Defining flood risk in a multivariate framework: Application on the Panaro watershed

Eleni Maria Michailidi and Baldassare Bacchi

DICATAM, Università degli studi di Brescia, Via Branze 42, 25123 Brescia, Italy

Correspondence to: E. M. Michailidi (e.michailidi@unibs.it)

Abstract. One of the most important tasks a hydrologist must face is to estimate the hydrological risk (i.e. probability) of a variable exceeding a certain threshold. This risk is often expressed in terms of a Return Period, T, and refers to the failure of the hydraulic structure which controls this variable. Sometimes the "structure" is simply the river embankments- the failure of which means their overtopping by the river. The widely adopted definition of T, in a problem regarding the maxima of hydro-
logical variables, is "the average time elapsing between two successive occurrences of an event exceeding a given magnitude of the natural variables".

Conventional approaches for the estimation of T involve a single natural variable (i.e. flood peak, maximum rainfall intensity, etc.) and its frequency analysis.

However, a univariate approach in complex problems ignores the effect of other significant variables leading to different risk levels for each quantity of interest and resulting in an inaccurate estimate of the risk- often wrongfully set equal to the risk of the hydrological event. For example, if one considers the flood inflow in a lake around which establishments are positioned, the variable to be investigated relating to risk assessment is the lake water level. The same water level may occur from very different flood hydrographs, even when the same initial water level and specific spillway characteristics are taken into account. We considered this a result of the interaction of three joint factors: the hydrograph's peak, volume and shape. Consequently, we apply a multivariate distribution framework (using copula functions) in order to find a region where all underlying events are assigned to the same risk- associated here to the maximum water level.

1 Introduction

1.1 Literature review

In recent years, the application of copula functions has facilitated overcoming the inadequacies of traditional multivariate distributions such as that the marginals must derive from the same distribution family and their parameters may define the dependence structure between the variables (Salvadori et al., 2007). Copulas are functions that combine marginal distributions to the joint cumulative distribution, therefore the latter is only indirectly affected by the choice of the marginals. So the practical problem of identification and estimation of the joint distribution is handled from two non-interwinding aspects; the dependence

structure of the set of variables and the marginal distributions. In the relatively recent literature there is a wide application of the copula functions to model hydrometeorological variables.

De Michele and Salvadori (2003) have utilised the Frank copula and marginals belonging to the heavy-tailed Generalised Pareto distribution to model the intensity and duration of extreme rainfall events in a basin in Liguria. Zhang and Singh (2007)
5 have analysed three rainfall variables (intensity, depth, and duration) from three sites in Louisiana in a bivariate framework and have concluded that the Gumbel-Hougaard copula performed better in simulating the rainfall depth and duration at one site, the Ali-Mikhail-Haq copula the intensity and depth for all three sites and the Frank copula the intensity and duration for all three sites and the depth and duration at two sites. Balistrocchi and Bacchi (2011) researched the dependence between rainfall volume, wet weather duration and interevent period for three time series recorded in different Italian climates and have
10 attained that the first pair can be modelled by the Gumbel-Hougaard copula while when considering the interevent dry period, the independence case cannot be rejected.

In other papers (Singh and Zhang, 2007; Ariff et al., 2012) the Frank copula was used to simulate the rainfall intensity and duration and to derive IDF curves which were later compared to the empirical IDF curves, derived from common practice. Both works concluded that the two methods are in agreement.

15 In recent literature, focusing on the flood natural variables, the use of copulas is increasing. Aronica et al. (2012) carried out a bivariate analysis of flood peaks and volumes, introducing also the hydrograph shape as an independent variable in a Monte Carlo framework to produce flood hazard maps. The Gumbel-Hougaard copula was deemed appropriate for the peak-volume pair. Balistrocchi et al. (2014) have studied the Panaro (which is also the study watershed for this research) and the Tagliamento watershed and modelled the flood peak, volume and duration pairs using the Gumbel-Hougaard copula, that fitted
20 well despite the various regional diversities. Candela et al. (2014) derived synthetic hydrographs through bivariate rainfall analysis (intensity-duration simulated with Frank copula), stochastic temporal pattern rain generation and a rainfall-runoff model. The Gumbel-Hougaard copula was adopted to fit the flood peaks and volumes resulting from the synthetic hydrographs and in order to generate peak-volume pairs and synthetic hydrographs through cluster statistics. Domeneghetti et al. (2013) studied the effect of the uncertainties of the rating curve in flood mapping and have considered the uncertainty of the flood
25 hydrograph through the Gumbel copula for the peak and volume.

