

# Chapter 8

## Copulas – New Risk Assessment Methodology for Dam Safety

Bastian Klein, Andreas H. Schumann, and Markus Pahlow

**Abstract** Consideration of a broad range of hydrological loads is essential for risk-based flood protection planning. Furthermore, in the planning process of technical retention facilities it is necessary to use flood events, which are specified by several characteristics (peak, volume and shape). Multivariate statistical methods are required for their probabilistic evaluation. Coupled stochastic-deterministic simulation may be applied to generate a runoff time series, since the required amount of data is generally not available. Even the effect of complex flood protection systems may be evaluated through generation of a data base by means of stochastic-deterministic simulations with subsequent statistical analysis of the individual hydrological load scenarios. Multivariate frequency analyses of correlated random variables are useful to specify these scenarios statistically. Copulas are a very flexible method to estimate multivariate distributions, because the marginal distributions of the random variables can differ. Here a methodology for flood risk assessment is presented which was applied in two case studies in Germany.

### Contents

8.1	Introduction . . . . .	150
8.2	Copula Theory . . . . .	151
8.2.1	Basic Principles of Copula Theory . . . . .	152
8.2.2	Archimedian Copulas . . . . .	154
8.2.3	Parameter Estimation . . . . .	154
8.2.4	Identification of the Appropriate Copula Model . . . . .	156
8.2.5	Bivariate Frequency Analysis . . . . .	158
8.3	Case Study 1: Risk Analysis for the Wupper Dam . . . . .	161
8.3.1	Study Area . . . . .	162
8.3.2	Stochastic-Deterministic Generation of Flood Events . . . . .	162

---

B. Klein (✉)

Institute of Hydrology, Water Resources Management and Environmental Engineering,  
Ruhr-University Bochum, Bochum, Germany  
e-mail: Klein@bafg.de

8.3.3	Bivariate Frequency Analysis of Annual Flood Peaks and Corresponding Volumes . . . . .	163
8.3.4	Evaluation of the Effect of the Wupper Dam on Flood Control . . . . .	169
8.4	Case Study 2: Unstrut River Basin . . . . .	171
8.4.1	Description of the River Basin . . . . .	172
8.4.2	Stochastic-Deterministic Generation of Flood Events . . . . .	173
8.4.3	Bivariate Frequency Analysis of Corresponding Flood Peaks at the Reservoir Sites . . . . .	174
8.4.4	Bivariate Frequency Analysis of the Annual Flood Peaks and the Corresponding Volumes . . . . .	177
8.4.5	Evaluation of the Effect of the Reservoir Strausfurt on Flood Control . . .	180
8.5	Conclusions . . . . .	182
	References . . . . .	183

## 8.1 Introduction

Recent flood events and the related drastic consequences indicate that it is important to consider 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 only a single design flood of a predefined return period. For single flood protection structures the different hydrological inflows can be generated stochastically (see e.g. De Michele et al., 2005; Klein and Schumann, 2006) or by rainfall-runoff modeling. In large catchments with interacting flood protection structures the efficiency in flood mitigation depends mainly on the spatial distribution of the flood event. To consider these aspects in the design phase spatially distributed precipitation events can be generated by coupling a spatially distributed stochastic rainfall generator with a distributed or semi-distributed rainfall-runoff model to generate hydrological scenarios (see e.g. Blazkova and Beven, 2004; Haberlandt et al., 2008). These hydrological scenarios have to be evaluated probabilistically by frequency analysis to provide a base for risk analyses of flood protection structures or systems.

In most cases, the return period of flood events is defined by a univariate analysis of a random variable such as the flood peak (Hosking and Wallis, 1997; Stedinger et al., 1993). However, technical flood retention depends on several coinciding random variables such as flood peak, volume, shape of the hydrograph and duration. Hence univariate frequency analysis may lead to an over- or underestimation of the risk associated with a given flood (Salvadori and De Michele, 2004). For the design of flood control structures the flood peak plays a fundamental role. However, for flood storages such as reservoirs and polders the flood volume is also very important and should be considered in design and planning. For example an event with a very large peak and a small volume could be stored in the flood control storage of a reservoir, whereas an event with a smaller peak but a larger volume may well lead to flood spillage.

If a dependency exists between the random variables it is advantageous to define the return period of a flood event by using joint probabilities for the flood characteristics that are relevant for the design.

## 8.2 Copula Theory

One reason for the fact that multivariate statistical models are rarely used for multivariate analysis of random variables in hydrology is that they require considerably more data for the estimation of a reliable model than the univariate case. Owing to the data requirements, also the application of multivariate analysis of correlated random numbers in practice is mainly reduced to the bivariate case.

Several multivariate distributions are available for multivariate frequency analysis. In 1976, Hiemstra derived the bivariate lognormal probability distribution for bivariate frequency analysis (Hiemstra et al. 1976). More recently bivariate distributions such as the bivariate exponential distribution (Singh and Singh, 1991), the bivariate normal distribution (Bergmann and Sackl, 1989; Goel et al., 1998; Yue, 1999), the bivariate gamma distribution (Yue, 2001; Yue et al., 2001) and the bivariate extreme value distribution (Shiau, 2003; Yue et al., 1999) have been applied to bivariate frequency analysis (e.g. for the correlated variables flood peak and flood volume or flood volume and flood duration). Sandoval and Raynal-Villasenor (2008) applied a trivariate extreme value distribution for frequency analysis.

The application of these multivariate distributions to model the dependency between the correlated random variables has a number of drawbacks. First of all, it is a precondition that the marginal distributions have to come from the same family of distributions or the random variables have to be transformed otherwise (e.g. normalized in the case of the bivariate normal distribution). This is not necessarily fulfilled by the dependent variables characterizing a given process. Furthermore, most of the available distributions are limited to the bivariate case and an extension to the multivariate case can not be derived easily. In addition, the dependencies between the random variables of these distributions are generally described by the Pearson's coefficient of linear correlation. However, in reality the dependency structure of the flood variables may not be Gaussian.

These problems can be avoided by applying copulas to model the dependencies between correlated random variables. A joint distribution of correlated random variables can be expressed as a function of the univariate marginal distributions using a copula. Copulas enable one to model the dependency structure independently from the marginal distributions. A large number of different copulas are available to describe this dependency. Therefore it is possible to build multidimensional distributions which differ in their marginal distributions.

Copulas have been extensively used in financial studies (see e.g. Cherubini et al., 2004; Embrechts et al., 2003), but more recently copulas also have been implemented in hydrology for multivariate analysis of hydrological random variables. Salvadori and De Michele (2004) provide a general theoretical framework for exploiting copulas to study the return periods of hydrological events. In several works copulas have been used for the bivariate analysis of the properties of flood events (Karmakar and Simonovic, 2008, 2009; Shiau et al., 2006; Zhang and Singh, 2006), for the bivariate frequency analysis of the properties of precipitation (De Michele and Salvadori, 2003; Evin and Favre, 2008; Zhang and Singh, 2007a) and for the analysis of droughts (Shiau et al., 2007). In De Michele et al. (2005) the dependency between peak and volume is described using the concept of copulas. In

this work the adequacy of dam spillways is tested by means of Monte-Carlo simulation for the generation of volume-peak combinations from copula and marginal distributions and their transformation into hydrographs. Favre et al. (2004) presented modeling results of bivariate extreme values using copulas for different case studies. Trivariate analyses of the dependency of random variables using copulas have been applied to model the properties of precipitation events (Grimaldi and Serinaldi, 2006a; Kao and Govindaraju, 2008; Salvadori and De Michele, 2006; Salvadori and De Michele, 2007; Zhang and Singh, 2007b) and for the trivariate flood frequency analysis of flood peak, volume and event duration (Grimaldi and Serinaldi, 2006b; Serinaldi and Grimaldi, 2007; Zhang and Singh, 2007c). In Renard and Lang (2006) the use of copulas for multivariate extreme value analysis is demonstrated with different case studies.

### 8.2.1 Basic Principles of Copula Theory

Only a short overview of the theory of copulas is presented here. More detailed discussions can be found in Joe (1997), Nelsen (1999), Salvadori and De Michele (2004) and Salvadori et al. (2007). A copula is a function which exactly describes and models the dependency structure between correlated random variables independently of their marginal distributions. It is defined as mapping  $C : [0, 1]^n \rightarrow [0, 1]$ , such that for all  $u_1, u_2, \dots, u_n \in [0, 1]$  the copula function  $C(u_1, u_2, \dots, u_n) = 0$  if at least one of the arguments is equal to zero and  $C(u_1, u_2, \dots, u_k, \dots, u_n) = u_k$  if all the arguments are 1 except  $u_k$ .

The link between the copula and the multivariate distribution is provided by the theorem of Sklar (1959):

$$F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = C[F_{X_1}(x_1), F_{X_2}(x_2), \dots, F_{X_n}(x_n)] \quad (8.1)$$

where  $F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$  is the joint cumulative distribution function (cdf) with the continuous marginals  $F_{X_1}(x_1), F_{X_2}(x_2), \dots, F_{X_n}(x_n)$  of the random variables.

The content of this chapter is concentrated on bivariate analysis of the random variables  $X$  and  $Y$  with the marginal distributions  $F_X(x)$  and  $F_Y(y)$  with a joint cumulative distribution function (cdf):

$$F_{X,Y}(x, y) = C[F_X(x), F_Y(y)]. \quad (8.2)$$

Because copulas are invariant for strictly increasing transformations of  $X$  and  $Y$  it is possible to use two uniformly distributed random variables  $U$  and  $V$  on  $[0, 1]$ , defined as  $U = F_X(x)$  and  $V = F_Y(y)$ .

Assuming continuous marginal distributions with the probability density functions  $f_X(x)$  and  $f_Y(y)$  the joint probability density function (pdf) becomes:

$$f_{X,Y}(x, y) = c[F_X(x), F_Y(y)] f_X(x) f_Y(y) \quad (8.3)$$

where  $c$  is the density function of the copula  $C$ :

$$c(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v}. \tag{8.4}$$

Three special cases of copulas are of particular importance:

1. The copula of the Fréchet-Hoeffding lower bound  $W$ :

$$W(u, v) = \max \{u + v - 1, 0\} \tag{8.5}$$

2. The copula of the Fréchet-Hoeffding upper bound  $M$ :

$$M(u, v) = \min \{u, v\} \tag{8.6}$$

3. The independence copula  $\Pi$ :

$$\Pi(u, v) = u \cdot v. \tag{8.7}$$

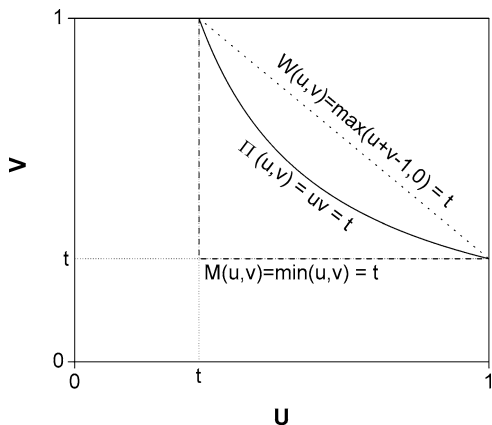
The copulas  $W$  and  $M$  are the general bounds for any copula  $C$  and any pair  $(u, v) \in [0, 1]^2$ :

$$W(u, v) = \max(u + v - 1, 0) \leq C(u, v) \leq M(u, v) = \min(u, v). \tag{8.8}$$

In Fig. 8.1 the level curves, which connect the points  $(u, v)$  where the value of the copula has the same value  $t$  for the three special cases,  $W(u, v) = t$ ,  $M(u, v) = t$  and  $\Pi(u, v) = t$  are illustrated.

Conditional distributions such as the conditional distribution of  $U$  for a given value of  $V = v$  can be calculated easily using copulas (other conditional laws see e.g. Salvadori and De Michele, 2004; Salvadori et al., 2007; Zhang and Singh, 2006):

$$P(U \leq u | V = v) = \frac{\partial}{\partial v} C(u, v). \tag{8.9}$$



**Fig. 8.1** Level curves of the special cases  $W(u, v)$ ,  $M(u, v)$  and  $\Pi(u, v)$

The empirical copula (Deheuvels, 1979) derived from a random sample  $(X_1, Y_1), \dots, (X_n, Y_n)$  of size  $n$  is formally defined by:

$$C_n(u, v) = \frac{1}{n} \sum_{i=1}^n 1 \left( \frac{R_i}{n+1} \leq u, \frac{S_i}{n+1} \leq v \right) \quad (8.10)$$

where  $R_i$  is the rank of  $X_i \in \{X_1, \dots, X_n\}$  and  $S_i$  is the rank of  $Y_i \in \{Y_1, \dots, Y_n\}$  of the sample and  $1(A)$  the indicator function of the set  $A$ .

A large variety of families of copulas described e.g. in Joe (1997) and Nelsen (1999) are available to model multivariate dependencies. There are generally three families of copulas that are applied in hydrological practice: the Archimedean copulas, the elliptical copulas (including the Gaussian, Student and Cauchy copulas) and the extreme-value copulas (including the Gumbel-Hougaard copula -the only Archimedean extreme-value copula-, the Galambos copula and the Hüsler-Reiss copula). The Archimedean copulas are widely used because they can be constructed easily and a large variety of copulas belongs to this family.

