J. Hydrol. Hydromech., 71, 2023, 4, 449–463
https://doi.org/10.2478/johh-2023-0027

sciendo Journal of Hydrology and Hydromechanics Institute of Hydrology SAS

The testing of a multivariate probabilistic framework for reservoir safety evaluation and flood risks assessment in Slovakia: A study on the Parná and Belá Rivers

Roman Výleta[1*], Peter Rončák[2], Anna Liová[1], Peter Valent[3], Tomáš Bacigál[4], Zoltán Gribovszki[5], Zuzana Danáčová[6], Peter Šurda[2], Justína Vitková[2], Kamila Hlavčová[1]

[1] Department of Land and Water Resources Management, Faculty of Civil Engineering, Slovak University of Technology in Bratislava, Radlinského 11, 810 05 Bratislava, Slovakia.
[2] Institute of Hydrology, Slovak Academy of Sciences, Dúbravská cesta 9, 841 04 Bratislava, Slovakia.
[3] Institute of Hydraulic Engineering and Water Resources Management, Technische Universität Wien, Vienna, 1040, Austria.
[4] Department of Mathematics and Constructive Geometry, Faculty of Civil Engineering, Slovak University of Technology in Bratislava, Radlinského 11, 810 05 Bratislava, Slovakia.
[5] Institute of Geomatics and Civil Engineering, Faculty of Forestry, University of Sopron, Bajcsy-Zsilinszky u. 4., 9400 Sopron, Hungary.
[6] Slovak Hydrometeorological Institute, Jeséniova 17, 833 15 Bratislava, Slovakia.
[*] Corresponding author. Tel.: + 421 232 888 727. E-mail: roman.vyleta@stuba.sk

Abstract: Intense floods represent a challenge to risk management. While they are multivariate in their nature, they are often studied in practice from univariate perspectives. Classical frequency analyses, which establish a relation between the peak flow or volume and the frequency of exceedance, may lead to improper risk estimations and mitigations. Therefore, it is necessary to study floods as multivariate stochastic events having mutually correlated characteristics, such as peak flood flow, corresponding volume and duration. The joint distribution properties of these characteristics play an important role in the assessment of flood risk and reservoir safety evaluation. In addition, the study of flood hydrographs is useful because of the inherent dependencies among their practice-relevant characteristics present on-site and in the regional records. This study aims to provide risk analysts with a consistent multivariate probabilistic framework using a copula-based approach. The framework respects and describes the dependence structures among the flood peaks, volumes, and durations of observed and synthetic control flood hydrographs. The seasonality of flood generation is respected by separate analyses of floods in the summer and winter seasons. A control flood hydrograph is understood as a theoretical/synthetic discharge hydrograph, which is determined by the flood peak with the chosen probability of exceedance, the corresponding volume, and the time duration with the corresponding probability. The framework comprises five steps: 1. Separation of the observed hydrographs, 2. Analysis of the flood characteristics and their dependence, 3. Modelling the marginal distributions, 4. A copula-based approach for modelling joint distributions of the flood peaks, volumes and durations, 5. Construction of synthetic flood hydrographs. The flood risk assessment and reservoir safety evaluation are described by hydrograph analyses and the conditional joint probabilities of the exceedance of the flood volume and duration conditioned on flood peak. The proposed multivariate probabilistic framework was tested and demonstrated based on data from two contrasting catchments in Slovakia. Based on the findings, the study affirms that the trivariate copula-based approach is a practical option for assessing flood risks and for reservoir safety.

Keywords: Copula-based approach; Flood risk; Flood seasonality; Trivariate analysis; Synthetic hydrographs.

INTRODUCTION

Resilient flood risk management needs reliable estimations of the design values for proposing structural mitigation measures such as sizing reservoir volumes and the spillways of dams, the height of flood levees, and planning areal zoning for residential areas at risk. Additionally, control values are also needed to test the safety of existing measures and structures. In practice, the most often used quantities of interest in flood risk estimations are the design or critical flood peaks (sometimes also their volume or duration). Classical frequency analysis has established a relation between the peak flow (or volume) and the respective frequency of exceedance, which describes the intensity of the hazards. However, univariate analysis ignores the dependence between the characteristics of the hydrological design and control variables, which may lead to inappropriate conclusions about the overall flood risk (Rizwan et al., 2019).

Therefore, univariate schemes need to be extended to multivariate ones for several risk-based design scenarios. In some cases, even the inclusion of an entire design flood hydrograph may be advantageous and necessary (Brunner et al., 2016a; Brunner, 2023), including measures of the probability of the whole design flood hydrograph rather than just the probabilities of its components (peaks, volumes and durations, separately). For example, specific hydraulic structures such as dams, bridges, and culverts can only be correctly sized when the shape of the whole hydrograph is accounted for (e.g., Škvarka et al., 2021). Considering only univariate design values may also not allow for a sufficient description of the multiple impacts of areal flooding. In some cases, such an insufficient description could lead to improper risk estimations since such impacts depend (e.g., in estimating flood damage) on combined hazardous effects related to various interdependent characteristics of the phenomenon (such as the peak flow, flood volume, duration, flow velocity, and the depth of the water).

Therefore, the common practice in many countries is to create synthetic hydrographs for design and control. Gadek et al. (2022) provide a comprehensive overview of such methods, which are divided into deterministic (mainly relying on rainfall-runoff modelling) and statistical methods. Drobot et al. (2021) provide a more detailed classification. Their statistical approaches contain parametric and nonparametric methods. Parametric approaches represent the shape of the hydrograph by an analytic expression controlled by a set of parameters. Probability distribution functions (their shape) are often in use as shapes of hydrographs. Pramanik et al. (2009) provided a comprehensive review of such studies and the distribution functions used. Non-parametric methods usually do not make explicit assumptions about the shape of the hydrograph, which is determined through aggregations of a set of observed hydrographs, as, e.g., in Le Clerc et al. (2003). Gadek et al. (2022) commented on and provided several sources in the literature on hydrographs, which can also respect the seasonality of the flood generation processes, as Brunner et al. (2017) show.

For conducting a multivariate risk analysis, all these methods should preferably include a joint probabilistic description of the characteristics of the design and control hydrograph. National laws and guidelines usually fix a given return period for dam designs (Requena et al., 2013), which usually refers to the flood peak associated with the entire hydrograph. A flood hydrograph can be described in several ways (O'Connor et al., 2014; Pandi, 2010) according to several interdependent characteristics. The flood peaks, total volume, volume above the base flow, total duration, timescales, compactness coefficient, rise-to-duration ratio, rainfall duration, season of occurrence, flood type, the existence of multi-peaks, various shape parameters, etc., could be considered (Drobot et al., 2021; Ganapathy et al., 2022; Yue et al., 2002). Interdependencies between different flood characteristics need to be considered, and the analysis needs to focus on such variables, which are of interest according to the particular application (Brunner, 2023). For a risk analysis, a multivariate distribution function of these variables as a joint distribution of their marginal distributions is preferred (Medeiro et al., 2010). Consequently, recent efforts have been devoted to joint probabilistic analyses of multiple flood characteristics or whole flood events (Xiao et al., 2009). Various classical multivariate statistical analysis methods were historically first proposed, and bi- and tri-variate probability distributions were tested. Rizwan et al. (2019) has provided a short review of these methods, which generally exhibit their limitations, such as the marginal distributions coming from identical statistical distributions. In nature, however, the most suitable marginal distributions of the constituent variables of a multidimensional analysis are usually not identical. Moreover, in the tri- and more variate cases, the mathematical formulations also become more complex (Rizwan et al., 2019).

These problems have recently been eliminated by connecting the differing marginal distributions by copula functions, which account for the dependence structure between differently distributed random variables in a probability space. They also provide a framework for estimating the joint probabilities (overall or conditional) based on a multivariate analysis. Nazeri Tahroudi et al. (2022), Tootoonchi et al. (2022), Größer and Okhrin (2022) provide comprehensive overviews of these developments and the methods and applications.

Various copula families, which usually reflect a bivariate dependence between peaks and volumes, have been reported in the literature for modelling floods. This limitation is mainly caused by the limited availability of higher-dimensional expressions for copula families. In recent years, vine copulas, which allow for overcoming these and other limitations, have been suggested as a solution (Tosunoglu et al., 2020). Vine copulas can cover flexible dependence structures by mixing bivariate copulas. This advantage over other copula families has drawn increased attention to the study of multivariate modelling, including that of floods (e.g., Brunner et al., 2019; Gómez et al., 2018; Jafry et al., 2022; Latif and Simonovic, 2022; Nazeri Tahroudi et al., 2022).

Bivariate return periods have often been used to quantify flood peaks and volumes (Brunner et al., 2016a, 2017; Carril-Rojas and Mediero, 2023; Requena et al., 2013; Rizwan et al., 2019); the trivariate case was tackled in Ganguli and Reddy (2013), Gräler et al. (2013), Grimaldi et al. (2016), Tosunoglu et al. (2020), Jafry et al., (2022), Latif and Simonovic (2022). Multivariate distributions enable an assessment of the probabilities of exceedance or return periods of a set of variables in a multivariate context. Whereas, in a univariate case, these are uniquely defined based on the distribution function, in a multivariate framework, we need to select one definition (out of several that preferably depend on the problem at hand); an overview is given in Gräler et al. (2013) and Brunner et al. (2016b).

Since various types of floods have different shapes and consequently exhibit different dependence structures between their peak discharges and flood volumes (Gaál et al., 2015; Grimaldi et al., 2016; Szolgay et al., 2015), it is recommended to respect the joint frequency analysis separately in the analysis of the potentially most hazardous flood types (Brunner et al., 2017; Hundecha et al., 2017).

In Slovakia (and, until 1993, in Czechoslovakia), the mandatory use of water management calculation procedures was prescribed by technical standards, which have gradually been abandoned since 1989. This change has allowed water managers and designers more flexibility in selecting calculation methods. However, it has also caused several problems that need to be addressed to ensure that the best available methods are adopted, and that all relevant data are used. Therefore, it has become necessary to develop and test a methodology that fulfills the objectives of the "Methodological Instruction of the General Director of the Water Section of the Ministry of the Environment of the Slovak Republic for the safety evaluation of dams and tailing ponds during flood loads as a part of technical and safety supervision" (Water Section, No.: 05/2020-4 Bratislava 18.11.2020, available at https://www.minzp.sk/voda/technicko-bezpecnostny-dohlad/). The methodological instruction, which is an integral part of the technical and safety supervision of waterworks, defines a control flood hydrograph as a theoretical/synthetic discharge hydrograph, which is determined by the flood peak with the selected probability of exceedance, the corresponding volume, and the duration with the corresponding probability. This hydrograph is used to critically test the safety of hydraulic structures.