A thorough study in a multivariate framework was carried out by Salvadori et al. (2011) who defined and compared three different approaches for the representation of the return period- not uniquely defined in the multidimensional case- and have suggested a method to select the design event. Following the latter, Gräler et al. (2013) applied the different definitions of the return period in a trivariate framework (discharge peak, volume and duration), implementing high-dimensional structures called vines. Additionally, Ganguli and Reddy (2013) applied fully nested Archimedean copulas on the aforementioned flood
30 variables using data of the Delaware River basin and have concluded that the results obtained by the various return periods may vary substantially. Salvadori et al. (2013) revisited the concept of the multivariate return period and introduced an alternative expression of it that possess an advantage over the others (details in Sect. 2.5).

De Michele et al. (2005) were the first (to our knowledge) to check the adequacy of a dam's spillway under a bivariate
35 hydrological load. They have done so by creating 1000 synthetic hydrographs- of specific peak and volume generated by a fitted

Gumbel copula- using the Nash model and routing them to obtain a frequency curve. Pinya et al. (2009) applied a multivariate copula framework (peak at different locations, volume and duration) to a catchment in Denmark in order to estimate the cdf of the sea water level at the stream outlet. Comparison of the results with a continuous river flow simulation of observed data shows a significant difference in the tails. Recently, Requena et al. (2013)- also using the Gumbel copula- directly associated the return period to the risk of dam overtopping and have compared the results with the ones obtained from the association of the return period to a natural probability of flood occurrence. Volpi and Fiori (2014) stressed out that the return period of a failure of a structure depends on the structure of interest and therefore the interaction between the hydrological loads and the structure should be taken into consideration. In particular, they illustrated a structure-based return period and compared it with the design event-based approach, applied on a theoretical structure.

Serinaldi (2014) was critical about the use of the appropriate return period, stating that its implementation should be based on the definition of risk in the case at hand and that every comparison between the different definitions is of little sense since they refer to different mechanisms of failure.

A comprehensive and regularly updated list of the publications that regard the use of copula can be found at the website of STAHY [1].

1.2 Objective of the study and outline of the paper

The main objective of this study is to state in a clear manner if it is possible, in a multivariate context, to define the return period (here abbreviated as RP) of an "event" which is expressed as a point in the positive R^d space. As stated in a lot of papers (partially referenced here) it is possible to define some sort of return period (e.g. T_{OR}, T_{AND}, T_{KEND}, T_{SKEND}). However, all these values are strongly different from the hydrological concept of the return period T which assesses the value of an interest variable with exceedance probability of one time in T years. When a surjective application from positive R^d to R^+ is established it becomes possible to identify a subset of positive R^d so that all points belonging to this subregion produce in R^+ values of interest with $RP > T$. Applying this concept to a design or a verification of a hydraulic structure we can refer to the return period as structure-oriented RP.

For this analysis, we consider as the interest variable, the maximum water level (MWL) in a reservoir of a flood routing dam. Therefore we expressed the RP in terms of probability of exceedance of this variable, since the risk of a given natural variable (e.g. rainfall height, intensity, flood peak etc.) translates into a different risk of failure for the structure of interest, due to the system's non-linearity.

Apart from the determination of the bivariate function of flood peak and volume, which relate to the MWL, the hydrograph shape was taken also into account. An intensive analysis demonstrated that if one assumes, in a random manner, a shape derived from the clustering of available real hydrographs in more than one "mean" dimensionless shape, the sub-region of the application of positive R^d to R^+ becomes not univocally identifiable.

