



Extensive spatio-temporal assessment of flood events by application of pair-copulas

M. Schulte and A. H. Schumann

Institute of Hydrology, Water Resources Management and Environmental Engineering, Ruhr-University Bochum, Bochum, Germany

Correspondence to: A. H. Schumann (andreas.schumann@rub.de)

Received: 25 March 2015 - Accepted: 25 March 2015 - Published: 11 June 2015

Abstract. Although the consequences of floods are strongly related to their peak discharges, a statistical classification of flood events that only depends on these peaks may not be sufficient for flood risk assessments. In many cases, the flood risk depends on a number of event characteristics. In case of an extreme flood, the whole river basin may be affected instead of a single watershed, and there will be superposition of peak discharges from adjoining catchments. These peaks differ in size and timing according to the spatial distribution of precipitation and watershed-specific processes of flood formation. Thus, the spatial characteristics of flood events should be considered as stochastic processes. Hence, there is a need for a multivariate statistical approach that represents the spatial interdependencies between floods from different watersheds and their coincidences. This paper addresses the question how these spatial interdependencies can be quantified. Each flood event is not only assessed with regard to its local conditions but also according to its spatio-temporal pattern within the river basin. In this paper we characterise the coincidence of floods by trivariate Joe-copula and pair-copulas. Their ability to link the marginal distributions of the variates while maintaining their dependence structure characterizes them as an adequate method. The results indicate that the trivariate copula model is able to represent the multivariate probabilities of the occurrence of simultaneous flood peaks well. It is suggested that the approach of this paper is very useful for the risk-based design of retention basins as it accounts for the complex spatio-temporal interactions of floods.

1 Introduction

The design of flood retention basins for small catchments usually focuses on the runoff of the main channel. The relevant analyses involve both the peak runoff and the corresponding flood volume. Both characteristics are important parameters, which decide if a flood detention measure is sufficient to meet the required protection targets. Consequently, both variables need to be examined together. The use of bivariate copulas offers one possibility to do this (e.g. Favre et al., 2004; De Michele et al., 2005). However, the design of technical flood detention is getting more complex with increasing area of the watershed and the related branching of the river network. Because of the superposition of flood peaks of different sub-basins or adjoining catchments, the risk of flooding and overloading of storage (-systems) may increase downstream which increases the complexity of the analysis. In addition, the distribution of precipitation and the watershed-specific processes have a huge impact on the resulting flood waves. The associated probability of the peak coincidence can be quantified by multivariate statistics. The more tributaries the river network consists of, the more variates have to be taken into account by the models. Copula models are suitable to account for a number of variates. Copulas in the context of flood coincidence have been used by Wang et al. (2009). They utilized bivariate Frank-copulas to generate pairs of peak discharges of nearby gauging stations. Kao and Chang (2012) demonstrated the influence of dams on the time series of runoff of coincidence-affected sites using bivariate Gauss-copulas. Chen et al. (2012) presented copula models of higher dimensions. They chose the 4-D-Gumbel-copula to estimate the probability of simultaneous floods at the Yangtze River in China and the Colorado River in the United States. In addition to the peaks, they modelled the timing of the flood events. Ghizzoni et al. (2010, 2012) go a step further concerning the dimensionality. Applying the *t*-copula with 18 dimensions, they investigate the flood coincidence in terms of risk analysis at the Mississippi River. In all of these studies, one copula family describes the dependence structure. This may be a disadvantage if the dependence structure between the considered parameters differs between the several variates. The implementation of paircopulas may resolve this problem.

This paper presents a case study that adopts copulas to represent the superposed peak discharges of three adjoining catchments. We use trivariate Archimedean copulas as well as pair-copulas to estimate the multivariate return periods of historical flood events, and compare the different copula types. The results suggest that joint return periods are indeed able to represent the spatio-temporal flood patterns within the river basin in a meaningful way.

2 Estimation of multivariate return periods via 3-Dand pair-copulas

The conventional flood statistics analyses the univariate probability that the peak value X exceeds a design value x. The case of X > x characterises a critical situation. In the multivariate case, we have to consider more than one variable, so the critical region is multi-dimensional. Downstream of a confluence, critical discharges may only occur if the tributaries cause a flood at the same time. The corresponding peak discharges of the adjoining rivers usually show positive statistical (non-linear) correlation, which is even higher for adjacent catchments. To describe the influence of these interdependencies copulas can be applied. They are able to relate the marginal distributions of the peaks while maintaining their dependence structure. Here, only a brief outline of the concept of copulas is given. More information, e.g. for parameter estimation or tests of goodness of fit, can be found in Nelsen (2006) and Joe (2014). Copulas are based on the theorem of Sklar (1959). It describes the linking of the copula function C with the univariate distribution functions of correlated random variables X_1, \ldots, X_d (Eq. 1). F_1, \ldots, F_d represent the associated marginal cumulative density functions.

