An assessment of the uncertainty of the extremity of flood waves with vine copulas

Roman VÝLETA[*], Kamila HLAVČOVÁ, Silvia KOHNOVÁ, Tomáš BACIGÁL, Anna LIOVÁ, Ján SZOLGAY

Flood hazard connected with the failure of hydraulic structures and flood risk associated with areas subject to flooding need to be estimated using flood hydrographs. Multivariate statistics of flood wave parameters enable quantitative conclusions about such flood hazards and risks. This study focuses on uncertainties in their estimates using unbounded and bounded marginal distributions of flood durations in the joint modelling of flood peaks, volumes, and durations with vine copulas. We have respected the seasonality of floods by distinguishing between the durations of summer and winter floods. We propose to use the bounded Johnson's SB distribution to represent the hydrological constraints associated with flood durations. The practical consequences of selecting various unbounded and bounded distributions for modelling flood durations for the joint overall and conditional probabilities of the exceedance of flood peaks, volumes, and durations were demonstrated on data from the Parná River in Slovakia. Differences in modelling joint probabilities due to the tail behaviours of the marginal distributions tested were found. Although these are not critical for practical applications, accepting upper and lower bounds as hydrological constraints improves the quality of the statistical models.

KEY WORDS: vine copulas, assessment of uncertainty, flood wave characteristics, conditional probability, flood hazard

Introduction

Managing flood risks relies, among other inputs, on hydrological design values, which are used for sizing hydraulic structures (reservoir volumes, spillways, flood levees, etc.), protecting residential zoning, and other areas at risk of flooding. The most recognised quantity in hazard estimation is the peak of a design flood, which is assessed by classical flood frequency analysis. Since specific hydraulic structures or flooding zones can only be correctly sized with a design input hydrograph, the knowledge of a such flood hydrograph (or at least of its selected parameters) is advantageous for a reliable design and adequate description of the impact of flooding events (Brunner et al., 2016a; Brunner, 2023; Kotaška and Říha, 2023; Škvarka et al., 2021). Another problem of univariate analysis is, that it does not consider the dependence between the hydrological control variables, e.g., flood peaks and volumes, which leads to inappropriate conclusions about the risks associated with their join impact (Rizwan et al., 2019). Interdependencies inherent among parameters of flood waves, which are of interest for a particular problem, must be accepted and assessed within a multivariate probabilistic framework (Brunner, 2023).

Recently efforts were therefore directed towards a multivariate probabilistic analysis of two or more flood wave characteristics (Xiao et al., 2009). Classical applications of bi- and tri-variate probability distributions were initially preferred. Rizwan et al. (2019) provided a short review of these methods. Unfortunately, their inherent limitations, such as the requirement that the marginal distributions of the constituent variables be identical, are challenging to meet in practice since the suitable marginal distributions of the multi-dimensional analysis usually differ. In addition, the mathematical formulations exceeding two dimensions are increasingly complicated (Rizwan et al., 2019).

Multivariate statistical analysis using the copula functions was introduced as a solution; it provides a convenient framework for estimating overall or conditional joint probabilities. Copulas allow for different marginal distributions of their components. These are connected by the copula function and which accounts for the dependence structure between them in the probability space. For a comprehensive overview of these developments, methods, and applications, see Nazeri Tahroudi et al. (2022); Tootoonchi et al. (2022); Größer and Okhrin (2022). Copula methods efficiently enable the assessment of joint and conditional probabilities of exceedance or return periods of a set of variables. Whereas, probabilities of exceedance in a univariate case, are uniquely defined, in a multivariate case we need to prefer one definition out of several possibilities for the practical problem given; an overview

is provided by Gräler et al. (2013) and Brunner et al. (2016b). In recent years, vine copulas have been suggested as convenient models of higher dimensions (Tosunoglu et al., 2020). They can establish flexible dependence structures by mixing components modelled by bivariate copulas.

Because of its practical convenience, the vine approach has recently received increased attention in studying multivariate hydrological analyses (e.g., Gómez et al., 2018; Brunner et al., 2019; Nazeri Tahroudi et al., 2022; Latif and Simonovic, 2022; Jafry et al., 2022), which has also led to increased efforts to conduct multivariate statistical analyses of entire flood events (e.g. Bačová Mitková et al., 2014; 2023; Ganguli et al., 2013; Requena et al., 2013; Rizwan et al., 2019). For example, Medeiro et al. (2010) described an ensemble of design flood hydrographs given by the peaks and volumes of peak-volume probabilistic spaces. A set of hydrographs with the same joint return period was proposed for a reservoir design, which leads to return periods for water levels for different flood peaks and volume combinations.

However, similarly to the case of design flood peaks, hydrologic flood wave characteristics and their respective parameters are inherently uncertain. The sources of the uncertainty are rooted in the length and quality of the hydrological records, the selection of uni- and multivariate statistical models, and the estimation of the parameters. To guarantee reliable flood wave estimates for designs, such uncertainties must be quantified and communicated to practitioners (Brunner et al., 2018; Liová et al., 2022).

Traditional uncertainty analyses of univariate design variables in hydrology are well established; they focus on many relevant aspects of the problem, such as the choice of the probability distributions, their parameter uncertainties, and comparisons of annual maxima sampling versus peak-over-threshold sampling (Beven and Hall, 2014; Kjeldsen et al., 2014). For a bivariate analysis, the problem becomes more complex. Brunner et al. (2018) presented a complex framework for quantifying and solving uncertainty issues, which could also be generalised to a multivariate case. The framework reflects, among other issues, problems of the length of the record, flood wave sampling, base flow separation, determination of marginal distributions and fitting of probability density functions, dependence modelling between the variables, and estimations of the joint and conditional design variable quantiles.

This study aims to invite the attention of practising risk analysts to one particular aspect of uncertainty associated with this new consistent probabilistic framework not tackled in previous studies. In the dependence modelling of peaks, volumes, and durations with vine copulas, we are only focusing on the consequences of respecting the hydrological constraints connected with modelling flood durations and propose to use the bounded Johnson distribution to describe them. While all sources of uncertainty mentioned above are acknowledged, we are ignoring here these uncertainties. Instead we pragmatically accept results of statistical testing when selecting the dependence structures among the peaks, volumes, and durations of the flood hydrographs

modelled by vine copulas their marginal distributions. We are also robustly respecting the hydrologic seasonality of floods by distinguishing between summer and winter floods since flood types of diverse origins have different shapes and consequently exhibit other dependence structures between the characteristics of flood waves (e.g., Gaál et al., 2015; Grimaldi et al., 2016; Szolgay et al., 2015). Finally, the practical consequences for the joint overall and conditional probabilities of the exceedance of flood peaks, volumes, and durations of choosing a bounded distribution for the flood durations are compared with the performance of distributions without an upper bound. The comparison was demonstrated on summer and winter flood data from the Parná River above the Horné Orešany reservoir in Slovakia.