## 8.2.2 Archimedean Copulas

Archimedean copulas are expressed by the copula generator  $\varphi : [0, 1] \rightarrow [0, \infty]$ , which is a convex, continuous decreasing function with  $\varphi(1) = 0$ :

$$C_\theta(u, v) = \varphi^{-1}[\varphi(u) + \varphi(v)]. \quad (8.11)$$

In Table 8.1 four widely used one parameter Archimedean copulas are presented with their corresponding generator  $\varphi(t)$ , where  $\theta$  is the parameter of the generator function.

All four copulas are applicable to positively correlated random variables but only the Ali-Mikhail-Haq and the Frank copula can model negatively correlated random variables as well. The Ali-Mikhail-Haq copula is not appropriate for highly correlated (both positive and negative) variables.

Besides one-parameter Archimedean copulas several two-parameter copulas are described in Nelsen (1999) and Joe (1997). The advantage of these two-parameter copulas is that they may be used to capture more than one type of dependencies. One parameter models the upper tail dependency and the other parameter models the lower tail dependency of the copula. As an example the two-parameter copula BB1 (as described in Joe (1997)), which has also the form of the Archimedean copula family (Eq. 8.11) is presented in Table 8.1.

## 8.2.3 Parameter Estimation

There are several methods employed to estimate the parameter  $\theta$  of the copula function. The two standard rank-based nonparametric measures of dependency,

**Table 8.1** Summary of various 2-dimensional Archimedean copulas and their generator (Joe, 1997; Nelsen, 1999)

Copula		$\varphi(t)$
Gumbel-Hougaard:		
$\exp \left\{ - \left[ (-\ln u)^\theta + (-\ln v)^\theta \right]^{1/\theta} \right\}$	$\theta \geq 1$	$(-\ln t)^\theta$
Ali-Mikhail-Haq:		
$\frac{uv}{1-\theta(1-u)(1-v)}$	$-1 \leq \theta < 1$	$\ln \frac{1-\theta(1-t)}{t}$
Frank:		
$-\frac{1}{\theta} \ln \left[ 1 + \frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1} \right]$	$\theta \neq 0$	$-\ln \left( \frac{e^{-\theta t}-1}{e^{-\theta}-1} \right)$
Cook-Johnson:		
$[u^{-\theta} + v^{-\theta} - 1]^{-1/\theta}$	$\theta \geq 0$	$t^{-\theta} - 1$
BB1		
$C_\theta(u, v) = \left\{ 1 + \left[ (u^{-\theta_1} - 1)^{\theta_2} + (v^{-\theta_1} - 1)^{\theta_2} \right]^{1/\theta_2} \right\}^{-1/\theta_1}$	$\theta_1 > 0, \theta_2 \geq 1$	$\varphi(t) = (t^{-\theta_1} - 1)^{\theta_2}$

Spearman’s  $\rho$  and Kendall’s  $\tau$  can be expressed solely in terms of the copula function. Therefore the parameter of the copula can be estimated from the measure of dependency of the bivariate sample. Kendall’s  $\tau$  can be expressed as:

$$\tau = \int_{[0,1]^2} C_\theta(u, v) dC_\theta(u, v) - 1 \tag{8.12}$$

which, for Archimedean copulas with a generator function  $\varphi$ , is reduced to

$$\tau = 1 + 4 \int \frac{\varphi(t)}{\varphi'(t)} dt, \tag{8.13}$$

where  $\varphi'$  is the derivative of  $\varphi$  with respect to  $t$ .

Another estimation method is the maximum pseudolikelihood method (Genest and Favre, 2007) which is also called canonical maximum likelihood. It can also be used if  $\theta$  is multidimensional. The method of maximum pseudolikelihood involves maximizing a rank-based log-likelihood function of the form:

$$l(\theta) = \sum_{i=1}^n \log \left[ c_\theta \left( \frac{R_i}{n+1}, \frac{S_i}{n+1} \right) \right], \tag{8.14}$$

where  $R_i$  is the rank of  $X_i \in \{X_1 \dots X_n\}$  and  $Y_i \in \{Y_1 \dots Y_n\}$  of the sample of size  $n$  and the copula density distribution  $c_\theta(u, v)$  after Eq. (8.4)

In Joe (1997) a parametric two-step procedure called the “inference from margins” or IFM method is recommended to estimate the parameter of the model. An estimate of  $\theta$  is obtained by maximizing the log-likelihood function

$$l(\theta) = \sum_{i=1}^n \log \{c_{\theta}[F_X(x_i), F_Y(y_i)] f_X(x_i) f_Y(y_i)\}. \quad (8.15)$$

In general estimation methods are based on a measure of dependency. The maximum pseudolikelihood estimation method estimates the parameters independently from the marginal distributions. They only model the rank-based dependency between the two random variables. In contrast the IFM depends on the estimation of the margins and therefore the parameters of the copula are unduly affected if the marginal distributions are wrong (see e.g. Kim et al., 2007).

## 8.2.4 Identification of the Appropriate Copula Model

Methods to identify the copula which provides the best fit to the observed data are required, since there is a large variety of different copula families available to model the dependency between random variables. Several methods exist to select the appropriate model.

### 8.2.4.1 Graphical Diagnostics

Possibly the most natural way for the identification of the appropriate copula is the graphical comparison between the scatter plot of the pairs  $[R_i/(n+1), S_i/(n+1)]$  (representing the empirical copula  $C_n$ ) from the bivariate data with pairs  $(U_j, V_j)$  from an artificial data set, generated from the copula  $C_{\theta n}$ , or by comparing the pairs of the bivariate data with the pairs  $(X_j, Y_j)$  generated from the copula  $C_{\theta n}$ . and transformed back into the original units using the marginal distributions  $F_X(x)$  and  $F_Y(y)$ . Algorithms to generate random pairs  $(U, V)$  from a copula are described e.g. in Whelan (2004). A simple simulation algorithm to generate random pairs  $(u, v)$  from the copula is given in Genest and Favre (2007):

1. Generate  $U$  from a uniform distribution on the interval  $(0,1)$ .
2. Given  $U = u$ , get  $V = Q_u^{-1}(U^*)$  from the inverse function of the conditional distribution according to Eq. (8.16), where  $U^*$  is another random number generated from the uniform distribution on the interval  $(0,1)$ .

$$Q_u(v) = P(V \leq v | U = u) = \frac{\partial}{\partial u} C(u, v). \quad (8.16)$$

Another option is the comparison between the parametric and nonparametric estimate of the probability  $K_C(t)$  (Genest and Rivest, 1993) that the value of the copula  $C(u, v)$  is smaller than or equal to  $t$  ( $0 < t \leq 1$ ):

$$K_C(t) = P[C(u, v) \leq t], 0 < t \leq 1. \quad (8.17)$$



Let

$$B_C(t) = \{(u, v) \in [0, 1]^2 : C(u, v) \leq t\}, 0 < t \leq 1 \quad (8.18)$$

where  $B_C(t)$  is the region within  $[0, 1]^2$  lying on, below and to the left of the level curve  $L_t$  with

$$L_t = \{(u, v) \in [0, 1]^2 : C(u, v) = t\}, 0 < t \leq 1 \quad (8.19)$$

of the joint distribution. The level curve is a function of the two variables  $u$  and  $v$  that connects points where the copula has the same value  $t$ . The probability distribution  $K_C(t)$  provides a unique probability measure of the set  $B_C(t)$  (see Salvadori and De Michele, 2004). For Archimedean copulas the parametric estimate of  $K_{C_\theta}(t)$  can be calculated easily using

$$K_{C_\theta}(t) = t - \frac{\varphi(t)}{\varphi'(t)} \quad (8.20)$$

and the level curve  $L_t$  as a function of  $u$ :

$$L_t(u) = \varphi^{-1}[\varphi(t) - \varphi(u)]. \quad (8.21)$$

The nonparametric estimate of  $K_{C_n}(t)$  can be calculated after Genest et al. (2006):

$$K_{C_n}(t) = \frac{1}{n} \sum_{i=1}^n 1(V_{in} \leq t); t \in [0, 1] \quad (8.22)$$

where

$$V_{in} = \frac{1}{n} \sum_{j=1}^n 1(x_j \leq x_i, y_j \leq y_i) = \frac{1}{n} \sum_{j=1}^n 1(R_j \leq R_i, S_j \leq S_i). \quad (8.23)$$

#### 8.2.4.2 Goodness-of-Fit Statistics

There are different possibilities to compare and evaluate the goodness of fit of the different possible copula functions to model the dependency of the random variables. The “Root Mean Square Error” (RMSE) goodness-of-fit measure is the mean squared difference between the empirical and the estimated distribution functions:

$$\text{RMSE} = \sqrt{E(x_c - x_0)^2} = \sqrt{\frac{1}{n - k} \sum_{i=1}^n [x_c(i) - x_0(i)]^2}, \quad (8.24)$$

where  $x_c(i)$  is the  $i$ -th computed value (here: value of the fitted theoretical distribution function),  $x_0(i)$  the  $i$ -th observed value (here: value of the empirical distribution function),  $k$  the number of model parameters used to obtain the calculated value and  $n$  the number of observations.

The Akaike’s information criterion AIC (Akaike, 1974) and the Bayesian information criterion BIC (Schwarz, 1978) are measures of the goodness of fit of an

estimated statistical model and can also be used for model selection. Both criteria depend on the maximized value of the likelihood function  $L$  for the estimated model. The AIC is defined as

$$\text{AIC} = (-2) \cdot \ln(L) + 2 \cdot k \quad (8.25)$$

where  $k$  is the number of model parameters and the BIC is defined as:

$$\text{BIC} = (-2) \cdot \ln(L) + k \cdot \ln(n) \quad (8.26)$$

where  $n$  is the sample size. Both criteria differ from each other with increasing sample size. According to these criteria, the model with the lowest value of the AIC and BIC is the best model.

The multidimensional Kolmogorov-Smirnov (KS) goodness-of-fit test uses (see Saunders and Laud, 1980), in a similar way as for the univariate KS-Test, the maximal distance between the empirical and the theoretical distribution function as a goodness-of-fit measure. Standard tables or bootstrap resampling can be applied to find the critical values for a significance level  $\alpha$  or the  $p$ -values for the goodness-of-fit test.

In Genest et al. (2006, 2009) and Genest and Favre (2007) several methods for formal goodness-of-fit testing of copulas are presented. Two of these methods use the difference between the empirical and theoretical distribution of  $K_C(t)$  (see Eq. 8.17) as test measure. The first methodology is based on the Cramér-von-Mises statistics and the second one on the Kolmogorov-Smirnov statistics. To estimate the critical values for a significance level  $\alpha$  and the  $p$ -value a bootstrap methodology is presented for the goodness-of-fit test of copulas. Another bootstrap methodology based on the Cramér-von-Mises statistics as goodness-of-fit test for copulas is described in Genest and Favre (2007) and Genest and Remillard (2008). There the difference between the empirical and theoretical copula is used as measure.

### 8.2.5 Bivariate Frequency Analysis

The joint distribution of Eq. (8.2) is the probability  $P(X \leq x, Y \leq y) = F_{X,Y}(x, y)$  that  $X$  and  $Y$  are less than or equal to the specific thresholds  $x$  and  $y$ , respectively. Given the assumption of total independence between the two random variables, the joint probability  $F_{X,Y}(x, y)$  becomes the product of the two individual probabilities  $F_X(x)$  and  $F_Y(y)$ :

$$P(X \leq x, Y \leq y) = F_{X,Y}(x, y) = F_X(x)F_Y(y). \quad (8.27)$$

The joint probability  $F_{X,Y}(x, y)$  is reduced to the univariate probability  $F_X(x)$  or  $F_Y(y)$  if the two random variables are completely dependent. In this case the random variables  $X$  and  $Y$  are functions of each other:

$$P(X \leq x, Y \leq y) = F_{X,Y}(x, y) = F_X(x) \text{ or } F_{X,Y}(x, y) = F_Y(y). \quad (8.28)$$

If there are dependencies, the probability that  $X$  and  $Y$  both exceed  $x$  and  $y$  is defined as

$$\begin{aligned} P(X > x \wedge Y > y) &= 1 - F_X(x) - F_Y(y) + F_{X,Y}(x, y) \\ &= 1 - F_X(x) - F_Y(y) + C[F_X(x), F_Y(y)]. \end{aligned} \quad (8.29)$$

In the case of independence between the two random variables the joint return period is the product of the two univariate return periods:

$$T_{X,Y}^{\wedge} = \frac{\mu_T}{1 - F_X(x) - F_Y(y) + F_X(x)F_Y(y)} = \frac{\mu_T}{(1 - F_X(x))(1 - F_Y(y))} = T_X T_Y, \quad (8.30)$$

where  $\mu_T$  is the mean interarrival time between two successive events ( $\mu_T = 1$  year, considering annual values).