The aim of this study is to therefore provide risk analysts with a consistent multivariate probabilistic framework for evaluating reservoir safety and assessing flood risks, which respects and describes the dependence structures among the flood peaks, volumes and durations of the observed and synthetic flood hydrographs. Additionally, the seasonality of flood generation is respected by a separate analysis of summer and winter floods. The framework consists of five steps: identification of floods and the separation of observed hydrographs; analysis of the dependence of flood characteristics for the maximum annual floods and maximum seasonal floods in the summer and winter seasons; modelling the marginal distributions of flood characteristics; modelling the joint distribution of flood peaks, volumes and durations; and the construction of synthetic flood hydrographs. The dependence modelling of flood peaks, volumes and durations is carried out with various types of non-parametric and parametric copulas. The flood risk assessments are described by hydrograph analyses and the conditional joint probabilities of the exceedance of flood volumes and durations conditioned on flood peaks. The properties of the synthetic hydrographs are also based on these probabilities. The evaluation of the reservoir safety and assessment of the flood risks has been demonstrated based on data from two contrasting catchments in Slovakia. The study not only provides suggestions for the development of a broader and more comprehensive flood risk estimation framework and reservoir safety evaluation in Slovakia, but also emphasizes the importance of moving beyond the classical univariate perspective commonly employed in practice. The proposed multivariate framework and copula-based approach offer valuable tools for risk analysts seeking a robust and accurate assessment of flood risks and reservoir safety.

MATERIAL AND METHODS

Data application - study area

The first pilot catchment discussed is located on the Parná River in the western part of Slovakia on the inflow into the Horné Orešany reservoir (at an altitude of 235 m a.s.l.). The riverhead of the Parná river is situated in the Little Carpathians; the catchment is fan-shaped; and the area of the catchment contributing to the reservoir is 45.59 km^2 (Figure 1a). The catchment is almost completely forested with dense deciduous forests (90%) and patches of clearings, transitional woodlands, and grasslands. The mean annual precipitation is 810 mm. The largest mean monthly precipitation totals can be observed in June and the smallest in April. The flood runoff regime is a mixture of rain and snowmelt floods, with the highest flows at the end of the winter in March: they are induced by the rapid melting of snow cover, which is often combined with steady rain. The lowest flows can be observed at the end of summer in the months of August, September, and October after long periods of drought caused by low precipitation in general.

The second catchment is situated on the Belá River with an outlet in Liptovský Hrádok (at an altitude of 629 m a.s.l.) in the northern part of Slovakia (Figure 1b). The Belá River is a typical alpine mountain river and is a significant right-hand tributary of the Váh River. The total stream length is 23.6 km, with a catchment area of 244.26 km^2. The Belá River basin is covered by coniferous forests (50%), transitional woodlands (30%), and natural meadows. The alluvial lower part of the basin is mainly covered by agricultural land. The mean annual precipitation is 1,150 mm. The amount of runoff also depends on the geological bedrock. About 70–90% of the total precipitation falls on crystalline rocks at higher elevations and only 50% at lower elevations on limestone. Low discharges mainly occur in February, when the groundwater resources are almost exhausted at the end of winter and in the autumn season due to low precipitation and high summer temperatures. The highest floods start when the snowpack begins to melt in April and continues to the second half of May. Flash floods can be observed in June and July.

The dataset used in the analysis was provided by the Slovak Hydrometeorological Institute (SHMI). In the late 1990s, SHMI undertook the task of standardizing the equipment used for automated monitoring of surface water conditions in Slovakia. Since then, water level data has been continuously measured and transmitted at 15-minute intervals (Danáčová et al., 2015). The SHMI only validates and reports hourly discharges; the daily data is also available for longer periods. Due to the size of the catchment area and the availability of data, hourly discharge data from the 5250 Parná – Horné Orešany and 5480 Belá – Liptovský Hrádok water-gauging stations for the period 1988 to 2019 were selected as the shortest measurement time step.

Methodology

The method of a consistent multivariate probabilistic framework for reservoir safety evaluations and flood risk assessments involves fitting statistical density functions to observed flood hydrographs, while considering the interdependence between the characteristics, i.e., the flood peak, volume and duration. This step is crucial because solely relying on a univariate frequency analysis cannot accurately assess flood risk-associated probabilities. To arrive at this methodology, an analysis was conducted using a copula model that accurately captures and describes the dependence structures among the peaks, volumes, and durations of the flood hydrographs. This process is divided into five steps, as illustrated in Figure 2.

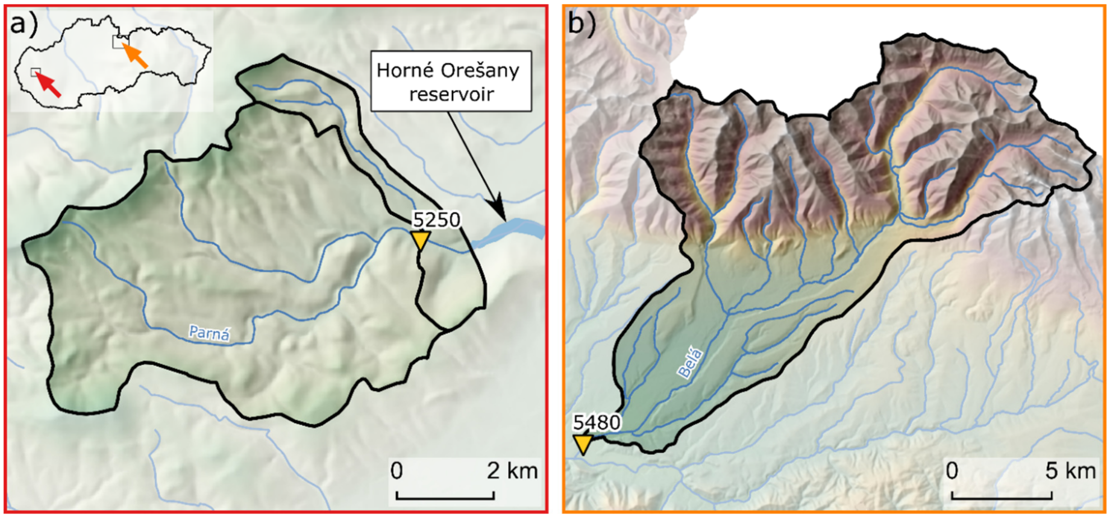

Fig. 1. Position of the two case study catchments: **a)** the Parná River at the inlet to the Horné Orešany water reservoir; **b)** the Belá River at Liptovský Hrádok. The yellow triangles represent the nearest discharge gauges to the catchment outlets.

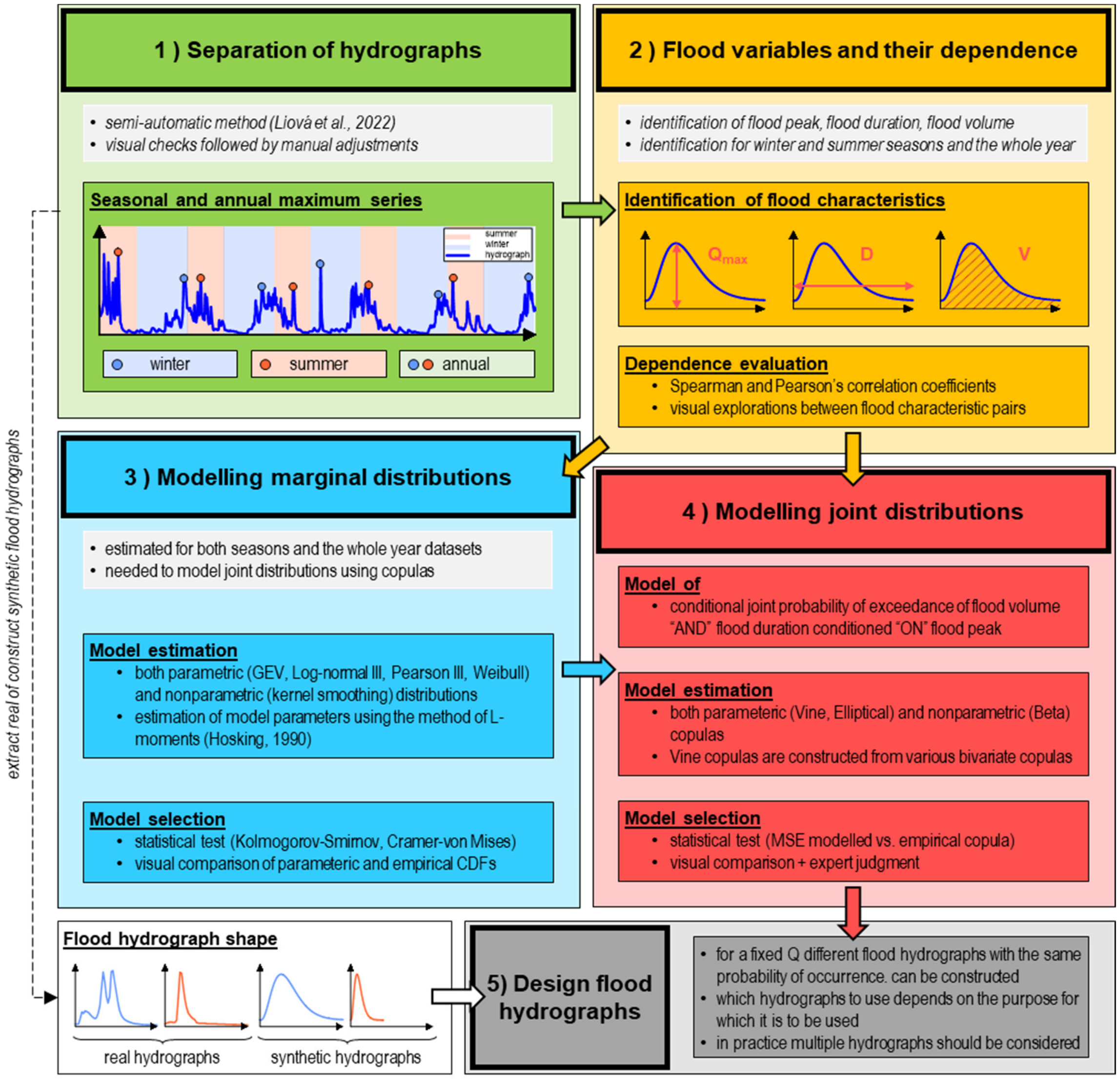

Fig. 2. Flowchart describing the main steps of the adopted procedure used to estimate flood hydrographs for the reservoir safety evaluations and flood risk assessments.

1. Separation of hydrographs

The identification and characterization of floods were performed for the annual and seasonal maximum discharges for both catchments. The maximum summer discharges are from June to October; the maximum winter discharges are from November to May. The flood hydrograph characteristics of the seasonal maximum and annual maximum series of the flood peaks (Q_{max}), flood durations (D) and flood volumes (V) were separated using a semi-automatic procedure that identifies the beginning and end of flood events, as characterized by the position of their peaks (see details in Liová et al. (2022)). The method then separates the baseflow using one of the pre-defined and frequently used approaches and, using a set of rules, tries to estimate the positions of the beginning and end of a given event. These rules and the parameters affecting their behavior were selected in such a way that the beginning of the event would be positioned at the point of an abrupt increase in the discharge departing from the baseflow. On the other hand, the end of the event is positioned at the point where the recession limb of the hydrograph approaches the baseflow. Numerous studies have already shown that it is not possible to define a set of rules that would be applicable to all types of catchment runoff (Giani et al., 2022; Oppel and Mewes, 2020; Thiesen et al., 2019). An important part of the method used was therefore composed of a manual control of every flood event. In this process the problematic events were consulted with the experts from SHMI and, if necessary, the positions of the beginning and end of the flood hydrographs were adjusted to comply with their understanding of flood-generating mechanisms and their experience as a body responsible for the preparation of flood hydrographs in practice. The method also enables accounting for auxiliary variables such as precipitation, which could be displayed together with the discharges and used as a guide to better position the beginning and the end of the flood hydrographs. The main reasons for the selection of this method for the identification of flood events were the extensive functionalities for the manual interventions in an otherwise automated process and the fact that this method has become a standard approach at SHMI. A total number of 33 separate flood events in each season were analysed.