Methods and material

Vine copulas for modelling the joint distribution of flood peaks, volumes, and durations

We are treating the respective flood peaks, volumes, and durations as elements X_i of the random vector $X = (X_1, X_2, X_3)$. Their (cumulative) distribution function (CDF) is defined as the probability of the (simultaneous) non-exceedance of:

$$F_i(x) = Pr(X_i \leq x) \text{ for } i \in (1,2,3) \tag{1}$$

$$F(x_1, x_2, x_3) = Pr(X_i \leq x_1 \wedge X_2 \leq x_2 \wedge X_3 \leq x_3) \tag{2}$$

while their survival function as a probability of the (simultaneous) exceedance,

$$\overline{F}_i(x) = Pr(X_i > x) \tag{3}$$

$$\overline{F}(x_1, x_2, x_3) = Pr(X_i > x_1 \wedge X_2 > x_2 \wedge X_3 > x_3) \tag{4}$$

The formula linking multivariate survival functions to CDF includes all the lower-dimensional marginals:

$$\overline{F}_{ijk} = 1 - F_i - F_j - F_k + F_{ij} + F_{ik} + F_{jk} - F_{ijk} \tag{5}$$
for $i, j, k \in \{1,2,3\}$ and $i \neq j \neq k$.

The joint probability distribution random variables are modelled by a combination of the marginal distributions of X_1, X_2, and X_3, which describe the individual stochastic behaviour of the elements, and a copula, which models their mutual relationship:

$$F(x_1, x_2, x_3) = C[F_1(x_1), F_2(x_2), F_3(x_3)] \tag{6}$$

$$f(x_1, x_2, x_3) = c[F_1(x_1), F_2(x_2), F_3(x_3)] \cdot f_1(x_1)f_2(x_2)f_3(x_3) \tag{7}$$

where F is the joint cumulative distribution function (CDF); F_1, F_2, and F_3 are the marginal CDFs; f_1, f_2, and f_3 are the corresponding densities; and C is a copula having uniform marginals defined on the unit hypercube $C: [0,1]^3 \rightarrow [0,1]$ (for the comprehensive details, see

Nelsen (2006). The joint CDF of the X_2, X_3 conditional on X_1 is then given by the formula:

$$F_{23|1}(x_2, x_3, x_1) =$$
$$\int_{-\infty}^{x_3} \int_{-\infty}^{x_2} c\big(F_1(x_1), F_2(s), F_3(t)\big) f_2(s) f_3(t)\, ds\, dt \quad (8)$$

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

Here, the vine copula approach is used to construct copula C by using a graphic vine tool (see, e.g., Asquith (2022); Czado and Nagler (2022)) and bivariate copulas as building blocks, so that a multivariate copula is obtained through conditioning. Since, in practice, the flood peak is usually considered the most critical flood characteristic, we have chosen this variable as being conditioned upon; then the trivariate density is given by:

$$f(x_1, x_2, x_3) = f_1(x_1) \cdot f_{2|1}(x_1, x_2) \cdot f_{3|12}(x_1, x_2, x_3)$$
$$= f_1(x_1) \cdot c_{12}[F_1(x_1), F_2(x_2)] \cdot f_2(x_2)$$
$$= c_{32|1}[F_{3|1}(x_3, x_1), F_{2|1}(x_2, x_1)] \cdot$$
$$c_{13}[F_1(x_1), F_3(x_3)] \cdot f_3(x_3) \quad (9)$$

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. Further, the 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, $k \in \{1,2,3\}$ and $i \neq j \neq k \neq i$. Finally, $F_{i|j}(x_i, x_j) = \partial C_{ij}[F_i(x_i), F_j(x_j)]/\partial F_j(x_j)$ is a conditional CDF of X_i given X_j, see Schirmacher and Schirmacher (2008) and Dissmann et al. (2013) for further details.

Bivariate copulas, which are often used in hydrology, such as those from 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 as building blocks of the vine copula. Their selection and estimation of the parameters were made using a sequential procedure based on the AIC criterion and the maximum likelihood method (Dissmann et al., 2013).

The marginal CDFs of flood volume and discharge were selected from several distribution functions often used in the hydrology of extremes, such as the Generalized Extreme Value (GEV), Pearson 3 (p3), Weibull 3 (wb3) and Log-normal 3 (ln3) distributions. The models of the marginal distributions of flood durations considered with two to five parameters were selected from distribution functions with a lower bound, such as Wakeby (wak), Log-normal 3 (ln3), Gamma (gam), Rayleigh (ray), Rice (rice), Exponential (exp), Govindarajulu (gov), and Generalized Pareto (gpa) distributions. The parameters were estimated using the L-moments (Hosking, 1990). Additionally, flood durations were also modelled by the bounded distribution of Johnson (1949), specifically Johnson's SB distribution (joh), which showed that a random variable X with an upper and lower bound could be transformed to

an approximately normal distribution with the density function given as:

$$f(x) = \frac{\delta}{\sqrt{2\pi}} \frac{\lambda}{(x-\xi)(\xi+\lambda-x)} e^{\left\{-\frac{1}{2}\left[y+\delta\ln\left(\frac{x-\xi}{\xi+\lambda-x}\right)\right]^2\right\}} \quad (10)$$

where $\xi < x < \xi + \lambda, \delta > 0, -\infty < y < \infty, \lambda > 0, \xi \geq 0$. The location parameter ξ, the range parameter λ, and shape parameters γ and δ were estimated by the maximum likelihood method.

The best-fitting distribution function for each flood variable was evaluated based on the CDF plots and the statistical Kolmogorov-Smirnov test (at 0.05 of the level of significance). The mean square error (MSE) between the observed and the simulated flood variables were calculated to evaluate the performance of the distribution functions selected. The statistical analysis and the visualizations were performed using the following R packages: Vine Copula (Nagler et al., 2022); ggplot2 (Wickham, 2016); metR (Campitelli, 2021); and cubature (Narasimhan et al., 2023).

Trivariate flood wave characteristics on the Parná River

The analysis was carried out on the Parná River inflow into the Horné Orešany reservoir located in western Slovakia (Fig. 1). The gauging station is situated above the reservoir on the Parná River at rkm 26.8 and has a basin area of 37.86 km². A more detailed description of the basin can be found in Liová et al. (2022). The discharge time series in hourly time steps and the maximum annual discharges used in the analysis were provided by the Slovak Hydrometeorological Institute (SHMI) for the period from 01.11.1988 to 31.12.2021. The precipitation and air temperature data in the daily time step, which were collected from the Dolné Orešany rainfall station and SHMI Modra-Piesok climatological station, were used for a better determination of the duration of the flood wave.