Under consideration of dependency the corresponding joint return period (labeled with the logical AND operator “ $\wedge$ ”) is expressed as

$$T_{X,Y}^{\wedge} = \frac{\mu_T}{P(X > x \wedge Y > y)} = \frac{\mu_T}{1 - F_X(x) - F_Y(y) + C[F_X(x), F_Y(y)]}. \quad (8.31)$$

This joint return period is always larger than the maximum of the univariate return periods  $T_X = \mu_T/(1 - F_X(x))$  and  $T_Y = \mu_T/(1 - F_Y(y))$  of the individual random variables.

The probability with which either  $X$  or  $Y$  exceed their respective thresholds  $x$  or  $y$  is defined as

$$P(X > x \vee Y > y) = 1 - F_{X,Y}(x, y) = 1 - C[F_X(x), F_Y(y)]. \quad (8.32)$$

Here the joint return period (labeled with the logical OR operator “ $\vee$ ”) is expressed as

$$T_{X,Y}^{\vee} = \frac{\mu_T}{P(X > x \vee Y > y)} = \frac{\mu_T}{1 - C[F_X(x), F_Y(y)]}. \quad (8.33)$$

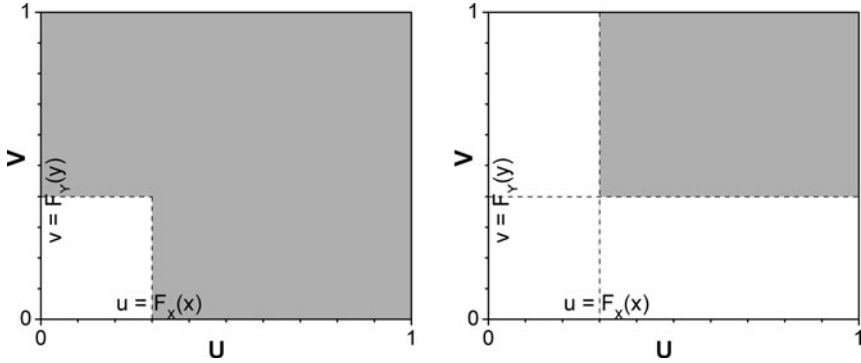
This joint return period is always smaller than the minimum of the two univariate return periods  $T_X$  and  $T_Y$  of the individual random variables:

$$T_{X,Y}^{\vee} \leq \min[T_X, T_Y] \leq \max[T_X, T_Y] \leq T_{X,Y}^{\wedge}. \quad (8.34)$$

In the case of independence of the two random variables this joint return period becomes:

$$T_{X,Y}^{\vee} = \frac{\mu_T}{1 - F_X(x)F_Y(y)} = \frac{T_X T_Y}{T_X + T_Y - \mu_T}. \quad (8.35)$$

In Fig. 8.2 the regions of  $[0,1]^2$  embodying the probability masses for the two cases: exceedance probability  $P(X > x \vee Y > y)$  -“OR” case- and  $P(X > x \wedge Y > y)$  -“AND” case- are illustrated as shaded areas.



**Fig. 8.2** Regions in  $[0,1]^2$  (shaded areas) embodying the probability masses of (a)  $P(X > x \vee Y > y)$  and (b)  $P(X > x \wedge Y > y)$

Conditional distribution functions and return periods can also be expressed easily with copulas (see e.g. Salvadori and De Michele, 2004; Salvadori et al., 2007; Zhang and Singh, 2006). For example the conditional distribution  $F_X(x|Y = y)$  of  $X$  for a given value of  $Y = y$  can be calculated using Eq. (8.9) with  $u = F_X(x)$  and  $v = F_Y(y)$ .

In Salvadori and De Michele (2004) another “secondary” return period  $\rho_t^\vee$  is defined to emphasize the difference to the “primary” return period  $T_{X,Y}^\vee B_C(t)$  in Eq. (8.18) is the region where the return period  $T_{X,Y}^\vee$  of all events is equal to or less than a threshold

$$\vartheta(t) = \frac{\mu_T}{(1 - t)}. \tag{8.36}$$

Therefore  $K_C(t)$  is the probability that an event with the return period  $T_{X,Y}^\vee \leq \vartheta(t) = \mu_T/(1 - t)$  occurs for any realization of the random process. This implies that the survival function of  $K_C(t)$ , defined as

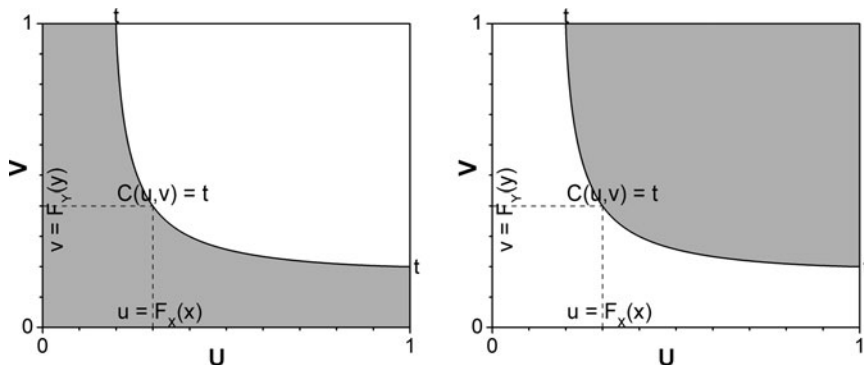
$$\bar{K}_C(t) = 1 - K_C(t), \tag{8.37}$$

is the probability that for any given realization of the random process an event with a return period  $T_{X,Y}^\vee > \vartheta(t) = \mu_T/(1 - t)$  occurs. The secondary return period  $\rho_t^\vee$  can be expressed as the reciprocal value of the survival function

$$\rho_t^\vee = \frac{\mu_T}{\bar{K}_C(t)} = \frac{\mu_T}{1 - K_C(t)}. \tag{8.38}$$

If a critical threshold  $\vartheta(t)$  is defined in the design stage for a flood protection structure, the secondary return period specifies the mean interarrival time of a critical event.

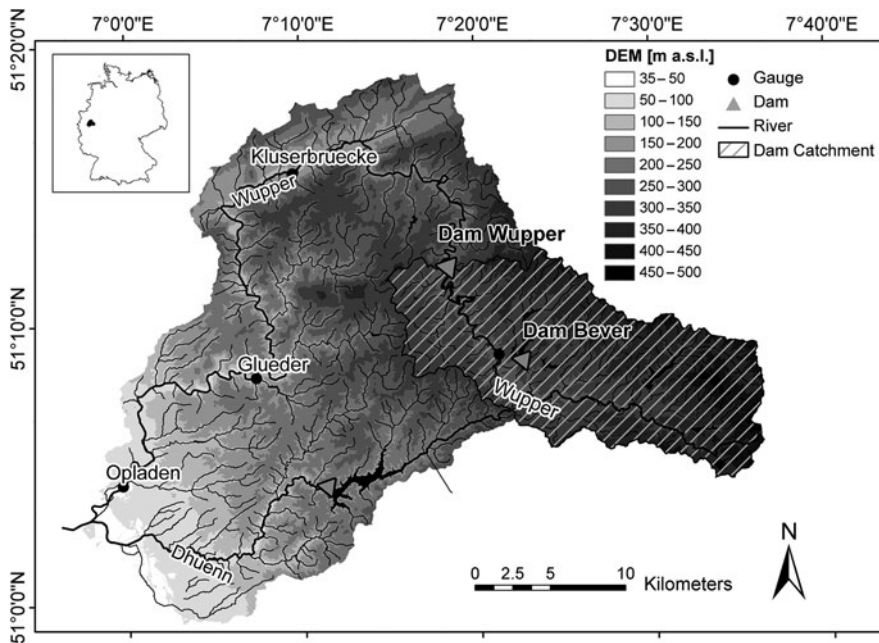
The regions in  $[0,1]^2$  embodying the probability masses for  $K_C(t) = P(C(u, v) \leq t)$  and  $\bar{K}_C(t) = P(C(u, v) > t)$  (shaded areas) are illustrated in Fig. 8.3.



**Fig. 8.3** Regions in  $[0,1]^2$  (shaded areas) embodying the probability masses of (a)  $K_C(t) = P(C(u, v) \leq t)$  and (b)  $\bar{K}_C(t) = P(C(u, v) > t)$

### 8.3 Case Study 1: Risk Analysis for the Wupper Dam

The first case study for an application of copulas within the framework of risk analysis of dams is concerned with the Wupper dam, located in the mid-western part of Germany (see Fig. 8.4). As mentioned before it is very important for the design of flood control structures to consider the flood volume besides the flood peak.



**Fig. 8.4** Topographical map of the watershed of the river Wupper

Therefore these two flood variables have been used in this case study for a bivariate frequency analysis of flood events. The flood events for the risk analysis were generated by coupling a continuous stochastic rainfall generator with a deterministic rainfall-runoff model. Using this methodology a synthetic 2000 year time series of discharge with a time step of 1 h was generated. Since the observed time series was too short for multivariate frequency analysis, these 2000 years has been used as a database for bivariate frequency analysis.

### 8.3.1 Study Area

The Wupper dam has a watershed of 212 km<sup>2</sup> and a mean inflow of 4.4 m<sup>3</sup>s<sup>-1</sup>. It is located in the catchment of the river Wupper with a total area of 813 km<sup>2</sup> (see Fig. 8.4). Apart from the flood control function it is used mainly for low water regulation in drought periods. The available flood control storage is seasonally variable. In the summertime between May and October no flood storage is allocated. From the beginning of November until the end of January a maximum flood storage of  $9.9 \times 10^6$  m<sup>3</sup> is provided. This flood storage is reduced continuously until the end of April. The flood spillway is a weir with a total width of 36 m and is controlled with fish belly gates. The maximum capacity of the spillway with open gates is 318 m<sup>3</sup>s<sup>-1</sup>. The two bottom outlets have a capacity of 88 m<sup>3</sup>s<sup>-1</sup> each.

In the watershed of the Wupper dam several smaller dams are located. From these dams particularly the dam Bever is important for flood control with a catchment area of 25.7 km<sup>2</sup> and an available flood control storage of  $5 \times 10^6$  m<sup>3</sup> in wintertime from November until the begin of February.

During a flood the Wupper dam is operated in dependency of the discharge at the control gage Kluserbruecke located downstream in the city of Wuppertal. The discharge at this gage is an indicator for the danger of flooding of the residential areas downstream. The uninfluenced subcatchment between the dam and the gage has an area of 125 km<sup>2</sup>. During a flood the controlled discharge at the gage should stay below 80 m<sup>3</sup>s<sup>-1</sup>. The outflow of the dam should not exceed 50 m<sup>3</sup>s<sup>-1</sup> to ensure this limit. During large floods the control level for the discharges at the gage Kluserbruecke is increased stepwise to 100, 150 and 190 m<sup>3</sup>s<sup>-1</sup>.

### 8.3.2 Stochastic-Deterministic Generation of Flood Events

In the first step the stochastic rainfall generator of Hundecha et al. (2009) was applied to generate continuous spatially distributed precipitation time series with a time step of 1 day and a length of 2000 years. In total 28 stations with daily precipitation with a minimum time series length of 30 years were used for model parameterization. In the second step the 2000 year time series of daily precipitation was disaggregated in hourly values using the stochastic models MuDRain (Koutsoyiannis, 2001) and HYETOS (Koutsoyiannis et al., 2003). The semi-distributed deterministic model NASIM (Hydrotec, 2007) has been used for

rainfall-runoff modeling. It was calibrated with observed rainfall-runoff data for the period 1. January 2001–31 December 2006. The operation of the dams in the region has been considered in NASIM. It has been shown that the statistical properties of the rainfall and the runoff were well reproduced (Petry et al., 2008).

### 8.3.3 Bivariate Frequency Analysis of Annual Flood Peaks and Corresponding Volumes

The annual flood peaks and corresponding volume were selected from the generated hourly inflow time series of 2000 years for the Wupper dam. The start of the surface runoff was marked by the abrupt rise of the hydrograph and the end of the flood runoff was identified by the flattening of the recession limb of the hydrograph. Between these two points the total volume was estimated for the analysis. In Fig. 8.5 the selection of the flood volume is illustrated.

The statistical dependencies between the two random variables flood peak and flood volume have been estimated by the statistical measures listed in Table 8.2, where particularly Pearson’s coefficient of linear correlation and the rank based measure of dependency Spearman’s  $\rho$  as well as Kendall’s  $\tau$  have been considered.