2. *Flood characteristics and their dependence*

To quantitatively assess the dependence of flood characteristics (Q_{max}, V, and D) for the maximum annual and seasonal floods in the summer and winter seasons, Pearson's and Spearman's correlation coefficients with the associated *p*-values for significance tests were estimated.

3. *Modelling marginal distributions*

Models of marginal distributions were selected from non-parametric kernel smoothing (ks) class and parametric families with three parameters frequently used in hydrological applications, i.e., the Generalized Extreme Value (gev), Pearson (pe3), Weibull (wei), and Log-normal (ln3) distributions. Their parameters were estimated using the L-moment method (Hosking, 1990). The fitting of the distribution function for each flood characteristic was statistically tested for propriety by the Kolmogorov-Smirnov (KS) and Cramer-von Mises (CvM) tests at a significance level of 0.05, and the superimposition of parametric and empirical cumulative distribution functions allowed us to visually examine the validity of the estimated models. Additionally, the mean square error (MSE) between the observed and simulated flood characteristics was calculated to assess and compare the performance of the selected distribution functions.

4. *Modelling the joint distribution of the flood peaks, volumes and durations*

To model the joint distribution of the flood peaks, volumes and durations, a flexible copula-based model of the joint probability distribution among all three characteristics was built so that it could be used to estimate the probability of their simultaneous exceedance. They are denoted as the elements of the random vector $X = (X_1, X_2, X_3)$ arranged in the same order as the above, i.e., peaks, volumes, and durations. The copula approach constructs the joint distribution by decomposing it into marginal distributions of the X components and into a copula. While the marginals describe the individual stochastic behavior of the random vector elements, the copula captures their mutual relationships.

The following relations formally represent the decomposition in terms of the cumulative distribution functions (CDF) for three-dimensional random vectors:

$$F(x_1, x_2, x_3) = C[F_1(x_1), F_2(x_2), F_3(x_3)] \tag{1}$$

$$f(x_1, x_2, x_3) = c[F_1(x_1), F_2(x_2), F_3(x_3)] f_1(x_1) f_2(x_2) f_3(x_3) \tag{2}$$

where F is the joint CDF; secondly, F_1, F_2 and F_3 are the marginal CDFs, and finally, copula C is a CDF defined on the unit hypercube, i.e., C: $[0,1]^3 \rightarrow [0,1]$, having uniform marginals. Similarly, f is a joint PDF with marginal densities f_1, f_2, f_3 and the copula density c. For comprehensive details on the fundamentals of the copula theory, see Nelsen (2006).

There are several modelling methods for copulas that are flexible enough to describe this dependence in more than two dimensions. One group consists of non-parametric models based on an empirical copula, namely the Bernstein empirical copula or rather its convenient special case, the beta empirical copula, see Segers et al. (2017) for further details. Another group contains parametric classes such as the copulas of elliptically contoured distributions (normal and t-copula), vine copulas, and hierarchical Archimedean copulas (see Okhrin et al. (2017) for an introduction). In this study, a class of vine copulas was also selected due to their good interpretability. The underlying pair-copula construction approach utilises a graphic vine tool with bivariate copulas as building blocks to obtain a multivariate copula density through conditioning (see Czado and Nagler (2022) for an overview). Because the number of possible factorizations increases rapidly with dimensionality, it is advantageous to represent the conditional chain as vines. In the simplest three-dimensional case, 3! = 6 possible density factorization chains exist. If the first element of a random vector (the peak discharge) is selected to be conditioned upon and the third (the duration) to depend on the others, then the joint probability density is factored as follows:

$$\begin{aligned} f(x_1, x_2, x_3) &= f_1(x_1) \cdot f_{2|1}\ (x_1, x_2) \cdot f_{3|12}(x_1, x_2, x_3) \\ &= f_1(x_1) \cdot c_{12}[F_1(x_1), F_2(x_2)] \cdot f_2(x_2) \cdot \\ c_{32|1}\big[F_{3|1}(x_3, x_1), &F_{2|1}(x_2, x_1)\big] \cdot c_{13}[F_1(x_1), F_3(x_3)] \cdot f_3(x_3) \end{aligned} \tag{3}$$

where $f_{i|j}(x_i, x_j) = f(x_i, x_j) / f_j(x_j)$ is the conditional density function of X_i, given X_j, and copula density c_{ij} couples X_i and X_j, while $c_{ij|k}$ couples the bivariate conditional distributions of $X_i|X_k$ and $X_j|X_k$, where i, j, $k \in \{1,2,3\}$, $i \neq j \neq k \neq i$. Finally, $F_{i|j}(x_i, x_j) = \partial C_{ij}\big[F_i(x_i), F_j(x_j)\big] / \partial F_j(x_j)$ is a conditional CDF of X_i, given X_j.

Graphically, the construction from the bivariate copulas can be shown as a hierarchy of vine trees (see Figure 3, left), where in the first (topmost) tree, the nodes represent characteristics coupled by copulas – edges, which turn into nodes within the second tree. See also Schirmacher and Schirmacher (2008) for an explanation.

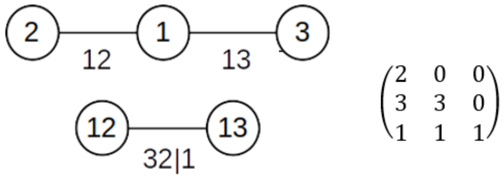

Fig. 3. Details of the trees of the vine copula (left) and the density factorization represented by a regular vine matrix (right).

The same density factorization as shown by trees can be formally represented by a regular vine matrix (Figure 3, right). The conditioning characteristic X_1 is coded by the last element in the first column, and the couplings are coded upward in all but the last column (left to right): 12, 32|1, 13; for details, see Dissmann et al. (2013).

As the building blocks of the vine copula, bivariate copulas from parametric classes often used in hydrological studies, such as the Archimedean class (the Clayton, Gumbel, Frank, Joe and BB1 to BB8 copula families with their survival counterparts), and copulas of elliptically contoured distributions (Gaussian and Student t-copula) were considered. Their selection and estimation are based on a sequential procedure using the AIC criterion and the maximum likelihood method (Dissmann et al., 2013).

The cumulative distribution function is defined as the probability of (simultaneous) non-exceedance, i.e.:

$$F_i\ (x) = Pr(X_i\ \leq\ x) \tag{4}$$

$$F_{ij}(x_i,\ x_j) = Pr(X_i\ \leq\ x_i\ \wedge\ X_j\ \leq\ x_j) \tag{5}$$

for $i, j \in \{1, 2, 3\}, i \neq j$, where $\wedge$ denotes the logical conjunction 'AND'. On the other hand, the survival function means (simultaneous) exceedance, i.e.:

$$\overline{F}_i(x) = Pr(X_i > x) \tag{6}$$

$$\overline{F}_{ij}(x_i, x_j) = Pr(X_i > x_i \wedge X_j > x_j) \tag{7}$$

However, there is generally not a simple conversion between the CDF and survival function, as shown below for the first two dimensions:

$$\overline{F}_i = 1 - F_i \tag{8}$$

$$\overline{F}_{ij} = 1 - F_i - F_j + F_{ij} \tag{9}$$

because the complement of CDF in a higher dimensional case is interpreted as a probability when either condition may hold, i.e., 1 – $F_{i,j}(x_i, x_j) = Pr(X_i \leq x_i \vee X_j \leq x_j)$, where $\vee$ stands for the logical conjunction 'OR'.

The formula for survival function in three dimensions is even more complicated and includes all the lower-dimensional marginals, i.e.,

$$\overline{F}_{ijk} = 1 - F_i - F_j - F_k + F_{ij} + F_{ik} + F_{jk} - F_{ijk} \tag{10}$$

for $i, j, k \in \{1, 2, 3\}$ and $i \neq j \neq k$.

After the joint distribution is built up, it is then straightforward to calculate the joint CDF of X_2, X_3 conditional on X_1, that is:

$$\begin{aligned}\overline{F}_{23|1}(x_2, x_3|x_1) &= \Pr(X_2 > x_2 \wedge X_3 > x_3 | X_1 = x_1) \\ &= \int_{x_3}^{\infty}\int_{x_2}^{\infty} f_{23|1}(s, t|x_1) dsdt \\ &= \int_{x_3}^{\infty}\int_{x_2}^{\infty} \frac{f(x_1, s, t)}{f_1(x_1)} dsdt \\ = \int_{x_3}^{\infty}\int_{x_2}^{\infty} c\,(F_1(x_1), F_2(s), F_3(t)) f_2(s) f_3(t)\, dsdt &\end{aligned} \tag{11}$$

although numerical integration is necessary, since the primitive function of the integrand is not available in a closed form.

In this study trivariate models based on the empirical nonparametric Bernstein (beta) copula and parametric copulas (Elliptical and R-vine copulas) for modeling the joint distribution of the flood peak, volume and duration were developed. The selection and estimation of the pair copulas (for the vine copula) was based on a sequential procedure using the AIC criterion and the maximum likelihood method. The validity of the estimated trivariate copula models was examined based on the goodness of fit using the MSE and visually on the Kendall (K) plots.

As mentioned earlier, the flood peak flow is considered to be the most significant variable for reservoir safety for a control flood hydrograph. Therefore, this study focuses on the event given by Equation (11), while the case is specifically denoted as V,D|Q_{max} here. The selection of the copula was determined by the setting of the flood peak (Q_{max}) as a root characteristic such that the conditioning of the flood volume (V)-flood duration (D) distribution by Q_{max} is implicitly included in the model. Using the copula-based approach, the isolines associated with the same joint conditional probabilities of exceedance of the flood volume and durations on the flood peak with the marginal probability of exceedance were constructed. The possibility for risk analysts is to work with an ensemble sampled according to the probability distribution. From there, they can choose a specific design realization or assess the reservoir safety based on the probabilities of exceedance curve or isoline.

All the calculations and visualizations were performed in R (R Core Team, 2022) with the help of the following packages: Vine Copula (Nagler et al., 2022), lmomco (Asquith, 2022), tidyverse (Wickham et al., 2019), metR (Campitelli, 2021) and cubature (Narasimhan et al. 2023).