The flood waves analysed differ significantly in their shape, volume, and duration. For example, winter flood waves have a longer duration and greater volume than summer flood waves, which often arise from storm events and are also associated with melting snow or a combination of snowmelt and rain. On the other hand, summer flood waves are slimmer in shape and have a shorter duration. For this reason, flood waves were analysed separately for the summer (June-October) and winter (November-May) seasons. The separation of the flood waves, base flow, and estimations of the flood wave characteristics have been processed by an algorithm used in the FloodSep method (Valent, 2019). Next, we focused on a more detailed analysis of the flood wave durations and their effect on estimating the joint probability of the exceedance of flood characteristics.

Fig. 2 presents the probability density functions and histograms of flood durations in the summer and winter seasons. We can observe a unimodal form of both histograms. The summer flood durations have a range from 1.08 to 8.08 days and a median of 3.5 days.
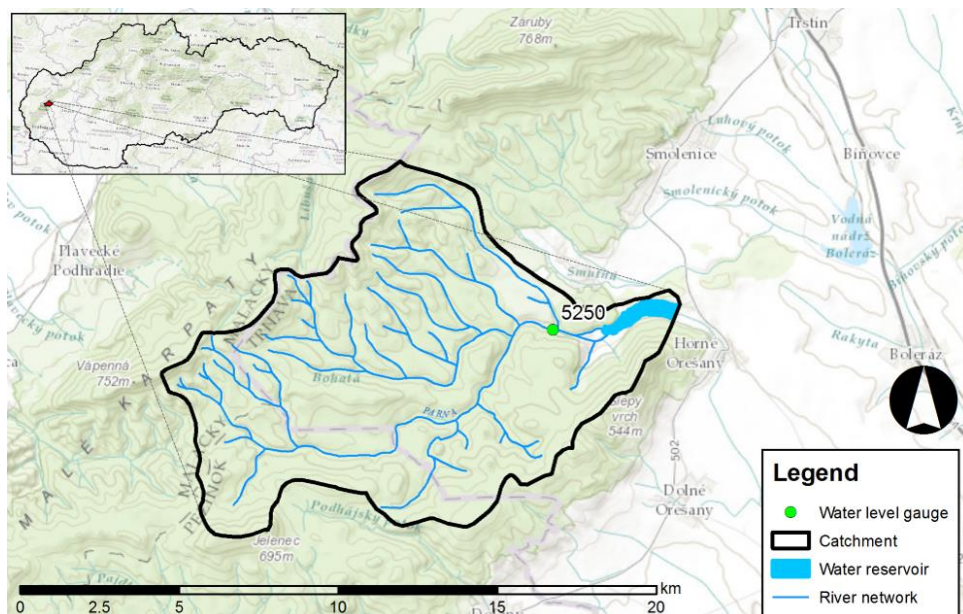
Fig. 1.   Location of gauging station on the Parná River and the water reservoir of Horné Orešany within Slovakia.
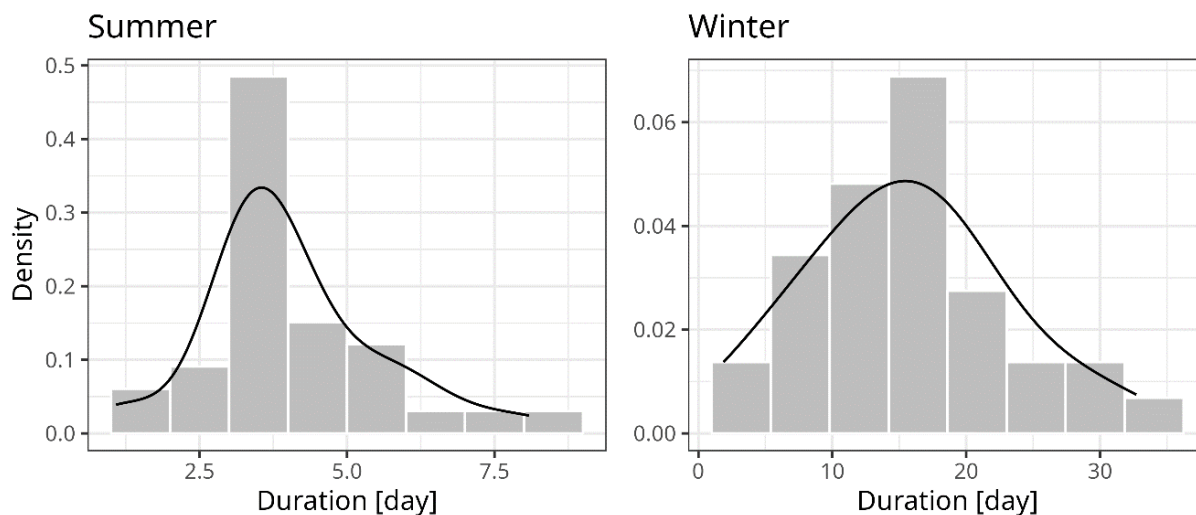

Fig. 2.   Empirical probability density functions for durations of floods in the summer (left) and winter (right) seasons on the Parná River.

The winter flood durations are considerably longer in a range from 1.88 to 32.70 days and a median of 15 days, and their histogram is left-skewed. These results already show how different the flood waves are in the seasons selected and confirm the importance of treating them separately.

Results

Estimation of the best-fitting distributions of the flood peaks, volumes and durations and defining marginals

Prior to the modelling, the peak discharges, volumes, and durations were extracted from each flood wave to form the distributions of the extreme values in the summer and winter seasons. The most appropriate distribution functions were then found for the peak discharges, the Generalized Extreme Value distribution (gev), and, for the flood volumes, the Pearson type III distribution (p3). As a means of comparison, the DVWK (1999) methodology, which provided similar results of the distribution fitting, was also used.

For the flood durations, the efficiency of several selected distribution models with a lower bound, which are frequently used in hydrological applications sensitive to extreme values, were evaluated in detail. The validity of

the estimated models was examined based on the goodness of fit (GoF) using the MSE (Fig. 3) and visually on the CDFs of the flood durations (Fig. 4). This was the rationale for adopting the best types of the distributions of the flood durations in the paper. According to the results, the three appropriate distribution functions of the flood durations were selected to assess the uncertainty of the extremity of the flood events. In Fig. 3, these are shown by colour as joh (red point), ln3 (green point) and rice (blue point) distribution.