In short, after we extracted the flood events from the continuous time runoff series, we modelled the co-dependence of flood peak and volume by a copula function and generated a certain amount of duples. From the historic series we identified typical

---

[1] http://www.stahy.org/Activities/STAHYReferences/ReferencesonCopulaFunctiontopic/tabid/78/Default.aspx

hydrograph shapes to obtain the synthetic hydrographs. Next, we routed them through the dam and we calculated the MWL for each. We repeated the routing process for the observed hydrographs and compared their MWL to the synthetic ones. In addition, the RP in terms of dam overtopping was compared to the RP associated to the natural variables (presented in Sect. 2.5). Finally, the same procedure was followed after clustering in only one "mean" shape. A general flowchart of the procedure is depicted in Fig. 1.

The structure of this paper is as follows. First, we introduce the methodology; that, includes the procedure for the extraction of events, the choice of marginal and joint distributions, the Monte Carlo framework and an overview of the return periods. Next, we present the study area and the data followed by the results of the analysis and the conclusions.

2 Methodology

2.1 Univariate flood frequency analysis

The annual maximum discharge peaks were extracted from the time series. For the calculation of the flood volume the episode's start and finish had to be well defined and multipeak events should be considered as one. Since the event separation procedure can be characterized as intuitive and well-established rules do not exist, heuristic criteria were applied. After careful examination of the time series at the considered basin, general criterias were established in order to define the start of the rising limb and the end of the recession limb. Also, consecutive peaks with an interarrival greater than the time of concentration were considered parts of a multipeak event.

The marginal distributions of flood peak and volume (after the baseflow removal) were selected taken into consideration the Bayesian information criterion (BIC) value- which gives similar results to the Akaike information criterion (AIC) value but has a preference towards more parsimonious models (Laio et al., 2009). Akaike (1974) and Schwarz (1978) have defined their criteria as

$$AIC = -2\log(L) + 2k \tag{1}$$

$$BIC = -2\log(L) + \log(n)k \tag{2}$$

where L is the maximum likelihood value of the marginals, k the number of the parameters and n the number of observations.

The parameters of the distributions were estimated by maximizing the likelihood function. Lastly, the Kolmogorov-Smirnov, the Anderson-Darling and the Cramér-von Mises tests were carried out to test the goodness-of-fit.

2.2 Hydrograph clustering

The shape of the design hydrograph is frequently claimed as an important factor in the design procedure and is related to the spatial and temporal rain distribution as well as the basin's shape and behaviour (Singh, 1997). Therefore typical hydrographs

Hydrology and
Earth System
Sciences

Open Access

Discussions

EGU

were determined using events from the time series. In particular, the events' hydrographs were clustered according to their characteristics and utilizing the methodology proposed by Dyck and Peschke (1995) which suggests the normalisation of the hydrograph (after the removal of the baseflow) by

$$Q_{norm} = (Q - q_{base})/(Q_{max} - q_{base}) \tag{3}$$

5

$$t_{norm} = t/t_{Qmax} \tag{4}$$

where Q_{max} the hydrograph's peak, q_{base} the base flow and t_{max} the time to peak, starting from the rising limb of the flood event.

Consequently, the normalised peak equals to one at time 1. All normalised hydrographs were extended to a common duration (for comparison purposes) and cluster analysis with the Ward method and Euclidean distances was implemented (Aronica et al., 2012).

2.3 Dependence structure between variables

The degree of relationship between pairs of variables was examined by measures of association. These include Kendall's τ, Spearman's ρ_S which express the existence or absence of concordance and Pearson's ρ_P which expresses linear dependence. The p value that corresponds to each coefficient was also calculated to test for independence, rejecting it if p is less than 0.05.

To graphically assess independence chi-plots and Kendall plots were generated. Chi-plots' patterns portray characteristics such as independence, complexity of dependence and existence of monotonicity up to some degree. Chi-plots display a measure of distance, λ_i, of each observation from the centre of the data against a measure of association between the marginals, χ_i. The values of the ranks of the data determine the shape of the graph. If a certain percentage, i.e. 95 % of the points, lie between the confidence line the two variables are considered independent, whereas positive and negative dependence is indicated when the points lie above the upper limit and below the lower limit, accordingly (Fisher and Switzer, 1985, 2001).