Pearson’s correlation coefficient provides a measure of linear dependency only and depends on the marginals, in contrast to the rank-based measure of dependency provided by Kendall’s and Spearman’s coefficient. The population value of

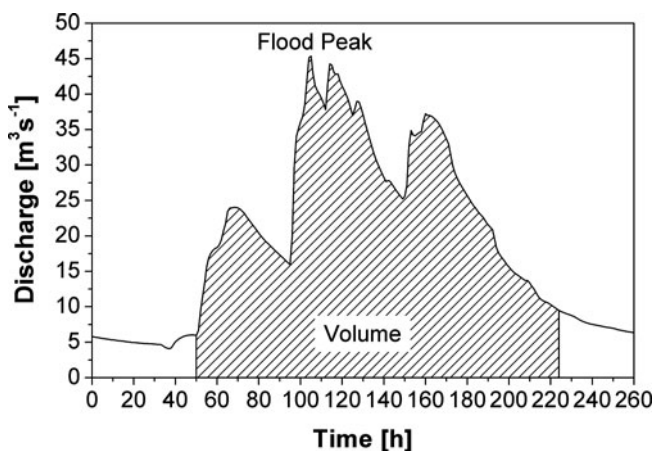


Fig. 8.5 Evaluation of the corresponding flood volume

Table 8.2 Measure of dependency for flood peaks and corresponding volumes for the Wupper dam

Sample size	Pearson’s $r$	Kendall’s $\tau$	Spearman’s $\rho$
2,000	0.546	0.379	0.541

the Pearson's correlation coefficient is always well defined. However, in some cases (e.g. for heavy-tailed distributions such as the Cauchy-distribution) a theoretical value of Pearson's correlation does not exist. Furthermore, Pearson's correlation coefficient is not robust. If the two variables are not linearly related the correlation is not well defined and outliers can strongly affect the correlation. On the other hand the rank based correlations such as Kendall's  $\tau$  and Spearman's  $\rho$  are robust since they are not dependent on the distributional assumptions. They can describe a wider class of dependencies and are resistant to outliers. The summary in Table 8.2 shows that there is a strong positive dependency between the two random variables.

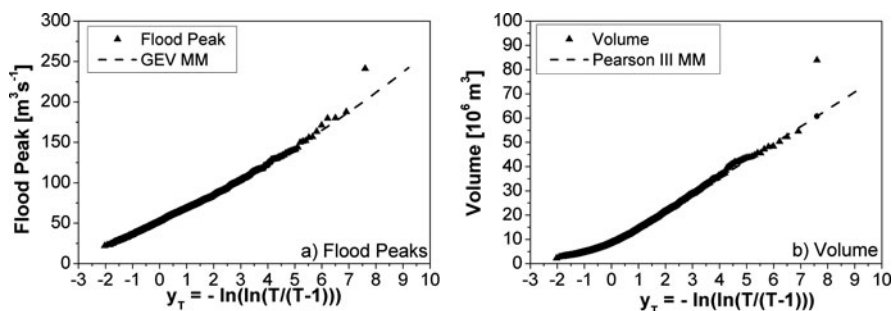
### 8.3.3.1 Marginal Distributions

The first step to build the bivariate probability distribution according to Eq. (8.2) consists in an estimation of the marginal distributions of the random variables. The Pearson type 3, log-Pearson type 3, Gumbel, Weibull, Gamma, Exponential, Generalized Pareto and Generalized Extreme Value (GEV) distributions were used for the univariate statistical analyses of the random variables. The goodness-of-fit of the different distributions and the different parameter estimation methods have been compared with the Kolmogorov-Smirnov and the Cramér-von-Mises  $n\omega^2$  tests. Because the focus of this chapter is the bivariate analysis the results of the univariate frequency analysis are not described here in detail.

These goodness-of-fit tests revealed that the flood peak can be described in the best way by the GEV distribution. The volume is described with the Pearson III distribution optimally. For parameter estimation the product moments were used. The fitted distributions are shown in Fig. 8.6.

### 8.3.3.2 Copula Estimation

The bivariate probability distribution (Eq. 8.2) has been constructed using the four Archimedian copulas Ali-Mikhail-Haq, Frank, Cook-Johnson (also known as Clayton copula) and Gumbel-Hougaard (Equations see Table 8.1), since those are the commonly used families of Archimedian copulas (see e.g. Shiau et al., 2006;



**Fig. 8.6** Comparison of the (a) annual flood peaks and the (b) corresponding flood volumes of the simulated inflows to the dam Wupper using the stochastic-deterministic approach and the fitted (a) GEV distribution and (b) Pearson III distribution



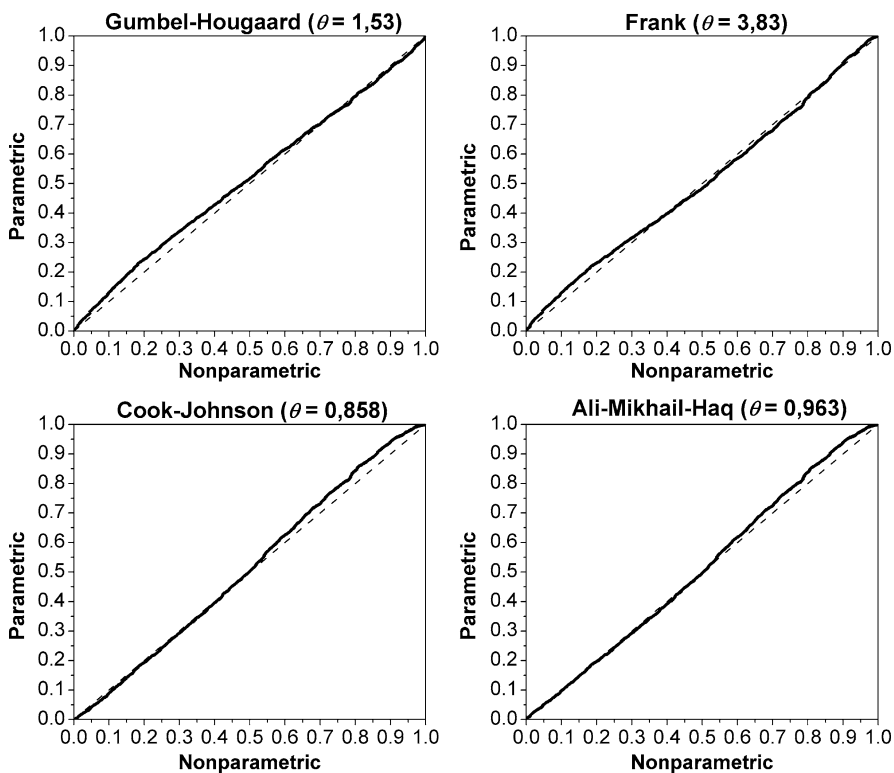
**Table 8.3** Parameters and goodness-of-fit measure AIC of the fitted copulas for the analysis of the flood peaks and the corresponding flood volumes at the dam Wupper

Copula	Estimate(s)	AIC
Gumbel-Hougaard	$\theta = 1.53$	-689
Frank	$\theta = 3.83$	-681
Cook-Johnson	$\theta = 0.858$	-631
Ali-Mikhail-Haq	$\theta = 0.963$	-658
BB1	$\theta_1 = 0.404; \theta_2 = 1.31$	-779

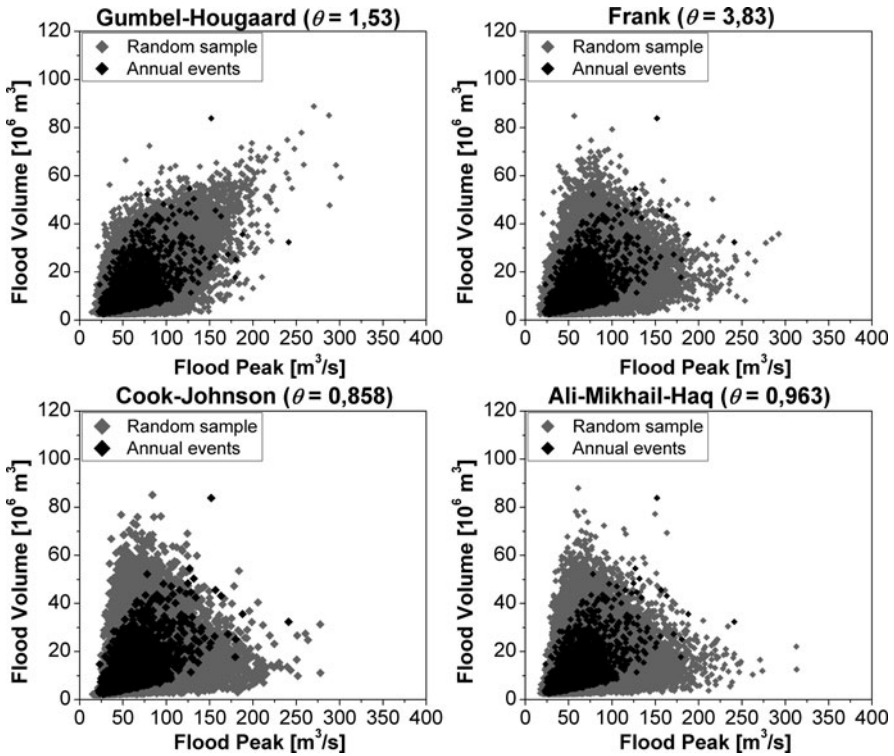
Zhang and Singh, 2006). The parameters have been estimated with the maximum pseudolikelihood method. The fitted parameters are summarized in Table 8.3.

In Fig. 8.7 the parametric and nonparametric values of  $K_C(t)$  are plotted for the four different copulas. A copula is considered as satisfactory if the plot shows a good agreement with the straight dashed line that passes through the origin at  $45^\circ$ .

As a further visual test, the margins of 100,000 random pairs  $(U_j, V_j)$ , 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. The scatter



**Fig. 8.7** Comparison of nonparametric and parametric estimations of  $K_C(t)$  for the fitted Archimedean copulas (The dashed  $45^\circ$ -line represents perfect agreement)



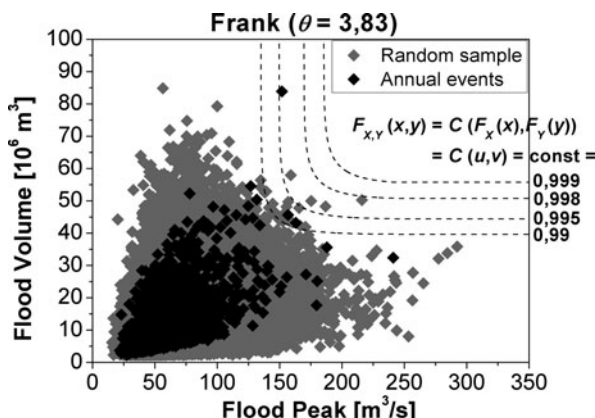
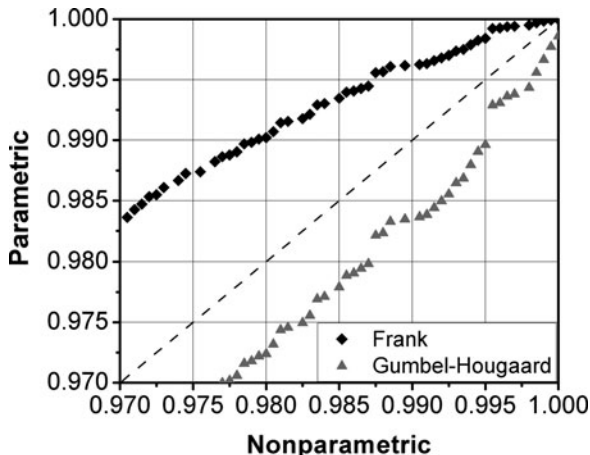
**Fig. 8.8** 100,000 random pairs (*grey dots*) of the annual flood peaks and the corresponding volumes at the dam Wupper chosen from the four Archimedean copulas and the events from the simulated time series (*black dots*)

plots of the four copulas are shown in Fig. 8.8. It can be seen that among these four Archimedean copulas only the Gumbel-Hougaard copula can describe the dependency structure of the sample adequately since only the random pairs generated from this particular copula (*grey dots*) follow the dependency structure of the sample data (*black dots*).

As an additional measure the Akaike's information criterion AIC (Eq. 8.25) was used to compare the goodness of fit of the different copula models and to select the appropriate model which models the dependency between the two random variables optimally. The best model provides the minimal value for this criterion.

Comparing the AIC values of the four Archimedean copulas in Table 8.3 and the parametric and nonparametric estimates it seems that the Gumbel-Hougaard copula and the Frank copula provide a similar goodness-of-fit to the database. Looking more closely at the comparison of the parametric and nonparametric estimation of  $K_C(t)$  for these two copulas in the region of large probabilities (Fig. 8.9) it is obvious that the Gumbel-Hougaard provides a better fit in this region. For the Frank copula the parametric estimate is much higher than the corresponding nonparametric estimate from the sample.