$$F(x_1, ..., x_d) = C[F_1(x_1), ..., F_d(x_d)]$$

= $P(X_1 \le x_1, ..., X_d \le x_d)$ (1)

Similarly to a univariate distribution, a copula can be taken as a multivariate distribution function. A huge number of copula functions exists, which can be classified into different families. One of them is the family of Archimedean copulas. They are based on their specific generation function ϕ , which is a continuous, convex and strictly monotonic decreasing function. The trivariate Archimedean copula has the form

$$C(u, v, w) = \varphi^{-1} \left[\varphi(u) + \varphi(v) + \varphi(w) \right],$$
(2)



Figure 1. Composition of Pair-Copulas.

where u, v and w stand for $F_1(x_1)$, $F_2(x_2)$ and $F_3(x_3)$. By use of its generation function, which can be described by $\varphi(t) = -\ln \left[1 - (1 - t)^{\theta}\right]$, we obtain the trivariate Joe-copula with parameter θ :

$$C_{\theta}(u, v, w) = 1 - \left\{ 1 - \left[1 - (1 - u)^{\theta} \right] \cdot \left[1 - (1 - v)^{\theta} \right] \right.$$

$$\left. \cdot \left[1 - (1 - w)^{\theta} \right] \right\}^{1/\theta}.$$
(3)

The combination of the flood peaks of the adjoining rivers characterises a flood event with regard to its spatial-temporal occurrence. The multivariate return periods T of the event is:

$$T_{U>u\wedge V>v\wedge W>w} = \frac{\mu}{P(U>u\wedge V>v\wedge W>w)} = \frac{\mu}{1-u-v-w+C(u,v)+C(v,w)-C(u,v,w)}.$$
 (4)

The value μ is the mean time between events. In case of annual maxima μ is equal to 1 (a). As Eq. (4) shows, the bivariate copula functions are needed for estimating the return periods.

Another possibility to construct copula models of higher dimensions is the use of pair-copulas (Joe, 1996; Aas et al., 2009). They are based on conditional bivariate copulas, which can be coupled by the concept of vines (Bedford and Cooke, 2002). Figure 1 shows the popular D-Vine. It is easy to imagine how a D-Vine can represent a network of three joining rivers. According to Joe (1996) the marginal conditional distribution functions are derived as

$$F_{u|v}(u|v) = \frac{\delta C_{11}(u,v)}{\delta F_v(v)},$$
(5)

$$F_{w|uv}(w|u,v) = \frac{\delta C_{21}(F_{w|v}(w|v), F_{u|v}(u|v))}{\delta F_{u|v}(u|v)}.$$
 (6)

In case of pair-copulas we estimated the required return periods by using a bootstrapping approach. We generated ten trivariate samples of 1 000 000 elements each by the fitted pair-copula model and obtained the return periods via frequency analysis.

3 Case study

Flood events in large river basins are composed of the contributions of a number of tributaries. The main intention of this case study is to estimate the variation of multivariate probabilities of these flood peaks and to compare the results of the



Figure 2. Position of gauging stations Wechselburg, Lichtenwalde and Nossen in the Mulde catchment.

two differing copula models. The case study illustrates the methodology for the Mulde catchment in eastern Germany, where the three streams Zwickauer Mulde, Zschopau and Freiburger Mulde merge (Fig. 2). We analysed the time series from the respective gauging stations Wechselburg (area: 2107 km²), Lichtenwalde (1575 km²) and Nossen (585 km²) for the years 1926-2012. All peak discharges of simultaneous flood events were selected if they exceeded the local 2.5fold mean runoff. The maximum time differences allowed for peaks of adjoining catchments to be considered the same event was one day. The resulting trivariate sample consists of 178 flood events measured at the three sites. Their interdependent flood peaks exhibit strong correlations. The corresponding values of Pearson's r, Spearman's ρ and Kendall's t for the possible pairs confirm this (Table 1). This finding confirms that simultaneous examination of the peak discharges is a useful exercise.

3.1 Determination of the univariate marginal distributions

In the next step, marginal distribution functions were estimated for the three univariate samples. The peak discharges at Wechselburg and Lichtenwalde were described by a log-Weibull distribution, those at Nossen by a generalized Extreme Value Distribution.