5. Construction of synthetic flood hydrographs

The isolines constructed in the previous step enable us to specify the flood risk represented by any annual maximum hydrograph defined by the flood peak of the selected probability of exceedance and to determine the joint conditional probability of its exceeding the flood duration 'AND' volume. Based on this analysis, a set of annual or seasonal control synthetic flood hydrographs can be constructed with the design maximum discharge, the associated volume and duration with the selected probability, and the typical shape of the flood hydrograph.

RESULTS

Evaluation of the dependence between the flood characteristics

The pair-wise association among the flood characteristics (Q_{max}, V, and D) for the maximum annual and maximum seasonal floods in the summer and winter seasons as well as the strength of the dependency was estimated using the Pearson and Spearman correlations. The significance of the dependence was estimated using *p*-values with a threshold of $p < 0.05$. The values of the Pearson and Spearman correlations and *p*-values are summarized in Tables 1a, b for both catchments.

The dependence between the flood characteristics for the annual, summer and winter seasons in the Parná – Horné Orešany catchment is shown in Table 1a. The highest dependence between the flood characteristics is observed in the summer season, i.e., 0.76 (0.86) for the flood peak-volume pair, 0.13 (0.41) for the flood peak-duration pair and 0.55 (0.72) for the volume-duration pair. The first value represents an estimate of the Pearson coefficient; the second one (in parentheses) is that of Spearman. A slightly lower dependence was estimated in the winter season, i.e., 0.78 (0.74) for the flood peak-volume pair, 0.20 (0.21) for the flood peak-duration pair, and 0.66 (0.74) for the volume-duration pair. The lowest dependence was estimated for the annual maximum floods, i.e., 0.26 (0.44), –0.18 (–0.17), and 0.69 (0.70), respectively. In all the seasons, the highest dependence was estimated for the flood peak-volume pair or the volume-duration pair, and the smallest dependence was evident for the flood peak-duration pair.

The dependence between the flood characteristics for the floods in the Belá – Liptovský Hrádok catchment is illustrated in Table 1b. The highest dependence is again for the summer season, i.e., 0.67 (0.76) for the flood peak-volume pair, 0.057 (0.18) for the flood peak-duration pair, and 0.64 (0.66) for the volume-duration pair. A lower dependence was estimated between the flood characteristics in the winter season, i.e., 0.33 (0.25) for the flood peak-volume pair, –0.03 (–0.09) for the flood peak-duration pair, and 0.86 (0.78) for the volume-duration pair. The smallest dependence was estimated for the annual floods, i.e., 0.28 (0.34), –0.16 (–0.055), and 0.81 (0.77), respectively. The highest dependence was estimated again for the flood peak-volume pair, and the lowest dependence was evident for the flood peak-duration pair.

The low correlation in the flood peak-duration pairs was probably caused by the fact that the annual maximum flood series are composed of seasonal flood events of various origins.

Table 1. Testing the dependence between the flood characteristics, i.e., the flood peak (Q_{max}) - flood volume (V) pair, the flood peak (Q_{max}) - flood duration (D) pair, and the flood volume (V) - flood duration (D) pair using the Pearson and Spearman correlations (at a 5 % significance level) in the annual and seasonal floods in the summer and winter seasons; the bracketed values represent the *p*-values of the estimate: **a)** Parná – Horné Orešany catchment; **b)** Belá – Liptovský Hrádok catchment.

a)

Season/Dependence measure		Pearson's *r*	Spearman's *ρ*
Annual floods	Q_{max} - V	0.2629 (0.1394)	0.4408 (0.0102)
	Q_{max} - D	–0.1782 (0.3211)	–0.1695 (0.3458)
	V - D	0.6879 (0.0000)	0.6965 (0.0000)
Summer floods	Q_{max} - V	0.7616 (0.0000)	0.8643 (0.0000)
	Q_{max} - D	0.1309 (0.4679)	0.4108 (0.0175)
	V - D	0.5545 (0.0008)	0.7234 (0.0000)
Winter floods	Q_{max} - V	0.7806 (0.0000)	0.7443 (0.0000)
	Q_{max} - D	0.1965 (0.2732)	0.2092 (0.2426)
	V - D	0.6678 (0.0000)	0.7350 (0.0000)

b)

Season/Dependence measure		Pearson's *r*	Spearman's *ρ*
Annual floods	Q_{max} - V	0.2823 (0.1115)	0.3392 (0.0534)
	Q_{max} - D	–0.1623 (0.3669)	–0.0547 (0.7626)
	V - D	0.8124 (0.0000)	0.7716 (0.0000)
Summer floods	Q_{max} - V	0.6689 (0.0000)	0.7588 (0.0000)
	Q_{max} - D	0.0569 (0.7532)	0.1794 (0.3179)
	V - D	0.6404 (0.0001)	0.6599 (0.0000)
Winter floods	Q_{max} - V	0.3300 (0.0607)	0.2530 (0.1554)
	Q_{max} - D	–0.0309 (0.8644)	–0.0934 (0.6051)
	V - D	0.8587 (0.0000)	0.7803 (0.0000)

For that purpose, we have also analyzed the two seasons separately. Despite this seasonal analysis, a lower dependence was also estimated in the winter season for the flood peak-duration pairs in the Liptovský Hrádok station. This could have an origin in a combination of flood events of various origins, e.g., snowmelt and rain-on-snow flood events. But an analysis of flood discharge series from different origins was not the aim of this study, but should surely be considered in the future. Another reason for the low correlation could also be hidden in the separation method of the flood duration, which was estimated by a subjective method based on expert knowledge and the experience of people in the hydrometeorological service.

The flood dependence analysis showed that the highest dependence between the flood characteristics was identified for the maximum summer and winter floods and the lowest for the annual maximum floods. Therefore, only the maximum floods selected in the summer and winter seasons were chosen to evaluate the multivariate unconditional and conditional probabilities of exceedance in the next part of the study.

In conclusion this simple comparison clearly supports the need for a separate analysis of the annual maximum values in characteristic seasons. It is acknowledged that for a comprehensive flood regime analysis and especially for both univariate and multivariate flood frequency analyses, a larger differentiation of flood types could be of advantage, especially when envisaging independent and identically distributed (IID) flood variables.

Selection of the marginal probability distributions for the flood peak, volume and duration series

A comprehensive evaluation was performed on several distribution models that are frequently used in hydrological applications sensitive to extreme values of flood characteristics. These models were evaluated for both catchments in the summer and winter seasons. The validity of the estimated models was assessed by examining the goodness of fit and visually comparing the CDFs of the flood characteristics (with a focus on the upper bound), see Figure 4a–d, which helped in the selection of a final model. Table 2a, b presents the performance measures of the Kolmogorov-Smirnov (KS) and Cramer-von Mises (CvM) parametric distribution functions of the flood characteristics, which were identified as the most appropriate in the summer and winter seasons for both catchments. The results of the statistical tests indicate that all the selected marginal distributions meet the criteria of being acceptable at the 95% confidence interval.

Based on the results of the MSE, the appropriate distribution functions for modelling the joint distribution of the flood peak, volume, and duration were selected and are marked in bold in Table 3a, b.

Table 2. Statistical tests, with a 95% confidence interval estimate, for the selected probability models for fitting the marginal distributions for the flood peak (Q_{max}), volume (V) and duration (D) in the summer and winter seasons: **a)** Parná – Horné Orešany catchment; **b)** Belá – Liptovský Hrádok catchment.

a)

Seasons/ Characteristics/ Marginal distributions			Statistical tests	
			KS	CvM
Summer floods	Q_{max}	ln3	0.737	0.904
	V	wei	0.950	0.926
	D	gev	0.634	0.581
Winter floods	Q_{max}	wei	0.975	0.998
	V	wei	0.904	0.905
	D	pe3	0.998	0.996

b)

Seasons/ Characteristics/ Marginal distributions			Statistical tests	
			KS	CvM
Summer floods	Q_{max}	gev	0.992	0.998
	V	wei	0.993	0.972
	D	pe3	0.971	0.971
Winter floods	Q_{max}	gev	0.407	0.619
	V	pe3	0.994	0.999
	D	wei	0.742	0.836

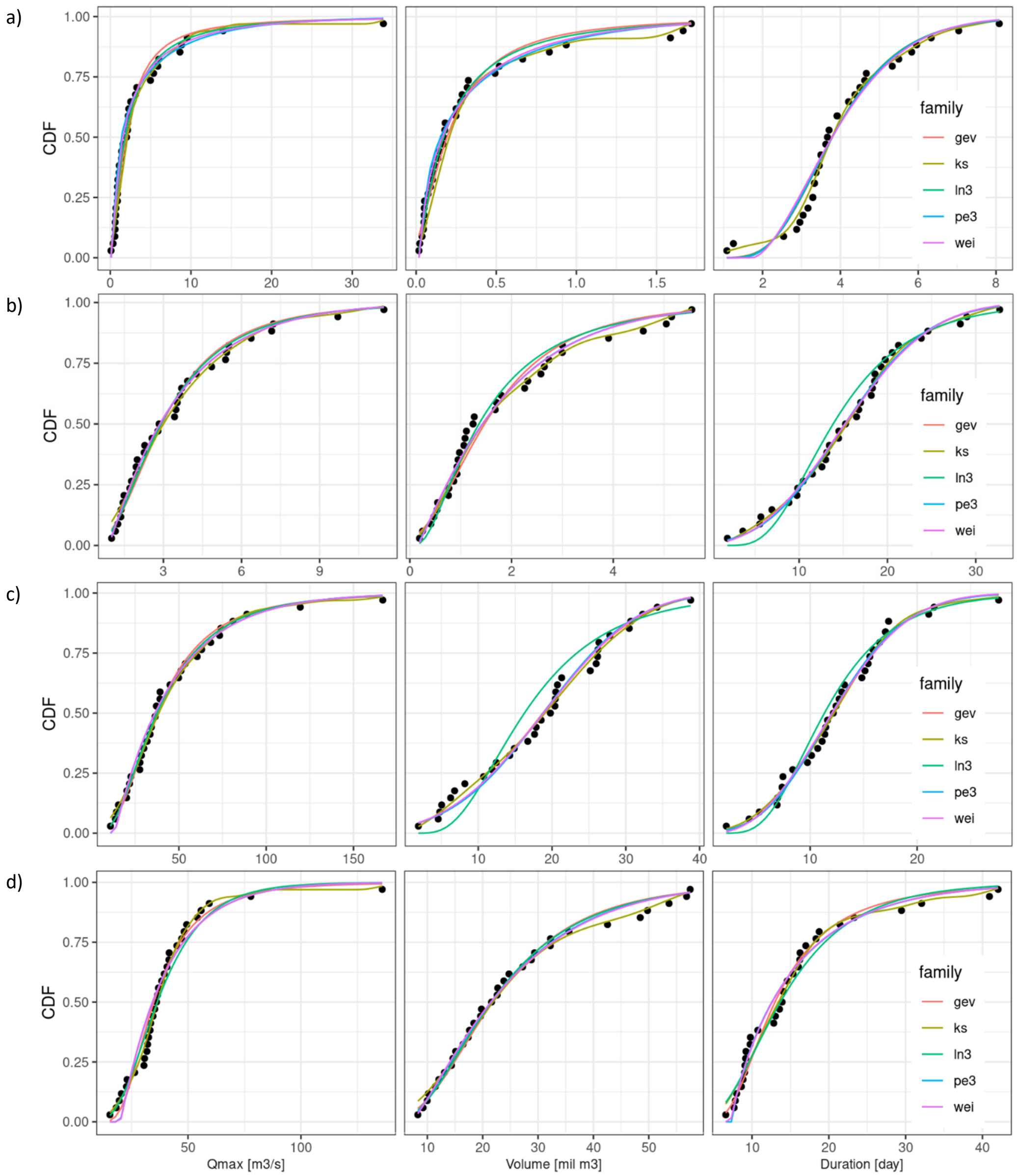

Fig. 4. Fitted marginal distributions for the flood peak (Q_{max}), volume and duration in the summer and winter seasons (from up to down) as CDFs of the selected types of distribution probability: **a, b)** Parná – Horné Orešany catchment; **c, d)** Belá – Liptovský Hrádok catchment.