**Fig. 8.9** Comparison of the nonparametric and parametric estimations of  $K_C(t)$  for the Frank and Gumbel-Hougaard copula in the region of large probabilities

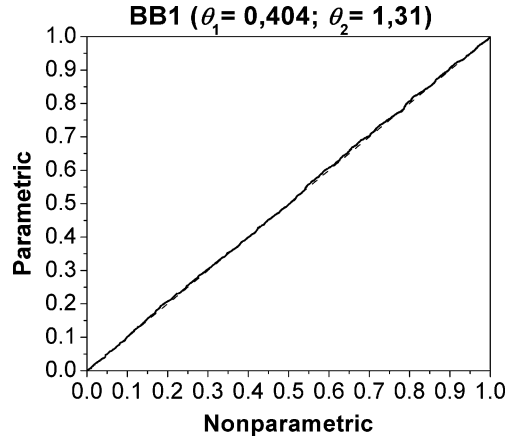


**Fig. 8.10** 100,000 random pairs (grey dots) of the annual flood peaks and the corresponding volumes at the dam Wupper chosen from the Frank copula, the events from the simulated time series (black dots) and level curves for different probabilities  $C(u, v) = t$

E.g. for the same value  $t$  the nonparametric value provides  $K_{C_n}(t) = 0.99$  and the parametric estimate of  $K_{C_\theta}(t)$  for the Frank copula has a value of approx. 0.996. If the random process is modeled with the Frank copula the probability of the occurrence  $P(C(u, v) \leq t) = K_{C_\theta}(t) = 0.996$  of an event with a value of the copula  $C(u, v) \leq t$  is much higher than the empirical probability derived from the dataset  $P(C(u, v) \leq t) = 0.99$ , which was used for parameter estimation of the model. This effect could be demonstrated for the 100,000 random pairs generated from the Frank copula in Fig. 8.10.

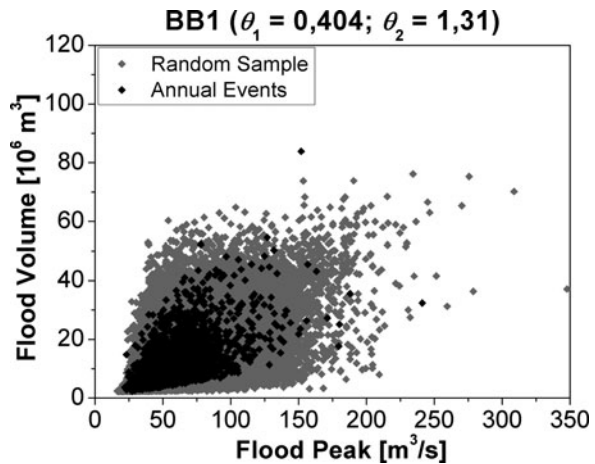
In the random process 18 values of the random pairs  $(X_j, Y_j)$  generated from the copula are above the level curve of  $C(u, v) = t = 0.99$ . The parametric estimate of  $K_{C_\theta}(0.99)$  for the fitted Frank copula is 0.99981. Hence generating 100,000 random

**Fig. 8.11** Comparison of nonparametric and parametric estimations of  $K_C(t)$  for the fitted BB1 copula (The dashed 45°-line represents perfect agreement)

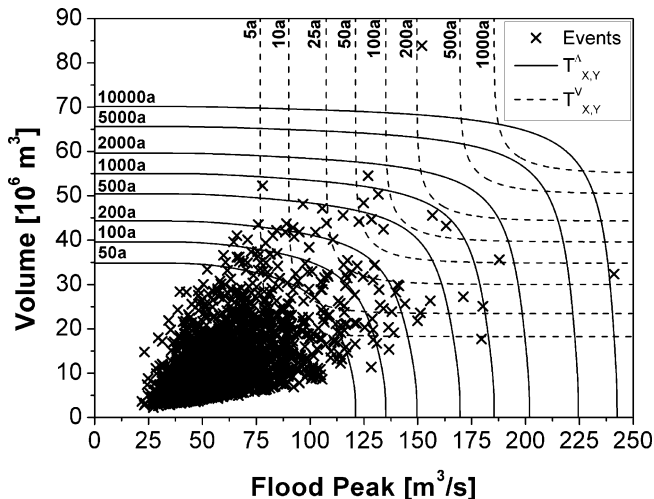


pairs from the copula theoretically 99981 of the pairs should be lying on, under or to the left of the level curve  $C(u, v) = t = 0.99$  and 19 random pairs above. In the sample with the size of 2,000 elements four pairs are above the level curve  $C(u, v) = t = 0.99$ . Therefore for 100,000 realisations of the random process 200 values should be above the level curve to model the random process adequately. This analysis has shown that the Frank copula can not model the dependency structure adequately in the region of large probabilities. Using the Frank copula only a few random pairs are generated where both variables have a small exceedance probability. This example shows that it is important to use different methods for model selection to choose the model with the best fit to the data. Using only one goodness-of-fit measure may result in selecting a wrong model.

None of the four considered copulas provides an acceptable fit to the database. Therefore the two parameter BB1 copula was analyzed besides the four one parameter Archimedian copulas. The comparison of the parametric and nonparametric values of  $K_C(t)$  are shown in Fig. 8.11 and the comparison between 100,000 randomly generated pairs from the BB1 copula and the database are shown in Fig. 8.12.



**Fig. 8.12** 100,000 random pairs (grey dots) of the annual flood peaks and the corresponding volumes at the dam Wupper chosen from the BB1 copula and the events from the simulated time series (black dots)



**Fig. 8.13** Joint return periods  $T_{X,Y}^{\vee}$  (exceeding  $x$  or  $y$ ) and  $T_{X,Y}^{\wedge}$  (exceeding  $x$  and  $y$ ) of the corresponding flood peak and flood volume and the annual events from the simulated time series (*black crosses*)

From these figures and the AIC values in Table 8.3 it became obvious that the BB1 copula provides the best fit to the database. Therefore it was selected for further analysis.

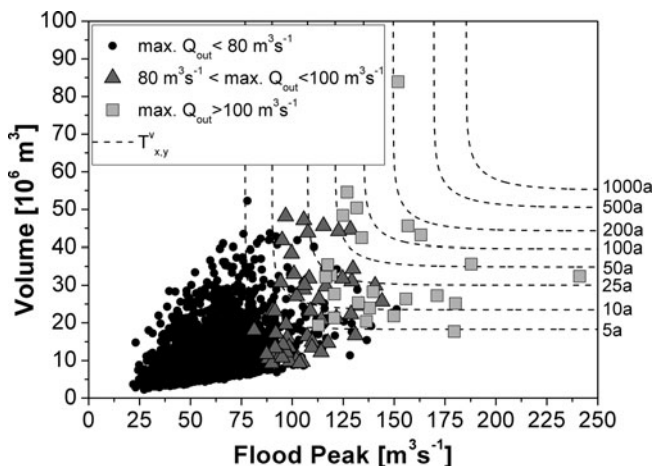
### 8.3.3.3 Bivariate Frequency Analysis

For the bivariate frequency of flood events the contours of the joint return periods  $T_{X,Y}^{\wedge}$  (for which  $x$  and  $y$  are exceeded) and  $T_{X,Y}^{\vee}$  (for which either  $x$  or  $y$  are exceeded by the respective random variables  $X$  and  $Y$ ) with respect to flood peak and corresponding volume and the annual events from the simulated time series are illustrated in Fig. 8.13.

### 8.3.4 Evaluation of the Effect of the Wupper Dam on Flood Control

The yearly floods from the simulated 2000 year time series of the inflow to the dam were used for the analysis of the efficiency of the Wupper dam in flood mitigation. The resulting maximum water levels and maximum releases for these events were calculated with the integrated storage module of the rainfall-runoff model NASIM considering the release rules of the dam with respect to the control gage Kluserbruecke. The discharge from the uninfluenced subcatchment area between the dam and the control gage was considered. As mentioned before it is important to consider flood volume and flood peak for the design and analysis of flood control structures. Here the joint probability of these two flood variables was used as probability measure for the single flood events.

Two different critical values for the release were considered in the analysis: (1) a first control level of  $80 \text{ m}^3\text{s}^{-1}$  and (2) a second critical control discharge



**Fig. 8.14** Maximum resulting outflows at the dam Wupper from the annual events from the simulated time series and the contours of the joint return periods  $T_{X,Y}^v$  (exceeding  $x$  or  $y$ )

of  $100 \text{ m}^3\text{s}^{-1}$  at the gage Kluserbruecke. If the releases of the dam exceed these values critical control levels are exceeded anyway independently from the runoff of the uninfluenced subcatchment between the dam and the gage. These two critical values were chosen to demonstrate the methodology to evaluate the effect of the Wupper dam on flood control. Though these values are defined as control levels at the gage Kluserbruecke both are not specifying critical flood conditions in this region. In Fig. 8.14 the critical events which exceed the respective thresholds are illustrated together with the contours of the joint return periods  $T_{X,Y}^v$  (for which either  $x$  or  $y$  are exceeded by the respective random variables  $X$  and  $Y$ ).

The contours of the joint return period  $T_{X,Y}^v$  can be used to define regions of different hydrological risk where an critical event occurs. The contours which define the boundaries of the different regions are selected in relation to the secondary return period  $\rho_t^v$  after Eq. (8.38). Thresholds  $\vartheta(t)$  for the boundaries were selected for which (in the statistical mean) every  $\rho_t^v = 25, 50, 100$  and  $200$  years an event with a joint return period  $T_{X,Y}^v > \vartheta(t)$  occurs. E.g. the value  $\vartheta(t)$  of the contour line for which in the random process every  $\rho_t^v = 200$  years an event with a return period  $T_{X,Y}^v > \vartheta(t)$  occurs (region situated to the right and above the contour line) is according to the inverse function of Eqs. (8.38) and (8.36)  $\vartheta(t) = 50$  years. The thresholds  $\vartheta(t)$  for the other secondary return periods  $\rho_t^v = 25, 50$  und  $100$  years are summed up in Table 8.4.

If the discharge value of  $80 \text{ m}^3\text{s}^{-1}$  is defined as a critical release of the Wupper dam (triangles and squares in Fig. 8.14) 81 out of the 2,000 events are specified as critical events. The ratios of critical events in the different regions defined by their boundaries in Table 8.4 are illustrated in Table 8.5.

In the region with  $T_{X,Y}^v > 8$  years 60% (45 of 76 events) of the events are critical events. The statistical return period that an event with a joint return period  $T_{X,Y}^v > 8$

**Table 8.4** Thresholds  $\vartheta(t)$ , for which every  $\rho_t^\vee$  years an event occurs with a joint return period  $T_{X,Y}^\vee > \vartheta(t)$

$\rho_t^\vee$ [a]	$\vartheta(t)$ [a]
200	50
100	26
50	14
25	8

**Table 8.5** Ratio of critical events for the regions depending on the joint return period  $T_{X,Y}^\vee$

$T_{X,Y}^\vee$	Number of events	Ratio of critical events	
		Critical release $> 80 \text{ m}^3\text{s}^{-1}$	Critical Release $> 100 \text{ m}^3\text{s}^{-1}$
$> 50 \text{ a}$	9	100% (9 events)	89% (8 events)
$26 \text{ a} < T_{X,Y}^\vee \leq 50 \text{ a}$	7	71% (5 events)	28% (2 events)
$14 \text{ a} < T_{X,Y}^\vee \leq 26 \text{ a}$	25	68% (17 events)	24% (6 events)
$8 \text{ a} < T_{X,Y}^\vee \leq 14 \text{ a}$	35	40% (14 events)	11% (4 events)
$< 8 \text{ a}$	1,924	2% (36 events)	0.2% (4 events)

years occurs in the random process is according to Table 8.4  $\rho_t^\vee = 25$  years. Hence the hydrologic risk is high that an event in this region is a critical event.

If a value of  $100 \text{ m}^3\text{s}^{-1}$  is defined as critical release (squares in Fig. 8.14) 24 out of the 2,000 events are specified as critical events. Eight out of nine events with a joint return period  $T_{X,Y}^\vee > 50$  years are critical events. The statistical return period that an event with  $T_{X,Y}^\vee > 50$  years occurs in the random process is  $\rho_t^\vee = 200$  years (see Table 8.4). In this region the hydrological risk is very high that an event is a critical event.

Measures to improve the efficiency of flood protection of the dam, e.g. optimized flood control or enlarged flood control storage, can be evaluated by comparing the change of the ratios of critical events in Table 8.5. For the example illustrated in Table 8.5 the flood protection of the dam is higher for a larger defined critical release.

It is obvious that the distribution of the critical events can not be described by using the univariate return period of flood peak or flood volume only. Hence it is very important to consider the joint probability of these two flood characteristics. Even with the joint probability of flood peak and flood volume the distribution of critical events can not be described completely because it depends also on other random variables such as the water level at the begin of the event, pre-events, runoff of the subcatchment, shape of the event, etc.

### 8.4 Case Study 2: Unstrut River Basin

The second case study for the application of copulas for the risk analysis of dams is the Unstrut River Basin. In contrast to the first case study examining only a single dam the flood protection system of the river Unstrut catchment is substantially

more complex. The flood control system of the river Unstrut has been investigated, optimized and extended through an integrated and interdisciplinary flood risk assessment instrument as part of an interdisciplinary research project (Nijssen et al., 2009). For the risk-oriented approach a large variety of different hydrological scenarios was generated by coupling a stochastic rainfall generator with a deterministic rainfall-runoff model. For risk analysis a probability has to be attributed to the individual hydrological scenarios. Copulas are used to set up multivariate probability distributions which consider more than one flood characteristic in the frequency analysis of the events (Klein et al., 2010). Two different bivariate probabilities were used for the risk-analysis:

The joint probability of flood coincidences at reservoir sites at two rivers is applied to consider possible superposition of floods from the main river and one tributary. It aims to describe the spatial heterogeneity of flood events within these analyses (Section 8.4.3).