The Log-normal 3 and Rice distributions represent those with only the lower boundary parameters. However, Johnson's SB distribution can estimate not only the lower bound but also the upper bound of flood durations. The upper bound of the flood durations in the summer season using Johnson's SB distribution at 12 days and in the winter summer season at 37 days was estimated.

Trivariate distribution modeling using the vine copulas and the uncertainty of extremity of the seasonal floods

Modeling of unconditional joint probability of the exceedance of flood characteristics

To analyze the joint probability of the exceedance of the characteristics of maximum summer and winter floods, we evaluated the joint probability of the exceedance of the flood peaks "AND" the flood volumes "AND" the flood durations denoted as $(1- F(Q_{max},V,D))$ for three various distribution functions the duration of floods (see Tables 1, 2).

The most significant differences in the joint probability of the exceedance of flood peak Q_{max}, flood volume V, and flood duration D for summer flood waves were found for those with the longest durations. For example, for the longest flood wave, i.e., No. 25, with a duration
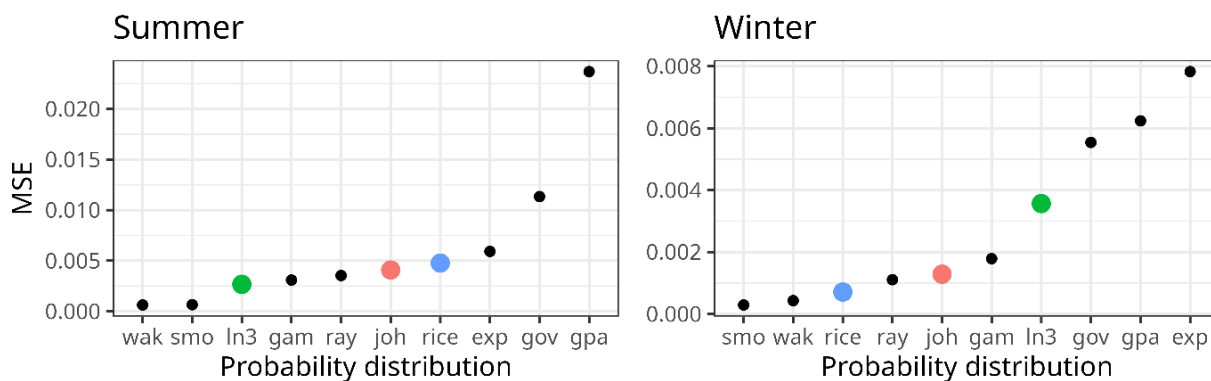

Fig. 3.   Performance of various probability models for fitting marginals distributions for flood durations in the summer (left) and winter (right) seasons on the Parná River. The three distribution functions of flood duration selected to assess the uncertainty of the extremity of flood events are marked by colour.
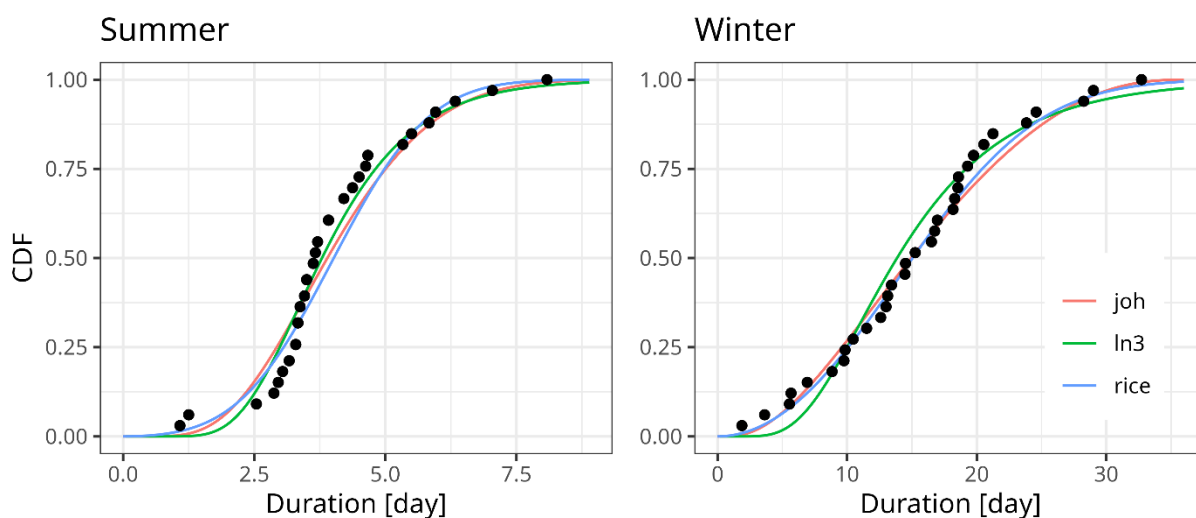

Fig. 4.   Fitted marginal distributions for the flood durations in the summer (left) and winter (right) seasons: CDFs of the three selected types of distribution probability on the Parná River.

Table 1. The joint probability of the exceedance of maximum summer floods on the Parná River (probability distribution of flood durations: joh – Johnson's SB, rice – Rice and ln3 - Log-Normal3 distribution)

No. of the wave	Q_{max} [m³ s⁻¹]	Volume [mil.m³]	Duration [day]	1-$F(Q_{max},V,D)$ joh	rice	ln3
1	1.39	0.18	4.38	0.278	0.29	0.252
2	0.56	0.05	3.04	0.661	0.681	0.741
3	0.67	0.11	3.92	0.403	0.424	0.379
4	1.11	0.25	5.33	0.154	0.144	0.135
5	0.88	0.05	3.46	0.527	0.566	0.509
6	0.35	0.05	3.17	0.642	0.670	0.633
7	2.29	0.32	3.33	0.269	0.273	0.266
8	3.28	0.25	4.21	0.184	0.19	0.171
9	**8.63**	**1.59**	**7.04**	**0.007**	**0.004**	**0.008**
10	5.37	0.66	5.83	0.041	0.035	0.037
11	6.02	0.49	3.92	0.098	0.101	0.095
12	0.70	0.05	3.29	0.585	0.609	0.577
13	2.01	0.17	3.29	0.358	0.365	0.352
14	5.89	0.17	2.54	0.140	0.141	0.142
15	0.09	0.02	2.96	0.764	0.763	0.749
16	0.88	0.09	3.67	0.458	0.481	0.436
17	1.37	0.15	3.38	0.417	0.430	0.409
18	1.58	0.10	3.5	0.412	0.430	0.401
19	2.14	0.29	4.67	0.191	0.195	0.172
20	0.60	0.11	5.96	0.104	0.083	0.094
21	1.09	0.14	3.92	0.375	0.396	0.351
22	4.98	0.83	5.50	0.052	0.048	0.047
23	**33.86**	**1.67**	**3.62**	**0.005**	**0.005**	**0.005**
24	0.93	0.05	3.50	0.533	0.546	0.491
25	**2.53**	**0.52**	**8.08**	**0.004**	**0.001**	**0.010**
26	13.99	0.94	4.50	0.023	0.024	0.022
27	3.12	0.33	4.62	0.153	0.156	0.139
28	2.17	0.18	3.71	0.302	0.315	0.295
29	0.89	0.08	2.88	0.588	0.609	0.588
30	9.54	0.28	3.33	0.063	0.065	0.063
31	2.22	0.04	1.08	0.415	0.443	0.411
32	8.83	1.72	6.33	0.011	0.009	0.011
33	0.59	0.02	1.25	0.805	0.787	0.716

