Slovak University of Technology in Bratislava, Faculty of Civil Engineering, Department of Land and Water Resources Management. This financial support is gratefully acknowledged. Furthermore, we would like to acknowledge the support of the ERC “Flood-Change” Advanced Grant and the Slovak Grant Agency under VEGA Project Nos. 1/0776/13 and 1/0710/15 for their financial support. The authors would like to thank Katarína Jeneiová, PhD., for her help with data processing.

## REFERENCES

- Aronica, G.T., Candela, A., Fabio, P., Santoro, M., 2012. Estimation of flood inundation probabilities using global hazard indexes based on hydrodynamic variables. *Phys. Chem. Earth*, 42–44, 119–129.
- Bacigál, T., 2013. Modelling dependence with multivariate Archimax (or any user-defined continuous) copulas. Package ‘acopula’, R package – CRAN.
- Bacigál, T., Mesiar, R., 2012. 3-dimensional Archimax copulas and their fitting to real data. In: Proc. 20th International conference on computational statistics COMPSTAT 2012. The International Statistical Institute, The Hague, The Netherlands, pp. 81–88.
- Báčová-Mitková, V., 2002. The relationship between volume of the flood wave and the time duration of flood events. *Acta Hydrologica Slovaca*, 13, 1, 165–174.
- Báčová-Mitková, V., 2011. Different approaches to the flood volumes estimation on the Bodrog River example. *Acta Hydrologica Slovaca*, 12, 2, 296–303. ISSN 1335-6291 (In Slovak.)
- Báčová-Mitková, V., Halmová, D., 2014. Joint modeling of flood peak discharges, volume and duration: a case study of the Danube River in Bratislava. *Journal of Hydrology and Hydromechanics*, 62, 3, 186–196. doi: 10.2478/johh-2014-0026.
- Balistracchi, M., Baldassarre, B., 2011. Modelling the statistical dependence of rainfall event variables through copula functions. *Hydrol. Earth Syst. Sci.*, 15, 6, 1959–1977.
- Bezák, N., Šraj, M., Mikoš, M., 2016. Copula-based IDF curves and empirical rainfall thresholds for flash floods and rainfall-induced landslides. *Journal of Hydrology*. doi:10.1016/j.jhydrol.2016.02.058.
- Cunnane, C., 1988. Methods and merits of regional flood frequency analysis. *J. Hydrol.*, 100, 269–290.
- Cunnane, C., 1989. Statistical distributions for flood frequency analysis. Operational Hyd. Rep. 33. World Meteorological Organization, Geneva, Switzerland.
- Dawdy, D., Griffis, V., Gupta, V., 2012. Regional Flood-Frequency Analysis: How We Got Here and Where We Are Going. *J. Hydrol. Eng.*, 17, 9, 953–959. doi: 10.1061/(ASCE)HE.1943-5584.0000584.
- De Michele, C., Salvadori, G., 2003. A generalized Pareto intensity-duration model of storm rainfall exploiting 2.-Copulas. *J Geophys Res.*, 108, D2, 4067.
- 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.
- Dupuis, D.J., 2007. Using copulas in hydrology; benefits, cautions and issues. *J. Hydrol. Eng.*, 12, 381–393.
- Favre, A.C., El Adlouni, S., Perreault, L., Thiémonge, N., a Bobeé, B., 2004. Multivariate hydrological frequency analysis using copulas. *Water Resour. Res.*, 40, W01101.
- Gaál, L., Szolgay, J., Kohnová, S., Hlavcova, K., Parajka, J., Viglione, A., Merz, R., Blöschl, G., 2015. Dependence between flood peaks and volumes: a case study on climate and hydrological controls. *Hydrological Sciences Journal*, 60, 6, 968–984. doi: 10.1080/02626667.2014.951361.
- Gaál, L., Viglione, A., Szolgay, J., Blöschl, G., Bacigál, T., Rogger, M., 2010. Bivariate at-site frequency analysis of simulated flood peak-volume data using copulas. In: CD Rom – EGU General Assembly; Vienna, EGU2010-13441.
- Genest, C., Favre, A.C., 2007. Everything you always wanted to know about copula modeling but were afraid to ask. *J. Hydrol. Eng. ASCE*, 12, 4, 347–368.

