4th International Symposium on Flood Defence:
Managing Flood Risk, Reliability and Vulnerability
Toronto, Ontario, Canada, May 6-8, 2008

Institute for Catastrophic Loss Reduction

PROBABILISTIC ANALYSIS OF HYDROLOGICAL LOADS TO OPTIMIZE THE DESIGN OF FLOOD CONTROL SYSTEMS

B Klein[1], M Pahlow[1], Y Hundecha[1], C Gattke[1], and A Schumann[1]
1. Institute of Hydrology, Water Resources Management and Environmental Engineering, Ruhr-University Bochum, Germany

ABSTRACT: Recent severe flood events prompt for considering a broad range of different hydrological scenarios in the design of flood protection structures such as flood control reservoirs and polders, instead of using the traditional approach of considering one single design flood of a predefined return period only. In this paper a method to categorize generated hydrological loads is presented, which is then applied for the analysis, optimization and extension of the flood control system of the Unstrut watershed in Mid-East Germany. The spatial structure of the flood events is analyzed by the joint probability of the inflow peaks of two dams located downstream of the northern and southern main tributaries. For the design of flood detention structures it is important to consider the flood volume apart from the flood peak. Therefore the joint probability of corresponding flood peaks and volumes are used to categorize the flood events. Hereby the copula-method is used for the construction of the bivariate distribution function.

Key Words: 2-copulas, Flood frequency, Probability distributions

1. INTRODUCTION

The experience of recent flood events shows that it is important to employ a broad range of different hydrological loads for the design of flood protection structures such as flood control reservoirs and polders, instead of using the traditional approach of considering one single design flood of a predefined return period only. In most cases the return period is defined by a univariate analysis of a random variable such as the flood peak. This could lead to an over- or underestimation of the risk associated with a given flood. Therefore it is advantageous to define the return period of an event by a joint probability of the relevant flood properties. For the flood design of dams e.g. the flood volume and the flood peak are important and therefore the return period should be defined by the joint probability of these two variables. In the past much attention has focused on univariate flood frequency analysis in terms of flood peaks (e.g. Hosking and Wallis, 1997; Stedinger et al., 1993). However, flooding is a multivariate event that is characterized by random variables such as flood peak, volume, shape and duration of the hydrograph. One reason why multivariate models are rarely used in hydrological analysis is that they require considerably more data than the univariate case. Therefore the application of multivariate analysis of correlated random numbers in practice is mainly reduced to the bivariate case. In this application data for bivariate analysis were generated stochastically. In the literature bivariate distributions are often applied to bivariate frequency analysis. For example Singh and Singh (1991) derived the bivariate exponential distribution for bivariate frequency analysis. The traditional multivariate analyses used in several studies has a number of drawbacks. Those are mainly applicable in the bivariate case and the random variables have to follow the same type of marginal distribution. However, in reality the flood variables do not have the same marginal distributions and the dependence structure is generally different from the Gaussian case, described by Pearson's correlation coefficient, because the variables do not follow a normal distribution in general.

These problems can be avoided using copulas to model the dependence between correlated random variables. A joint distribution of correlated random variables can be expressed as a function of the univariate marginal distributions. This function is called copula. Copulas are able to model the dependence structure independently from the marginal distributions. Several models are available to model this dependence. Therefore it is possible to build multidimensional distributions with different margins. In the following we shall mainly concentrate on the bivariate case. Recently, copulas have been used for bivariate frequency analysis in hydrology (e.g. Salvadori and DeMichele, 2004). In this paper copulas are applied for bivariate frequency analysis of flood peak and the corresponding volume and for the analysis of coincidences of flood peaks from two tributaries within a river basin.

2. COPULAS

In order to express a bivariate distribution function for the two correlated random variables X and Y with their univariate marginal cumulative density functions $F_X(x)$ and $F_Y(y)$ the link between copulas and the joint distribution is provided by the theorem of Sklar (1959)

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

where $F_{X,Y}(x,y)$ is the joint cumulative distribution function CDF of the random variables. Because copulas are invariant under strictly increasing transformations of X and Y we may concentrate on the two uniformly distributed random variables U and V on [0,1], defined as $U = F_X(x)$ and $V=F_Y(y)$. C is the copula function C: $[0,1]^2$ ➔ [0,1] such that C=0, if at least one argument is zero and C(u,1)=u and C(1,v)=v. Joe (1997) and Nelsen (2006) summarize several types of copulas, such as Archimedian copulas.

