DE GRUYTER
OPEN

Contributions to Geophysics and Geodesy Vol. 46/3, 2016 (155–178)

Similarity of empirical copulas of flood peak-volume relationships: a regional case study of North-West Austria

Ladislav GAÁL[1,2], Ján SZOLGAY[1], Tomáš BACIGÁL[3], Silvia KOHNOVÁ[1], Kamila HLAVČOVÁ[1], Roman VÝLETA[1], Juraj PARAJKA[4], Günter BLÖSCHL[4]

[1] Slovak University of Technology, Faculty of Civil Engineering, Department of Land and Water Resources Management, Radlinského 11, 810 05 Bratislava, Slovak Republic; e-mail: ladislav.gaal@stuba.sk

[2] MicroStep-MIS, s.r.o., Čavojského 1, 841 04 Bratislava 4, Slovak Republic

[3] Slovak University of Technology, Faculty of Civil Engineering, Department of Mathematics and Descriptive Geometry, Radlinského 11, 810 05 Bratislava, Slovak Republic

[4] Vienna University of Technology, Institute for Hydraulic and Water Resources Engineering, Karlsplatz 13/223, A-1040 Vienna, Austria

Abstract: This paper analyses the bivariate relationship between flood peaks and corresponding flood event volumes modelled by empirical copulas in a regional context in the North-West of Austria. Flood data of a total of 69 catchments in the region are analysed for the period 1976–2007. In order to increase the sample size and the homogeneity of the samples for the statistical analysis, 24872 hydrologically independent flood events were isolated and assigned to one of three flood process types: synoptic floods, flash floods or snowmelt floods in contrary to the more traditional engineering approach of selecting annual maxima of flood peaks and corresponding flood volumes. The first major part of the paper examines whether the empirical peak-volume copulas of different flood process types are statistically distinguishable, separately for each catchment. The results indicate that the empirical copulas of flash floods tend to be different from those of the synoptic and snowmelt floods in the target region. The second part examines how similar are the empirical flood peak-volume copulas between catchments for a given flood type. For the majority of catchment pairs, the empirical copulas of all flood types are indeed statistically similar. The flash floods show the largest degree of spatial heterogeneity. It is concluded that there is merit in treating flood types separately and in pooling events of the same type in a region when analysing and estimating flood peak-volume dependence copulas; however, the sample size of the analysed events is a limiting factor in spite of the introduced event selection procedure.

Key words: Flood types, regionalisation, flood peaks, flood volumes, copulas, bivariate distributions

doi: 10.1515/congeo-2016-0011

1. Introduction

The relationship of flood peaks and the flood volumes associated with these is interesting both from a theoretical and engineering perspective. From the theoretical point of view it can be regarded as a statistical fingerprint of catchment response. Bivariate frequency distributions of flood peaks vs. flood event volumes are also needed for a range of practical purposes in hydrology including retention basin design and for identifying flood hazard zones. Because of the larger number of degrees of freedom relative to bivariate distributions, the estimation of these from observed peak-volume pairs, however, is associated with substantial uncertainty. For fitting, say, a GEV distribution to observed flood peaks alone one usually requires on the order of 50 and more years of data when one is interested in estimating the 100-year flood (*DWA, 2012*), although this also depends on the required reliability. Nevertheless, in fitting a bivariate distribution one would need at least 50×50 or more years of data to obtain the same uncertainty, as there are parameters for two marginal distributions to estimate (peaks and volumes) as well as some parameters for the dependence between peaks and volumes. Observed flood records of this length are rarely available, thus this deficiency poses a serious challenge to estimating bivariate distributions.

A regional pooling procedure that is usual for univariate distributions (where this approach has a large tradition, see, e.g., *Dalrymple, 1960; Hosking and Wallis, 1997; IH, 1999; Kohnová and Szolgay, 1999; Gaál et al., 2008*) have similarly been applied to bivariate distributions (*Chebana and Ouarda, 2007, 2009; Requena et al., 2016*). These approaches make usually use of formalised catchment similarity metrics based on catchment attributes. However, little is known about the similarity of empirical peak-volume relationships between catchments and between different flood processes, since most of the recent research on peak-volume relationships revolved around copulas for individual catchments (e.g., *Favre et al., 2004; Shiau et al., 2006; Zhang and Singh, 2006; Genest and Favre, 2007; Poulin et al., 2007; Karmakar and Simonovic, 2009; Chowdhary et al., 2011; Ben-Aissia et al., 2012; Reddy and Ganguli, 2012; Bačová-Mitková, 2012; Ganguli and Reddy, 2013; Sraj et al., 2014; Bačová-Mitková and Halmová, 2014*). These studies examined parameter estimation methods and the suitability of particular copula types on the basis of a single or a few catchments, but neither in a spatial context nor in the light of flood generation processes.

The paper is aimed at examining the dependence between flood peaks and corresponding flood volumes from an empirical regional perspective in order to potentially assist in pooling catchments into groups of similar behaviour. The peak-volume dependence is framed in terms of empirical copulas. In order to go beyond purely statistical analyses and understand the process controls, the paper adopts the concept of flood process types where the rainfall-runoff events are classified according to their generating mechanisms. Specifically, the paper addresses two science questions: (i) *How similar are the peak-volume dependence structures of different flood types for a given catchment*? (ii) *How similar are the peak-volume dependence structures between catchments for a given flood type?*

The paper builds on the flood typology of *Merz and Blöschl (2003)* and classifies rainfall-runoff events into synoptic, flash and snowmelt floods. It also builds on the study of *Gaál et al. (2012)* who identified storm type, antecedent soil moisture and geology as the main factors controlling the flood event time scale (the ratio of volume and peak) in Austria, and on *Gaál et al. (2014)* who found that the consistency of the peak-volume relationship in terms of the Spearman's rank correlation coefficient is closely related to the consistency of the climatic drivers. More recently, *Szolgay et al. (2015, 2016)* aimed at analysing the formal suitability of various copula models of flood peaks and flood volumes, with a particular focus on flood generating seasons (summer and winter floods) and processes (synoptic floods, flash floods or snowmelt floods) with the goal of going beyond the statistics alone in the choice of the copula models from the engineering application perspective. *Szolgay et al. (2015, 2016)* have also included a preliminary analysis of the similarity of empirical copulas, which is examined in depth in the current paper. The analyses in this paper are performed on a high quality flood database in the North-West of Austria, consisting of 69 catchments and a total of 24872 flood events.

2. Methods

Two-dimensional copulas are functions that allow the modelling of the dependence structure between two stochastic variables. They split the problem of constructing a bivariate probability distributions into two parts: (i) the marginal one-dimensional distribution functions and (ii) the dependence structure, which can be studied and estimated separately and then rejoined to form a joint distribution function (for details and hydrological applications see, e.g., *Genest and Favre, 2007*).