- Genest, C., Nešlehová, J., Quessy, J.F., 2012. Tests of symmetry for bivariate copulas. *Annals of the Institute of Statistical Mathematics*, 64, 4, 811–834.
- Genest, C., Rémillard, B., Beaudoin, D., 2009. Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Math. and Economics*, 44, 2, 199–213.
- Giustarini, L., Camici, S., Tarpanelli, A., Brocca, L., Melone, F., Moramarco, T., 2010. Dam spillways adequacy evaluation through bivariate flood frequency analysis and hydrological continuous simulation. In: *Proc. World Environmental and Water Resources Congress 2010*. American Society of Civil Engineers (ASCE), Reston, Virginia, USA, pp. 2328–2339.
- Goel, N.K., Seth, S.M., Chandra, S., 1998. Multivariate modeling of flood flows. *ASCE, J. Hydraul. Eng.*, 124, 146–155.
- Groupe de recherche en hydrologie statistique (GREHYS), 1996. Presentation and review of some methods for regional flood frequency analysis. *J. Hydrol.*, 186, 63–84.
- Gudendorf, G., Segers, J., 2010. Extreme-value copulas. In: *Jaworski, P., Durante, F., Härdle, W.K., Rychlik, T. (Eds.): Copula Theory and its Applications*. Springer, Berlin Heidelberg, pp. 127–145.
- Chowdhary, H., Escobar, L.A., Singh, V.P., 2011. Identification of suitable copulas for bivariate frequency analysis of flood peak and flood volume data. *Hydrology Research*, 42, 2–3, 193–216. doi:10.2166/nh.2011.065.
- ICPDR, 2009. <http://www.icpdr.org/main/activities-projects/danube-river-basin-management-plan-2009>
- Kendall, M.G., 1955. *Rank Correlation Methods*. Hafner Publishing Co., New York.
- Laio, F., Ganora, D., Claps, P., Galeati, G., 2011. Spatially smooth regional estimation of the flood frequency curve (with uncertainty). *J. Hydrol.*, 408, 67–77. doi:10.1016/j.jhydrol.2011.07.022.
- Ljung, G.M., Box, G.E.P., 1978. On a measure of lack of fit in time series models. *Biometrika*, 65, 297–303.
- Mann, H. B., 1945. Nonparametric tests against trend. *Econometrica: Journal of the Econometric Society*, 13, 3, 245–259.
- Mediero, L., Jimenez-Alvarez, A., Garrote, L., 2010. Design flood hydrographs from the relationship between flood peak and volume. *Hydrol. Earth Syst. Sci.*, 14, 2495–2505.
- Mediero, L., Kjeldsen, T., 2014. Regional flood hydrology in a semi-arid catchment using a GLS regression model. *J. Hydrol.*, 514, 158–171. doi:10.1016/j.jhydrol.2014.04.007.
- Nelsen, R.B., 2006. *An introduction to copulas*. Lecture notes in statistics. 2<sup>nd</sup> ed., Springer, New York.
- Pekárová, P., Halmová, D., Bačová Mitková, V., Miklánek, P., Pekár, J., Škoda, P., 2013. Historic flood marks and flood frequency analysis of the Danube River at Bratislava, Slovakia. *J. Hydrol. Hydromech.*, 61, 4, 326–333.
- Pekárová, P., Onderka, M., Pekár, J., Miklanek, P., Halmová, D., Škoda, P., Bačová Mitková, V., 2008. Hydrologic scenarios for the Danube River at Bratislava. Key Publishing, Ostrava, 160 p., <http://www.ih.savba.sk/danubeflood>.
- R Core Team, 2014. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: <http://www.R-project.org>.
- Reddy, M.J., Ganguli, P., 2012. Bivariate flood frequency analysis of Upper Godavari River flows using Archimedean copulas. *Water Resour. Manag.*, 26, 3995–4018. doi:10.1007/s11269-012-0124-z.
- Salvadori, G., De Michele, C., 2004. Frequency analysis via copulas: theoretical aspects and applications to hydrological events. *Water Resour. Res.*, 40, 12. doi: 10.1029/2004WR003133.
- Salvadori, G., De Michele, C., 2010. Multivariate multiparameter extreme value models and return periods: A copula approach. *Water Resour. Res.*, 46, 10. doi: 10.1029/2009WR009040.
- Shiau, J.T., Wang, H.Y., Tsai, C.T., 2006. Bivariate frequency analysis of floods using copulas. *J. Am. Wat. Resour. Assoc.*, 42, 6, 1549–1564.
- Singh, K., Singh, V.J., 1991. Derivation of bivariate probability density functions with exponential marginals. *Stoch. Hydrol. Hydraulics*, 5, 1, 55–68.
- Sommerwerk, N., Baumgartner, C., Blösch, J., Hein, T., Ostojic, A., Paunovic, M., Schneider-Jacoby, M., Siber, R., Tockner, K., 2009. Chapter 3: Danube River basin. In: *Tockner, K., Robinson, T.C., Uehlinger, U. (Eds.): Rivers of Europe*. Academic Press, London, pp. 59–112.
- Sraj, M., Bezak, N., Brilly, M., 2015. Bivariate flood frequency analysis using the copula function: a case study of the Litija station on the Sava River. *Hydrol. Process.*, 29, 235–248.
- Szolgay, J., Gaál, L., Bacigál, T., Kohnová, S., Hlavčová, K., Výleta, R., Blöschl, G., 2016. A regional look at the selection of a process-oriented model for flood peak/volume relationships. In: *Proceedings of IAHS*, 373, Copernicus Publications, Göttingen, Germany, pp. 61–69. doi: 10.5194/piahs-373-1-2016.
- Szolgay, J., Gaál, L., Kohnová, S., Hlavčová, K., Výleta, R., Bacigál, T., Blöschl, G., 2015. A process-based analysis of the suitability of copula types for peak-volume flood relationships. In: *Proc. IAHS*, 370, Copernicus Publications, Göttingen, Germany, pp. 183–188. doi: 10.5194/piahs-370-183-2015.
- Szolgay, J., Kohnová, S., Bacigál, T., Hlavčová, K., 2012. Proposed flood: joint probability analysis of maximum discharges and their pertaining volumes. *Acta Hydrologica Slovaca*, 13, 2, 297–305.
- Tawn, J.A., 1988. Extreme value theory: Models and estimation. *Biometrika*, 75, 397–415.
- Willems, P., 2009. A time series tool to support the multi-criteria performance evaluation of rainfall-runoff models. *Environmental Modelling and Software*, 24, 311–321.
- Yue, S., Ouarda, T.B.M.J., Bobee, B., 2001. A review of bivariate Gamma distribution for hydrological application. *J. Hydrol.*, 246, 1–4, 1–18.
- 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. Bivariate rainfall frequency distributions using Archimedean copulas. *J. Hydrol.*, 332, 93–109.