Table 3. Performance of various probability models using the MSE for fitting marginal distributions for the flood peak (Q_{max}), volume (V) and duration (D) in the summer and winter seasons (selected distribution functions are marked in bold): **a)** Parná – Horné Orešany catchment; **b)** Belá – Liptovský Hrádok catchment. The selected marginal distributions were marked in bold.

	Seasons/Characteristics		Mean Square Error (MSE)				
			gev	ks	ln3	pe3	wei
a)	Summer floods	Q_{max}	0.00168	0.00317	**0.00114**	0.00441	0.00150
		V	0.00173	0.00379	0.00079	0.00189	**0.00109**
		D	**0.00255**	0.00102	0.00267	0.00293	0.00338
	Winter floods	Q_{max}	0.00136	0.00099	0.00099	0.00052	**0.00050**
		V	0.00219	0.00105	0.00178	0.00130	**0.00130**
		D	0.00058	0.00033	0.00357	**0.00054**	0.00061
b)	Summer floods	Q_{max}	**0.00058**	0.00060	0.00052	0.00111	0.00111
		V	0.00101	0.00071	0.00730	0.00109	**0.00097**
		D	0.00083	0.00067	0.00269	**0.00080**	0.00085
	Winter floods	Q_{max}	**0.00256**	0.00065	0.00230	0.00468	0.00475
		V	0.00077	0.00049	0.00057	**0.00045**	0.00045
		D	0.00190	0.00139	0.00219	0.00157	**0.00156**

Note: non-parametric kernel smoothing (ks) distribution and parametric families with three parameters: Generalized Extreme Value (gev), Pearson (pe3), Weibull (wei) and Log-normal (ln3) distributions.

Estimation of the copula model and modelling joint distributions of the flood variables

The Bernstein (beta) empirical copula, two elliptical Normal (nor) and Student's t (tco) copulas, and vine copulas (vi1, vi2, vi3) were chosen to model the trivariate flood characteristics. Model vi1 represents a type of vine copula that makes the joint modelling of flood volume and duration conditional on the flood peak the most transparent. Similarly, the types vi2 and vi3 set the other two variables (flood volume and duration, respectively) as conditioning. The best matching copula was examined visually, based on a goodness of fit Cramér–von Mises (CvM) statistical test, and the final copula was selected using the MSE.

The adequacy of the selected copula functions was evaluated using the CvM test applied at a 0.05 significance level. The selected statistical test could reject none of the copulas tested. To assess the graphic dependence between the flood variables (Q_{max}, V and D), the K-plots are shown (Figure 5a, b).

A fully positive dependence between the variables is when the events are on the x = y line (the straight line corresponds to a co-monotonicity copula). To the contrary, the upper black curve corresponds to independence. The black dots represent the empirical version. In this example, all the models are below the black curve, thus indicating a positive dependence. Most of them have been fitted in this study to check the performance; a class of vine copulas (namely type vin1) was finally chosen due to its good interpretability for the purposes of this study for which the CvM test results are presented (see Table 4a, b).

The copulas studied had values of the estimated MSE ranging from 0.000861 to 0.0015 for the summer (for the winter season from 0.0012 to 0.00322) in the Parná – Horné Orešany catchment. In the Belá – Liptovský Hrádok catchment, they ranged from 0.000769 to 0.00153 in the summer season (for the winter season, from 0.000711 to 0.001140). The graphic results are consistent with the conclusions drawn from the calculation of MSE.

For a better visual illustration, a set of random a size = 1000 samples was generated from the vine (vi1) copula and transformed back into their original units using corresponding marginal quantile functions and compared with the flood characteristics observed (Q_{max}, V and D) as shown in Figure 6a–d.

From these plots, it can be observed that the vine copula (vi1) with the models of marginals are performing satisfactorily, as the random pairs generated from this copula (gray dots) adequately overlap with the dependence pattern of the sample data (black dots).

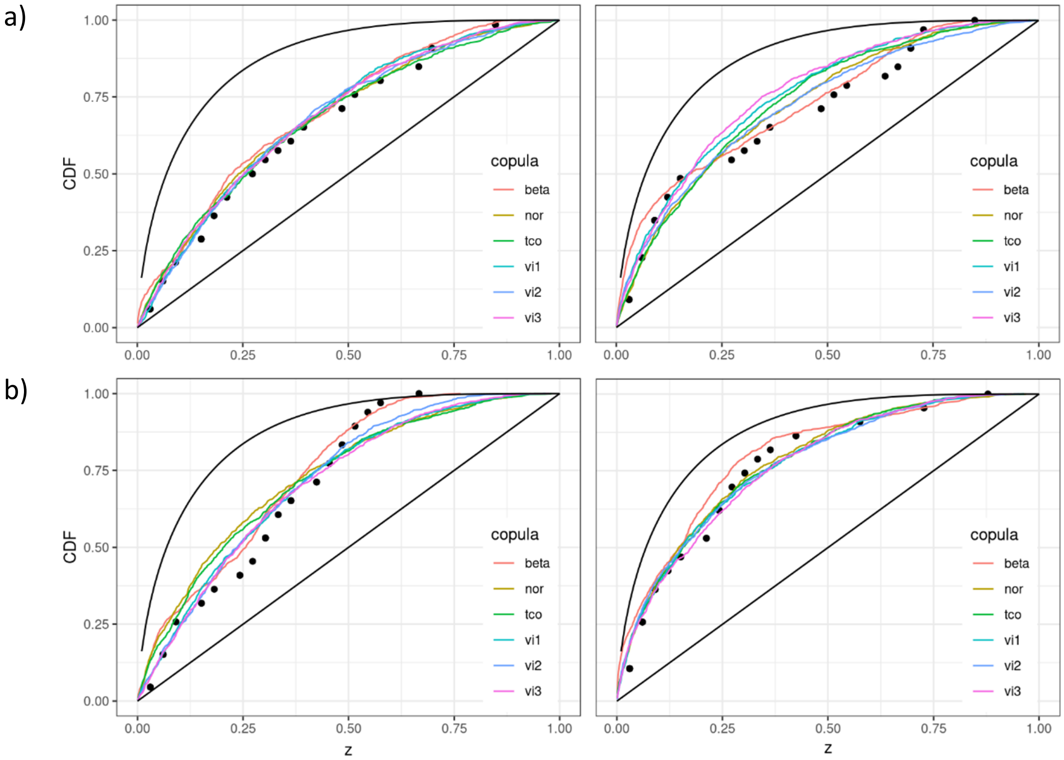

Fig. 5. The performance assessment of the copulas using the graphic dependence between the flood variables (K-Plots) in the summer and winter seasons (from left to right): **a)** Parná – Horné Orešany catchment; **b)** Belá – Liptovský Hrádok catchment.

Table 4. The Cramér–von Mises (CvM) test at a 0.05 significance level, for the selected class of vine copula (vin1) models in the summer and winter seasons; the bracketed values represent the *p*-values of the estimate: **a)** Parná – Horné Orešany catchment; **b)** Belá – Liptovský Hrádok catchment.

a)

Seasons	Statistical test
	CvM
Summer floods	0.5795 (0.7780)
Winter floods	0.3739 (0.8480)

b)

Seasons	Statistical test
	CvM
Summer floods	0.3752 (0.8120)
Winter floods	0.2070 (0.8380)

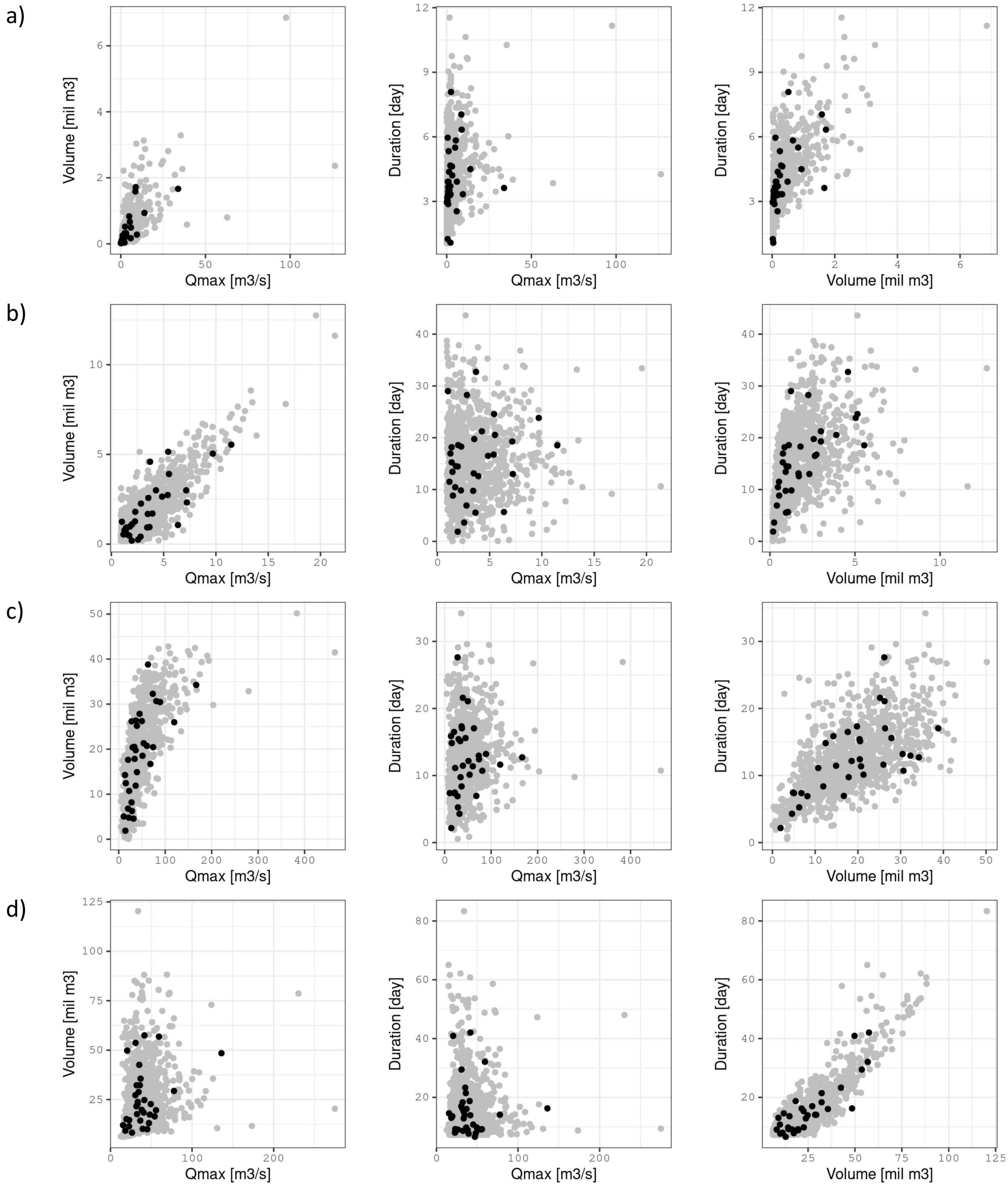

Fig. 6. Scatter plots of the observed vs. 1000 simulated samples of the flood variable pairs from the vine (vi1) copula in the summer and winter seasons (from up to down): **a, b)** Parná – Horné Orešany catchment; **c, d)** Belá – Liptovský Hrádok catchment. The black dots represent the observed samples, while the gray dots represent the samples generated.