Formally, a bivariate copula function can be written as:

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

where F_X and F_Y are respective marginal distribution functions of random variables X and Y which, in this paper, stand for flood peak and flood event volume, respectively and C is the copula. Usually, the marginal distribution functions F_X and F_Y can be modelled in the traditional way, but here we assume that these are not known. Therefore, they are estimated on the basis of the observations of the random variables (X_i, Y_i), $i = 1, ...n$, using a corresponding empirical distribution function (sometimes referred to as plotting position formula):

$$F_X(x) = \textstyle\sum_i \mathbf{1}(X_i \leq x)/(n+1) \tag{2}$$

and similarly for F_Y. The observed data transformed by empirical distribution function are called pseudo-observations and are defined as $U_i = F_X(X_i)$ and $V_i = F_Y(Y_i)$, $i = 1, ...n$. A two-dimensional extension of (2) using pseudo-observations:

$$C_n(u,v) = \textstyle\sum_i \mathbf{1}(U_i \leq u)\mathbf{1}(V_i \leq v)/(n+1) \tag{3}$$

is called empirical copula C_n. To investigate homogeneity in the peak-volume relationship (i.e., in pairs of flood peaks and event volumes) we test for equality of their empirical copulas by the approach of *Remillard and Scaillet (2009)* using a Cramér-von Mises test statistic. The null hypothesis H_0 states that empirical copulas C_{n1} and C_{n2} of two distinct peak-volume random vectors with observations (X_i^1, Y_i^1), $i = 1, ...n_1$, and (X_i^2, Y_i^2), $i = 1, ...n_2$, come from the same – yet unknown – bivariate distribution. The probability distribution of the test statistic is unknown and needs to be obtained by bootstrap simulations. The result of the similarity test is a p-value, which is the percentage of how many simulations of the test statistic (under H_0) exceeds the estimator from observations. The test is performed here with the help of the R package *TwoCop* (*Remillard and Plante, 2012*). Since the test statistic, in principle, represents a distance, an intuitive (and not statistically strictly correct) approach of comparing similarity was adopted here. Since the lower values (smaller departure from null hypothesis statement) lead to larger p-values, we use these in turn as a measure of similarity here.

The preliminary results of a comparison of empirical copulas indicated that the similarity testing is sensitive to differences in the sample size between C_{n1} and C_{n2}. Therefore, we also subsampled the data to obtain samples of the same sizes to which the test was applied. The subsampled datasets were used to both compare the empirical copulas of different catchments and the empirical copulas of different flood process types.

3. Study region and data

Due to different hydrological, climatological and geological settings there is a wide variety of flood generation mechanisms across Austria (e.g., *Parajka et al., 2010; Gaál et al., 2012*) which complicates the analyses of flood peak-volume relationships. In order to reduce this complexity, we restricted our analysis to a geographically more limited area (approximately 20000 km^2), namely to the Northern Lowlands region in the North-West of Austria. The elevation range is moderate, from about 400 up to 1500 m a.s.l. The western parts of the region are influenced by the air masses flowing from the Atlantic area from the West or North-West. The annual precipitation amounts are between 500 to 1500 mm, and their decreasing trend from the West to the East is recognisable. Floods occur both in the summer and winter seasons. The summer floods have origin in synoptic weather systems or localised convective events (flash floods). Winter floods are usually caused by rain-on-snow processes when antecedent snowmelt saturates the soils and relatively low rainfall intensities may then result in significant floods. Snowmelt floods without rain contributions also occur but are less important due to the relatively low elevations.

Runoff data from 69 catchments of North-West Austria region were used in this paper (Fig. 1). The catchment areas range from 10.6 to 444.3 km^2 (median: 74.6 km^2), while the mean catchment elevations are from 342 to 888 m a.s.l. (median: 570 m a.s.l.). The runoff records cover the period from 1976 to 2007 and the time resolution of the data is 1 hour.

In the very first step of the study, the rainfall-runoff events were identified using a several-step automated procedure (e.g., *Merz and Blöschl, 2003, 2009; Merz et al., 2006*). The most important steps of this algorithm were (i) the adoption of the recursive digital filter of *Chapman and Maxwell (1996)*, which separated the base flow and direct runoff, and (ii) identification of the start and the end of the individual rainfall-runoff events based on a number of criteria that were related to the values and ratios of the direct runoff and the base flow at the beginning of the event, at the time of the culmination and at the end of the event. For the 69 catchments in the target region, the algorithm identified 24872 flood events (an average of ∼360 events per catchment).

The flood type classification of *Merz and Blöschl (2003)* served as one of the cornerstones of the study. Nevertheless, the original definition of 5 process types was modified by *Gaál et al. (2014)* by merging similar categories. The three new classes of flood types consist of synoptic floods (originally long-rain and short-rain floods), flash floods (no change in the classification) and snowmelt floods (originally rain-on-snow floods and snowmelt floods). Note that while for several previous studies focused on Austria (e.g., *Merz et al., 2006; Merz and Blöschl, 2009*) only the events associated with the annual maxima of flood peaks were classified, in this paper we classified all flood events in the database for North-West Austria.

The flood event types were identified by selecting snowmelt events first. This was carried out using information on snow cover depth in different elevation zones of the individual catchments (*Parajka et al., 2007*). From the remaining events flash floods were selected, if the event occurred between May and September and event duration was of no more than 5 hours. The remaining flood events were classified as synoptic flood events. Note that the occurrence of synoptic flood types was not restricted to the summer period only; they could occur throughout the whole year. A flood event was considered hydrologically independent from the previous event if it began at least 7 days after the end of the previous event or it followed a period of at least 7 days without significant rainfall (≤ 0.1 mm/h). The classification procedure and all analyses in this paper were performed on the whole flood database consisting of a total of 24872 flood events.