Received 1 December 2015  
Accepted 12 September 2016

**APPENDIX**

All the copula families used in our study are expressed analytically (Bacigál, 2013):

The selected copula class	Family	Copula $C$ , generator $\varphi$ or dependence function $A$	Parameters
Archimedean copulas	Ali-Mikhail-Haq	$\varphi(t) = \log\left(\frac{1-(1-t)p}{t}\right)$	$p \in [-1,1)$
	Clayton	$\varphi(t) = t^{(-p)} - 1$	$p \in [-1, \infty) \setminus \{0\}$
	Frank	$\varphi(t) = -\log\left(\frac{\exp(-pt)-1}{\exp(-p)-1}\right)$	$p \in \mathbb{R} \setminus \{0\}$
	Joe	$\varphi(t) = -\log(1-(1-t)^p)$	$p \geq 1$
	Gumbel	$\varphi(t) = (-\log t)^p$	$p \geq 1$
Extreme-value copulas	Hüsler-Reiss	$A(t) = t\Phi\left[\frac{1}{p} + \frac{p \log\left(\frac{t}{1-t}\right)}{2}\right] + (1-t)\Phi\left[\frac{1}{p} - \frac{p \log\left(\frac{t}{1-t}\right)}{2}\right]$	$p \geq 0$
	Galambos	$A(t) = 1 - (t^{-p} + (1-t)^{-p})^{-1/p}$	$p \geq 0$
	Tawn	$A(t) = 1 - p_2 + (p_2 - p_1)t + (p_1 t^{p_0} + p_2^{p_0} (1-t)^{p_0})^{1/p_0}$	$0 \leq p_1, p_2 \leq 1$ $1 \leq p_0$
Other copulas	normal	$C(u, v) = \Phi_p(\Phi^{-1}(u), \Phi^{-1}(v))$	$p \in [-1,1]$
	Plackett	$C(u, v) = \frac{[1 + (p-1)(u+v)] - \sqrt{[1 + (p-1)(u+v)]^2 - 4uvp(p-1)}}{2(p-1)}$	$p > 0$
	FGM	$C(u, v) = uv[1 + p(1-u)(1-v)]$	$p \in [-1,1]$
Note: $\Phi^{-1}$ is a quantile function of the standard normal distribution, and $\Phi_p$ is a joint CDF of the bivariate standard normal distribution with parameter $p$ (Pearson correlation coefficient).			