The joint probabilities of the flood peak and corresponding flood volume at the two reservoirs sites are used to characterize the flood retention in both reservoirs completely. The flood volume is very important and therefore its probability should be considered in the risk-based analysis in addition to the probability of the peak (Section 8.4.4).

### ***8.4.1 Description of the River Basin***

The flood prone catchment of the river Unstrut has a watershed of 6,343 km<sup>2</sup> and is situated in Mid-East Germany (see Fig. 8.15). The catchment has a variable topographic structure with elevations ranging from 104 to 982 m a.s.l. Two mountainous regions are located in the area: the Harz Mountains in the North and the Thuringian Forest Mountains in the South of the basin. Two large flood control reservoirs, the reservoir Kelbra with a catchment area of 664 km<sup>2</sup> and a flood control storage of 35.6 × 10<sup>6</sup> m<sup>3</sup> and the flood detention reservoir Straussfurt with a catchment area of 2,044 km<sup>2</sup> and a flood control storage of 18.6 × 10<sup>6</sup> m<sup>3</sup> were built to control the two main tributaries. A large polder system with a total storage volume of approximately 50 × 10<sup>6</sup> m<sup>3</sup> is located downstream of these two flood control reservoirs.

The reservoir Straussfurt is operated with seasonally varying flood control storage. During winter time a storage of 18.6 × 10<sup>6</sup> m<sup>3</sup> and in summer a storage of 12.7 × 10<sup>6</sup> m<sup>3</sup> is allocated for flood control. The maximum capacity of the operating outlets is approx. 300 m<sup>3</sup>s<sup>-1</sup> and the flood spillway is a crested weir with a length of 270 m.

The reservoir Kelbra is operated with seasonally varying flood control storage as well. In winter a storage of 35.6 × 10<sup>6</sup> m<sup>3</sup> and in summer a storage of 23.3 × 10<sup>6</sup> m<sup>3</sup> is available for flood control. The maximum capacity of the operating outlets is 244 m<sup>3</sup>s<sup>-1</sup>. Due to this large capacity of the operating outlets the reservoir has no flood spillway. The control gages for the combined control of the reservoirs are Bennungen, Oldisleben and Wangen. During a flood the discharges at



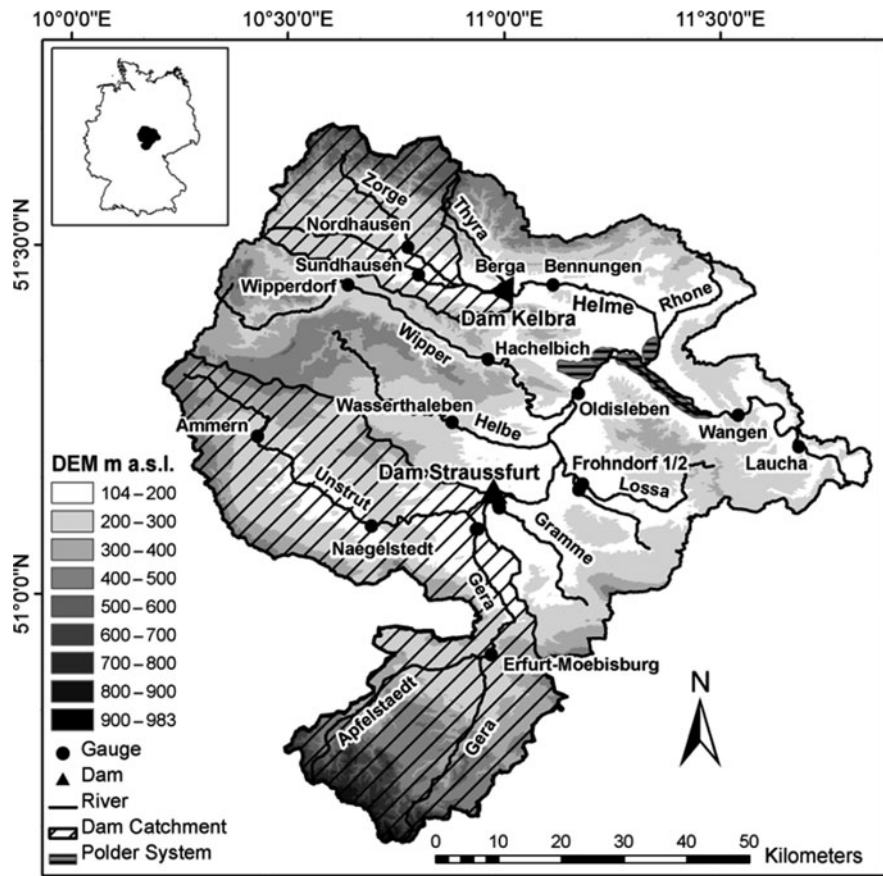


Fig. 8.15 Topographical map of the watershed of the river Unstrut, shown are important gages in the watershed, the polder system and the dams Kelbra and Strausfurt with their respective subcatchments

the gage Bennungen must not exceed  $30 \text{ m}^3\text{s}^{-1}$ , at gage Oldisleben  $131 \text{ m}^3\text{s}^{-1}$  and at the gage Wangen  $152 \text{ m}^3\text{s}^{-1}$ .

### 8.4.2 Stochastic-Deterministic Generation of Flood Events

A long-term daily discharge time series of 10,000 years was generated by coupling a stochastic spatially distributed daily rainfall generator (Hundecka et al., 2009) with a hydrological rainfall-runoff model following the concept of the well-known Swedish model HBV (Lindström et al., 1997). For the parameter estimation of the stochastic rainfall generator 122 stations with observed daily values from 1961 to 2003 and for the calibration of the continuous rainfall-runoff model on daily basis

discharge series from 1991 to 1996 were used. It has been shown that the statistical properties of the daily rainfall and the runoff were well reproduced (Hundecha et al., 2008, 2009). Due to computational limitations it wasn't possible to build up a continuous rainfall-runoff model on hourly basis. Flood events have been selected from the generated daily discharge time series of 10,000 years. For these events the daily precipitation was disaggregated into hourly values using the models MuDRain (Koutsoyiannis, 2001) and HYETOS (Koutsoyiannis et al., 2003) and used as input for an event-based hourly hydrological rainfall-runoff model following the same concept as the continuous rainfall-runoff model on daily basis. The starting and boundary conditions for the event-based model were derived from the continuous model. For the calibration of the event-based rainfall runoff model only two flood events (April 1994 und December/January 2002/03) were available. The operation of the reservoirs has been implemented in the rainfall-runoff models.

A representative sample of the population of possible flood events is required to evaluate a flood control system. Here six different return periods were considered ( $T = 25, 50, 100, 200, 500, 1,000$  years). The risk is not only related to the flood peak, since different hydrograph shapes (e.g. multi-peak floods or floods with long duration and large volume) will result in different risk levels. The selection of hydrological scenarios was therefore further supported by cluster analysis of historical events to determine typical hydrograph shapes and volumes. In total five different hydrological scenarios were selected for each return period at the reference gage Strausfurt, which encompass various spatial distributions of precipitation. Furthermore, one event with a return period of the peak of more than 1,000 years was considered. Hence in total 31 hydrological scenarios were selected from the entire 10,000 years discharge time series. A more detailed description of the event selection and the generation of the flood events can be found in Schumann (2009).

### ***8.4.3 Bivariate Frequency Analysis of Corresponding Flood Peaks at the Reservoir Sites***

Since the observed time series was too short for multivariate frequency analysis and was influenced by the control of the reservoirs and river construction works in the sixties the generated synthetic 10,000 year time series has been used for the bivariate frequency analysis shown here.

To account for the impact of different tributaries in the risk analysis, it is crucial to characterize the spatial distribution of the flood events due to the topographic structure of this particular catchment, where flooding originates mainly from mountainous regions upstream of the two reservoirs. From the analysis of the historical flood events, it was found that particularly three different rainfall distributions have to be considered in the risk-analysis: (1) heavy precipitation in the mountainous region in the North/North-West, (2) heavy precipitation in the mountainous region in the South and (3) concurrent heavy precipitation in both mountainous regions.

The inflow gages of the two reservoirs (see Fig 8.15) were henceforth used as reference locations to quantify the probability of occurrence of the 31 hydrological

**Table 8.6** Measure of dependency for corresponding flood peaks at reservoirs Strausfurt and Kelbra

Sample size	Pearson's $r$	Kendall's $\tau$	Spearman's $\rho$
12,768	0.72	0.48	0.67

scenarios that were selected from the generated 10,000 year time series for the risk analysis. These two reservoirs represent the upper boundary of the flood control system of the Unstrut catchment. The remaining flood protection structures are located downstream of these two reservoirs. The joint probability between the two flood peaks at the inflow gage of the two reservoirs was assessed by copulas to describe the overall probability of occurrence of flood events for the entire region in terms of probabilities of coincidences of a particular flood event at both tributaries. Hence the corresponding annual flood peaks had to be selected for further analysis. The annual flood peaks of the inflow time series to the reservoir Strausfurt and corresponding flood peaks of the inflow time series to the reservoir Kelbra have been determined. In cases where the identified flood peak at the reservoir Kelbra did not coincide with the annual flood peak at this reservoir, the annual flood peak at dam Kelbra and the corresponding flood peak at dam Strausfurt were added to the sample. In 2,768 years of the generated 10,000 years of discharge the annual flood peaks of the two inflow gages did not occur at the same time, and therefore the total selected sample size was 12,768.

The statistical dependencies between the two random variables have been estimated by the statistical measures listed in Table 8.6. The summary shows that there is a strong positive dependency between the two random variables.

#### 8.4.3.1 Marginal Distributions

It was found that both variables can be described in the best way with the GEV distribution. The two parameter estimation methods L-Moments and product of moments provided nearly identical goodness of fit to the flood peaks at the two reservoirs. However, using graphical diagnostics the parameter estimation method of L-Moments (Hosking and Wallis, 1997) gave a better fit to the sample data for the distribution of the flood peaks at Strausfurt, while the method of product of moments provided a better fit for flood peaks at Kelbra, particularly in the region with small exceedance probabilities.

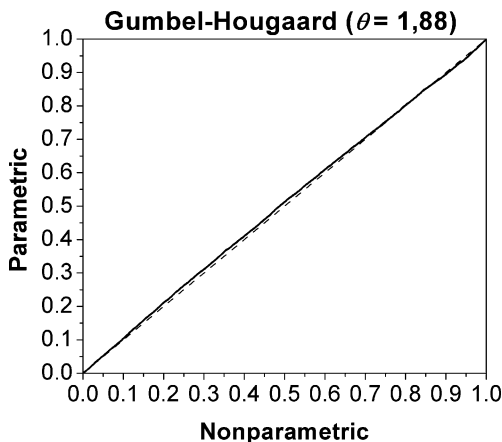
#### 8.4.3.2 Copula Estimation

Only the Gumbel-Hougaard family was applicable to describe the dependency structure of the data among the four commonly applied Archimedian copulas mentioned in Section 8.3.3. This copula was chosen for the analysis, since the dependency structure between the flood peaks and the volumes at the dam Wupper is akin to the dependency structure between the two corresponding flood peaks. However, in addition to the Gumbel-Hougaard copula, the two-parameter copula BB1 was used for the construction of the bivariate distribution function. The parameters have been

**Table 8.7** Parameters and goodness-of-fit measure AIC of the fitted copulas for the analysis of the corresponding flood peaks at the two reservoirs

Copula	Estimate (s)	AIC
Gumbel-Hougaard	$\theta = 1.88$	-8,254
BB1	$\theta_1 = 0.114; \theta_2 = 1.79$	-8,303

**Fig. 8.16** Comparison of nonparametric and parametric estimations of  $K_C(t)$  for the fitted Gumbel-Hougaard copula

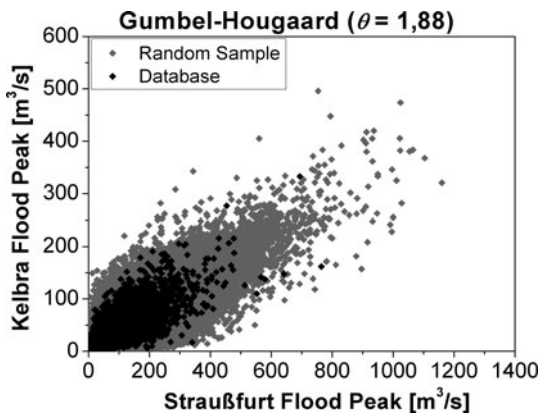


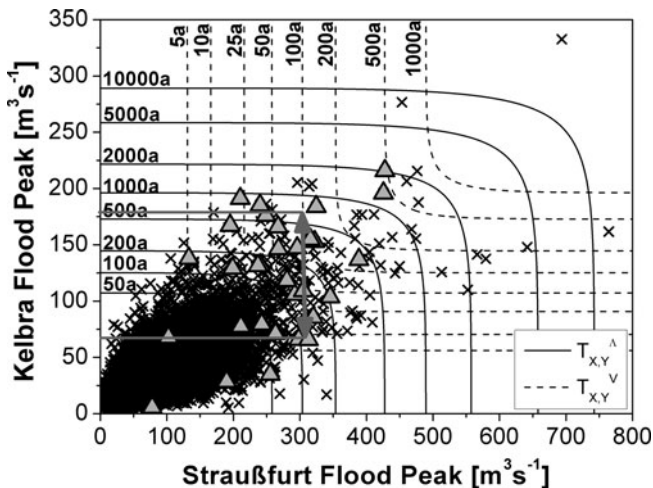
estimated with the maximum pseudolikelihood method. The fitted parameters are summarized in Table 8.7.

