Flood risk mitigation by reservoirs – application of multivariate statistical methods for risk assessments downstream

MARKUS SCHULTE & ANDREAS SCHUMANN
Institute of Hydrology, Water Resources Management and Environmental Engineering, Ruhr University Bochum, D-44780 Bochum, Germany
markus.schulte-3@rub.de

Abstract A flood is an extreme hydrological event which is characterized by several correlated variables (flood peak, volume and duration). The efficiency of flood reservoirs is strongly affected by the event-specific combination of the two correlated characteristics flood peak and flood volume. As flood events with unusual combinations of these characteristics can be more critical than a single design flood, growing emphasis is given to multivariate frequency analysis to consider this uncertainty. Here a new approach to assess the efficiency of a reservoir for downstream flood risk mitigation which considers interactions of tributaries is discussed. It is shown that coincidences of floods which are originating from different sub-basins have a large impact on the efficiency of a reservoir. As the interaction of tributaries depends strongly on the spatial and temporal distributions of rainfall, the event-specific flood risk mitigation by a reservoir stays uncertain. This uncertainty can be described by multivariate statistical analyses. Here it is shown how multivariate statistics can be used to locate reservoirs within river basins under consideration of their capacity to mitigate floods downstream.

Key words flood risk; multivariate statistic; copulas; coincidence

INTRODUCTION

In the past, flood risk management was based on the promise to protect people against negative consequences of floods up to a certain level, specified by a design flood. A design flood was defined as a limit up to which a flood could be controlled completely by technical measures. By definition, a failure of the system was to be expected only in such cases where the design flood is overtopped. Under these assumptions design floods with very small probabilities had to be used. Flood-prone people got the impression that the risk could be neglected as the exceedence probabilities of design floods are often small. There are several critical aspects in this traditional approach. One of them consists in the insufficient specification of design floods. Traditionally, univariate distributions are used to describe them by probabilities of their peaks. The peak is the most important characteristic of a flood event with regard to the hydraulic capacity of river profiles. The volume and duration are very important for protection measures based on flood retention. Often these characteristics are specified by fixed relationships between peaks, volumes and shapes of hydrographs.

The experience of floods at the beginning of the 21st century has shown that 100% safety cannot be achieved by technical means. The remaining risk of failures demands improved considerations of multiple and interacting hydrological loads. Unusual and unexpected combinations of flood characteristics have to be considered. It is easy to specify events which are most critical for existing or planned reservoirs, but more difficult to estimate their probabilities. Here we have to consider the joint distribution of several correlated random variables. Several approaches were developed to apply multivariate statistics. The limitations of these approaches resulted mainly from the need to use the same type of marginal distributions to combine them into bivariate distributions and to apply the Pearson's correlation (e.g. Hiemstra, 1976; Bergmann & Sackl, 1989; Goel *et al.*, 1998; Yue, 2001). These limitations were vanquished by introducing the theory of copulas into hydrology. Several publications discuss applications of copulas for multivariate flood frequency analysis (e.g. Favre *et al.*, 2004; Grimaldi & Serinaldi, 2006; Zhang & Singh, 2007; Salvadori & De Michele, 2010). Especially design floods for reservoirs are more and more specified by means of multivariate statistics to consider the dependencies between peak and volume (e.g. De Michele *et al.*, 2005; Grimaldi & Serinaldi, 2006; Klein *et al.*, 2010). With these applications the question of the probabilistic meaning of copulas became evident. The basic concept of return periods especially needs clarification. There are several approaches to estimate multivariate quantiles and return periods and to identify suitable design events (Chebana & Ouarda, 2009; Salvadori *et al.*, 2011; Volpi & Fiori, 2012) which will be discussed in the next section.

Beside the theoretical problem of the probabilistic meaning of copula-based return periods, here the potential application of copulas for spatial flood problems is discussed. One example is flood mitigation by a reservoir where the downstream effects has to be specified. The impact of a flood reservoir depends here on the coincidences of flood waves originating from tributaries. The changes of flood characteristics have to be comprised at the points of confluence. In this case the interactions of tributaries and their differences with regard to the flood characteristics, but also the timing of flood peaks have to be considered. After a short introduction to the applications of copulas for flood risk assessments, some of the typical questions in planning of flood mitigation by reservoirs are discussed.