3. BIVARIATE FREQUENCY ANALYSIS

The joint distribution of Equation [1] is the probability $P(X \le x, Y \le y) = F_{X,Y}(x,y)$ that X and Y are less or equal than the specific thresholds x and y.

Given the assumptions of total independence between the two random variables the joint probability becomes the product of the two individual probabilities $F_X(x)$ and $F_Y(y)$.

If there are dependencies the probability of both exceeding x **and** y is defined as

[2] $P(X \ge x \wedge Y \ge y) = 1 - F_X(x) - F_Y(y) + F_{X,Y}(x,y) = 1 - F_X(x) - F_Y(y) + C\left[F_X(x), F_Y(y)\right]$,

with the corresponding joint return period in which x **and** y are exceeded being expressed as

[3] $T^{\wedge}_{X,Y} = \dfrac{1}{P(X \ge x \wedge Y \ge y)} = \dfrac{1}{1 - F_X(x) - F_Y(y) + C\left[F_X(x), F_Y(y)\right]}$.

The probability of X and Y exceeding the thresholds x **or** y is defined as

[4] $P(X \ge x \vee Y \ge y) = 1 - F_{X,Y}(x,y) = 1 - C\left[F_X(x), F_Y(y)\right]$,

with the joint return period in which either x **or** y are exceeded written as

$$[5] \quad T^{\vee}_{X,Y} = \frac{1}{P(X \geq x \vee Y \geq y)} = \frac{1}{1 - C\left[F_X(x), F_Y(y)\right]}.$$

4. APPLICATION

The flood prone watershed of the river Unstrut is situated in Mid-East Germany with a catchment area of 6343 km² (see Figure 1). Due to recent flooding with severe damage the flood control system is analyzed, optimized and extended through an integrated and interdisciplinary flood risk assessment instrument. To consider the complexity different hydrological scenarios were selected to encompass a large variety of possible flood events for analysis of the flood control structures. For the multi-criteria risk-analysis the events have to be categorized by probability. In total 30 generated events with return periods of the peak discharges between 25 and 1000 years and one event with a return period larger than 1000 years were selected. Two different bivariate probabilities are used for the risk-analysis:

1. The joint probability of coincident floods at dam sites of the two main tributaries to include the spatial variability of the events in the analysis.

2. The joint probability of the flood peak and corresponding volume at the two reservoirs. For the analysis of flood detention systems the flood volume is very important and therefore its probability should be considered in the risk-based analysis as well as the peak.