Kendall plots preserve some of the desirable properties of the chi-plots, like their reliance on ranks, but also give a better understanding of the nature of dependence. They plot the measure of concordance, H_i against its expected value under the null hypothesis of independence, $W_{i:n}$. The distance from the diagonal line is an indication of the greatness of the dependence. If the points lie above the diagonal line positive dependence is present and if they lie below negative. The lack of association is depicted with a straight diagonal line and in the case of comonotonicity the points will follow the curve or the horizontal axis for Kendall's $\tau = 1$ and $\tau = -1$, accordingly (Genest and Boies, 2003).

2.4 Multivariate analysis using copula functions

Copulas are functions that describe the dependence structure between variables independently of the choice of marginal distributions. The joint distribution functions and the marginals are linked by Sklar's theorem (Sklar, 1959):

$$F(x_1,...,x_d) = C(F_1(x_1),...,F_d(x_d)) \tag{5}$$

for all $\mathbf{x} \in R^d$, where the F_i are the marginals of F and C the copula function.

Theoretical background is included in Sklar (1959) and Nelsen (2013) as well as in the more hydrologically-oriented publication of Salvadori et al. (2007).

First, a copula function was selected, according to the BIC, and its goodness-of-fit was tested for the peak/volume pair utilising the Kolmogorov-Smirnov and the Cramér-von Mises test for Archimedean bivariate copulas based on Kendall's process investigated by Genest and Rivest (1993); Wang and Wells (2000). P values for these tests are calculated using a parametric bootstrap procedure.

Since we are dealing with extremes we also assessed the tail dependence of the observations. The tail dependence coefficient was calculated for the observations and was compared with the theoretical coefficient of the chosen copula. We preferred the estimator proposed by Frahm et al. (2005), expressed as

$$\hat{\lambda}^{CFG} = 2 - 2\exp[\frac{1}{n}\sum_{i=1}^{n}\log\{\frac{\sqrt{\log\frac{1}{u_i}\log\frac{1}{v_i}}}{\log\frac{1}{\max(u_i,v_i)^2}}\}] \tag{6}$$

over the one proposed by Schmidt and Stadtmüller (2006) since the latter varies greatly in base of threshold selection. Here, n is the sample size, $u_i = F_1(x_{1,i})$ and $v_i = F_2(x_{2,i})$.

Next, 10000 pairs of values were generated using that copula and their fitted parameters.

2.5 Defining the return period

The definition of the return period in a multivariate case can be approached through various ways. Salvadori et al. (2011, 2013) have included an overview of these approaches for the multivariate framework, the choice of which should be based on the engineer's interest (Serinaldi, 2014). Two widely used joint RP's are the so-called OR and AND return periods that are defined by the following equations :

$$T_{OR} = \frac{\mu_T}{1 - C(u,v)} \tag{7}$$

where μ_T is the mean interarrival time between two consecutive occurrences of (X_1, X_2) (in this case $\mu_T = 1$ year). The OR alludes to the probability $P[X_1 > x_1 \vee X_2 > x_2]$. Consequently T_{OR} refers to a specific value of the $C(U,V)$ that an infinite number of pairs (u,v) satisfy. From the inverse of the marginal CDFs we obtain:

$$x_{1,OR} = F_1^{-1}(u) \quad , \quad x_{2,OR} = F_2^{-1}(v) \tag{8}$$

Similarly, the AND return period is defined as:

$$T_{AND} = \frac{\mu_T}{1 - F_1 - F_2 + C(u,v)} \tag{9}$$

The AND alludes to the probability $P[X > x \wedge Y > y]$.

Salvadori et al. (2016) noted that the OR RP can be optimal for estimating the flood risk for example, at the confluence of two rivers and the AND RP for scenarios where the combined effect of two or more variables can be damaging. Dung

et al. (2015) stated that flood volume can be an equally governing factor with the peak in the inundation process- taking as an example events from the Mekong Delta in Vietnam- and utilised the AND RP for their risk analysis.

In the OR and AND case realisations of the same critical level do not always yield the same dangerous region. So Salvadori et al. (2011) included an additional definition called secondary RP or Kendall's RP expressed as:

$$T_{KEND} = \frac{\mu_T}{1 - K_C(t)} = \frac{\mu_T}{1 - P(C(U,V) \leq t)} \tag{10}$$

where K_C is the Kendall's measure and t is a copula level curve.