3.2 Selection of copulas

The fitting of several trivariate Archimedean copulas via the pseudo-likelihood method (Genest et al., 1995) showed the best goodness-of-fit for the Gumbel-Hougaard and Joe copulas. Because of the better performance in the test of Genest and Rivest (1993), we finally chose the Joe-copula for the statistical model of coinciding flood events. The superposi-



Figure 3. Randomized trivariate samples of 1000 elements overlayed with the observed simultaneous peak values, 1st row: trivariate Joe-Copula, 2nd row: Pair-Copula (HQ stands for peak discharge).

tion of a copula-generated trivariate sample and the observed flood peaks in the first row of Fig. 3 shows that the choice of Joe-copula is justified. The scatter plot reproduces the shape of the measured data and their interdependencies. As can be seen from Eq. (4) we also need some bivariate copulas for estimating multivariate return periods. Therefore, we reduced the 3-D-model to the three possible bivariate cases. The type of copula and the parameterisation were retained for consistency.

The construction of the pair-copula followed the D-Vine composition. The gauging station Lichtenwalde served as the linking variate (variable v in Fig. 1) because of its spatial location. While the trivariate Archimedean copula specifies the three variables by only one function, the paircopula is composed of three bivariate copula functions. Consequently, the model can reproduce the interdependencies in a more detailed way. The fitting of the copulas Wechselburg–Lichtenwalde (C_{WL}) , Lichtenwalde–Nossen (C_{LN}) and Wechselburg–Nossen|Lichtenwalde $(C_{WN|L})$ was executed by the R-package CDVine (Brechmann and Schepsmeier, 2013). The data suggest that the copulas of Joe (C_{WL}) , Gumbel (C_{LN}) and Frank $(C_{WN|L})$ are the most realistic ones. The copulas on the first level (C_{WL} and C_{LN}) both show upper tail dependence. The existence of upper tail dependence is typical for flood data. In the bottom row of Fig. 3 it is obvious, that the pair-copula also reproduces the dependence of the peaks very well. It does so even better than the 3-D-Joe-copula because the variance in the lower range is smaller.

	Pearson's r	Spearman's ρ	Kendall's τ
Wechselburg-Lichtenwalde	0.82	0.67	0.49
Lichtenwalde-Nossen	0.83	0.75	0.56
Wechselburg-Nossen	0.70	0.69	0.51

Table 1. Bivariate correlation coefficient of the coinciding peak values at the gauging stations Wechselburg, Lichtenwalde and Nossen.

Table 2. Return periods based on trivariate Joe-Copula and Pair-Copula for the simultaneous peak discharges at the sites Wechselburg, Lichtenwalde und Nossen for selected flood events; the last column shows the univariate return periods at the gauging station Golzern based on the official flood statistics. The results of the multivariate model are highlighted in bold font.

Flood	$Q_{\text{Wechselburg}}$ (m ³ s ⁻¹)	$Q_{\text{Lichtenwalde}} \pmod{(\text{m}^3 \text{ s}^{-1})}$	Q_{Nossen} (m ³ s ⁻¹)	T _{3-D-Joe-Copula} (a)	T _{Pair-Copula} (a)	T _{Golzern} (a)
Aug 2002	1000	1250	690	556	627	471
Aug 2010	521	136	54	26	26	6
June 2013	1010	774	312	257	274	172

3.3 Frequency analysis, evaluation and comparison of the results

Basin wide flood events always differ in the spatial distribution of the runoff contributions. Therefore, the multivariate probabilities differ, even if the runoff below the confluence (here gauging station Golzern) is similar between flood events. Table 2 specifies the peak values of the last three extreme flood events in the river basin. This shows the relations among the events and, by including the catchment areas, the corresponding core area. So the event in August 2002 had its focus especially in the eastern part of the catchment whereas eight years later the focus was clearly in the western part. Using both copula models we estimated the corresponding return periods. In addition, we determined the univariate return periods of the resulting runoff in Golzern by use of the official local gauge statistic. The table indicates that, overall, the multivariate return periods are higher than the univariate ones. This is because the copula models include the probabilities of the individual catchments and their combination whereas the univariate statistics only relates to the total runoff downstream of the confluence. The spatial composition of the flood peaks are not part of the univariate distribution function. The flood event of August 2010 is a case in point. About 75 % of the total runoff originates in the catchment of the Zwickauer Mulde. This spatial heterogeneity can not be considered in the univariate flood statistic for the gauge Golzern where the peak value was $697 \text{ m}^3 \text{ s}^{-1}$. As the spatial distribution of the flood-causing rainfall (and as a result of the runoff) was unusual, this event has a significantly smaller univariate than multivariate return period. The multivariate return period considers the probability of the occurrence of a certain combination of floods from different tributaries. Thus, the composition of this event has a smaller probability than the peak which could result from several combinations. The composition or the spatial-temporal distribution

of the flood event in 2010 therefore affected the multivariate probability much stronger than the marginal distributions of peak discharges did. Accordingly, the multivariate return period of the peak values is almost four times the univariate return period of the aggregated discharge. With the exception of the 2002 flood event, both copula models give almost the same return periods.