The regional distribution of the percent flood types for each catchment is shown in Fig. 1 and in Table 1. There is a clear prevalence of synoptic

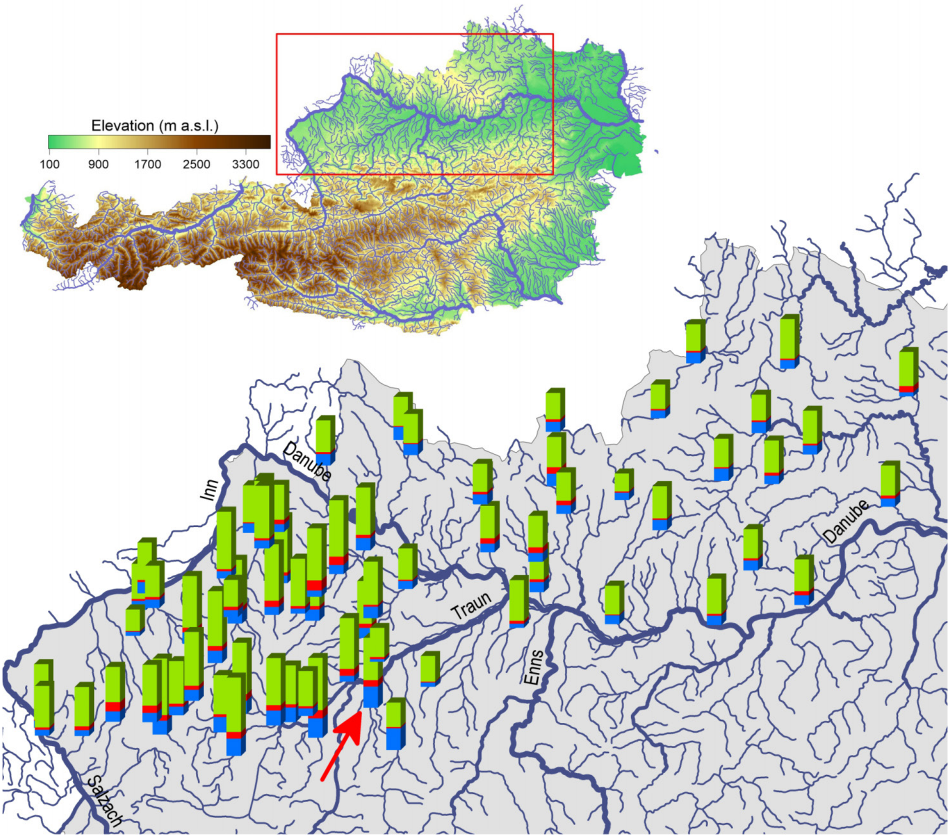

Fig. 1. Map of the study region in North-West Austria with charts of percent flood types of individual catchments as indicated by the colours. Size of the stack charts corresponds to total number of events per catchment. Green / red / blue colours denote synoptic / flash / snowmelt floods. The red arrow points to the example catchment St. Au (Mittlere Au) at the stream Laudach which is introduced in more detail in Section 4.1.

floods (74% of all events). Snowmelt and flash floods constitute 19% and 7% , respectively. Locally, the percentage of the flood types ranges between 51 and 74% for the case of synoptic floods, and 8 and 46%, and 2 and 17% for the snowmelt and flash floods, respectively. The relatively small number of snowmelt floods is due to the modest elevations in the region. The small number of flash floods is partly related to the sizes of the gauged catchments (median of 74.6 km^2). If data for smaller catchments were available,

Table 1. Number of independent flood events for the respective processes in the 69 catchments.

	Minimum	Maximum	Median	Total
Synoptic floods	123	396	257	18481 (74.3%)
Flash floods	6	70	21	1709 (6.9%)
Snowmelt floods	18	150	63	4682 (18.8%)
All floods	181	549	337	24872 (100.0%)

the percentage of flash floods would likely increase. Overall, there are no clear patterns of the percent flood types within the region. In fact, the region was selected in a way to minimise the spatial trends of the types to allow for a consistent regional evaluation of copulas.

The total number of events identified does show a spatial pattern. It is largest in the Southwest and smaller in the Northwest and the Northeast. These differences are related to the spatial distribution of the typical time scales of floods that have been defined by *Gaál et al. (2012)* as the ratio of flood volume and peak flow. *Gaál et al. (2012, Table 3)* found median flood time scales of 25.0, 36.3 and 41.9 hrs of (maximum annual) long-rain floods in the three areas that roughly correspond to the current target region. The shorter the flood time scales, the larger the number of independent events that can be identified. The differences in the flood time scales are related to differences in both climate and catchment response characteristics (*Gaál et al., 2012*).

4. Results

4.1. A comparison of empirical copulas for different flood types locally

First we were interested in whether different flood types, for the same catchment, were distinguishable in terms of their empirical peak-volume copulas. Each catchment was examined separately and the flood data samples of process types were compared pairwise, i.e., synoptic floods vs. snowmelt floods, synoptic floods vs. flash floods, and flash floods vs. snowmelt floods. The comparison of the empirical copulas was carried out in two ways. First, the full data sets were compared. We, however, observed that p-values of the similarity test were sensitive to the relative sample sizes; i.e., pairs of empirical copulas were often constructed on the basis of a large amount of data of one process type (say synoptic floods), but a low number of other flood type (say flash floods). Therefore, in a second turn of the evaluation the data were resampled in order to get exactly the same number of floods for the three flood types at each catchment. As the flash floods always were the least frequent flood type, the synoptic and snowmelt floods were resampled to the number of flash floods at each catchment. In each case, 10 subsamples were selected that were similar to the entire sample in terms of Kendall's τ. The final p-value for the given catchment and the given flood process pair was then estimated as the median p-value of the 10 subsamples. Herein we show both sets of results.

Empirical distribution functions of the p-values are shown in Fig. 2 for the three pairs of processes. The p-values are used here as a measure of distance/similarity between the empirical data and not in their original statistical sense (nor are these interpreted here in a statistical sense). This intuitive approach allows ranking the strength of similarity according to the measure of the distance between the pairs of empirical dependence structures of flood peaks and volumes.

Each distribution function in Fig. 2 consists of 69 values which is the number of catchments. In both cases, i.e., for the full samples and subsamples, the p-values for the comparison of synoptic floods and snowmelt floods are larger than those for other process pairs. This means that the synoptic and the snowmelt floods are more similar to each other (in terms of the distance between their empirical copulas) than the other process pairs. In other words, the flash floods tend to be dissimilar from both synoptic and snowmelt floods. This is partly related to their slightly higher Kendall's τ (not shown here).

The effect of the sample size can be illustrated in the shape of the distribution of p-values. The cumulative distribution functions of p-values from the comparison of the full data samples (Fig. 2, top) have concave shape. At the significance level of $\alpha = 0.1$, the empirical copulas of flash and synoptic floods (flash and snowmelt floods / synoptic and snowmelt floods) are significantly different in 25 (15 / 6) cases, which yields 36.2% (21.7% / 8.7%) of the all 69 sites.