of 8.08 days, the values of the joint probability of the distribution of Q_{max}, V and D under various marginal distributions of D were 0.004 for Johnson's SB, 0.001 for Rice, and 0.01 for the Log-Normal 3 distribution; for the flood wave No. 9 with a duration of 7.04 days, the values of the joint probability of exceedance were 0.007, 0.004 and 0.008. However, these two flood waves with the longest D were not significant in either their peaks or volumes. On the contrary, for the flood wave with the highest flood peak and volume, i.e., with a duration of 3.62 days, the joint probability of exceedance was the same as for all three marginal distributions of D, i.e., 0.005.

For winter flood waves, the most significant differences in the joint probability of exceedance were again found for flood wave No. 11 with the longest duration of 32.7 days; the probability of exceedance was 0.002 for Johnson's SB, 0.003 for Rice, and 0.007 for the Log-Normal 3 marginal distribution of D. Here, the differences were also found for the most significant winter floods. Flood wave No. 18, with the highest flood peak and volume and a relatively long duration of 18.54 days, had a value of the probability of exceedance of 0.007 for the Johnson marginal distribution of D, 0.006 for Rice, and 0.005 for the Log-Normal3 marginal distribution. The second largest flood wave in flood peak and volume, i.e., No. 21, with a duration of 23.8 days, had joint probabilities of exceedance of 0.005, 0.005, and 0.004.

Table 2.   The joint probability of the exceedance for maximum winter floods on the Parná River (probability distribution of flood duration: joh – Johnson's SB, rice – Rice and ln3 - Log-Normal3 distribution)

No. of the wave	Q_{max} [m³ s⁻¹]	Volume [mil.m³]	Duration [day]	$1 - F(Q_{max}, V, D)$		
				joh	rice	ln3
1	1.49	0.55	8.83	0.623	0.635	0.63
2	1.99	1.11	18.58	0.237	0.225	0.187
3	1.77	0.98	14.50	0.37	0.375	0.328
4	5.51	3.91	20.54	0.037	0.035	0.030
5	3.48	1.68	13.12	0.213	0.217	0.194
6	3.53	2.58	19.75	0.097	0.091	0.077
7	3.43	0.93	9.75	0.299	0.302	0.291
8	5.43	5.15	24.58	0.013	0.012	0.012
9	2.83	2.26	28.25	0.028	0.026	0.034
10	1.45	0.93	13.42	0.439	0.429	0.393
11	**3.70**	**4.59**	**32.71**	**0.002**	**0.003**	**0.007**
12	2.28	1.80	18.29	0.189	0.184	0.155
13	1.95	1.09	14.46	0.347	0.350	0.307
14	1.25	0.77	16.96	0.350	0.340	0.286
15	1.16	0.54	11.5	0.575	0.589	0.568
16	3.93	1.70	12.58	0.190	0.194	0.174
17	7.16	3.00	19.29	0.026	0.025	0.020
18	**11.47**	**5.54**	**18.54**	**0.007**	**0.006**	**0.005**
19	1.36	0.86	18.17	0.299	0.287	0.239
20	1.01	1.24	29.00	0.040	0.040	0.057
21	**9.70**	**5.04**	**23.83**	**0.005**	**0.005**	**0.004**
22	5.39	2.73	16.75	0.070	0.069	0.058
23	7.22	2.32	13.00	0.048	0.049	0.044
24	1.71	0.47	10.46	0.547	0.558	0.532
25	4.26	3.00	21.25	0.059	0.054	0.048
26	2.56	0.25	3.62	0.564	0.556	0.546
27	4.85	2.64	16.50	0.088	0.087	0.074
28	2.26	1.27	9.83	0.401	0.417	0.397
29	1.37	0.76	15.25	0.402	0.404	0.337
30	1.95	0.19	1.88	0.718	0.701	0.752
31	6.37	1.06	5.67	0.104	0.104	0.111
32	2.79	0.41	6.92	0.457	0.454	0.460
33	3.64	0.95	5.54	0.323	0.317	0.318

Modeling of the conditional probability of the exceedance of flood characteristics

In the next step, the uncertainties of the selection of the marginal distribution functions of the flood durations were demonstrated by the joint conditional probability of the exceedance of V and D on the 100-year flood peak Q_{100}, written as $S(V,D/Q_{100})$. From the marginal distribution of the flood peaks, the design discharge with the probability of exceedance 0.01 (100-year flood) was estimated as 32.76 m³ s⁻¹. The results of the conditional probability of exceedance are illustrated by the isolines of $S(V,D/Q_{100})$ in Fig. 5.