To show how copulas can be applied to estimate the probability of flood coincidence, to evaluate the remaining risk and to support the decision of reservoir location, a case study is presented below. The coincident flood peaks of two joining tributaries were analysed under the condition of a reservoir installed in one of the watersheds. The bivariate probabilities of the three cases: (a) without reservoir, (b) reservoir in watershed A, and (c) reservoir in watershed B can be compared. The application demonstrates how assessments of reservoir efficiencies depend on the targets of flood mitigation. It is shown that not the absolute size of a reservoir is important, but the relationship between flood probabilities and reservoir effectiveness to reduce flood peaks. In addition it becomes apparent that the copulas include both the assumed reservoir capacity and control, and that the results depend strongly on the threshold of non-damaging discharge downstream. Finally the paper presents a brief summary and some conclusions.

ESTIMATION OF EXCEEDENCE PROPABILITIES WITH COPULAS

The hydrological flood risk in the univariate case is specified by the probability that the random variable X exceeds a critical design value x. Critical situations are characterized by X > x. In the bivariate case, where we have to consider two variables, the dangerous region is two-dimensional. Events where two variables exceed critical thresholds (X > x; Y > y) can be constructed in two different ways (Salvadori & De Michele, 2004):

– the "OR" case is given when at least one of the components exceeds a prescribed threshold (X > x v Y > y)
– the "AND" case where both components exceed their thresholds simultaneously (X > x ʌ Y > y)

If a design flood has to be specified, e.g. by joint return periods of flood volume and flood peak, there are two ways to estimate it: under the condition that events cause damage if either peak or volume exceeds a certain magnitude (logical "OR") and if both, volume and peak, must exceed a certain magnitude (logical "AND") (Shiau, 2003).

The method of copulas is based on the theorem of Sklar (1959). It expresses the connection between the copula-function *C* and the bivariate distribution functions of two correlated random variables *X* and *Y*, which can be described separately by their (univariate) marginal cumulative density functions $F_X(x)$ and $F_Y(y)$:

$$F_{X,Y}(x,y) = C\left[F_X(x), F_Y(y)\right] = P(X \le x, Y \le y) \tag{1}$$

This equation describes the probability that both characteristics stay below *x* and *y* within one unit of time (here mostly one year). The probability (and the corresponding joint return period) with which either *X* **or** *Y* exceed their thresholds is defined in equations (2) and (3).

$$P_{X>x \vee Y>y} = 1 - F_{X,Y}(x,y) = 1 - C\left[F_X(x), F_Y(y)\right] \tag{2}$$

$$T_{X>x\vee Y>y} = \frac{1}{P(X>x\vee Y>y)} = \frac{1}{1-C\left[F_X(x),F_Y(y)\right]} \tag{3}$$

The second case (the probability with which X **and** Y both exceed x and y) is defined in equation (4). It specifies the inter-arrival times between two such successive events.

$$P_{X>x\wedge Y>y} = 1-F_X(x)-F_Y(y)+F_{X,Y}(x,y) = 1-F_X(x)-F_Y(y)+C\left[F_X(x),F_Y(y)\right] \tag{4}$$

$$T_{X>x\wedge Y>y} = \frac{1}{P(X>x\wedge Y>y)} = \frac{1}{1-F_X(x)-F_Y(y)+C\left[F_X(x),F_Y(y)\right]} \tag{5}$$

This return period is higher or equal to the maximum of the two univariate return periods T_X and T_Y. As becomes evident from equations (3) and (5), the return periods are estimated as the inverse of the exceedence probabilities. Thus the critical regions can be specified, as shown in Fig. 1.