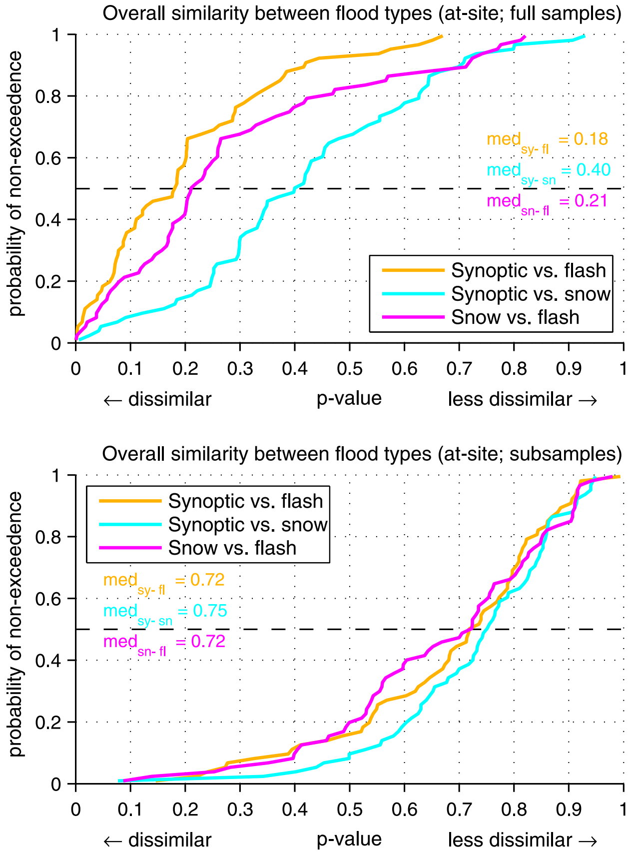

Fig. 2. Cumulative distribution functions of p-values that measure how similar the empirical copulas of different flood types are for each catchment. Plots on the top refers to full samples, while plots at bottom show p-values for subsamples. Medians (med) are also indicated. The copulas of synoptic floods and snowmelt floods are slightly more similar than the other process pairs.

On contrary, in the subsample approach, the test only compares empirical copulas constructed on the basis of the same number of data pairs at each site. As a result, the distributions of the p-values are shifted to the right, with definitely convex shapes (Fig. 2, bottom), which indicates that the significance of the former results is practically lost. There are only 2 cases where the empirical copulas were found to be significantly different (one case for the comparison of synoptic and snowmelt floods, and one for snowmelt vs. flash floods).

The features described above may also be illustrated and commented on the example of the Au (Mittlere Au) / Laudach catchment shown in Fig. 3 where we show the flood hydrographs for each process type and the peak-volume relationships in a unit square for all events, and similar relationships for resampled subsamples where each sample size equals to that of flash floods. When the full data samples are used, the comparison of syn-

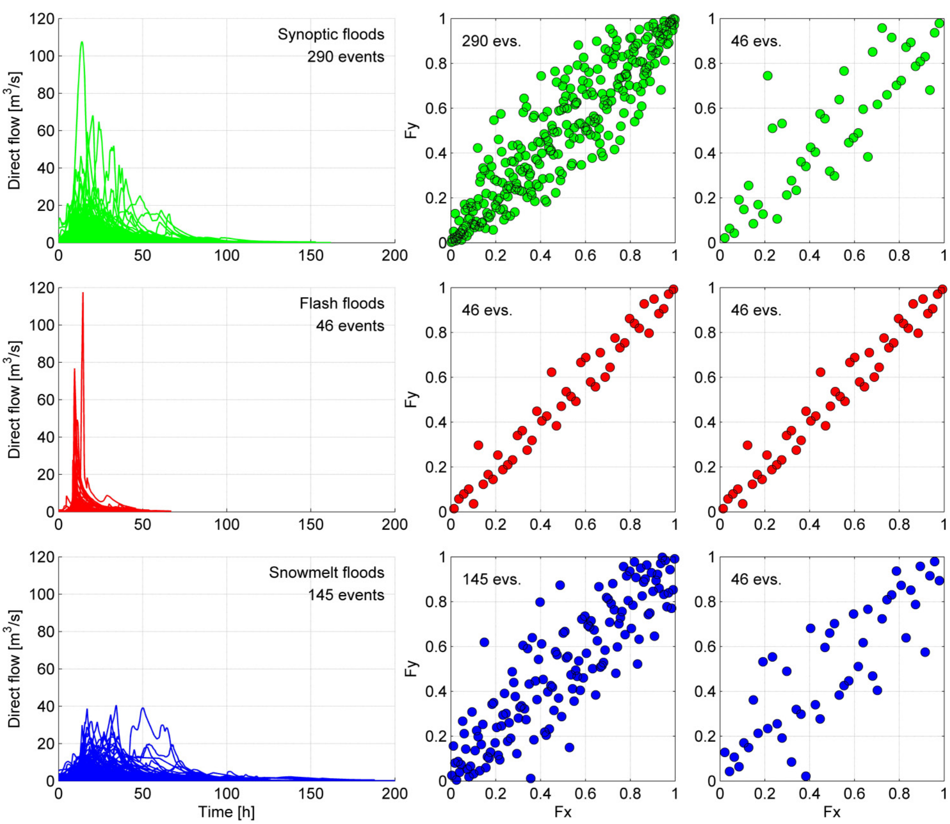

Fig. 3. Hydrographs of observed synoptic, flash and snowmelt floods (left) and peak-volume relationships (middle and right) for Au (Mittlere Au) / Laudach (70.3 km^2 of catchment area). F_x and F_y are the pseudo-observations of the peaks and event volumes, respectively. The middle column shows peak-volume relationships in a unit square for all events, while in the right column, relationships for resampled subsamples are shown where each sample size equals to that of flash floods.

Table 2. p-values of the pairwise comparison of empirical copulas constructed either on the basis of full samples or subsamples, for the catchment of Au (Mittlere Au) / Laudach, shown in Fig. 3. Cases rejected at the significance level of $\alpha = 0.1$ are printed in italics.

	synoptic vs. flash floods	snowmelt vs. flash floods	synoptic vs. snowmelt floods
Full samples	*0.000*	*0.001*	0.731
Subsamples	0.221	0.139	0.863

optic and flash floods gives a p-value of zero (Table 2), indicating rejection of the null hypothesis (i.e., the processes are different in terms of their empirical copulas). The same holds true for the comparison of snowmelt and flash floods ($p = 0.001$). On the other hand, the comparison of synoptic and snowmelt processes gives a high p-value (0.731), i.e., one cannot reject the null hypothesis about the equality of empirical copulas (Table 2). Clearly, both synoptic and snowmelt floods show quite variable hydrograph shapes and durations, which result in a higher degree of spread of the points along the diagonal of the scatter plots, but less so for the flash floods (Fig. 3, middle column).

Nevertheless, as soon as subsamples are used to test the similarity of empirical copulas, slightly different results are obtained. The p-values from the comparison of flash vs. synoptic (0.221) and flash vs. snowmelt floods (0.139) are still lower than the p-value from the comparison of snowmelt and synoptic floods (0.863), but the results are not significant at the level of $\alpha = 0.1$ (Table 2). This is also illustrated in the right column of Fig. 3. A visual comparison of the subsample scatter plots suggests that the synoptic and snowmelt floods are not as different from the flash floods as in the case of full data sets; therefore, the statistical test is unable to come to significant conclusions.