From the results illustrated by the isolines of the joint probability of the exceedance of V and D, it can be seen that for summer floods, the differences in the joint probabilities of exceedance $S(V,D/Q_{100})$ due to various marginal distributions of D are visible for flood waves with durations longer than approximately 2 days. With the decreasing of the joint probability of the exceedance of V and D, the flood durations from which the differences are evident are increasing. While for the joint probability of exceedance $S(V,D/Q_{100})$ of 0.9, it is approximately 2 days; for $S(V,D/Q_{100}) = 0.7$, it is 3 days; and for $S(V,D/Q_{100}) = 0.3$, it is 4 days. We can see similar results in the winter season, but the flood durations from which the differences in $S(V,D/Q_{100})$ are visible are longer. Because $S(V,D/Q_{100}) = 0.9$, the flood durations are approximately 2.5 days; for $S(V,D/Q_{100}) = 0.7$, they are approximately 5 days; and for $S(V,D/Q_{100}) = 0.3$; they are approximately 10 days.
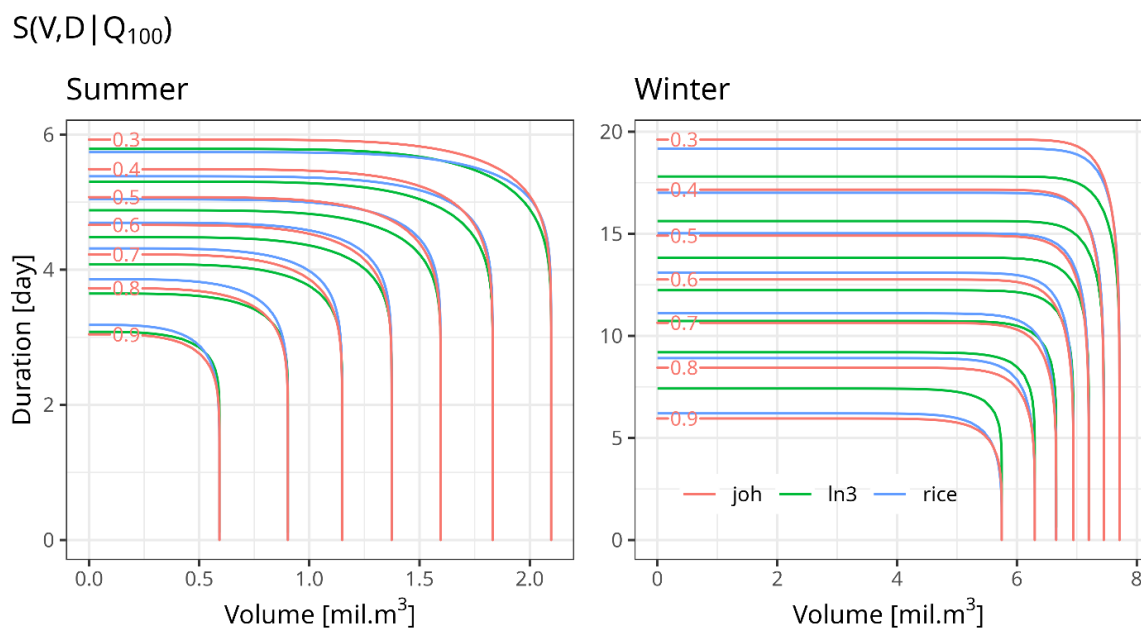
$S(V,D|Q_{100})$

Fig. 5.   *The isolines of the joint conditional probability of the exceedance of the flood duration with various probability distributions "AND" the volume of the 100-year flood peak in the summer (left) and winter (right) seasons on the Parná River.*

Discussion

Besides factors that cause uncertainties in frequency analysis, which are well reflected in the univariate methods (the length of the record, the annual maxima versus the peak- over-threshold sampling, the choice of the probability distributions, and their parameter uncertainties), in the multivariate case additional uncertainty sources occur, such as the sampling of flood waves (including base flow separation), determination of the marginal distributions, and fitting of multivariate probability density functions, dependence modelling between the variables, and estimation of the joint and conditional design variable quantiles (Beven and Hall, 2014; Kjeldsen et al., 2014).

This study aimed only at attracting the attention of practising analysts to one particular hydrological aspect of the uncertainty not tackled in previous studies. We acknowledge the importance of all uncertainty sources listed and treated in detail in Brunner et al. (2018). However, as is usual in a practical hydrological design exercise, we intentionally left out their assessments and pragmatically implemented the results of the basic statistical testing when selecting the appropriate models for the frequency and dependence analysis of the peaks, volumes, and durations of the flood hydrographs observed by vine copulas.

The effect of baseflow considerations on the beginning and end of a flood was not evaluated here, too. We could also have discussed the options for the wave selection, such as the direct runoff separation or taking the entire hydrograph. Instead, the results from the FloodSep method described in Liová et al. (2022) were implemented. Nevertheless, we acknowledge the immense importance of using hydrologically sound and practically relevant flood wave separation methods for the outcomes of the analyses of volumes and durations.

We also restricted the analysis of the typology of flood generation in the pilot catchment only to the basic seasonality of the floods by distinguishing between summer and winter floods. The discharge waves analysed differ significantly in their shape, volume, and duration, which can be seen, e.g., on the bimodal shape of the histogram of the annual maximum flood durations in Fig.6. Winter waves have longer durations associated with melting snow or combinations of snowmelt and rain. Their volumes are more significant than those of summer waves, which often arise from storm events and are slimmer in shape and have a shorter duration. For this reason, the flood waves were analysed separately for the summer (June–October) and winter (November–May) seasons, reaching a median of 3.5 days duration in the summer and 15 days in the winter seasons.

The unimodal frequency histograms of the respective summer and winter flood durations showed that such a simplification might be acceptable in the given catchment. Although the seasonal histograms of flood durations are unimodal, we can also see that they are skewed (see Fig. 2), which could indicate that not all floods may have the same origin; e.g., in the winter, a pattern composed of a mix of flood events caused by snowmelt, prolonged rains, and rain on snow events could be hidden in a histogram. In such and similar cases in basins with a richer flood typology, respecting more flood types (short rain, long rain, rain on snow etc., see Gaál et al. (2015)) than in our method may be required since flood types of diverse origins have

different shapes and consequently exhibit other dependence structures between the parameters of flood waves (Gaál et al., 2015).

We only focused in this analysis on the consequences of respecting or neglecting the obvious hydrological constraints connected with flood durations. To model a flood duration coming from a finite interval, we proposed to use the bounded Johnson's SB distribution to describe them. The SB Johnson distribution proved to be flexible. When respecting its domain of applicability defined through its third and fourth moments, see Johnson (1949) and Parresol (2003), it also enables the use of hydrologically determined upper and lower bounds. Practical consequences for

the joint overall and conditional probabilities of the exceedance of the flood peaks, volumes, and durations of choosing a bounded distribution for the flood durations were compared with the performance of distributions without an upper bound, which is a hydrologically erroneous (but often used) assumption, see Figs. 4 and 7.