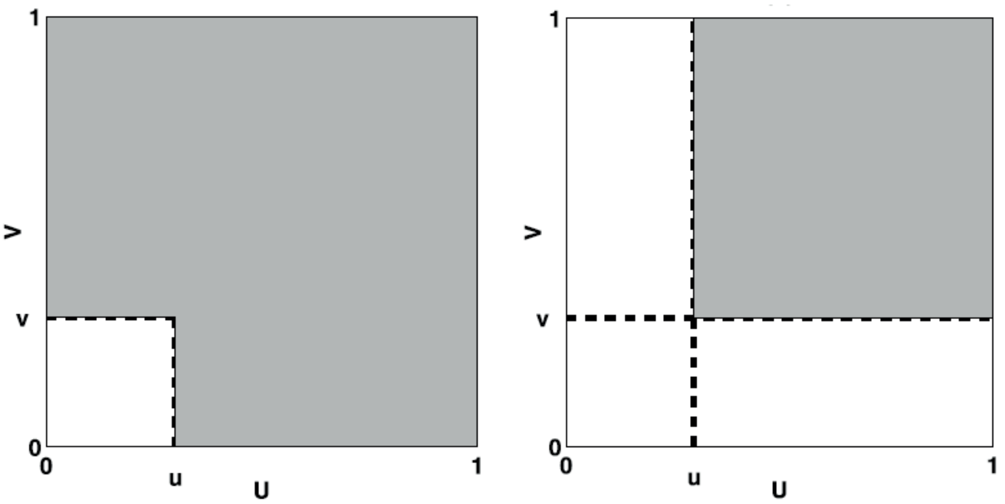

Fig. 1 Differences of critical regions, at the left the case "OR" (x > x* v y > y*), at the right the case "AND" (x > x* ʌ y > y*) (Salvadori & De Michele, 2004).

If a range can be specified by critical design values x* and y* two different regions exists: one where both variables are exceeding these critical values (x > x* ʌ y > y*) and the other one where only one of these variables are exceeding a threshold (x > x* ⊕ y > y*).

Contour lines of the return periods $T^{\wedge}$ and $T^{\vee}$ separate the subcritical from the supercritical region (Fig. 2). Though Salvadori & De Michele (2004) pointed out that for risk analysis another return period should be computed, namely the secondary return period in the OR-case. The problem they fixed is that every pair on the isoline has a different subcritical area. But equation (3) is based on equation (1), which specified the non-exceedence probability associated with the subcritical area. So they suggest the determination of a return period in the OR-case that "provides the average time between the occurrence of two supercritical events" whereas the primary return period "predicts that a critical event is expected to appear once in a given time interval" (Ganguli & Reddy, 2013). For further information of computing secondary return periods, see Salvadori *et al.* (2007). Secondary return periods can be used in applications of bivariate peak and volume observations, where either a certain peak value or a certain volume leads to an overtopping of the spillway. In the subsequent case study bivariate copulas of flood peaks are analysed, describing a problem that is more linked with the AND-case, so secondary return periods are neglected below.

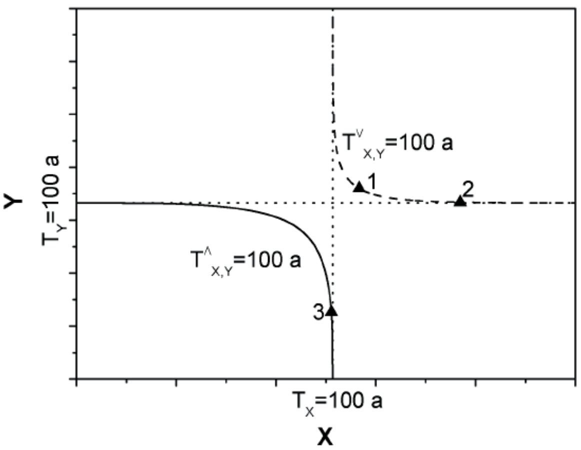

Fig. 2 Contour lines of the return periods $T^{\wedge}$ and $T^{\vee}$ (Klein, 2009).