Overall, the analysis suggests that to arrive at meaningful results, even more data is needed that we were able to prepare. Such larger samples may be gained from longer series of high resolution discharge data (temporal resolution 1 hour and less) by using rainfall-runoff models with measured or simulated precipitation/temperature inputs. Both options, however, were out of the scope of this study. Based on the data we have, we can state, that in North-West of Austria flash floods are more often distinguishable from the synoptic and snowmelt floods in terms of their empirical peak-volume copulas than the synoptic and the snowmelt floods from each other. In other words, the empirical copulas of synoptic floods and snowmelt floods are slightly more similar than the other process pairs.

4.2. A comparison of empirical copulas for each flood type regionally

Next we were interested in whether different catchments, for the same flood type, were distinguishable in terms of their empirical peak-volume copulas. Similar to the above analysis, the flood data were analysed for the full samples and resampled in a way that, for a given catchment, the numbers of synoptic, flash and snowmelt floods were the same. Each distribution of p-values in Fig. 3 consists of 2346 values which is the number of catchment pairs $(69 \times 68/2)$.

Two features of Fig. 4 (top) indicate that the less similar flood types are the synoptic floods. First, the synoptic flood types are associated with the lowest median of p-values, and second, the green line representing the synoptic floods has the leftmost position, at least for the smaller p-values.

If we look at the results from the traditional statistical point of view, i.e., want to make a yes/no decision on the basis whether the p-values exceed the significance level $\alpha = 0.1$, we arrive at similar conclusion: the largest probability of being the pairs of empirical copulas dissimilar appears in the case of the synoptic floods. *Szolgay et al. (2016)* explained this feature by the larger variability of flood hydrograph shapes and by the mixture of long-rain and short-rain floods. Nevertheless, the question still remains that to what extent can be these features attributed to the types of floods only or are they influenced by the effect of the sample size? The position of the three CDFs in Fig. 4 indicates that what we see may not be purely the effect of the sample size at least for the synoptic floods (largest samples). If it was so, the lines would be arranged in the order of the mean (median) sample size for the individual flood processes, i.e., synoptic – snowmelt – flash floods (Fig. 5).

The analysis in the bottom part of Fig. 4 is not based on the full samples, but on the resampled subsets for the synoptic and snowmelt floods (note that the samples of flash floods are taken with no change). It can be seen that the empirical copulas of flash floods are the least similar between catchments.

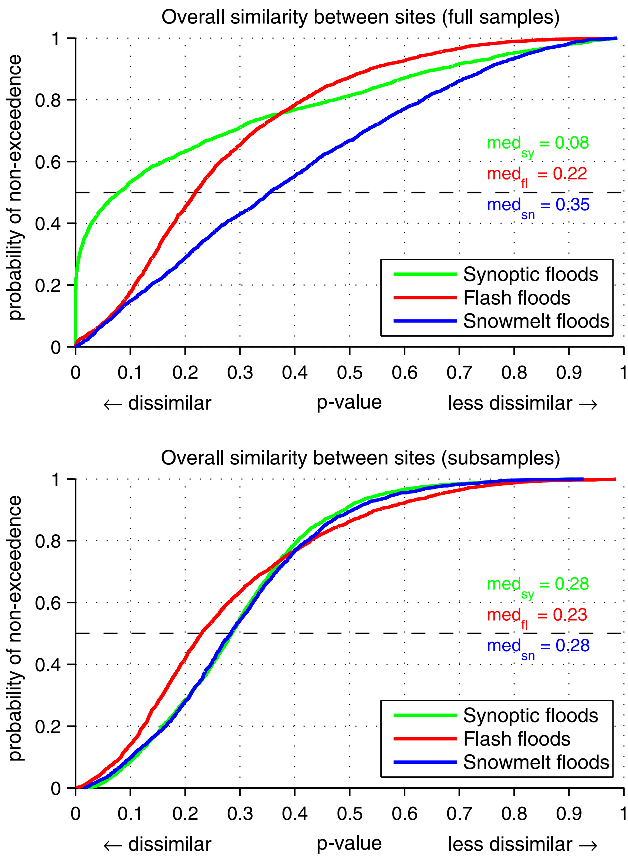

Fig. 4. Cumulative distribution functions of p-values that measure how similar the empirical copulas of different pairs of catchments are for each flood type. Medians (med) are also indicated in the case of full (top) and the resampled (bottom) data.

It is indicated by (i) the lowest median of p-values, and (ii) the leftmost position of the red curve, at least for the smaller p-values.

Again, if we look at the results from the traditional statistical point of view, i.e., from the perspective of a yes/no decision on the basis whether the p-values exceed the significance level $\alpha = 0.1$, the conclusion is similar: the largest probability of being the pairs of empirical copulas dissimilar appears in the case of the flash floods (17.7%). The corresponding frequencies for the synoptic and snowmelt floods are 8.6% and 10.4%, respectively. This

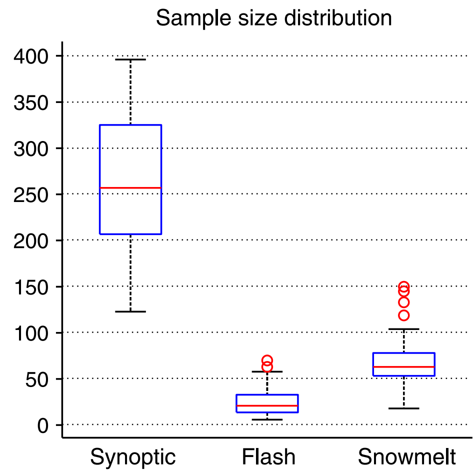

Fig. 5. Sample sizes of the respective processes.

means that the difference between the process types is small, although the flash floods do show slightly less spatial homogeneity. Please note that even though the large data samples for synoptic and snowmelt floods have been cut down by resampling, we still compare different sample sizes within the region, which ranges between 7 and 72. Such differences in sample size also can indicate higher likelihood that the blanket test classifies the empirical copulas as being different, i.e., more frequent occurrence of lower p-values. Overall, the analysis suggests that most catchment pairs, for a given flood type, are not distinguishable in terms of their empirical peak-volume copulas. We believe that this is to be attributed to the effect of the low sample size.

5. Discussion and conclusions

This paper has addressed two science questions; both of them will be discussed in more detail in the following paragraphs.