Differences in the modelling of joint probabilities due to the different tail behaviors of the marginal distributions tested were found. Although these are not located in critical regions of joint probabilities for practical applications, accepting hydrological constraints as upper and lower bounds improves the statistical model's quality and is recommended.
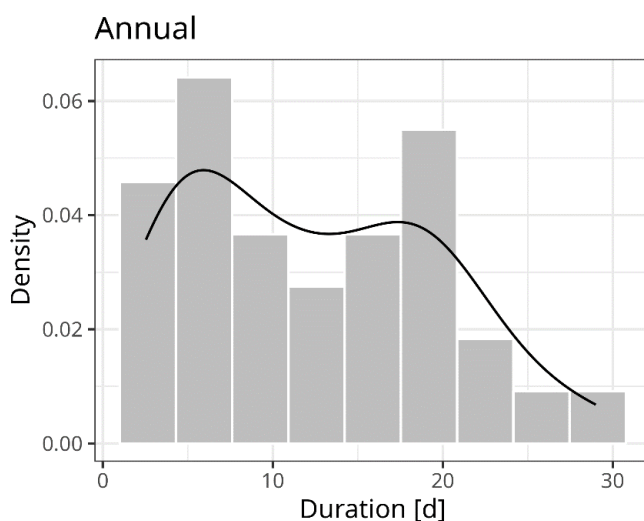

Fig. 6. Bimodal empirical probability density function for flood durations associated with annual maximum discharges on the Parná River.
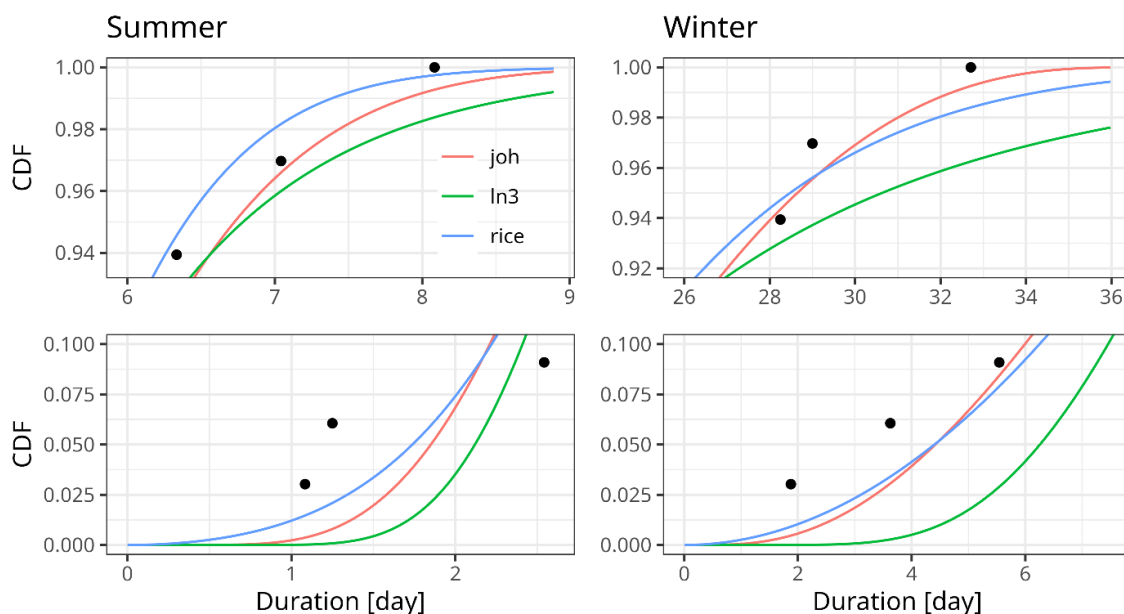

Fig. 7. Details of upper and lower bounds of fitted marginal distributions for flood durations in the summer (left) and winter (right) seasons on the Parná River: CDFs of three selected types of distribution probabilities.

Conclusion

The most recognized quantity in flood hazard estimations is the peak of the design flood. Univariate frequency analysis is used conventionally to assess its value and the associated uncertainties of its estimation. However, specific design tasks require considering the impact of the flood hydrograph, usually described by a set of hydrological control variables, e.g., the flood peaks, volumes or durations. Only by considering the dependence among these variables in the hazard estimation may lead to appropriate conclusions about the risks associated with the joint impact of these in their various combinations. The inherent interdependencies among flood wave shape control variables must be quantified within multivariate statistical assessment frameworks, which can also account for the uncertainties of the estimates (though in a more complex way than in a univariate case). Besides uncertainty-causing factors, which are reflected in univariate cases (such as problems involving the annual maxima versus peak- over-threshold sampling, the choice of the probability distributions, and their parameter uncertainties), additional issues such as flood wave sampling, base flow separation, fitting of multivariate probability density functions, the dependence modelling between the variables, and estimation of the joint and conditional probabilities of the design variable quantiles have to be considered.

As previously mentioned, this study only intended to attract the attention of practising analysts to the uncertainty associated with the joint probabilities assessed within the vine copula probabilistic framework. These must be addressed in practice and in connection with statistical flood duration models. Therefore, our analysis did not consider the statistical sources of uncertainties but illustrated those originating in hydrological sources, which could require attention. These are mostly related to the sampling of flood waves and include the knowledge of flood generation processes (short rains, prolonged rains, rain on snow, snowmelt, etc.), the estimation of the flood volume (direct runoff vs. a whole hydrograph), and the determination of the duration of the flood waves.

Regarding which flood generation processes to consider, using annual maxima in such an analysis may not meet the IID requirements in general. Instead, we recommend to consider the basic seasonality of the flood regime separately and to look within the seasons at a stepwise approach at the frequency histograms of the durations associated with respective flood types of the events sampled (unimodal vs. multimodal, symmetrical vs. skewed, etc.) and the practical goals of the design or risk analysis task. For example, frequency histograms of the respective durations of flood types may support a decision made on an acceptable grouping of sparser flood types (or eventually leaving them out) in the given catchment and for the specific task.

In a decision on the selection of the flood volumes for the analysis, more practical aspects could be preferred and a task-related choice performed. For example, in sizing the retention volumes of multipurpose reservoirs or flood zoning, a whole hydrograph could be required for the design. To only consider direct runoff for flood detentions may not be correct since the detention volume is also filled with the flood's baseflow. On the contrary, direct runoff may be preferred when looking at a regime of event-based rainfall-runoff relationships.

Determining the duration of a flood is connected to the decisions to be made on its volume, and it is not a straightforward task, especially when the whole hydrograph is taken as a flood wave. This determination is especially complicated in the winter season for snowmelt and mixed floods. Methods ranging from baseflow separations up to snowmelt and rainfall runoff modelling may be required for a hydrologically plausible solution. In any case, an upper and lower bound must be considered. The marginal distribution of the duration should preferably respect these bounds since, in the dependence modelling of peaks, volumes and durations, it affects the joint and conditional probabilities of the exceedance associated with any combined hazards. Our case study showed differences in the outcomes of modelling the joint probabilities due to the different tail behaviours of the marginal distributions tested. Although these may sometimes only be critical for practical applications, accepting hydrological constraints as upper and lower bounds improves the statistical model's quality. One computationally and conceptually acceptable option is the four-parameter bounded Johnson's SB distribution, as was illustrated in our case study.