Figure 2 shows that the isolines have a curvature, the section where the bivariate density offers its highest values. Chebana & Ouarda (2009) and Volpi & Fiori (2012) suggest defining a region in the bivariate space, which contains all reasonable pairs of values for rating purposes. The possibility to distinguish between a proper and a naive part in a copula could be very useful, but could be neglected when both variables have a strong correlative connection, because in this case the fitted copula should generate proper values more-or-less anyway.

Planning of flood mitigation by reservoirs

Flood prevention by reservoirs depends on the relationship between storage capacity and flood event. Subject to the reservoir capacity small floods can be absorbed completely whereas huge flood events are mitigated. If the reservoir's effort is too small it leads to an overtopping of the spillway until the reductions of peak flows can be neglected. Generally we distinguish between a retention basin ("wet pond") and a detention basin ("dry pond"), and in addition if a reservoir is regulated or not.

A flood reservoir has to be adapted for the upstream catchment: normally a larger catchment requires a larger reservoir. This is the main condition for effective downstream flood protection. But the effect of a reservoir on flood mitigation is decreasing with a growing intermediate catchment as floods from tributaries reach the channel. These flood events from tributaries can occur when the upper reservoir retains a wave as well as when there no flood event exists in the regulated catchment. In both cases the spatio-temporal distribution of the precipitation is a huge random factor for the resulting flood event. Besides this spatial heterogeneity it has to be taken into account that an (overloaded) reservoir could affect the downstream flood situation negatively when the peak is delayed. So a delayed flood peak could superpose with a flood peak of a downstream tributary. To consider this risk of interaction, the flood peaks should be viewed in a multivariate way. To locate a planned reservoir within a river basin, the decision variable is based on observed downstream discharges. This discharge is composed of the joining tributaries. Attention should be paid to the fact that one specific discharge value can result out of many combinations of runoff from different tributaries.

So it is essential to combine the marginal distributions of both tributaries under consideration of their correlation. To handle this problem copulas are an appropriate instrument. They can estimate the probability for a certain pair of values. Thus it is easy to evaluate the change of a marginal due to a (planned) reservoir. The modified copula, based on the new bivariate sample, specifies the modified probabilities. The decision maker can compare the resulting return periods for different reservoir locations and capacities according to the downstream flood mitigation targets. A sample application for such a case is given in the following case study.

Case study

The main intention of the case study is to estimate the variation of bivariate probabilities of flood peaks from two joining tributaries on condition that a reservoir is installed in one of the watersheds changing the local flood peaks. Generally the aim is to reduce the flood risk downstream of the point of confluence. The flood peaks from both tributaries are correlated as can be seen on the left in Fig. 3. Here, two watersheds "A" and "B" are used with areas around 140 km^2 which are located in estern Germany. The corresponding correlation coefficients by Pearson, Kendall and Spearman are 0.89, 0.65 and 0.82.

The downstream discharge results from an addition of peaks originating from both watersheds. A certain value could result from a multitude of discharge combinations of the two tributaries. Figure 3 (right) displays this fact for a critical peak discharge of 100 m^3/s. The corresponding return periods in the proper range of values vary from 40 to 60 years.