(i) The first science question was *how similar are the peak-volume dependence structures of different flood types for a given catchment?* The analysis suggests that, when empirical copulas for different flood processes are compared locally, flash floods are more often distinguishable from synoptic and snowmelt floods than are synoptic and snowmelt floods from each other. These findings along with the results of the goodness-of-fit test of copula types in *Szolgay et al. (2015, 2016)* suggest that there is value in the concept of process-based analysis of peak-volume relationship. Rather than treating the entire data set of flood-peak volume pairs as realisations of the same process, copulas should be fitted to the subsets representing different process types. This also could lead to the concept of mixed distributions in the bivariate case which has received considerable attention in univariate flood frequency analysis (e.g., *Waylen and Woo, 1982; Hirschboeck et al., 2000; Sivapalan et al., 2005*). In this case, one would not only assume the marginals to be mixed distributions (e.g., *Karmakar and Simonovic, 2009*) but also the dependence structure to be mix of different copulas, each of them representing different flood processes (see, e.g., *Hu, 2006* for an example of mixed copulas).

It is further important to note that sample size plays a key role in deciding whether empirical copulas can be regarded as different or not. In Fig. 5, the sample sizes of the respective processes are shown. The difference between the individual flood types (particularly between the synoptic and flash floods) is striking. Therefore, in a local testing of equality of empirical copulas, two datasets with considerably different sample sizes were often compared. The preliminary results of comparison of empirical copulas indicated that the resulting p-values were influenced the size of both samples. We observed that if the sample sizes were different (for instance, we compare 30 vs. 100 pairs of Q-V, which is a realistic situation), the test tended to indicate dissimilarity even if the rank correlations of the samples were of a similar magnitude. Therefore, in order to eliminate the sample size effect, we decided to limit the size of the data samples at each catchment by drawing subsamples. On the other hand, the results of the analysis of the intuitive similarity measure constructed on the basis of test of *Remillard and Scaillet (2009)* showed decreased ability of subsamples to yield regionally and locally interpretable results (Fig. 3).

It is therefore concluded that analyses like this should rather be based on data sets (i) that are sufficiently large (several tens or better over hundreds of events), and (ii) which have comparable sizes for different flood types. In the current analysis, the samples of flash floods did not meet any of the suggested criterions, but for the future, it is advised to analyze data sets that are more balanced in terms of sample size.

Furthermore, it also should be noted that in spite of the extension of the sample size by the methods used in this paper, some spatial pooling or data extension by simulation will be required in practical applications in the future in order to limit the sampling uncertainty, since even the enlarged dataset used herein was not able to lead to truly meaningful generalisations and recommendations.

(ii) The second science question was *how similar are the peak-volume dependence structures between catchments for a given flood type?* The analysis again suggests that, when for a given flood type we do not have sufficient data (despite of the proposed method to extend the sample size of the flood peak and volume pairs by selecting all possible events), the majority of catchment pairs is not distinguishable in terms of their empirical peak-volume copulas at $\alpha = 0.1$. For synoptic (snowmelt / flash) floods, 46.6% (85.2% / 82.3%) of the total number of 2346 pairs is not distinguishable when the full samples are compared (Fig. 4, top). In the case of subsamples, the figures are 91.3% (90.2% / 85.9%), respectively (Fig. 4, bottom).

To examine the sample size effect, we performed numerical experiments with the synoptic floods. First, at each catchment, we draw random samples of size of $N = 120$ (the lowest number of synoptic floods in the dataset is 123) from the total number of synoptic floods, and we performed the same analysis, i.e., compared the empirical copulas for the subsamples. The experiment was repeated 10 times. The results are presented in Fig. 6. It can be seen that the general features of the subsamples (particularly in the region of p-values between 0.0 and 0.1) are similar to the green line corresponding to the real observations.

A similar experiment was repeated with random samples of size of $N = 50$, with 100 repetitions. The results are presented in Fig. 7.

The results of the experiments with the sample size of synoptic floods suggest that the discriminating power of the intuitive method of comparison of similarity of empirical copulas diminishes with the decrease of the sample size. This conclusion leads to the question, whether the shapes of the cumulative distributions of our similarity measure do sufficiently represent differences in the flood generating processes or whether are, to some extent, corrupted by the effect of the sample size distribution of the respective processes. This question is important to examine, since in our case the sample

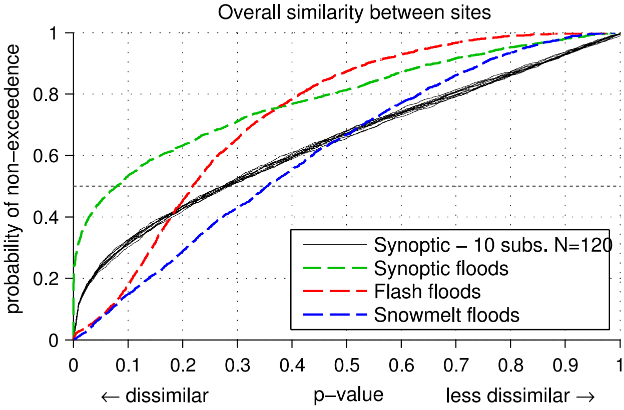

Fig. 6. Cumulative distribution functions of p-values that measure how similar the empirical copulas of different pairs of catchments are for each flood type for the original data (thick and dashed coloured lines) and the 10 subsamples of synoptic floods, each subsampled to $N = 120$ data pairs (thin black lines appearing close to each other).

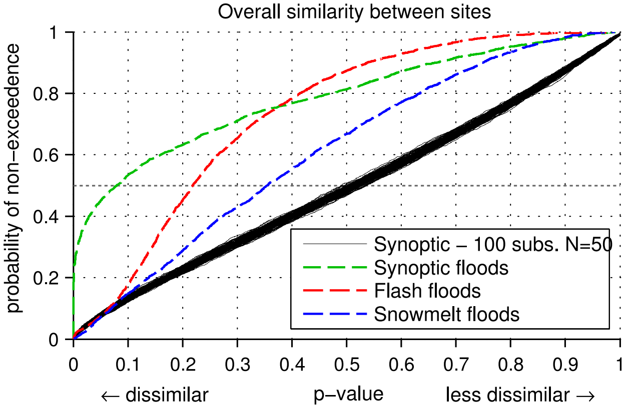

Fig. 7.Cumulative distribution functions of p-values that measure how similar the empirical copulas of different pairs of catchments are for each flood type for the original data (thick and dashed coloured lines) and the 100 subsamples of synoptic floods, each subsampled to $N = 50$ data pairs (thin black lines appearing close to each other).

size distribution is also influenced by the frequency of the respective events in the region.

In the experiments so far, the synoptic flood were subsampled to a predefined constant number at each catchment. Nevertheless, in the following experiments, the size of the subsamples will be different. At each catchment, we draw random samples of synoptic floods of size that equals to those of the snowmelt floods. The experiment was repeated 10 times. The results are presented in Fig. 8. Next, at each catchment, we draw random samples of synoptic and snowmelt floods of size that equals to those of the flash floods. Then, we performed the same analysis, i.e., compared the empirical copulas for the subsamples. The experiment was, again, repeated 10 times. The results are shown in Fig. 9.