The latter definition implies that some joint events that lie in the safe region may have larger marginal values than events on or above the design level. Salvadori et al. (2013) have introduced the Survival Kendall RP that yields a bounded safe region, containing multivariate events with limited marginals.

$$T_{SKEND} = \frac{\mu_T}{1 - \overline{K}_C(t)} = \frac{\mu_T}{1 - P(\hat{C}(1-U, 1-V) > t)} \tag{11}$$

where $\overline{K}_C$ is the Kendall's survival function and $\hat{C}$ is the survival copula for which $\hat{C}(u,v) = u + v - 1 + C(1-u, 1-v)$.

Salvadori et al. (2016) suggested that both Kendall RP's are useful for a preliminary risk assessment analysis to understand what can be expected regarding the joint probability of occurrences since these RP's don't identify beforehand the contribution of each variable to the risk.

Volpi and Fiori (2014) associated the RP with the structure of interest by relating the structure design parameter to the hydrological load through the function $Z = g(X_1, X_2)$. Consequently the structure-oriented RP would be expressed as:

$$T_{STR} = \frac{\mu_T}{1 - F_Z(z)} \tag{12}$$

where F_Z is the probability distribution function of the derived variable Z. In this case the structure function is very complex and therefore the whole analysis must be based on Monte Carlo simulations.

3 Case study

We have focused our interest on the Panaro catchment- an important influent of Po river in Northern Italy. In particular, the watershed under investigation is composed by the Panaro river itself, the Scoltenna and the Leo tributaries with an outlet at the Panaro dam (Fig. 2), that occupies an area of 876 km^2. The Panaro tributary has its source at Monte Cimone (2.165 m a.s.l.) and flows into Po, at Bondeno; It takes its name at the valley of Montespecchio after converging with the Leo and Scoltenna streams, that constitute the upper part of the river network. The hydrographic network of the watershed shows a low degree of hierarchy, indicating an evolving state which is also evident by the existence of torrential dynamic phenomena (Autorità di bacino del fiume Po, 2006).

The influence of snowfall is negligible due to the modest land elevation and the majority of rainfall events occur seasonally (September-April). The average precipitation height ranges between 700 and 2000 mm/year (Autorità di bacino del fiume Po, 2006).

Hydrology and
Earth System
Sciences
Discussions
Open Access
EGU

The basin's permeability is low and therefore the runoff is influenced little by water infiltration. In fact, the study basin consists mostly of sandstones and silicatic alternating sequences (44 % of total area) and marls and clay (34 %).

The Panaro dam is a concrete gravity dam (150 m in length), located near the city of Modena and constructed for flood mitigation purposes. The hydraulic system consists of two reservoirs, a principal on the river course and a secondary at the right river bank, and a series of levees that enclose them. The crest of the principal levees are at 44.85 m a.s.l.. The reservoirs can hold in total 23.66 hm^3 up to the spillway's crest at 41.1 m a.s.l.. There are also nine discharge outlets at the bottom of the dam that ensure constant flow to the downstream valley.

The available flood data included a 52-year discharge series (1936-1943,1945 and 1946 were missing) with an hourly time interval. The hydrological characteristics of the study basin as well as a summary of the available data are briefly presented in Table 1 and 2.

4 Results

4.1 Selection of marginals and hydrograph clustering

Various theoretical distribution were fitted (Fig. 3a) to the ecdf of the discharge peaks and the appropriate distribution (i.e. Inverse Gaussian with parameters shown in Table 3) was chosen based on the BIC value and the tail's behaviour and after being tested by the Kolmogorov-Smirnov, the Anderson-Darling and the Cramér-von Mises goodness-of-fit test (Table 4). The p values were much greater than 0.05 so the hypothesis of this distribution could not be rejected. We fitted the distributions using maximum likelihood estimation. The fitted Birnbaum-Sanders and Lognormal gave results very similar to the Inverse Gaussian, with a higher BIC value. The very heavy tail of the Generalised Extreme Value (GEV) distribution appears to overestimate the peak in the upper quantile (5%) and consequently, was not preferred. The interest is not focused on very high quantile estimates (>99%), but on a correct definition of the risk in the range of a return period comparable with the observation period. These results were corroborated when we used discharge peak data from the missing years of our study period. Unfortunately these data included only information about the peak and not about the temporal evolution of the flood.