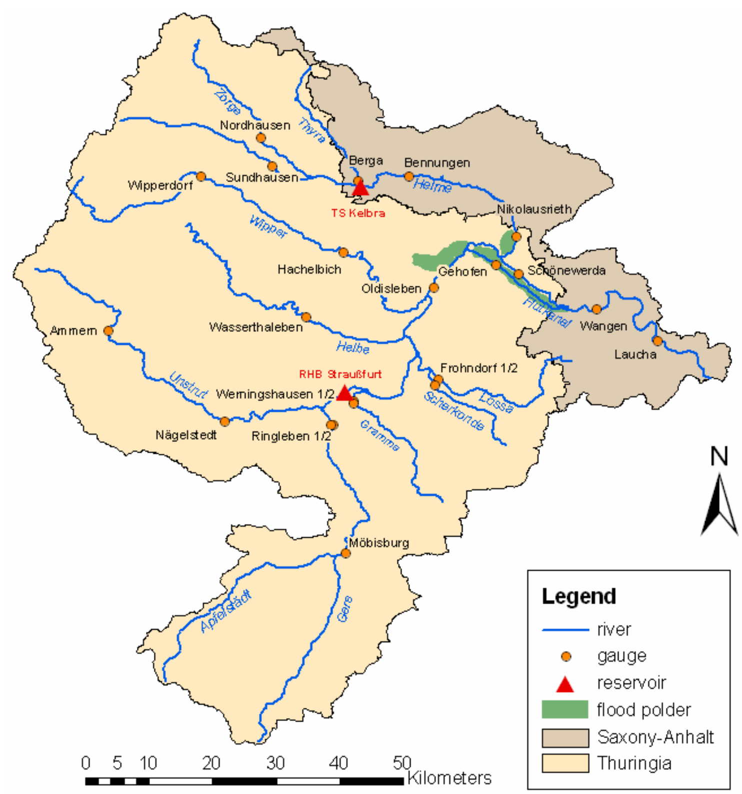

Figure 1: Map of the watershed of the river Unstrut

4.1 Generation of the flood events

Selecting a risk-based approach for the analysis, optimization and extension of the flood control system a broad range of possible hydrological scenarios should be considered. The data base in the river basin was limited to an observation period of some decades, influenced by construction of reservoirs since the 1960's. To extend this time series and to provide the information needed to estimate copulas a long-term discharge time series of 10000 years was generated. This long-term time series was derived from coupling a stochastic rainfall generator with an adapted semi-distributed rainfall-runoff-model following the HBV concept. Events for the risk-analysis are selected and categorized through evaluation of their statistical properties.

4.2 Bivariate analysis of the corresponding annual flood peaks

It is important to characterize the spatial distribution of the flood event because of the topographic structure of the catchment with flooding originating in the mountainous regions upstream of the two dams. Three flood situations should be considered in the risk-analysis: (1) precipitation in the mountainous region in the North, (2) precipitation in the mountainous region in the South and (3) precipitation in both mountainous regions.

The inflow gages of the two flood reservoirs situated within the two main tributaries are used as reference location to quantify the probability. These two dams represent the upper boundary of the flood control system of the Unstrut watershed. All other flood protection structures are located downstream of these two dams. To describe the overall probability for the entire region in terms of the different spatial distributions of the flood event, the joint probability between the two flood peaks from the inflow to the two dams is described by copulas. The annual flood peaks of the two gages and the corresponding flood peak of the other gage were chosen for the analysis. In 2767 years from the 10000 year time series the annual flood peaks of the two gages did not occur at the same time. Therefore the sample size is 12767 for a time series of 10000 years. The first step for the bivariate frequency analysis is the estimation of the marginal distributions. It was found that both variables follow a Generalized Extreme Value (GEV) distribution.

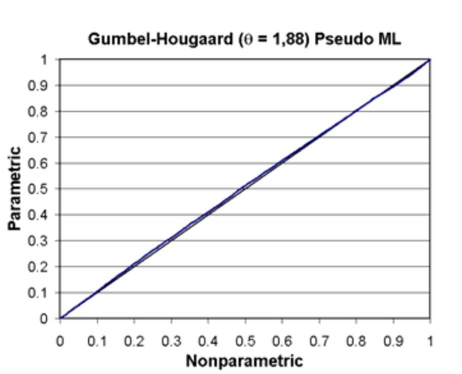

Figure 2: Comparison of nonparametric and parametric estimations of K(t) for the fitted Archimedian Gumbel-Hougaard-copula

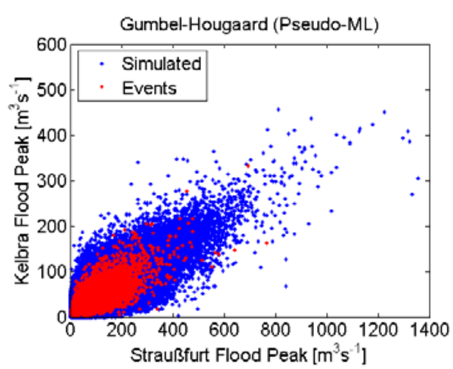

Figure 3: Simulated random sample of size 1000000 chosen from the Gumbel-Hougaard copula and the events from the generated time series

The next step is to identify the appropriate copula to obtain the joint distribution of the two variables, since various copula models can potentially be used to model the dependence between the two random variables. Genest and Rivest (1993) described a procedure for identification by the comparison of the parametric and the nonparametric estimate of

[6] $K(t) = \left\{(u,v) \in [0,1]^2 : C(u,v) \le t\right\}$.

The nonparametric estimate of K could be calculated e.g. after Genest and Rivest (1993).

Furthermore, the analysis revealed that among the four mainly in hydrology applied Archimedian copulas Ali Mikhail Haq, Frank, Cook-Johnson and Gumbel-Hougaard copula (Nelson, 2006), the Gumbel-Hougaard represents best the bivariate distribution of the correlated annual flood peaks of the two gages (see Figure 2). As a further test, the margins of 1000000 random pairs (u_i,v_i) chosen from the copula and transformed back into the original units using the marginal distributions $F_X(x)$ and $F_Y(y)$ are compared with the sample values (x_i,y_i). In Figure 3 the scatter plot of the Gumbgel-Hougaard copula is shown. It can be seen that the Gumbel-Hougaard copula could model the dependence structure of the sample quite well.