Evaluation of the conditional probability of exceedance

The flood regime for the maximum summer and winter floods with the maximum flood peak in both catchments were evaluated. Table 5a, b presents the marginal probabilities of the exceedance for the flood peak Q_{max} ($\overline{F}_{Q_{100}}$)), volume V ($\overline{F}_V$), and duration D ($\overline{F}_D$). Additionally, the joint conditional probabilities of exceedance for the two flood variables (V and D), which were conditioned on the flood peak with an exceedance probability, viz., $\overline{F}_{V,D|Q_{max}}$, were computed using Equation (11). These probabilities are depicted for four selected flood events with the maximum flood peak in the summer and winter seasons at the Parná – Horné Orešany and Belá – Liptovský Hrádok catchments.

It can be seen that flood peak (Q_{max}) of the maximum summer and winter floods in both catchments have the lowest marginal probability of exceedance of 0.01, 0.018, 0.014, and 0.006. The higher probabilities of exceedance can be seen for flood volume (V), with values of 0.033, 0.037, 0.058 and 0.082. The highest marginal probabilities of exceedance are evident for the flood duration (D) and also for the conditional probability of exceedance for volume (V) and duration (D), which are conditioned on the flood peak (Q_{max}), i.e., 0.466, 0.325, 0.484 and 0.124.

Next, the vine copula approach was employed to construct isolines representing the same joint conditional probabilities of the exceedance of the flood volume and flood durations, given the flood peak, with a marginal probability of exceedance of 0.01 (corresponding to a 100-year flood), viz. $\overline{F}_{V,D|Q_{100}}$ for the summer and winter seasons in both catchments was constructed (see Figure 7a, b). The constructed isolines enable specifying the flood risk represented by the annual seasonal maximum flood peak and determining its joint conditional probability of the exceedance of the flood duration and volume. The designer or flood risk analytics may also be equipped with a set of design flood hydrographs with diverse shapes, volumes, and durations for a selected design discharge (in this study, 100-year flood peak) with the same joint conditional probability of the exceedance for the flood risk analysis. The available variations are depicted through over 400 randomly generated (gray) dots, corresponding to pairs representing flood volumes and durations using a bivariate distribution (see Figure 7a, b).

For a sample example to compare with historical floods, historical flood events were selected with the following flood characteristics at the Parná – Horné Orešany catchment (historical flood from 07.06.2011): annual maximum summer flood peak, $Q_{max} = 33.8$ m^3/s; flood volume, V = 1.67 mil. m^3; and flood duration, D = 3.60 days. Similarly, at the Belá – Liptovský Hrádok catchment, the flood characteristics are Q_{max} = 183.8 m^3/s; V = 35.75 mil. m^3; and D = 12.71 days (historical flood from 29.06.1958). The historical (black dots) and synthetic (red dots) flood hydrographs represented by their variables (V and D) in the summer season for both catchments were created based on the assumption that the maximum peak flow of each hydrograph had a marginal probability of exceedance of 0.01 (see Figure 7a, b). From the marginal distribution of the peak flows, the design discharge with the marginal probability of exceedance of 0.01 (corresponding to a 100-year flood) was estimated as 33.3 m^3/s in the summer season for the Parná – Horné Orešany catchment and 183.3 m^3/s for the Belá – Liptovský Hrádok catchment. For the construction of the design synthetic flood hydrographs, it was necessary to determine the shape of the flood in addition to the peak, volume, and duration of the flood. The shape of the design synthetic flood hydrographs was derived from the fragments of the observed flood events in the dataset (the hydrographs were centered on the peak position), which have been simplified to maintain the monotonicity of their rising and falling limbs (unimodal to multimodal hydrographs). When design synthetic flood hydrographs are constructed, the important parameter is the percentile, which affects the shape of the representative hydrograph (see Liová et al., 2022). The task of this study was not to deal with the shape of the flood hydrographs, which would allow a designer to choose the appropriate critical shape of the flood design for the given purpose. In this study, the 50% shape percentile for a representative hydrograph was used so that the design of the synthetic hydrographs reproduces the properties and variability of the original shapes of the observed hydrographs. For a comparison with historical floods, one variant of synthetic hydrograph was selected from the generated data (gray dots), representing the joint conditional probabilities of exceedance of the flood volume and duration of 0.5, conditioned on the flood peak with an exceedance probability of 0.01 (corresponding to a 100-year flood) but with a different flood volume and duration for the summer season in both catchments (see Table 6, Figures 7a, b, and 8a, b).

Table 5. Marginal probability of the exceedance for the flood peak (Q_{max}), volume (V), and duration (D), and the conditional probability of exceedance for volume (V) and duration (D), conditioned on the flood peak (Q_{max}) for the selected flood events with the maximum flood peak: a) Parná – Horné Orešany catchment; b) Belá – Liptovský Hrádok catchment.

a)	Seasons/Characteristics	Q_{max}	V	D	$\overline{F}_{Q_{max}}$	$\overline{F}_V$	$\overline{F}_D$	$\overline{F}_{V,D\|Q_{max}}$
		[m^3/s]	[mil. m^3]	[day]				
	Summer floods	33.86	1.67	6.30	0.010	0.033	0.552	0.466
	Winter floods	11.47	5.54	18.54	0.018	0.037	0.328	0.325
b)	Summer floods	166.8	34.26	12.71	0.014	0.058	0.451	0.484
	Winter floods	136.0	48.46	16.25	0.006	0.082	0.331	0.124

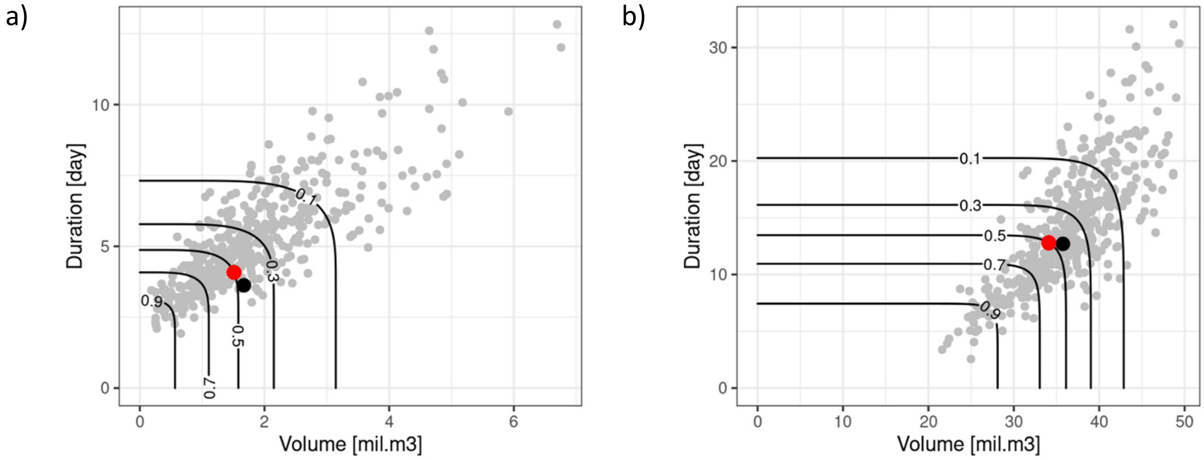

Fig. 7. The isolines of the joint conditional probability of exceedance of the flood volume and duration, conditioned on the flood peak with an exceedance probability of 0.01 (corresponding to a 100-year flood) in the summer season: **a)** Parná – Horné Orešany catchment; **b)** Belá – Liptovský Hrádok catchment. The black dots represent the historical flood hydrographs, while the red dots represent selected variant of synthetic hydrographs. The gray dots represent over 400 pairs randomly generated from the bivariate conditional distribution.

Table 6. The basic flood characteristics of the selected synthetic design and historical flood hydrographs with a joint conditional probability of the exceedance of volumes and durations (approximately 0.5), conditioned on the approximate 100-year flood peak (exceedance probability of 0.01) in the summer season: a) Parná – Horné Orešany catchment; b) Belá – Liptovský Hrádok catchment.

	Hydrograph/Characteristics	Summer floods		
		Q_{100}	V	D
		[m^3/s]	[mil. m^3]	[day]
a)	Historical (07.06.2011)	33.8	1.670	3.60
	Synthetic	33.3	1.503	4.12
b)	Historical (29.06.1958)	183.8	35.75	12.71
	Synthetic	183.3	34.02	12.80

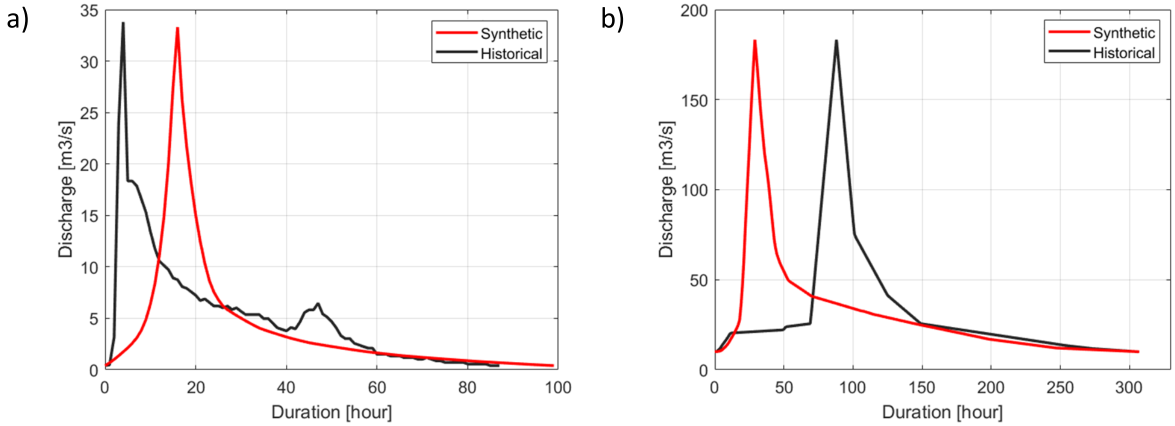

Fig. 8. Examples of the selected design synthetic (red) and historical flood hydrographs (black) with a joint conditional probability of the exceedance of volumes and durations (approximately 0.5), conditioned on the approximate 100-year flood peak (exceedance probability of 0.01) in the summer season: **a)** Parná – Horné Orešany catchment; **b)** Belá – Liptovský Hrádok catchment.