Both experiments (summarized in Figs 8 and 9) resulted in similar outcomes. In the first one, the shapes CDFs of the subsampled synoptic data closely follow the shape of the CDF of the snowmelt floods which served as the basis for the resampling of synoptic data (Fig. 8). A similar conclusion holds true for the resampled synoptic and snowmelt floods where the sample size distribution of the flash floods served as the muster (Fig. 9). In other words, the results largely reflects the effect of the regional sample size distribution, and are less conclusive regarding the differences in flood generating processes. This problem cannot be solved in our case with real world data, since the sample size distribution of the respective processes is influenced by the frequency of the respective events in the region.

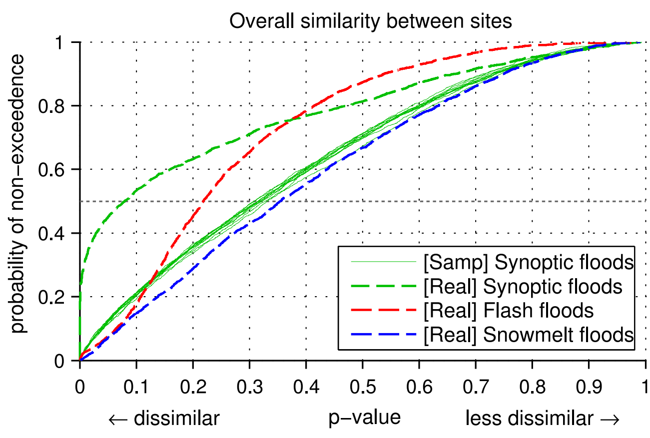

Fig. 8. Cumulative distribution functions of p-values that measure how similar the empirical copulas of different pairs of catchments are for each flood type for the original data (thick and dashed coloured lines denoted in legend as 'Real') and the 10 subsamples of synoptic floods, each subsampled to size that equals to those of the snowmelt floods (thin green lines appearing close to each other and denoted in legend as 'Samp').

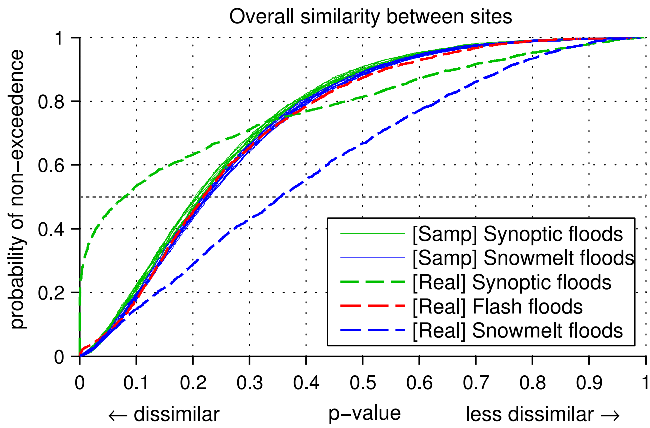

Fig. 9. Cumulative distribution functions of p-values that measure how similar the empirical copulas of different pairs of catchments are for each flood type for the original data (thick and dashed coloured lines denoted in legend as 'Real') and the 10 subsamples of synoptic and snowmelt floods, each subsampled to size that equals to those of the flash floods (thin coloured lines appearing close to each other and denoted in legend as 'Samp').

These experiments with subsamples lead to the conclusion that the shapes of the cumulative distributions of our intuitively introduced similarity measure may satisfactorily represent differences in the flood generating processes only in a case of sufficiently large sample size, as it seem so be in the case of synoptic floods in the target region. However, we also have to take into consideration that the synoptic floods in the target dataset include both short- and long-rain floods and this process inhomogeneity may increase the likelihood of the detection of differences between empirical copulas based on local differences in the ratio of respective type of events.

The differences in empirical copula types between flood processes that can be recognized visually in Fig. 3 do suggest that there is merit in treating the flood types separately when analysing and estimating flood peak-volume copulas. However, the sample sizes gained by our data extension method was not sufficient to prove that, nor did the proposed intuitive similarity measure proved enough discriminating power given the limiting sample size. Please note that in this study, the median sample sizes were 257 events per catchment for synoptic floods, 21 for flash floods and 63 for snowmelt floods. The usual sample size in the traditional engineering approach using annual maxima would be 31 in our case. These findings again support the need of a regional process-based pooling of catchments or the application of simulation-based data construction/extension. Similar to the univariate case of flood peaks, flood events in a region that are associated with the same process type could be pooled in order to increase the sample size (*Merz and Blöschl, 2003*) or the dataset could be extended by using simulation techniques (e.g., *Gräler et al., 2013*). A pooled analysis or the use of simulation may increase the sample size significantly beyond our sample sizes. Future research could than examine how empirical copulas are similar and what copula types are appropriate for these larger samples.

Acknowledgements. We would like to thank the project Flood change (ERC Advanced Grant, 291152) and the FWF (project No. P 23723-N21) for financial support. This research was also supported by the Slovak Research and Development Agency under contract No. APVV 0496-10, by the Slovak Grant Agency VEGA under the project No. 1/0776/13. The support of the Competence Center for SMART Technologies for Electronics and Informatics Systems and Services, ITMS 26240220072, funded by the Research & Development Operational Programme from the ERDF is acknowledged. The research is a contribution to the Working Groups 'Understanding Flood Changes' of the Panta Rhei Initiative of the International Association of Hydrological Sciences (IAHS).

References

Bačová-Mitková V., 2012: The relationship between volume of the flood wave and the time duration of flood events (Vzájomný vzťah objemu a dĺžky trvania povodňových vĺn). Acta Hydrologica Slovaca, **13**, 1 165–174.

Bačová-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`.

Ben-Aissia M.-A., Chebana F., Ouarda T. B. M. J., Roy L., Desrochers G., Chartier I., Robichaud É., 2012: Multivariate analysis of flood characteristics in a climate change context of the watershed of the Baskatong reservoir, Province of Québec, Canada. Hydrological Processes, **26**, 130–142, `doi: 10.1002/hyp.8117`.

Chapman T. G., Maxwell A. I., 1996: Baseflow separation—comparison of numerical methods with tracer experiments. In: 23rd Hydrology and Water Resources Symposium: Water and the Environment, Natl. Conf. Publ., 96/05, Inst. of Eng., Barton, A.C.T., Australia, pp. 539–545.

Chebana F., Ouarda T. B. M. J., 2007: Multivariate L-moment homogeneity test. Water Resources Research, **43**, W08406, `doi: 10.1029/2006WR005639`.