Two parameter copulas introduced by Joe (1997) were analyzed besides the four Archimedian copulas. The two-parameter copula BB1 fitted best and was therefore chosen for a comparison with the Gumbel-Hougaard copula.

The advantage of two parameter copulas is that they might be used to capture more than one type of dependence. One parameter models the upper tail dependence (in the case of BB1-copula the parameter δ) and the other parameter models the lower tail dependence. The parameters are estimated by the maximum pseudolikelihood method. The comparison of the parametric and nonparametric values of K(t) shown in Figure 4 and the scatter plot in Figure 5 suggests, that the BB1 fits the dependence structure of the data well. The parametric values of K(t) for the BB1 copula are calculated using numerical methods.

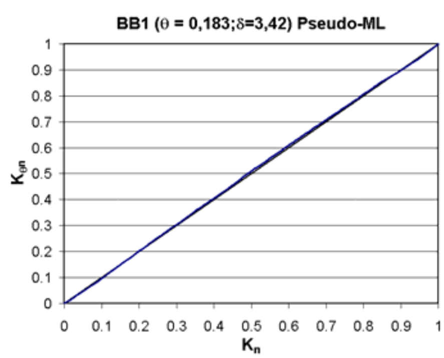

Figure 4: Comparison of nonparametric and parametric estimations of K(t) for the fitted BB1-copula

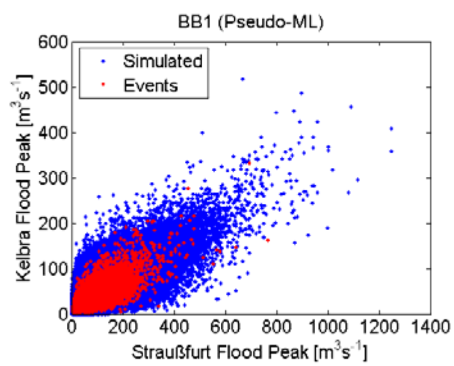

Figure 5: Simulated random sample of size 1000000 chosen from the BB1-copula and the events from the generated time series

Using the multidimensional goodness-of-fit test following Kolmogorov-Smirnov for both copulas the null hypothesis that the flood peaks are drawn from the corresponding copula are not rejected at the significant level of 5 percent. The Gumbel-Hougaard copula is chosen for further bivariate frequency analysis because the KS-value is smaller than the corresponding value of the BB1 copula.

Figure 6 illustrates both the joint return periods $T^{\wedge}_{X,Y}$ in which x and y are exceeded and the joint return periods $T^{\vee}_{X,Y}$ in which either x or y are exceeded. The chosen events represent a large variety of different hydrological scenarios. Using the selected events (red triangles) with a flood peak of about 100 year return period at the reservoir Straußfurt for the design of flood protection structures, the corresponding return periods of the flood peaks at the dam Kelbra range between 10 years and 500 years.

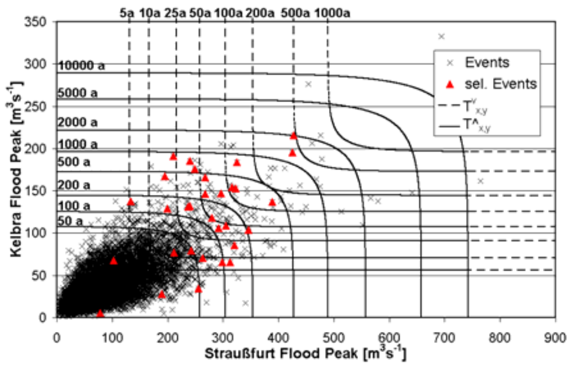

Figure 6: Joint return periods of the annual flood peaks of the two gages $T^{\vee}_{X,Y}$ (exceeding x or y) and the $T^{\wedge}_{X,Y}$ (exceeding x and y)

4.3 Bivariate analysis of the annual flood peaks and the corresponding volumes

The flood peak plays a fundamental role in the design of flood control structures. Therefore it is often used to assign a return period for an event. However, for flood detention structures such as dams and polders the flood volume is also very important. Therefore the hydrologic risk could be underestimated by using only the flood peak as probability measure. For example an event with a very large peak and a small volume could be stored in the flood control storage of a dam and an event with a smaller peak but a larger volume could lead to flood spillage. Generally, flood peak and flood volume are two statistically dependent variables. Here the joint return period of the corresponding flood peak and volume is used to assign a return period to the flood events at the reservoir Straußfurt. From the 10000 years of data the corresponding annual flood peaks and flood volume are selected for yearly floods. To estimate the flood volume the start of the surface runoff is marked by the abrupt rise of the hydrograph and the end of the flood runoff can be identified by the flattening of the recession limb of the hydrograph. Between these two points the total volume was used for the analysis.