Figures 8a, b show that the flood hydrographs have approximately the same 100-year peak and the same joint conditional probability of exceedance of the flood volume and duration. However, not only their shape but also the flood volume and duration may vary, which is crucial to consider in reservoir safety and flood risks. The chosen variant of the joint conditional probability of exceedance of the flood volume and duration for the control task is the choice of the designer or risk analyst. This information can also be helpful for the design of flood control structures, reservoirs, spillways etc., where the design of flood hydrographs is sought.

DISCUSSION

The types of sources of uncertainties rooted in local and regional hydrological analyses of flood hydrograph properties should be discussed more in the literature. Zhang et al. (2021) and Brunner (2023) provide comprehensive overviews. Here we mainly focused on such issues as how to define a flood's duration. It is recognised that the allocation of the starting and ending points of a flood event on the hydrograph and base flow separation are the keys to characterising the duration of a flood (Pandi, 2010; Zhang et al., 2021). Additionally, the flood volumes need to be delineated using baseflow separation methods. A baseflow primarily originating from groundwater is a critical streamflow component, although reliably estimating it is loaded with significant difficulties. Furthermore, an overabundance of approaches exists ranging from recession curve analyses to digital filters and conceptual rainfall-runoff models (see, e.g. Zhang et al., 2021; Cheng et al., 2022), which ideally need to be selected and adapted according to local runoff generation conditions and also the particular flood types analysed. Therefore, both flood hydrograph identifications and flood characteristic extractions are often preferred to be performed manually (Zhang et al., 2021). When selecting and using an approach for this study, we concluded that due to the complexity of the flood event generation and flood hydrograph typologies, it is difficult to address the problem without human intervention, especially in the case of multiple flood events over a more extended period, rain on snow events, and prolonged snowmelt (Gaál et al., 2015).

Therefore, a subjective widespread and often applied conceptual graphic estimation method was adopted. At the beginning of a flood, a sudden increase in the hydrograph, and at the end of the phenomenon, the flattening of the flow recession (the return to baseflow), was taken. We acknowledge the subjective decisions involved in the method. Therefore, the decision also considered the causative rainfall as a supporting variable. However, the air temperature data for snowmelt floods were not used. Simplifying approaches and also diverse algorithmic conceptual models have also been used in similar studies (e.g., Brunner et al., 2019; Gaál et al., 2015; Mediero et al., 2010).

In conclusion, the selection of the flood volume for an analysis of dependence needs to be done with respect to the objectives of the application for which it is needed. The delineation of the flood duration has more influence on the outcomes of the analysis. The causative drivers (rainfall, snowmelt or both) may guide the analyst to fix the duration, and a runoff generation model may also help to arrive at a plausible solution. One must see that when prolonging the duration of a falling limb, the volume values do not change significantly. However, the probabilistic properties of the triplet peak flow, flood volume, and flood duration may change. In this respect, exploring an upper bound for the distribution function of the durations should not be excluded.

In any case, the outcome is reflected in the dependence structure, and this uncertainty would need to be considered when generalizations (e.g., joint or conditional probabilities) are drawn concerning the local or regional flood regime or in a risk analysis.

In risk analyses and risk-based decision-making, it is essential to quantify the uncertainties associated with separately and jointly estimating the respective flood characteristics. Regarding the methodology of the multivariate analysis, several recommendations exist that can provide end-users with the necessary requirements, statistical assumptions, and consequential limitations of conducting a multivariate frequency analysis with copulas (e.g., Genest and Favre, 2007; Gräler et al., 2013). Even when following these recommendations, uncertainty concerning the model can be introduced because several copula models could not be rejected in the two catchments as expected (see, e.g., Szolgay et al., 2015). Sampling uncertainty was also introduced because the copula parameters could only be estimated based on the samples since the parent population was unknown (Brunner et al., 2017). This study has not addressed these problems in detail because, in this preliminary flood regime analysis, a pragmatic straightforward approach was followed, which could be acceptable for practical applications. This approach was followed, where the recommended statistical tests and MSE value was used in all the cases to select an appropriate model for practical reasons.

The choice of the probability of exceedance of flood characteristics in the case of conditional probability was discussed, e.g., in Brunner et al. (2016b), since several joint events have the same probability of exceedance and thus lie on an isoline. Several options were proposed in the literature as to how to choose the most likely value of these for a practical application (see a short review in Brunner et al. (2016a)). Such a choice, however, introduces uncertainty because one could choose another pair of design characteristics on the isoline instead of the one with the highest likelihood.

However, it is acknowledged that an uncertainty assessment framework as guidance for explicit estimates of uncertainties needs to be added to the analysis in the future. For a practical application, the introduction of national engineering standards could also help fix rules and procedures to overcome some crucial aspects of the problem (Blöschl and Merz, 2008; DWA 2012; LfU BW, 2005; Lorenz et al., 2011).

CONCLUSIONS

In risk analysis and risk-based decision-making for flood hazard mitigation, it is essential to quantify the exceedance probabilities (and uncertainties) associated with the individual flood characteristics. Besides the established univariate framework, recent advances in multivariate frequency analysis have opened up new perspectives and challenges. In this study, we have not focused on the problems associated with the multivariate statistical modelling apparatus itself in detail. Instead, we aimed to explore their potential in flood hydrograph and flood regime analysis. An aspect of the evaluation of the design flood hydrograph was touched on. In the multivariate analysis itself, a pragmatic straightforward approach was followed, which could be acceptable for practical applications when included in national engineering standards. However, an extended uncertainty assessment framework as guidance for explicit estimates of uncertainties would need to be added to the analysis. We have avoided using return periods, which, while still a central notion in practice, may be considered a concept rooted in past empirical univariate flood frequency analysis. Instead, we preferred using probabilities that do not refer to a temporal measure but focus on the chance of occurrence over a long-time perspective alone. We also recommend continuing risk evaluations along this line in practical applications.

This paper outlined a framework for assessing the hydrological sources of uncertainties associated with using multivariate frequency analysis. Our results indicate that the hydrological issues limit the successful application of multivariate analysis tools in comparative hydrology more than the tools themselves. Several drawbacks occurred concerning the hydrological problems of conducting a comparative multivariate frequency analysis. First, the definition of a flood hydrograph and, the extraction of its parameters (especially the most frequent volume and duration) in the current practice, cannot be considered as sufficiently addressing several questions. These, however, have directly influenced the outcomes of the multivariate analysis and the interpretation of the flood regime based on them. Even when an analysis of hydrographs is based on flood types, the duration of events with one large and multiple minor peaks cannot be uniquely determined by classical recession or digital filtering analysis. Multi-peak events such as floods with minor "aftershock" precipitation belonging to the same meteorological frontal driver and prolonged snowmelt flood caused by the same warmer air mass may have a longer out-phasing component, which prolongs the duration without significantly influencing the flood volume. The question of an upper bound for durations in a given physiographic setting is also still open (but is certainly not unrealistic in nature).

The volume and duration of a flood hydrograph are also a question to decide when conducting multivariate flood analysis. The choice may represent a non-negligible source of uncertainty in the outcomes of a multivariate analysis in specific physiographic settings, thereby directly influencing the outcomes and interpretations. It is difficult to suggest an appropriate approach from a hydrological perspective due to the vast set of potential analytical exercises. The best approach is related to the problem envisaged. The decision is more manageable when conducting a design exercise (e.g., when the hydrograph volume for designing a reservoir for flood regulation is needed, the baseflow should be a part of it).

Respecting the need for subjective expert-based interactions when dealing with the above-mentioned problems could bring a partial solution. However, the need for a more objective approach, such as a process analysis using rainfall-runoff modelling (e.g., Földes et al., 2022; Hu et al., 2021; Pekárová et al., 2021), is apparent. From a longer perspective, starting to differentiate flood generation processes within a flood type (e.g., long or short rain floods) may lead to a more advanced process typology, which must be respected when grouping IID events for analysis.

Studying floods in a multivariate frequency framework is also challenging because observational records with adequate short-time steps still need to become more available. This also holds for design hydrograph construction and multivariate evaluations, since synthetic design hydrographs are inherently uncertain due to the limited record lengths and sampling uncertainties for characteristic shapes and types.

The hydrological uncertainties in comparative multivariate hydrological analysis and the probabilistic assessment of synthetic design hydrographs could be most effectively reduced by enlarging the sample size and considering additional information such as historical floods and extending records by rainfall-runoff modelling. The above-mentioned uncertainties must be respected and communicated.

Acknowledgements. This work was supported by the Slovak Research and Development Agency under Contract No. APVV 19-0340, No. APVV 20-0374 and the VEGA Grant Agency No. VEGA 1/0577/23. The participation of Zoltán Gribovszki was supported by the OTKA Grant No. 143972SNN.

REFERENCES

Asquith, W., 2022. lmomco---L-moments, censored L-moments, trimmed L-moments, L-comoments, and many distributions. R package version 2.4.7.

Blöschl, G., Merz, R., 2008. Bestimmung von Bemessungshochwässern gegebener Jährlichkeit–Aspekte einer zeitgemäßen Strategie. Wasserwirtschaft, 11, 12–18.

Brunner, M., Seibert, J., Favre, A., 2016a. Bivariate return periods and their importance for flood peak and volume estimation. WIREs Water, 3, 819–833. https://doi.org/10.1002/wat2.1173

Brunner, M., Vannier, O., Favre, A., Viviroli, D., Meylan, P., Sikorska, A., Seibert, J., 2016b. Flood volume estimation in Switzerland using synthetic design hydrographs - a multivariate statistical approach. In: Proc. 13th Congress INTERPRAEVENT 2016, Luzern, pp. 468–476. https://doi.org/10.5167/uzh-124430

Brunner, M., Viviroli, D., Sikorska, A., Vannier, O., Favre, A., Seibert, J., 2017. Flood type specific construction of synthetic design hydrographs. Water Resources Research, 53, 1390–1406. https://doi.org/10.1002/2016WR019535

Brunner, M., Furrer, R., Favre, A., 2019. Modeling the spatial dependence of floods using the Fisher copula. Hydrology and Earth System Sciences, 23, 107–124. https://doi.org/10.5194/hess-23-107-2019

Brunner, M., 2023. Floods and droughts: a multivariate perspective on hazard estimation: a multivariate perspective on hazard estimation. Hydrology and Earth System Sciences Discussions, 1–26. https://doi.org/https://doi.org/10.5194/hess-2023-20

Campitelli, E., 2021. metR: tools for easier analysis of meteorological fields: tools for easier analysis of meteorological fields. R Package version 0.13.0. https://doi.org/10.5281/zenodo.2593516

Carril-Rojas, D., Mediero, L., 2023. Bivariate analysis with synthetic hydrograph shapes for hydrological dam safety assessment. Environ. Sci. Proc., 25, 2.