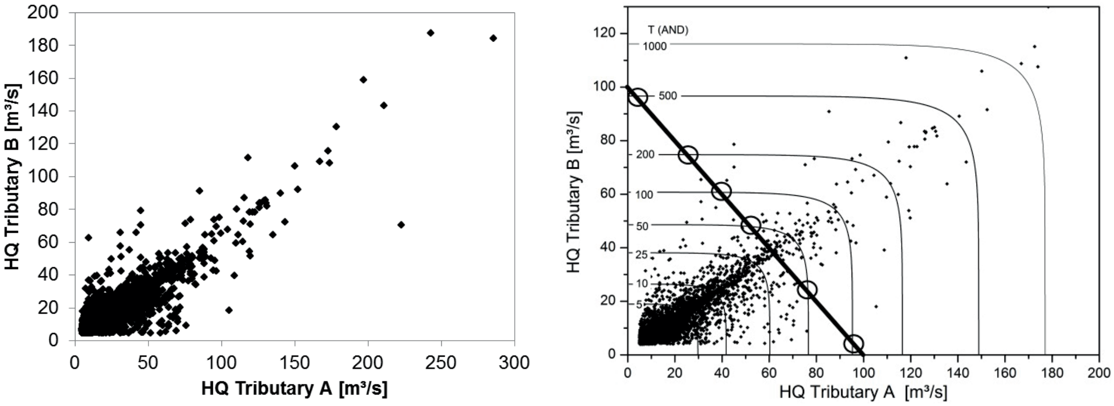

Fig. 3 Left: Simultaneous flood peak values in joining tributaries; right: Critical range and possible combinations of joining discharges for a downstream threshold of 100 m^3/s.

The large sample results from simulations using a stochastic simulator of spatially distributed rainfall (24 hourly events) in combination with a calibrated hydrological model. Only events whose peak values exceed the 2.5-fold mean discharge were selected for further consideration. So an initial sample of 4447 pairs of simultaneous peak values exists. Due to a reservoir in one of the two watersheds, the corresponding univariate sample would be affected. Not only the distribution of flood peaks at this site would be changed, but also the correlation between the flood peaks of both tributaries would be affected. A preliminary design has shown that the flood reservoirs would reduce the peaks proportional to the peak of the inflow and decrease the peak with a return period of 20 years, e.g. by 50%. Higher flood peaks would be reduced less. Discharges smaller than a certain threshold (tributary A: 10 m^3/s, tributary B: 8 m^3/s) would be reduced to an outflow of 7 or 5 m^3/s, respectively. No reduction of the inflowing flood peak can be expected for events with flood peaks with return periods above 160 years. But the effect of debasing can be almost neglected at return periods above 100 years. The mitigation for peak values was estimated from reservoir simulations and is visualized by distribution functions in Fig. 4. Both fictional reservoirs operate in a similar way, but because of the higher peaks in tributary A this location is more effective absolutely. As an intermediate step we have three bivariate samples: the initial sample without reservoir, sample A (reservoir in catchment A) and sample B (reservoir in catchment B). As shown in Fig. 4, marginal distributions were estimated for all univariate samples. The fictional reservoirs do not influence the upper tail of the marginals. So we use composite distribution functions for the modified samples. Here the lower ranges of values are represented by Pearson type III distributions. To describe the upper range we go back to the marginal distributions of the initial sample (Log-Pearson III Distribution).

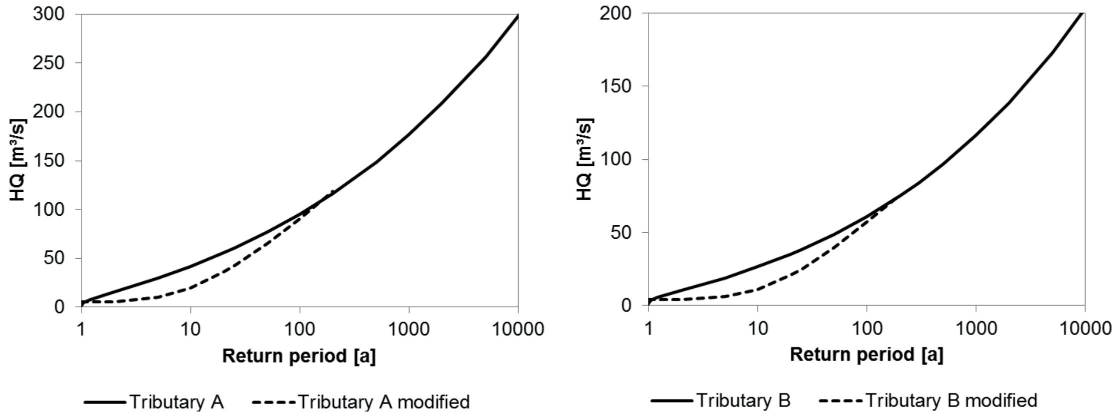

Fig. 4 Marginal distribution functions for the (modified) peak values at tributary A and B

Table 1 Correlation coefficients of the three bivariate samples.