The goodness-of-fit of both copulas was nearly identical (see e.g. the AIC values in Table 8.7). Hence the Gumbel-Hougaard copula, having less parameters than the BB1 copula, was selected for further analysis. For brevity only the results for the analysis of the Gumbel-Hougaard copula are shown here.

In Fig. 8.16 the parametric and nonparametric values of  $K_C(t)$  are plotted for the Gumbel-Hougaard copula. If the plot is in good agreement with the straight dashed line that passes through the origin at 45°, then the copula is considered satisfactory. In Fig. 8.17 the margins of 1,000,000 random pairs  $(U_j, V_j)$ , 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. Both figures show that

**Fig. 8.17** 1,000,000 random pairs (grey dots) of the corresponding annual flood peaks at the two reservoirs chosen from the Gumbel-Hougaard copula and the events from the simulated time series (black dots)





**Fig. 8.18** Joint return periods  $T_{X,Y}^{\vee}$  (exceeding  $x$  or  $y$ ) and  $T_{X,Y}^{\wedge}$  (exceeding  $x$  and  $y$ ) of the annual flood peaks at the two reservoirs, the annual events from the simulated time series (black crosses) and the selected hydrological scenarios (grey triangles)

the Gumbel-Hougaard copula models the dependency structure between the two random variables adequately.

### 8.4.3.3 Bivariate Frequency Analysis

Figure 8.18 illustrates the contours of the joint return periods  $T_{X,Y}^{\wedge}$  (“OR”-case) as well as the joint return periods  $T_{X,Y}^{\vee}$  (“OR”-case) with respect to the corresponding flood peaks at the two reservoirs.

The selected 31 events represent a large variety of different hydrological scenarios. Interestingly, using for example the selected events with a flood peak of a  $\sim 100$  year return period at the reservoir Straussfurt for the design of flood protection structures, the corresponding return periods of the selected flood peaks at the reservoir Kelbra range between 10 and 500 years. This additional information can be used for the planning and design of spatial distributed flood control structures in general as it is shown in Chapter 12.

### 8.4.4 Bivariate Frequency Analysis of the Annual Flood Peaks and the Corresponding Volumes

As mentioned before, for flood storage facilities such as reservoirs and polders it is important to consider the flood volume besides the flood peak in frequency analyses. Hence the joint return period of the corresponding flood peak and volume is used to assign a return period to the flood events at the reservoirs Straussfurt and Kelbra. Corresponding values of annual flood peaks and flood volumes were selected from 10,000 years of generated data. The corresponding flood volume to the annual flood peaks is selected according to Fig. 8.5. The estimated parameters

**Table 8.8** Measure of dependency for the peak-volume datasets at the Unstrut watershed

Sample size	Pearson's $r$	Kendall's $\tau$	Spearman's $\rho$
Corresponding flood peaks and volumes at reservoir Straussfurt 10,000	0.81	0.66	0.85
Corresponding flood peaks and volumes at reservoir Kelbra 10,000	0.74	0.57	0.76

which were applied to describe the dependency between flood peak and volume are given in Table 8.8. As before, there is a strong positive dependency between the corresponding flood peaks and volumes at the two reservoirs. For brevity, only the results of the analysis at the reservoir Straussfurt are presented here, since the dependency structures of the corresponding flood peaks and volumes at the two reservoirs were found to be similar.

**8.4.4.1 Marginal Distributions**

As for the analysis of the corresponding flood peaks at the two reservoirs, the Generalized Extreme Value (GEV) distribution was chosen as marginal distribution for the annual flood peaks at Straussfurt and Kelbra. For the corresponding flood volumes the GEV distribution provided the best fit using the method of product of moments as parameter estimation method.

**8.4.4.2 Copula Estimation**

As before the Gumbel-Hougaard and the BB1 copula were chosen for the analysis. The fitted parameters and the goodness-of-fit measure are listed in Table 8.9.

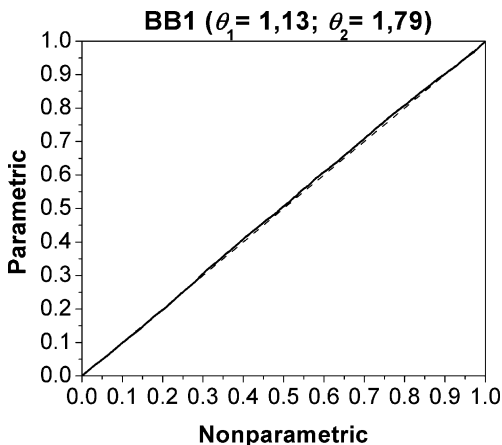
For both reservoirs the BB1 copula provided the better fit for the structure of dependency of the two random variables. Therefore it was chosen for further analysis.

In Fig. 8.19 the parametric and nonparametric values of  $K_C(t)$  are plotted for the BB1 copula and in Fig. 8.20 the margins of 1,000,000 random pairs  $(U_j, V_j)$ , 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. Both figures

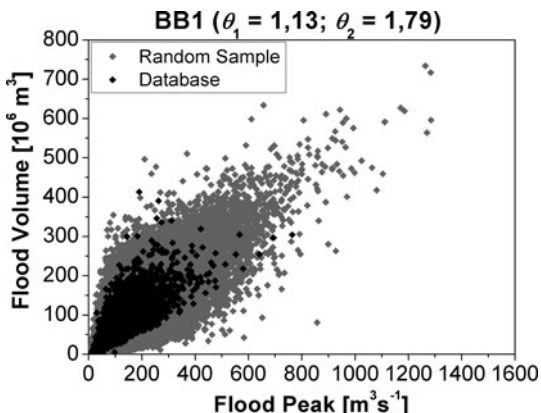
**Table 8.9** Parameters and goodness-of-fit measure AIC of the fitted copulas for the analysis of the flood peaks and the corresponding flood volumes at the dam Straussfurt and Kelbra

Copula	Estimate (s)	AIC
Corresponding flood peaks and volumes at reservoir Straussfurt		
Gumbel-Hougaard	$\theta = 2.6$	-11,787
BB1	$\theta_1 = 1.13; \theta_2 = 1.79$	-13,725
Corresponding flood peaks and volumes at reservoir Kelbra		
Gumbel-Hougaard	$\theta = 2.12$	-8,256
BB1	$\theta_1 = 0.986; \theta_2 = 1.52$	-9,818

**Fig. 8.19** Comparison of nonparametric and parametric estimations of  $K_C(t)$  for the fitted BB1 copula (The dashed 45°-line represents perfect agreement)



**Fig. 8.20** 1,000,000 random pairs (grey dots) of the annual flood peak and the corresponding flood volume chosen from the BB1 copula at the reservoir Strausfurt and the events from the simulated time series (black dots)

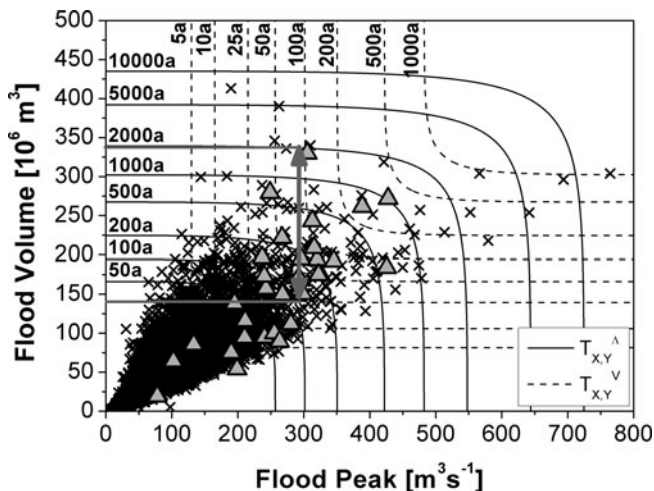


show that the BB1 copula models the dependency structure between the two random variables adequately.

### 8.4.4.3 Bivariate Frequency Analysis

Figure 8.21 illustrates the contours of the joint return periods  $T_{X,Y}^{\wedge}$  (“AND”-case) and  $T_{X,Y}^{\vee}$  (“OR”-case) with respect to the annual flood peak and the corresponding volume, the annual events from the simulated time series and the selected events (triangles). Again, using the events with a return period of the flood peak around 100 years as an example, the corresponding return periods of the flood volume range between 25 and 2,000 years. It is therefore recommended to use the joint probabilities for a detailed description of flood events instead of using the univariate probability of the flood peak only.





**Fig. 8.21** Joint return periods  $T_{X,Y}^{\vee}$  (exceeding  $x$  or  $y$ ) and  $T_{X,Y}^{\wedge}$  (exceeding  $x$  and  $y$ ) of the flood peaks and corresponding flood volumes at the reservoir Strausfurt, the annual events from the simulated time series (*black crosses*) and the selected hydrological scenarios (*grey triangles*)

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 Strausfurt, as well as the joint return periods of the flood peak at the reservoirs Strausfurt and Kelbra. The selected events are displayed in Figs. 8.18 and 8.21 (shown as triangles). By using this approach, the spatial distribution between the two main tributaries is considered and the crucial aspect of using different combinations of flood peak and volume can be taken into consideration in flood risk analyses. Those data are in turn used for assessment of the flood control system, which hereby integrates the spatial component of probability.

### 8.4.5 Evaluation of the Effect of the Reservoir Strausfurt on Flood Control

The efficiency of the reservoir Strausfurt in flood mitigation is evaluated using the 31 selected hydrological scenarios. In Fig. 8.22 the maximal resulting water levels at the reservoir Strausfurt 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.

Here a water level exceeding 150.3 m a.s.l. is defined as a critical hydrologic event, since it is known that the corresponding outflow of more than  $200 \text{ m}^3 \text{ s}^{-1}$  would cause severe damage downstream. With a joint return period  $T_{X,Y}^{\vee}$  (exceeding  $x$  or  $y$ ) greater than 50 years (gray area in Fig. 8.22), all considered events are critical events according to this definition. Hence the hydrological risk is very high for these events. The mean interarrival time of an event lying in the region of very



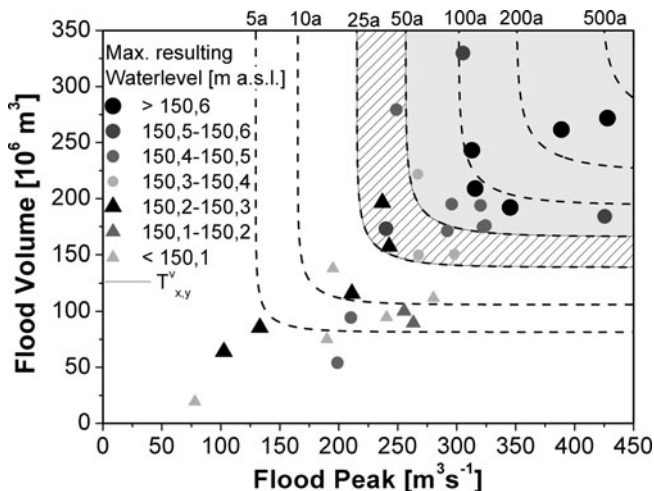


Fig. 8.22 Maximum resulting water levels at the reservoir Straussfurt from the different hydrological scenarios and the contours of the joint return periods  $T_{X,Y}^{\nabla}$  (exceeding  $x$  or  $y$ )

high hydrological risk, calculated with the secondary return period (Eq. 8.38) is up to 110 years. For joint return periods between 25 and 50 years (black shaded area in Fig. 8.22) five of the seven events which were considered are critical events. In this region the risk is high that an event with a return period  $T_{X,Y}^{\nabla}$  (exceeding  $x$  or  $y$ ) between 25 and 50 years is a critical event. The return period of an event lying in this region  $25 \text{ a} < T_{X,Y}^{\nabla} < 50 \text{ a}$  can be calculated to 105 years using following formula

$$\lambda_{t_1 t_2}^{\nabla} = \frac{\mu_T}{P(t_1 < C(u, v) \leq t/2)} = \frac{\mu_T}{K_C(t_2) - K_C(t_1)} \tag{8.39}$$

with

$$t = 1 - \frac{\mu_T}{T_{X,Y}^{\nabla}}. \tag{8.40}$$

In the region with a joint return period  $T_{X,Y}^{\nabla}$  smaller than 25 years only two out of 12 events will result in substantial damages. The risk that events with these joint probabilities may cause damage is relatively low. Not all critical events can be identified by the joint return period  $T_{X,Y}^{\nabla}$  such, as e.g. the two critical events lying in the region with  $T_{X,Y}^{\nabla} < 25 \text{ a}$ , because the distribution of the critical events depends additionally on other flood characteristics (shape of hydrograph, antecedent conditions, flood storage at the onset of the event etc.). But not all random variables can be considered in the multivariate frequency analyses and therefore this analysis is reduced to the two most important random variables flood peak and corresponding volume.