As in the analysis of the corresponding flood peaks at the two dams the Generalized Extreme Value GEV distribution using L-moments as parameter estimation method was chosen as marginal distribution for the annual flood peak at Straußfurt. The flood volume was fitted by a GEV distribution using the method of product moments as parameter estimation method.

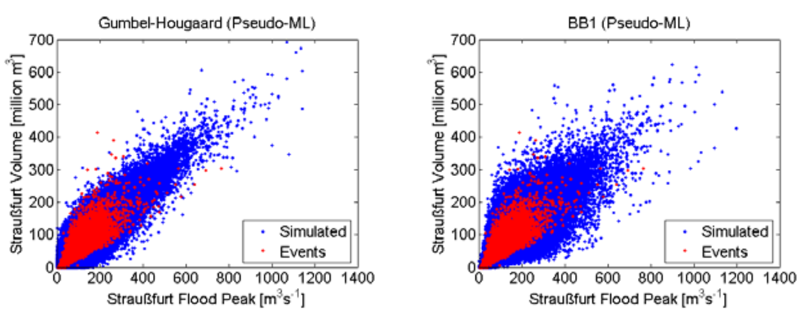

Figure 7: Simulated random sample of size 1 000 000 chosen from BB1 Copula and Gumbel-Hougaard copula

From the four mainly in hydrology applied Archimedian copulas mentioned in section 2 only the Gumbel-Hougaard family is applicable for the dependence structure of the data. Because the dependence structure between the flood peaks and the volume is similar to the dependence structure between the two corresponding flood peaks only the Gumbel-Hougaard copula was chosen for the analysis. Besides the Gumbel-Hougaard copula the two-parameter copula BB1 is used for the construction of the bivariate distribution function. The parameters are estimated using the method of maximum pseudolikelihood. Comparing the parametric and nonparametric values K(t) from the two copulas the BB1 copula provides a better fit to the data.

In Figure 7 it is shown that for the Gumbel-Hougaard copula some of the data points lie outside of the generated range of 1000000 randomly generated pairs (x_i,y_i). In contrary to this all values are within the simulated range of the BB1 copula. The BB1 copula is chosen for further bivariate frequency analysis since it provides the best fit to the data.

Figure 8 illustrates both the joint return periods with respect to corresponding annual flood peak and volume $T^{\wedge}_{X,Y}$ for which x and y are exceeded and the joint return periods $T^{\vee}_{X,Y}$ for which either x or y are exceeded. Again using the events with a flood peak of about 100 year return period for the flood design, the corresponding return periods of the flood volume range between 25 years and 2000 years. Therefore it is important to use the joint probabilities instead of using the univariate probability of the flood peak.

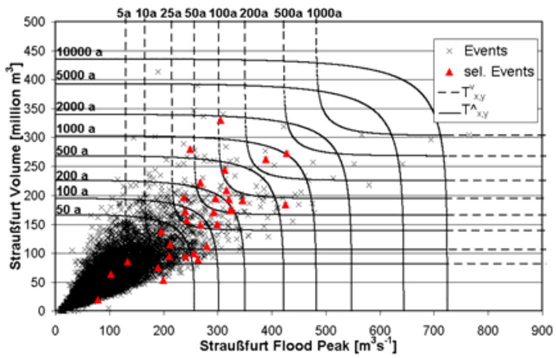

Figure 8: Joint return periods of the corresponding flood peak and volume $T^{\vee}_{X,Y}$ (exceeding x or y) and the $T^{\wedge}_{X,Y}$ (exceeding x and y)