	Pearson's r	Kendall's τ	Spearman's ρ
Initial sample	0.89	0.65	0.82
Sample A	0.86	0.66	0.82
Sample B	0.84	0.66	0.87

The interdependencies are characterized by copula functions. Although the reservoirs modified the bivariate samples, the correlation between the peak values is still distinctive as Table 1 pointed out. Fitting several different copula functions, the type of BB1 copula (Joe, 1997) emerged as the appropriate one in all three cases.

It must be emphasized that within the cut range (the range of peak values which are cut down by the reservoir's regulation to a constant value of 7 or 5 m^3/s), the primal relationship between the watersheds is lost. To give an example: in tributary A both peak values of 7.2 and 9.8 m^3/s are reduced by the flood control to 7 m^3/s. But a value of 9.8 m^3/s correlates with higher flood peaks in tributary B than a value of 7.2 m^3/s. So the rank correlation could be disordered and the copula cannot be adjusted sufficiently. Because of this, the probability of such a (censored) event has to be estimated by the initial copula, keeping in mind that the absolute value of the modified sample must cut down to the respective target value. So the mutual link between the tributaries would be conserved: an original higher peak value from the censored sample is still connected to an (averaged) higher peak value in the unaffected tributary. Figure 5 shows the scatter plot of the modified samples and a random sample by the fitted copula. In the following examination bivariate return periods based on the AND-case (equation (5)) are used, because the certain critical target value is made up of the discharge of both tributaries.

Based on these models the impacts of reservoirs on peaks at the downstream confluence were analysed. In the case study this was done for two different flow targets (discharges of 100 and 130 m^3/s) with the assumptions mentioned above. The results are presented in Fig. 6. There are many different combinations of flood peaks in both tributaries which are exceeding the critical thresholds. The results demonstrate that the choice of the target downstream discharge has an impact on the changes of return periods (not on the absolute discharge!). As high flood peaks from tributary B are less probable than those from tributary A, the bivariate return periods for the higher target discharge is always higher for the case "reservoir in B". If the critical value is lower (e.g. $Q_{critical}$ = 100 m^3/s) the reservoir in A outperforms the reservoir in B for peak values from catchment A above 65 m^3/s. Above this threshold the (smaller) discharge values originating from the non-modified watershed B are more probable and the higher runoff peaks from watershed A are less probable (due to impact of the local reservoir), thus the marginal distribution of catchment A dominates the joint distribution at this range. This is different from the case with a higher target (e.g. $Q_{critical}$ = 130 m^3/s). Here, the corresponding values from tributary A for small (realistic) peak values from channel B, are already in such range, that the local reservoir A would affect the peaks considerably. Accordingly the return periods assimilate for all reviewed combinations.

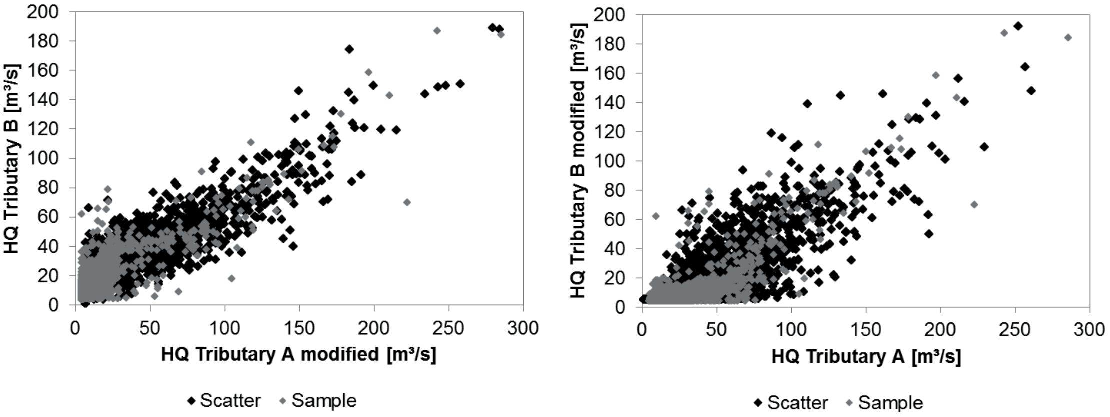

Fig. 5 Modified bivariate samples and random samples generated by fitted copulas.