Czado, C., Nagler, T., 2022. Vine copula based modeling. Annual Review of Statistics and Its Application, 9, 453–477. https://doi.org/10.1146/annurev-statistics-040220-101153

Danáčová, Z., Poórová, J., Blaškovičová, L., Liová, S., 2015. Instrumentation for surface water quantity monitoring and discharge measurements by ADCP. Acta Hydrologica Slovaca, 16, Thematic issue, 3–12.

Dissmann, J., Brechmann, E., Czado, C., Kurowicka, D., 2013. Selecting and estimating regular vine copulae and application to financial returns. Computational Statistics & Data Analysis, 59, 52–69. https://doi.org/10.1016/j.csda.2012.08.010

Drobot, R., Draghia, A., Ciuiu, D., Trandafir, R., 2021. Design floods considering the epistemic uncertainty. Water, 13. https://doi.org/10.3390/w13111601

DWA, 2012. Merkblatt DWA-M 552: Ermittlung von Hochwasserwahrscheinlichkeiten. Deutsche Vereinigung für Wasserwirtschaft, Abwasser und Abfall e. V.

Földes, G., Labat, M., Kohnová, S., Hlavčová, K., 2022. Impact of changes in short-term rainfall on design floods: Case study of the Hnilec River Basin, Slovakia. Slovak Journal of Civil Engineering, 30, 68–74. https://doi.org/10.2478/sjce-2022-0008

Gaál, L., Szolgay, J., Kohnová, S., Hlavčová, 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, 968–984. https://doi.org/10.1080/02626667.2014.951361

Gadek, W., Baziak, B., Tokarczyk, T., Szalińska, W., 2022. A novel method of design flood hydrographs estimation for flood hazard mapping. Water, 14, 12, 1856. https://doi.org/10.3390/w14121856

Ganapathy, A., Hannah, D., Agarwal, A., 2022. Flood classification based on hydrograph characteristics. Authorea Preprints.

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

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

Giani, G., Tarasova, L., Woods, R.A., Rico-Ramirez, M.A., 2022. An objective time-series-analysis method for rainfall-runoff event identification. Water Resources Research, 58, 2, e2021WR031283.

Gómez, M., Ausín, M., Domínguez, M., 2018. Vine copula models for predicting water flow discharge at King George Island, Antarctica. Stochastic Environmental Research and Risk Assessment, 32, 2787–2807. https://doi.org/10.1007/s00477-018-1599-9

Gräler, B., van den Berg, M., Vandenberghe, S., Petroselli, A., Grimaldi, S., De Baets, B., Verhoest, N., 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. https://doi.org/10.5194/hess-17-1281-2013

Grimaldi, S., Petroselli, A., Salvadori, G., De Michele, C., 2016. Catchment compatibility via copulas: A non-parametric study of the dependence structures of hydrological responses. Advances in Water Resources, 90, 116–133. https://doi.org/10.1016/j.advwatres.2016.02.003

Größer, J., Okhrin, O., 2022. Copulae: An overview and recent developments. WIREs Computational Statistics, 14. https://doi.org/10.1002/wics.1557

Hosking, J., 1990. L-moments: Analysis and estimation of distributions using linear combinations of order statistics. Journal of the Royal Statistical Society: Series B (Methodological), 52, 105–124. https://doi.org/10.1111/j.2517-6161.1990.tb01775.x

Hu, C., Ran, G., Li, G., Yu, Y., Wu, Q., Yan, D., Jian, S., 2021. The effects of rainfall characteristics and land use and cover change on runoff in the Yellow River basin, China. Journal of Hydrology and Hydromechanics, 69, 1, 29–40. https://doi.org/10.2478/johh-2020-0042

Hundecha, Y., Parajka, J., Viglione, A., 2017. Flood type classification and assessment of their past changes across Europe. Hydrology and Earth System Sciences Discussions, 1–29. https://doi.org/https://doi.org/10.5194/hess-2017-356

Cheng, S., Tong, X., Illman, W., 2022. Evaluation of baseflow separation methods with real and synthetic streamflow data from a watershed. Journal of Hydrology, 613, Part A, 128279. https://doi.org/10.1016/j.jhydrol.2022.128279

Jafry, N., Suhaila, J., Yusof, F., Mohd Nor, S., Alias, N., 2022. Preliminary study on flood frequency analysis in Johor River basin using vine copula. Proceedings of Science and Mathematics, 7, 52–55.

Latif, S., Simonovic, S., 2022. Parametric vine copula framework in the trivariate probability analysis of compound flooding events. Water, 14, 14, 2214. https://doi.org/10.3390/w14142214

Le Clerc, S., Sauquet, E., Lang, M., 2003. Scaling properties of flood hydrographs and their use to derive design flood hydrographs. WIT Transactions on Ecology and the Environment, 60.

LfU BW, 2005. Festlegung des Bemessungshochwassers für Anlagen des technischen Hochwasserschutzes. Leitfaden. Karlsruhe, 92 p.

Liová, A., Valent, P., Hlavčová, K., Kohnová, S., Bacigál, T., Szolgay, J., 2022. A methodology for the estimation of control flood wave hydrographs for the Horné Orešany reservoir. Acta Hydrologica Slovaca, 23, 52–61. https://doi.org/10.31577/ahs-2022-0023.01.0006

Lorenz, P., Gattermayr, W., Kölbl, C., Krammer, C., Maracek, K., Mathis, C., Moser, J., Schatzl, R., Wiesenegger, H., Wimmer, M., Lorenz, P. (Eds.), 2011. Leitfaden: Verfahren zur Abschätzung von Hochwasserkennwerten, 113 p.

Mediero, L., Jiménez-Álvarez, A., Garrote, L., 2010. Design flood hydrographs from the relationship between flood peak and volume. Hydrology and Earth System Sciences, 14, 2495–2505. https://doi.org/10.5194/hess-14-2495-2010

Nagler, T., Schepsmeier, U., Stoeber, J., Brechmann, E., Graeler, B., Erhardt, T., 2022. VineCopula: Statistical Inference of VinCopulas. R package version 2.4.4.

Narasimhan, B., Johnson, S., Hahn, T., Bouvier, A., Kiêu, K., 2023. cubature: Adaptive Multivariate Integration over Hypercubes. R package version 2.0.4.6.

Nazeri Tahroudi, M., Ramezani, Y., De Michele, C., Mirabbasi, R., 2022. Trivariate joint frequency analysis of water resources deficiency signatures using vine copulas. Applied Water Science, 12, 67. https://doi.org/10.1007/s13201-022-01589-4

Nelsen, R., 2006. An Introduction to Copulas. 2nd ed. Springer, New York.

O'Connor, K., Goswami, M., Faulkner, D., 2014. Flood Studies Update, Technical Research Report: Vol. III - Hydrograph Analysis.

Okhrin, O., Ristig, A., Xu, Y. F., 2017. Copulae in high dimensions: an introduction. Applied Quantitative Finance, 247–277. https://doi.org/10.1007/978-3-662-54486-0_19

Oppel, H., Mewes, B., 2020. On the automation of flood event separation from continuous time series. Frontiers in Water, 2, 18. https://doi.org/10.3389/frwa.2020.00018

Pandi, G., 2010. The analysis of flood waves. Aerul si Apa. Componente ale Mediului, 35–44.

Pramanik, N., Panda, R., Sen, D., 2009. Development of design flood hydrographs using probability density functions. Hydrological Processes, 24, 415–428. https://doi.org/10.1002/hyp.7494

Pekárová, P., Mészáros, J., Miklánek, P., Pekár, J., Siman, C., Podolinská, J., 2021. Post-flood field investigation of the June 2020 flash flood in the upper Muráň River basin and the catastrophic flash flood scenario. Journal of Hydrology and Hydromechanics, 69, 3, 288–299. https://doi.org/10.2478/johh-2021-0015

R Core Team, 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing.

Requena, A., Mediero, L., Garrote, L., 2013. A bivariate return period based on copulas for hydrologic dam design: accounting for reservoir routing in risk estimation. Hydrology and Earth System Sciences, 17, 3023–3038. https://doi.org/10.5194/hess-17-3023-2013

Rizwan, M., Guo, S., Yin, J., Xiong, F., 2019. Deriving design flood hydrographs based on copula function: A case study in Pakistan. Water, 11. https://doi.org/10.3390/w11081531

Segers, J., Sibuya, M., Tsukahara, H., 2017. The Empirical Beta Copula. Journal of Multivariate Analysis, 155, 35–51. https://doi.org/10.1016/j.jmva.2016.11.010

Schirmacher, D., Schirmacher, E., 2008. Multivariate dependence modeling using pair-copulas. Technical report.

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 the International Association of Hydrological Sciences, 370, 183–188. https://doi.org/10.5194/piahs-370-183-2015

Škvarka, J., Bednárová, E., Miščík, M., Uhorščák, Ľ., 2021. The Domaša reservoir in the spectrum of climate change. Slovak Journal of Civil Engineering, 29, 9–15. https://doi.org/10.2478/sjce-2021-0009

Thiesen, S., Darscheid, P., Ehret, U., 2019. Identifying rainfall-runoff events in discharge time series: a data-driven method based on information theory. Hydrology and Earth System Sciences, 23, 2, 1015–1034. https://doi.org/10.5194/hess-23-1015-2019

Tootoonchi, F., Sadegh, M., Haerter, J., Räty, O., Grabs, T., Teutschbein, C., 2022. Copulas for hydroclimatic analysis: A practice-oriented overview. WIREs Water, 9, 2, e1579. https://doi.org/10.1002/wat2.1579

Tosunoglu, F., Gürbüz, F., İspirli, M., 2020. Multivariate modeling of flood characteristics using Vine copulas. Environmental Earth Sciences, 79, 459. https://doi.org/10.1007/s12665-020-09199-6

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019. Welcome to the Tidyverse. Journal of open source software, 4, 43, 1686. https://doi.org/10.21105/joss.01686

Xiao, Y., Guo, S., Liu, P., Yan, B., Chen, L., 2009. Design flood hydrograph based on multicharacteristic synthesis index method. Journal of Hydrologic Engineering, 14, 1359–1364.

Yue, S., Ouarda, T., Bobée, B., Legendre, P., Bruneau, P., 2002. Approach for describing statistical properties of flood hydrograph. Journal of Hydrologic Engineering, 7, 147–153.

Zhang, Q., Zhang, L., She, D., Wang, S., Wang, G., Zeng, S., 2021. Automatic procedure for selecting flood events and identifying flood characteristics from daily streamflow data. Environmental Modelling & Software, 145, 105180. https://doi.org/10.1016/j.envsoft.2021.105180

Received 3 March 2023
Accepted 16 August 2023