Chebana F., Ouarda T. B. M. J., 2009: Index flood-based multivariate regional frequency analysis. Water Resources Research, **45**, 10, W10435, doi: 10.1029/2008WR007490.

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.

Dalrymple T., 1960: Flood frequency analysis, U.S. Geol. Surv. Water Supply Pap., 1543A, 80 pp.

DWA, Deutsche Vereinigung für Wasserwirtschaft, Abwasser und Abfall, 2012: Ermittlung von Hochwasserwahrscheinlichkeiten. Guidelines DWA-M. 552, Hennef, Germany.

Favre A.-C., El Adlouni S., Perreault L., Thiémonge N., Bobée B., 2004: Multivariate hydrological frequency analysis using copulas. Water Resources Research, **40**, W01101, doi: 10.1029/2003WR002456.

Gaál L., Kyselý J., Szolgay J., 2008: Region-of-influence approach to a frequency analysis of heavy precipitation in Slovakia. Hydrology and Earth System Sciences, **12**, 3, 825–839, doi: 10.5194/hess-12-825-2008.

Gaál L., Szolgay J., Kohnová S., Hlavčová K., Parajka J., Viglione A., Merz R., Blöschl G., 2014: 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., Szolgay J., Kohnová S., Parajka J., Merz R., Viglione A., Blöschl G., 2012: Flood timescales: Understanding the interplay of climate and catchment processes through comparative hydrology. Water Resources Research, **48**, 4, W04511, doi: 10.1029/2011WR011509.

Ganguli P., Reddy M. J., 2013: Probabilistic assessment of flood risks using trivariate copulas. Theoretical and Applied Climatology, **111**, 341–360. doi: 10.1007/s00704-012-0664-4.

Genest C., Favre A.-C., 2007: Everything you always wanted to know about copula modeling but were afraid to ask. Journal of Hydrologic Engineering, **12**, 4, 347–368, doi: 10.1061/(ASCE)1084-0699(2007)12:4(347).

Gräler B., van den Berg M. J., Vandenberghe S., Petroselli A., Grimaldi S., De Baets B., Verhoest N. E. C., 2013: Multivariate return periods in hydrology: a critical and practical review focusing on synthetic design hydrograph estimation. Hydrology and Earth System Sciences, **17**, 1281–1296, doi: 10.5194/hess-17-1281-2013.

Hirschboeck K. K., Ely L., Maddox R. A., 2000: Hydroclimatology of meteorologic floods. In Wohl E. (Ed.): Inland Flood Hazards: Human, Riparian and Aquatic Communities, pp. 39–72, Cambridge University Press, New York.

Hosking J. R. M., Wallis J. R., 1997: Regional Frequency Analysis: An Approach Based on L-moments, Cambridge University Press, Cambridge, UK.

Hu L., 2006: Dependence patterns across financial markets: a mixed copula approach. Applied Financial Economics, **16**, 10, 717–729, doi: 10.1080/09603100500426515.

IH, Institute of Hydrology, 1999: Flood Estimation Handbook. Institute of Hydrology: Wallingford, UK.

Karmakar S., Simonovic S. P., 2009: Bivariate flood frequency analysis. Part 2: A copula-based approach with mixed marginal distributions. Journal of Flood Risk Management, **2**, 32–44. doi: 10.1111/j.1753-318X.2009.01020.x.

Kohnová S., Szolgay J., 1999: Regional estimation of design summer flood discharge in small catchments of northern Slovakia. In: Gottschalk L., Olivry C., Reed D., Rosbjerg D. (Eds.): Hydrological Extremes: Understanding, Predicting, Mitigating, pp. 265-268. IAHS publ. 255, IAHS Press Wallingford.

Merz R., Blöschl G., Parajka J., 2006: Spatio-temporal variability of event runoff coefficients. Journal of Hydrology, **331**, 3–4, 591–604, doi: 10.1016/j.jhydrol.2006.06.008.

Merz R., Blöschl G., 2003: A process typology of regional floods. Water Resources Research, **39**, 12, 1340, doi: 10.1029/2002WR001952.

Merz R., Blöschl G., 2009: A regional analysis of event runoff coefficients with respect to climate and catchment characteristics in Austria. Water Resources Research, **45**, 1, W01415, doi: 10.1029/2008WR007163.

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, doi: 10.1016/j.jhydrol.2010.05.015.

Parajka J., Merz R., Blöschl G., 2007: Uncertainty and multiple objective calibration in regional water balance modelling – Case study in 320 Austrian catchments. Hydrological Processes, **21**, 435–446, doi: 10.1002/hyp.6253.

Poulin A., Huard D., Favre A.-C., Pugin S., 2007: Importance of tail dependence in bivariate frequency analysis. Journal of Hydrologic Engineering, **12**, 4, 394–403, doi: 10.1061/(ASCE)1084-0699(2007)12:4(394).

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, doi: 10.1007/s11269-012-0124-z.

Remillard B., Plante J.-F., 2012: TwoCop: Nonparametric test of equality between two copulas. R package version 1.0 (19-10-2012). http://cran.r-project.org/web/packages/TwoCop.

Remillard B., Scaillet O., 2009: Testing for equality between two copulas. Journal of Multivariate Analysis, **100**, 377–386, doi: 10.1016/j.jmva.2008.05.004.

Requena A. I., Chebana F., Mediero L., 2016: A complete procedure for multivariate index-flood model application. Journal of Hydrology, **535**, 559–580, doi: 10.1016/j.jhydrol.2016.02.004.

Shiau J.-T., Wang H.-Y., Tsai C.-T., 2006: Bivariate flood frequency analysis of floods using copulas. Journal of the American Water Resources Association, **42**, 6, 1549–1564, doi: 10.1111/j.1752-1688.2006.tb06020.x.

Sivapalan M., Blöschl G., Merz R., Gutknecht D., 2005: Linking flood frequency to long-term water balance: incorporating effects of seasonality. Water Resources Research, **41**, W06012, doi: 10.1029/2004WR003439.

Sraj M., Bezak N., Brilly M., 2014: Bivariate flood frequency analysis using the copula function: a case study of the Litija station on the Sava River. Hydrological Processes, **29**, 2, 225–238, doi: 10.1002/hyp.10145.

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. Proceedings of IAHS, **370**, 183–188, doi: 10.5194/piahs-370-183-2015.

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. Proceedings of IAHS, **373**, 61–67, doi: 10.5194/piahs-373-61-2016.

Waylen P., Woo M., 1982: Prediction of annual floods generated by mixed processes. Water Resources Research, **18**, 4, 1283–1286, doi: 10.1029/WR018i004p01283.

Zhang L., Singh V. P., 2006: Bivariate flood frequency analysis using the copula method. Journal of Hydrologic Engineering, **11**, 150–164, doi: 10.1061/(asce)1084-0699(2006)11:2(150).