Flood reduction measures such as an optimized control or enlarged flood storage could be analyzed by the change of the critical events in the defined regions.

## 8.5 Conclusions

A large variety of different hydrological loads can be generated by coupling a spatially distributed stochastic rainfall generator with a rainfall-runoff model. Especially in large catchment areas with several interacting flood protection structures it is important to consider the spatial distribution of floods. For the risk analysis the different hydrological loads have to be categorized probabilistically. In this chapter a methodology to categorize flood events for risk analysis and risk-orientated design of dams and flood control systems using copulas is presented. The advantage of using copulas is the possibility to apply a multivariate distribution function with different univariate marginal distributions. Joint return periods and also conditional return periods can be derived easily for hydrologic scenarios from copulas.

Results from two different case studies were presented: a single dam located at the river Wupper was analyzed in the first case study and a flood protection system with two dams and a flood polder system – the river Unstrut – were analyzed in the second case study. Different flood characteristics were used to derive probabilities for the risk analysis. For single flood control structures the flood volume was considered besides the flood peak to estimate the joint probability. For large river basins it is also important to consider the spatial distribution of the flood events in the risk analysis. Therefore, to reassign an overall probability of a flood event, which has different probabilities in the different tributaries, the joint return period of the corresponding flood peaks in the main tributaries can be obtained using copula analysis. For the example of the case study for the Unstrut river the flood peaks of the inflows to the two reservoirs located at two main tributaries were used for the bivariate frequency analysis to characterize the flood risk for the confluence.

Another advantage using copulas is the possibility to identify critical events for flood protection structures such as reservoirs and to evaluate the effect of flood control of these reservoirs. Using the joint return period  $T_{X,Y}^{\vee}$ , for which the random variables flood peak ( $X$ ) or corresponding flood volume ( $Y$ ) exceed the respective thresholds  $x$  or  $y$ , thresholds can be identified to evaluate the hydrological risk of critical events.

In general observed time series are too short for multivariate frequency analysis. By generating a synthetic discharge time series through coupling a stochastic rainfall generator with a deterministic rainfall-runoff model the database can be extended for the multivariate analysis. It must be noted that a simulated database can not replace observed records. On the contrary observations are required for reliable parameter estimation of the simulation model and for the verification of the results.

Copula analysis can provide valuable information to decision makers. Such information is more adequate for flood control problems than traditional return period analysis. Consideration of multiple flood characteristics is essential for risk-based planning.

## References

- Akaike H (1974) A new look at statistical model identification. *IEEE Trans Automatic Control* AC19(6):716–723
- Bergmann H, Sackl B (1989) Determination of design flood hydrographs based on regional hydrological data. In: Kavvas ML (ed) *New directions for surface water modelling*. IAHS Publ. no. 181, Wallingford
- Blazkova S, Beven K (2004) Flood frequency estimation by continuous simulation of subcatchment rainfalls and discharges with the aim of improving dam safety assessment in a large basin in the Czech Republic. *J Hydrol* 292(1–4):153–172
- Cherubini U, Luciano E, Vecchiato W (2004) *Copula methods in finance*. Wiley, Chichester
- Deheuvels P (1979) Empirical dependence function and properties – nonparametric test of independence. *Bull De La Classe Des Sci Acad R De Belg* 65(6):274–292
- De Michele C, Salvadori G (2003) A generalized pareto intensity-duration model of storm rainfall exploiting 2-Copulas. *J Geophys Res Atmospheres* 108(D2):ACL 15–1
- De Michele C, Salvadori G, Canossi M, Petaccia A, Rosso R (2005) Bivariate statistical approach to check adequacy of dam spillway. *J Hydrol Eng* 10(1):50–57
- Embrechts P, Lindskog F, McNeil AJ (2003) Modelling dependence with copulas and applications to risk management. In: Rachev ST (ed) *Handbook of heavy tailed distributions in finance*. Elsevier, North-Holland, Amsterdam
- Evin G, Favre AC (2008) A new rainfall model based on the Neyman-Scott process using cubic copulas. *Water Resour Res* 44(3):1–18
- Favre AC, El Adlouni S, Perreault L, Thiémonge N, Bobee B (2004) Multivariate hydrological frequency analysis using copulas. *Water Resour Res* 40(1):1–12
- Genest C, Favre AC (2007) Everything you always wanted to know about copula modeling but were afraid to ask. *J Hydrologic Eng* 12(4):347–368
- Genest C, Remillard B (2008) Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Annales de l'Institut Henri Poincaré – Probabilités et Statistique* 44(6):1096–1127
- Genest C, Rivest LP (1993) Statistical-Inference procedures for Bivariate Archimedean Copulas. *J Am Stat Assoc* 88(423):1034–1043
- Genest C, Quessy JF, Remillard B (2006) Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scand J Stat* 33(2):337–366
- Genest C, Remillard B, Beaudoin D (2009) Goodness-of-fit tests for copulas: a review and a power study. *Insur Math Econ* 44(2):199–213
- Goel NK, Seth SM, Chandra S (1998) Multivariate modeling of flood flows. *J Hydraulic Eng* 124(2):146–155
- Grimaldi S, Serinaldi F (2006a) Design hyetograph analysis with 3-copula function. *Hydrol Sci J-J Des Sci Hydrologiques* 51(2):223–238
- Grimaldi S, Serinaldi F (2006b) Asymmetric copula in multivariate flood frequency analysis. *Adv Water Resour* 29(8):1155–1167
- Haberlandt U, Ebner von Eschenbach AD, Buchwald I (2008) A space-time hybrid hourly rainfall model for derived flood frequency analysis. *Hydrol Earth Syst Sci* 12(6):1353–1367
- Hiemstra LAV, Zucchini WS, Pegram GGS (1976) Method of finding family of runhydrographs for given return periods. *J Hydrol* 30(1–2):95–103
- Hosking JRM, Wallis JR (1997) *Regional frequency analysis: an approach based on L-moments*. Cambridge University Press, Cambridge
- Hundecha Y, Pahlow M, Klein B, Schumann A (2008) Development and test of a stochastic rainfall generator for the generation of hydrologic load scenarios (in German). *Forum für Hydrologie und Wasserwirtschaft, Heft 23.06, Beiträge zum Tag der Hydrologie 2008, Hannover*, pp 34–41
- Hundecha Y, Pahlow M, Schumann A (2009) Modeling of daily precipitation at multiple locations using a mixture of distributions to characterize the extremes. *Water Resour Res* 45(W12412): 1–15

- Hydrotec (2007) Niederschlags-Abfluss-Modell – NASIM – Version 3.6.0. Programmdokumentation. Hydrotec Ingenieurgesellschaft für Wasser und Umwelt mbH, Aachen
- Joe H (1997) Multivariate models and dependence concepts. Chapman and Hall, New York
- Kao SC, Govindaraju RS (2008) Trivariate statistical analysis of extreme rainfall events via the Plackett family of copulas. *Water Resour Res* 44(2):1–19
- Karmakar S, Simonovic SP (2008) Bivariate flood frequency analysis: Part 1. Determination of marginals by parametric and nonparametric techniques. *J Flood Risk Manag* 1: 190–200
- Karmakar S, Simonovic SP (2009) Bivariate flood frequency analysis. Part 2: a copula-based approach with mixed marginal distributions. *J Flood Risk Manag* 2:32–44
- Kim G, Silvapulle MJ, Silvapulle P (2007) Comparison of semiparametric and parametric methods for estimating copulas. *Comput Stat Data Anal* 51(6):2836–2850
- Klein B, Schumann A (2006) Generation of multi-peak hydrographs for the flood design of dams (in German). *Forum für Hydrologie und Wasserwirtschaft*, Heft 15.06, Beiträge zum Tag der Hydrologie 2006, Band 2, München, pp 255–266
- 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 Eng* 15(5). doi: 10.1061/(ASCE)HE.1943-5584.0000204
- Koutsoyiannis D (2001) Coupling stochastic models of different time scales. *Water Resour Res* 37(2):379–392
- Koutsoyiannis D, Onof C, Wheeler HS (2003) Multivariate rainfall disaggregation at a fine timescale. *Water Resour Res* 39(7):1–18
- Lindström G, Johansson B, Persson M, Gardelin M, Bergström S (1997) Development and test of the distributed HBV-96 hydrological model. *J Hydrol* 201(1–4):272–288
- Nelsen RB (1999) *An introduction to copulas*. Springer, New York, NY
- Nijssen D, Schumann A, Pahlow M, Klein B (2009) Planning of technical flood retention measures in large river basins under consideration of imprecise probabilities of multivariate hydrological loads. *Nat Hazards Earth Syst Sci* 9(4):1349–1363
- Petry U, Hundecha Y, Pahlow M, Schumann A (2008) Generation of severe flood scenarios by stochastic rainfall in combination with a rainfall runoff model. *Proceedings of the 4th International Symposium on Flood Defense*, 6–8 May, Toronto, ON
- Renard B, Lang M (2006) Use of a Gaussian copula for multivariate extreme value analysis: Some case studies in hydrology. *Adv Water Resour* 30(4):897–912
- Salvadori G, De Michele C (2004) Frequency analysis via copulas: theoretical aspects and applications to hydrological events. *Water Resour Res* 40(12):1–17
- Salvadori G, De Michele C (2006) Statistical characterization of temporal structure of storms. *Adv Water Resour* 29(6):827–842
- Salvadori G, De Michele C (2007) On the use of copulas in hydrology: theory and practice. *J Hydrol Eng* 12(4):369–380
- Salvadori G, De Michele C, Kotegoda NT, Rosso R (2007) *Extremes in nature: an approach using copulas*. Series: Water Science and Technology Library, vol 56. Springer, Dordrecht
- Sandoval CE, Raynal-Villasenor J (2008) Trivariate generalized extreme value distribution in flood frequency analysis. *Hydrol Sci J-J Des Sci Hydrologiques* 53(3):550–567
- Saunders R, Laud P (1980) The multidimensional kolmogorov goodness-of-fit test. *Biometrika* 67(1):237–237
- Schumann A (2009) Abschlussbericht des BMBF-Verbundforschungsvorhabens “Integrative Nutzung des technischen Hochwasserrückhalts in Poldern und Talsperren am Beispiel des Flussgebiets der Unstrut”. Schriftenreihe Hydrologie/Wasserwirtschaft Nr. 24, Lehrstuhl für Hydrologie, Wasserwirtschaft und Umwelttechnik, Ruhr-Universität Bochum, Bochum
- Schwarz G (1978) Estimating the dimension of a model. *Ann Stat* 6(2):461–464
- Serinaldi F, Grimaldi S (2007) Fully nested 3-copula: procedure and application on hydrological data. *J Hydrol Eng* 12(4):420–430

- Shiau JT (2003) Return period of bivariate distributed extreme hydrological events. *Stoch Environ Res Risk Assess* 17(1–2):42–57
- Shiau JT, Feng S, Nadaraiah S (2007) Assessment of hydrological droughts for the Yellow River, China, using copulas. *Hydrol Processes* 21(16):2157–2163
- Shiau JT, Wang HY, Tsai CT (2006) Bivariate frequency analysis of floods using copulas. *J Am Water Resour Assoc* 42(6):1549–1564
- Singh K, Singh VP (1991) Derivation of bivariate probability density-functions with exponential marginals. *Stoch Hydrol Hydraul* 5(1):55–68
- Sklar A (1959) Fonctions de répartition à n dimensions et leurs marges. *Publ Inst Stat Univ Paris* 8:220–231
- Stedinger JR, Vogel RM, Foufoula-Georgiou E (1993) Frequency analysis of extreme events. In: Maidment DR (ed) *Handbook of hydrology*. McGraw-Hill, New York, NY
- Whelan N (2004) Sampling from Archimedean copulas. *Quant Finance* 4(3):339–352
- Yue S (1999) Applying bivariate normal distribution to flood frequency analysis. *Water Int* 24(3):248–254
- Yue S (2001) A bivariate gamma distribution for use in multivariate flood frequency analysis. *Hydrol Processes* 15(6):1033–1045
- Yue S, Ouarda T, Bobee B (2001) A review of bivariate gamma distributions for hydrological application. *J Hydrol* 246(1–4):1–18
- Yue S, Ouarda T, Bobee B, Legendre P, Bruneau P (1999) The Gumbel mixed model for flood frequency analysis. *J Hydrol* 226(1–2):88–100
- Zhang L, Singh VP (2006) Bivariate flood frequency analysis using the copula method. *J Hydrol Eng* 11(2):150–164
- Zhang L, Singh VP (2007a) Bivariate rainfall frequency distributions using Archimedean copulas. *J Hydrol* 332(1–2):93–109
- Zhang L, Singh VP (2007b) Gumbel-Hougaard copula for trivariate rainfall frequency analysis. *J Hydrol Eng* 12(4):409–419
- Zhang L, Singh VP (2007c) Trivariate flood frequency analysis using the Gumbel-Hougaard copula. *J Hydrol Eng* 12(4):431–439