After examining each event of the time series in Panaro, we applied empirical criteria for the event separation- successive peaks with an interarrival time of less than seven times the time of concentration are considered parts of a multipeak event, the episode's start occurs when the inclination of the rising limb supersedes a threshold for a certain number of intervals and the end when it falls below a threshold for the same number of intervals.

Then we calculated the volume of each event. Similarly to the marginal distribution selection for the discharge peaks, we chose the Rayleigh distribution (Table 3) based on the smallest BIC value and the modelling of the upper tail (Fig. 3b). The Birnbaum-Sanders and Inverse Gaussian are much more heavy-tailed and therefore yield increased volume values in the upper quantile (5%) and consequently, were not preferred. The goodness of fit tests permitted our selection since the p values exceeded 0.05 (Table 4).

Next, the hydrographs were classified in 4 clusters whose shapes are depicted in Fig. 4a, left. The characteristic shapes include hydrographs with one and two peaks and with abrupt or gradual recession limbs with specific probabilities of occurrence. While respecting these probabilities, 10000 cluster numbers were generated.

4.2 Copula selection

We implemented a two-variate frequency analysis on the hydrograph variables, flood peak and volume, and studied their dependence structure using copula functions. Initially, we calculated Kendall's, Spearman's (rank-based measures of association) and Pearson's coefficients (Table 5), whose high values indicated the existence of strong positive dependence between the considered variables. The majority of the points in the chi-plot lies in the upper area, thus suggesting a positive dependence (Fig. 5). We obtained similar results from the Kendall plot (Fig. 6); the majority of the points were located above the diagonal, which is also a sign of positive dependence.

Among all the copula distributions that were tested (Gauss, Gumbel, Student t, Frank, Clayton, Joe, BB1, BB6, BB7, BB8)- whose parameters were estimated with maximum likelihood estimation- the Gumbel (Table 6) and the Gaussian copula yielded similar BIC and AIC values. This similarity is a consequence of the medium sample size (set of 52 values). Since the Gumbel copula has been extensively used in the past to model peak/volume pairs, not only for the Panaro basin but also for other basins worldwide (see Sect. 1.1) it has been preferred over the Gaussian (see additionally Balistrocchi and Bacchi (2011)). We based also our selection on the fact that the Gumbel copula has an upper tail dependence coefficient allowing the modelling of extreme events; on the contrary, the Gaussian copula does not. The Gumbel's upper tail dependence coefficient (λ^C=0.643) approximates the empirical non-parametric ($\hat{\lambda}^{CFG}$=0.646) and satisfies the assumption of an extreme value copula. However, the available data for the tail dependence analysis are scarce so the coefficient can only be used qualitatively, as an indication of modelling tail dependence, and not as a model selection tool. It should be noted that the assumption that the dataset is characterised by upper tail dependence is based on past research and is somewhat intuitive; the empirical non-parametric coefficient suffers from bias, as mentioned in an extensive research of Serinaldi et al. (2015), and tends to indicate upper tail-dependence even if it does not exist. Serinaldi et al. (2015) proposed alternatives that are satisfactorily unbiased, but as it is logical when dealing with upper quantiles, they require large datasets in order to make a safe inference.

The goodness-of-fit test has shown that the hypothesis of the selected copula construction cannot be rejected (p values in Table 7 greater than 0.05). In Fig. 7 the scatter plot prior to the rank transformation, the scatterplot after the rank transformation- both indicating positive dependence- and the contours of the copula's cdf are depicted after the rank transformation of the data and the randomisation of the ties.

4.3 Return period estimation

Following the generation of 10000 pairs of flood peak and volume values from the selected copula and their random assignment to a specific cluster group, the typical hydrographs were rescaled and routed through the Panaro dam in order to obtain the maximum water level reached during each event.

In addition, the observed hydrographs were routed through the reservoir and their corresponding water levels were compared to the levels of the synthetic ones (Fig. 8). The results appear to be in accordance, with the only exceptions being the points corresponding to the 13, 18 and 26-year RPs for which the difference in the MWL reaches up to 1.50 meters. This frequency curve corresponds also to the frequency curve of the peak discharge downstream.