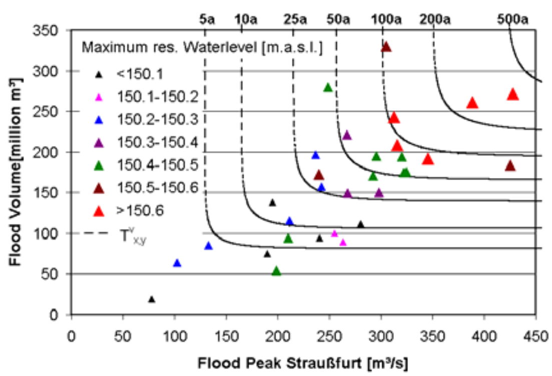

Figure 9: Maximum resulting waterlevels at the reservoir Straußfurt from the different hydrological scenarios and the $T^{\vee}_{X,Y}$ (exceeding x or y) isolines

A broad range of events has been selected from the time series and categorized based on the analysis of joint return periods of the flood-volume at the reservoir Straußfurt as well as the joint return periods of the flood peak at the reservoir Straußfurt and the dam Kelbra. In Figure 6 and 8 the selected events are displayed (red triangles). In this way the spatial distribution between the two main tributaries is considered and the very important aspect of using different flood peak and volume combinations is accounted for in the flood risk analysis. Those data are in turn used for assessment of the flood control system which hereby integrates the spatial component of probability. In Figure 9 the maximum resulting water levels at the reservoir Straußfurt are shown to demonstrate the effects of the different hydrological loads on the flood protection structures depending on the combination of flood peak and flood volume. We define water levels exceeding 150.3 m a.s.l. as a critical hydrologic event, since it is known that the corresponding outflow of more than 200 m³/s causes severe damage downstream. With a joint return period $T^{v}_{X,Y}$ (exceeding x or y) larger than 40 years all considered events are critical events according to this definition. Hence risk is high for this region. For joint return periods between 25 years and 40 years 3 of the 5 events considered are critical events. In the region with a joint return period $T^{v}_{X,Y}$ smaller than 25 years only 2 out of 12 events will result in substantial damage. The risk that events with these joint probabilities may cause damage is therefore low.

5. CONCLUSIONS

A methodology to categorize flooding based on copulas is presented in this work. The advantage of using copulas is that a bivariate distribution function with different univariate marginal distributions can be constructed to describe the two random variables. The dependence between the two random variables is described by the copula. As a test case copulas were used to derive probabilities for risk analysis in the Unstrut catchment. Two different probabilities were derived for a long-term daily discharge time series of 10000 years. To reassign an overall probability of a flood event which has different probabilities in the two main tributaries, the joint return period of the corresponding flood peaks of two reference stations located downstream of the two main tributaries is derived by copula analysis. In the flood design of flood detention structures it is important to consider the flood volume besides the flood peak. Therefore a bivariate distribution function is constructed to describe the relationship between peak and volume. With these probabilities it is feasible evaluate damage and cost for different hydrological scenarios using a risk assessment instrument. For parameter estimation the maximum pseudolikelihood method has been used. The Gumbel-Hougaard copula seems to provide the best fit for the relationship between the two corresponding flood peaks of the different tributaries and for the relationship between flood peak and flood volume the two parameter BB1 copula has been selected. From these copulas joint return periods could be easily assigned to selected hydrologic loads. Depending on the hydrologic design problem the joint return period exceeding both x and y or the joint return period exceeding either x or y is relevant.

6. REFERENCES

Genest, C. and Rivest, L.-P. 1993. Statistical inference procedures for bivariate Archimedian copulas. *Journal of the American Statistical Association* 88 (423): 1034-1043.

Hosking, J.R.M. and Wallis, J.R. 1997. Regional frequency analysis: an approach based on L-moments. Cambridge University Press, Cambridge, U.K.

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

Nelsen, R.B. 2006. An Introduction to Copulas. Springer-Verlag, New York

Salvadori, G. and DeMichele, C. 2004. Frequency analysis via copulas: Theoretical aspects and applications to hydrological events. *Water Resources Research* 40 (W12511):1-17

Singh, K. and Singh, V.P. 1991. Derivation of bivariate probability density functions with exponential marginals. *Stochastic Hydrology and Hydraulics* 5: 55-68

Sklar, A. 1959: Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Stat. Univ. Paris* 8: 229-231

Stedinger, J.R., Vogel, R.M. and Foufoula-Georgiou, E. 1993. Frequency Analysis of Extreme Events. In: *Handbook of Hydrology*, Maidment, D. R. (Editor). McGraw-Hill Inc., New York