Acknowledgement

Reference

Asquith, W. (2022): lmomco–L-moments, censored L-moments, trimmed L-moments, L-comoments, and many distributions. R package version 2.4.7. https://cran.r-project.org/web/packages/lmomco/lmomco.pdf

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 Hydromechanics, 62, 186–196. DOI: https://doi.org/10.2478/johh-2014-0026

Bačová Mitková, V., Halmová, D., Pekárová, P., Miklánek, P. (2023): The Copula Application for Analysis of the Flood Threat at the River Confluences in the Danube River Basin in Slovakia. Water, 15(5), 984. DOI: https://doi.org/10.3390/w15050984

Beven, K., Hall, J. (2014): Applied uncertainty analysis for flood risk management. World Scientific. DOI: https://doi.org/10.1142/p588

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. DOI: https://doi.org/https://doi.org/10.5194/hess-2023-20

Brunner, M., Furrer, R., Favre, A. (2019): Modeling the spatial

dependence of floods using the Fisher copula. Hydrology and Earth System Sciences, 23(1), 107–124. DOI: https://doi.org/10.5194/hess-23-107-2019

Brunner, M. I., Sikorska, A. E., Furrer, R., Favre, A. C. (2018): Uncertainty assessment of synthetic design hydrographs for gauged and ungauged catchments. Water Resources Research, 54(3), 1493–1512. DOI: https://doi.org/10.1002/2017WR021129

Brunner, M., Seibert, J., Favre, A. (2016a): Bivariate return periods and their importance for flood peak and volume estimation. WIREs Water, 3(6), 819–833. DOI: 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. 13th Congress INTERPRAEVENT 2016, Luzern, 30 May 2016 – 2 June 2016, 468-476. DOI: https://doi.org/10.5167/uzh-124430

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

Czado, C., Nagler, T. (2022): Vine Copula Based Modeling. Annual Review of Statistics and Its Application, 9(1), 453–477. DOI: https://doi.org/10.1146/annurev-statistics-040220-101153

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. DOI: https://doi.org/10.1016/j.csda.2012.08.010

DVWK 101 (1999): Wahl des Bemessungshochwassers. Empfehlung zur Berechnung der Hochwasser-wahrscheinlichkeit. DVWK Schriften, Heft 101. Verlag Paul Parey, Hamburg.

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(6), 968-984. DOI: https://doi.org/10.1080/02626667.2014.951361

Ganguli, P., Reddy, M. (2013): Probabilistic assessment of flood risks using trivariate copulas. Theoretical and Applied Climatology, 111(1–2), 341–360. DOI: https://doi.org/10.1007/s00704-012-0664-4

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(10), 2787–2807. DOI: 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(4), 1281–1296. DOI: 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. DOI: https://doi.org/10.1016/j.advwatres.2016.02.003

Größer, J., Okhrin, O. (2022): Copulae: An overview and recent developments. WIREs Comp. Statistics, 14(3). DOI: 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(1), 105–124. DOI: https://doi.org/10.1111/j.2517-6161.1990.tb01775.x

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. https://science.utm.my/procscimath/files/2022/06/52_55_Naqibah_Aminudin_Jafry.pdf

Johnson, N. (1949): Systems of frequency curves generated by methods of translation. Biometrika, 36(12), 149–176. DOI: https://doi.org/10.2307/2332539

Kjeldsen, T., Lamb, R., Blazkova, S. (2014): Uncertainty in flood frequency analysis. Applied Uncertainty Analysis for Flood Risk Management, 153–197. DOI: https://doi.org/10.1142/p588

Kotaška, S., Říha, J. (2023): Dam incidents and failures- cases in the Czech Republic. Slovak Journal of Civil Engineering, 31(1), 22–33. DOI: https://doi.org/10.2478/sjce-2023-0003

Latif, S., Simonovic, S. (2022): Parametric Vine Copula Framework in the Trivariate Probability Analysis of Compound Flooding Events. Water, 14(14). DOI: https://doi.org/10.3390/w14142214

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(1), 52–61. DOI: https://doi.org/10.31577/ahs-2022-0023.01.0006

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(12), 2495-2505. DOI: 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. https://CRAN.R-project.org/package=VineCopula

Narasimhan, B., Johnson, S., Hahn, T., Bouvier, A., Kiêu, K. (2023): cubature: Adaptive Multivariate Integration over Hypercubes. R package version 2.0.4.6. https://CRAN.R-project.org/package=cubature

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(4). DOI: https://doi.org/10.1007/s13201-022-01589-4

Nelsen, R. (2006): An introduction to copulas, (2nd Ed.), Springer.

Parresol, B. (2003): Recovering parameters of Johnson's SB distribution (Vol. 31): US Department of Agriculture, Forest Service, Southern Research Station. DOI: https://doi.org/10.2737/SRS-RP-31

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(8), 3023–3038. DOI: 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(8). DOI: https://doi.org/10.3390/w11081531

Schirmacher, D., Schirmacher, E. (2008): Multivariate dependence modeling using pair-copulas. Technical report. https://www.soa.org/4938dc/globalassets/assets/files/resources/essays-monographs/2008-erm-symposium/mono-2008-m-as08-1-schirmacher.pdf

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. DOI: 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(2), 9–15. DOI: https://doi.org/10.2478/sjce-2021-0009

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). DOI: 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(19). DOI: https://doi.org/10.1007/s12665-020-09199-6

Valent, P. (2019): Floodsep: user´s manual, SvF STU in Bratislava.

Wickham, H. (2016): ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag NY. https://ggplot2.tidyverse.org

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(12), 1359-1364. DOI: https://doi.org/10.1061/(ASCE)1084-0699(2009)14:12(1359)

Assoc. Prof. Ing. Roman Výleta, PhD. ([*]corresponding author, e-mail: roman.vyleta@stuba.sk)
Prof. Ing. Kamila Hlavčová, PhD.
Prof. Ing. Silvia Kohnová, PhD.
Ing. Anna Liová
Prof. Ing. Ján Szolgay, PhD.
Department of Land and Water Resources Management
Faculty of Civil Engineering
Slovak University of Technology in Bratislava
Radlinského 11
810 05 Bratislava
Slovak Republic

Assoc. Prof. Ing. Tomáš Bacigál, PhD.
Department of Mathematics and Descriptive Geometry
Slovak University of Technology in Bratislava
Radlinského 11
810 05 Bratislava
Slovak Republic