In Fig. 9 it is noted that events assigned to the same hydrograph shape are clustered together and events with the same return period but in different cluster groups can differ in the peak by 8 % when considering almost the same volume and in the volume by 27 % when considering almost the same discharge. This variability prohibits the clear definition of a region where

5 all multivariate events produce risk lower than an assigned value.

When looking at the effect of the hydrograph shape of a specific peak (i.e. $\pm$ 0.25 %) on MWL it seems that the least favorable is shape 1 and most favorable shape 3 and 4 (Fig. 10a). However in the case of constant peak and volume this comparison does not make much sense since when denormalising, in order to preserve the desired volume (after the baseflow removal) and peak, the peak time must assume small values; as a result the recession limb will be shortened. For this reason,

10 under these conditions, the differences in the recession limbs of the typical shapes will be annulled. This is visible in Fig. 10 a, where shape 4 seems to have the shortest recession limb despite belonging to the group with the longest one (Fig. 4a) and in the cases of groups 1 and 2 whose shapes are very similar but for the recession limb's duration, their differences may be indiscernible. The only safe inference that can be made is that the hydrograph shape's role is secondary in comparison with the peak's.

15 When clustering in one "mean" group instead of four the change detected in the frequency curve becomes acceptable (maximum difference is 0.37 m) (Fig. 8). In this case, the desired region can be defined (Fig. 11) but depends on the hydrological regime as well as the hydraulic and geometric characteristic of the structure of interest. Discharge peak is a more defining factor in MWL than flood volume, since for the same return period the maximum difference between the peaks is 11% and between the flood volumes 26% (Fig. 11) for $RP = 20$ years. The secondary axes correspond to the univariate RP for each

20 variable. It is evident that for an event with a certain RP (i.e. 20 years) the univariate RP can span from 15 to 27 years for peak and from 7 to 34 years for volume. For $RP = 10$ and 50 years the results are similar (Fig. 11). Of course, the conclusion that peak is more significant than volume in MWL is structure-dependent. Requena et al. (2013) demonstrated this by changing the spillway setting; when the reservoir volume was increased (under constant spillway length) the defining factor was the peak but when the spillway length (under constant volume) was increased the MWL was influenced more by volume.

25 In order to highlight the importance of choosing the appropriate return period expression we have included the comparison between the various definitions. As stated also by Serinaldi (2014), the choice of the proper RP definition (OR, AND, KEND, SKEND or structure-oriented) depends on the problem at hand and the definition of risk of the case study. It is logical that the five approaches yield completely different results (Fig. 8).

5 Conclusions

30 In this paper we evaluated the risk of reaching specific water levels in the reservoir of the Panaro dam by implementing a multivariate Monte Carlo analysis on three key-components; flood peak, volume and hydrograph shape. We modelled the first two by applying a copula bivariate function, while we considered the latter independent and we modelled it through cluster analysis and Monte Carlo simulations. Next, we compared the maximum water levels from the synthetic events with the observed.

Results have shown that there is an agreement between the simulations and the observations and therefore copula functions and cluster analysis can serve as a valuable tool for risk estimation. Even though some hydrograph shapes give more elevated levels than others, a clear comparison between them cannot be drawn. However the hydrological regime permits the consider-

5 ation of a mean hydrograph- not changing significantly the frequency curve- thus simplifying the process and making possible the definition of a risk region. This risk region corresponds only to the MWL in the reservoir and not to flood variables.

The results of this structure-oriented approach are case specific and therefore the outline of this research can only serve as a methodology that can be applied to other areas and not as a stepping-stone to generalise the findings (e.g. the risk region). Understanding the mechanisms of failure of our problem is crucial and can guide our decision regarding the return period

10 definition which differentiates greatly the results.

Needless to say, the importance of data availability for the areas of interest is immense in order to validate this procedure locally and reduce the uncertainty that surrounds a multivariate analysis.

Data availability

The dataset used in this research is available upon request from the corresponding author.

15 The authors declare that they have no conflict of interest.

References

Akaike, H.: A new look at the statistical model identification, Automatic Control, IEEE Transactions on, 19, 716–723, doi:10.1109/TAC.1974.1100705, 1974.