This study shows that both multivariate copula approaches estimate very similar return periods. This indicates that both of them can be adopted for the multivariate statistical assessment of flood events in large river basins. Although the trivariate Joe-copula only has one parameter, it seems not to be worse than the pair-copula, at least not in this application. In addition, the effort of estimating the return periods via 3-D-Archimedean copula is minor. However, the pair-copula should provide better fits to the data because of its more detailed structure and because of considering conditional bivariate dependencies. The copula-generated random samples in Fig. 3 demonstrate that this is the case. The scatter plots generated by the pair-copula show a less distinctive variation in the lower range than the 3-D-Joe-copula.

4 Summary and conclusions

The application of trivariate copula models shows that they are able to estimate the multivariate probabilities of the occurrence of simultaneous flood peaks. They quantify the dependencies of the variates among each other and, consequently, capture the probability of infrequent spatial combinations of extreme events in the resulting return periods. Because of this they provide an suitable instrument for the spatial assessment of flood events within a river basin. The generation of random samples from the copula models suggests that the pair-copula gives a better fit to the data than the trivariate Archimedean Copula because of its smaller variation. The shape of the trivariate distribution seems to be reproduced more realistically. Flood design applications may benefit from this property.

References

- Aas, K., Czado, C., Frigessi, A., and Bakken, H.: Pair-copula constructions of multiple dependence, Insurance: Mathe. Econom., 44, 182–198, 2009.
- Bedford, T. and Cooke, R. M.: Vines a new graphical model for dependent random variables, The Annal. Stat., 30, 1031–1068, 2002.
- Brechmann, E. C. and Schepsmeier, U.: Modeling Dependence with C- and D-Vine Copulas: The R Package CDVine, J. Stat. Softw., 52, 1–27, 2013.
- Chen, L., Singh, V. P., Shenglian, G., Hao, Z., and Li, T.: Flood Coincidence Risk Analysis Using Multivariate Copula Functions, J. Hydrol. Eng., 17, 742–755, 2012.
- De Michele, C., Salvadori, G., Canossi, M., Petaccia, A., and Rosso, R.: Bivariate Statistical Approach to Check Adequacy of Dam Spillway, J. Hydrol. Eng., 10, 50–57, 2005.
- Favre, A.-C., El Adlouni, S., Perreault, L., Thiemonge, N., and Bobée, B.: Multivariate hydrological frequency analysis using copulas, Water Resour. Res., 40, W01101, doi:10.1029/2003WR002456, 2004.
- Genest, C. and Rivest, L.-P.: Statistical Inference Procedures for Bivariate Archimedean Copulas, J. Am. Stat. Assoc., 88, 1034– 1043, 1993.

- Genest, C., Ghoudi, K., and Rivest, L.-P.: A semiparametric estimation procedure of dependence parameters in multivariate families of distributions, Biometrika, 82, 543–552, 1995.
- Ghizzoni, T., Roth, G., and Rudari, R.: Multivariate skew-t approach to the design of accumulation risk scenarios for the flooding hazard, Adv. Water Resour., 33, 1243–1255, 2010.
- Ghizzoni, T., Roth, G., and Rudari, R.: Multisite flooding hazard assessment in the Upper Mississippi River, J. Hydrol., 412–413, 101–113, 2012.
- Joe, H.: Families of m-Variate Distributions with Given Margins and m(m-1)/2 Bivariate Dependence Parameters, Lecture Notes-Monograph Series (28), 120–141, 1996.
- Joe, H.: Dependence Modeling with Copulas, Chapman & Hall/CRC, Boca Raton, Florida, 2014.
- Kao, S.-C. and Chang, N.-B.: Copula-Based Flood Frequency Analysis at Ungauged Basin Confluences: Nashville, Tennessee, J. Hydrol. Eng., 17, 790–799, 2012.
- Nelsen, R. B.: An Introduction to Copulas, 2nd Edn., New York, NY Springer New York, 2006.
- Sklar, A.: Fonctions de rèpartition à n dimensions et leurs marges, Publ. Inst. Stat. Univ. Paris (8), 229–231, 1959.
- Wang, C., Chang, N.-B., and Yeh, G.-T.: Copula-based flood frequency (COFF) analysis at the confluences of river systems, Hydrol. Process., 23, 1471–1486, 2009.