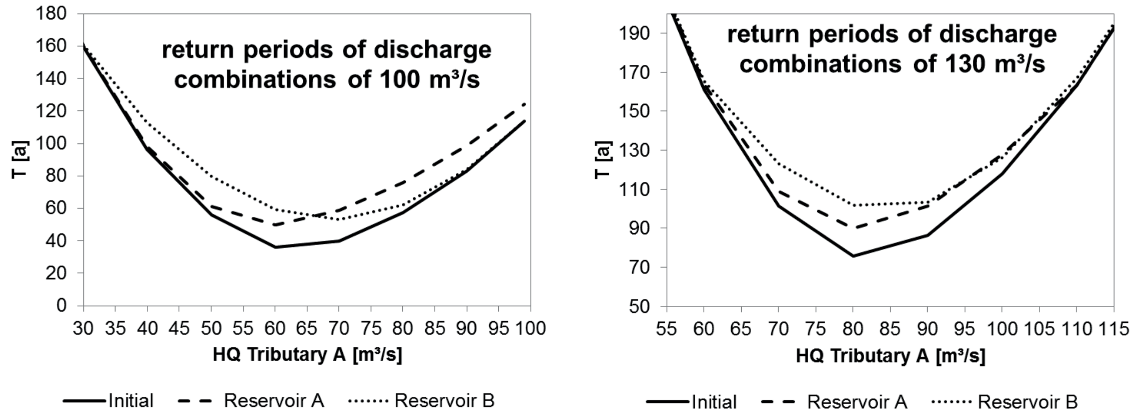

Fig. 6 Return periods of different combinations of flood peaks for downstream target discharges of 100 and 130 m^3/s; the discharge from tributary B follows from $HQ_A + HQ_B = 100$ and $HQ_A + HQ_B = 130$ m^3, respectively.

The study demonstrates a crossover of return periods, depending on the reservoirs' capacities and regulations. Determined by the interactions of the planned reservoirs with their watersheds the modifications of return periods for certain downstream target values could strongly differ. The resulting joint probability distribution of summarized flood peaks is, in different sections, dominated by a single marginal distribution only. This is especially the case if a single watershed contributes only a small flood rate (or a very large flood rate) to a point of confluence. The decisive marginal is the distribution which determines the smaller univariate probability for a certain combination of values. In the range of crossover, the capacity and regulation of the reservoirs become the determining factors. Here a small reservoir with an optimized regulation could be able to operate more effectively than a larger one. Furthermore it could be that the tributary with smaller floods (offering the smaller reservoir) is dominant for risk mitigations downstream even if a reservoir there is less effective in its absolute flood reduction. The probability for the occurrence of a combined event could be much smaller simply because the local peak values at this site are much smaller in general.

SUMMARY AND CONCLUSIONS

The applications of copulas in reservoir planning are often limited to peak–volume relationships of design floods. As reservoirs modify the spatial characteristics of flood events, coincidences of floods from tributaries should also be considered by means of multivariate statistics. This study demonstrated this option by an investigation of variations of bivariate return periods resulting from installing a reservoir in one of two confluent tributaries. To consider the present correlation and to estimate the bivariate probabilities, the technique of copulas was introduced. It becomes apparent that there exists a range of crossover of the return periods, depending of the reservoirs capacities and regulations. Due to the critical downstream target value either the one or the other reservoir could reduce the exceedence probabilities more effectively. This results from the multivariate characteristic of the critical target value which can be built-up out of many possible combinations. It could be demonstrated that the complexity of interactions between watersheds requires the extensive consideration of diverse combinations of superposed peak discharges. The statistical assessment of these multiple loading cases demands a consideration of the existing inter-dependencies. Here copulas are useful tools. A decision maker, who is interested in the adequate reservoir location, could use the introduced procedure to balance between different options. However, it has to be mentioned that the interpretation of the results becomes more challenging. A lot of additional conditions have to be considered, e.g. the flood events forming the basis of this statistic are caused by a multitude of possible spatio-temporal distributions of precipitation. Nevertheless the method presented here is one proper way to support the search for suitable sites and thus to reduce the remaining risk.

REFERENCES

Bergmann, H. & Sackl, B. (1989) Determination of design flood hydrographs. In: *New Directions for Surface Water Modeling* (Proc. Baltimore Symp., May 1989), 261–269. IAHS Publ. 181. IAHS Press, Wallingford, UK.

Chebana, F., & Ouarda, T. B. M. J. (2009) Index flood-based multivariate regional frequency analysis. *Water Resour. Res.* 45(10).

De Michele, C., Salvadori, G., Canossi, M., Petaccia, A. & Rosso, R. (2005) Bivariate statistical approach to check adequacy of dam spillway. *J. Hydrol. Engng ASCE* 10(1), 50–57

Favre, A.-C., El Adlouni, S., Perreault, L., Thiemonge, N. & Bobee, B. (2004) Multivariate hydrological frequency analysis using copulas. *Water Resour. Res*. 40 (1).

Ganguli, P. & Reddy, M. J. (2013) Probabilistic assessment of flood risks using trivariate copulas. *Theoretical and Applied Climatology* 111, 341–360.

Goel, N. K., Seth, S. M. & Chandra, S. (1998) Multivariate modeling of flood flows. *J. Hydrol. Engng ASCE* 124(2), 146–155.

Grimaldi, S. & Serinaldi, F. (2006) Design hyetograph analysis with 3-copula function. *Hydrol. Sci. J.* 51(2), 223–238.

Hiemstra, L. A. V., Zucchini, W. S. & Pegram, G. G. S. (1976) Method of finding family of run hydrographs for given return periods. *J. Hydrol.* 30(1-2), 95–103.

Joe, H. (1997) *Multivariate Models and Dependence Concepts*. Chapman and Hall, New York.

Klein, B. (2009) Ermittlung von Ganglinien für die risikoorientierte Hochwasserbemessung von Talsperren. Schriftenreihe Hydrologie/Wasserwirtschaft (24), Ruhr-Universität Bochum.

Klein, B., Pahlow, M., Hundecha, Y. & Schumann, A. (2010) Probability analysis of hydrological loads for the design of flood control systems using copulas. *J. Hydrol. Engng ASCE* 15(5), 360–369.

Salvadori, G. & De Michele, C. (2004) Frequency analysis via copulas: theoretical aspects and applications to hydrological events. *Water Resour. Res*. 40(12).

Salvadori, G., De Michele, C. *et al.* (2007) Extremes in nature: an approach using copulas. *Water Science and Technology Library*, 56, Springer, Dordrecht.

Salvadori, G. & De Michele, C. (2010) Multivariate multiparameter extreme value models and return periods: a copula approach. *Water Resour. Res*. 46(10).

Salvadori, G., De Michele, C. & Durante, F. (2011) On the return period and design in a multivariate framework. *Hydrol. Earth System Sci.* 15(11), 3293–3305.

Shiau, J. T. (2003) Return period of bivariate distributed extreme hydrological events. *Stochastic Environmental Research and Risk Assessment* 17(1–2), 42–57.

Sklar, A. (1959) Fonctions de reparttion à n dimensions et leura Marges. *Publ. Inst. Stat. Univ. Paris* 8, 220–231.

Volpi, E. & Fiori, A. (2012) Design event selection in bivariate hydrological frequency analysis. *Hydrol. Sci. J.* 57(8), 1506–1515.

Yue, S. (2001) A bivariate gamma distribution for use in multivariate flood frequency analysis. *Hydrol. Processes* 15(6), 1033–1045

Zhang, L. & Singh, V. P. (2006) Bivariate flood frequency analysis using the copula method. *J. Hydrol. Engng ASCE* 11(2), 150–164.