5 Akima, H., A., G., Petzoldt, T., and Maechler, M.: akima: Interpolation of irregularly spaced data, http://CRAN.R-project.org/package= akima, r package version 0.5-11, 2013.

Ariff, N. M., Jemain, A. A., Ibrahim, K., and Wan Zin, W. Z.: IDF relationships using bivariate copula for storm events in Peninsular Malaysia, Journal of Hydrology, 470–471, 158–171, doi:http://dx.doi.org/10.1016/j.jhydrol.2012.08.045, http://www.sciencedirect.com/science/article/pii/S002216941200724X, 2012.

10 Aronica, G. T., Candela, A., Fabio, P., and Santoro, M.: Estimation of flood inundation probabilities using global hazard indexes based on hydrodynamic variables, Physics and Chemistry of the Earth, Parts A/B/C, 42–44, 119–129, doi:http://dx.doi.org/10.1016/j.pce.2011.04.001, http://www.sciencedirect.com/science/article/pii/S1474706511000568, 2012.

Autorità di bacino del fiume Po: Caratteristiche del bacino del fiume Po e primo esame dell'impatto ambientale delle attività umane sulle risorse idriche, Report, 2006.

15 Balistrocchi, M. and Bacchi, B.: Modelling the statistical dependence of rainfall event variables through copula functions, Hydrol. Earth Syst. Sci., 15, 1959–1977, doi:10.5194/hess-15-1959-2011, http://www.hydrol-earth-syst-sci.net/15/1959/2011/, 2011.

Balistrocchi, M., Ranzi, R., and Bacchi, B.: Multivariate Statistical Analysis of Flood Variables by Copulas: Two Italian Case Studies, in: 3rd IAHR Europe Congress, 2014.

Brechmann, E. C. and Schepsmeier, U.: Modeling Dependence with C- and D-Vine Copulas: The R Package CDVine, Journal of Statistical 20 Software, 52, 27, doi:10.18637/jss.v052.i03, https://www.jstatsoft.org/index.php/jss/article/view/v052i03, 2013.

Candela, A., Brigandì, G., and Aronica, G. T.: Estimation of synthetic flood design hydrographs using a distributed rainfall–runoff model coupled with a copula-based single storm rainfall generator, Nat. Hazards Earth Syst. Sci., 14, 1819–1833, doi:10.5194/nhess-14-1819-2014, http://www.nat-hazards-earth-syst-sci.net/14/1819/2014/, 2014.

De Michele, C. and Salvadori, G.: A Generalized Pareto intensity-duration model of storm rainfall exploiting 2-Copulas, Journal of Geo-25 physical Research: Atmospheres, 108, n/a–n/a, doi:10.1029/2002JD002534, http://dx.doi.org/10.1029/2002JD002534, 2003.

De Michele, C., Salvadori, G., Canossi, M., Petaccia, A., and Rosso, R.: Bivariate Statistical Approach to Check Adequacy of Dam Spillway, Journal of Hydrologic Engineering, 10, 50–57, doi:doi:10.1061/(ASCE)1084-0699(2005)10:1(50), http://ascelibrary.org/doi/abs/10.1061/%28ASCE%291084-0699%282005%2910%3A1%2850%29, 2005.

Domeneghetti, A., Vorogushyn, S., Castellarin, A., Merz, B., and Brath, A.: Probabilistic flood hazard mapping: effects of uncertain bound-30 ary conditions, Hydrol. Earth Syst. Sci., 17, 3127–3140, doi:10.5194/hess-17-3127-2013, http://www.hydrol-earth-syst-sci.net/17/3127/2013/, 2013.

Dung, N. V., Merz, B., Bárdossy, A., and Apel, H.: Handling uncertainty in bivariate quantile estimation – An application to flood hazard analysis in the Mekong Delta, Journal of Hydrology, 527, 704–717, doi:http://dx.doi.org/10.1016/j.jhydrol.2015.05.033, http://www.sciencedirect.com/science/article/pii/S0022169415003832, 2015.

35 Dyck, S. and Peschke, G.: Grundlagen der Hydrologie, Verlag für Bauwesen, Berlin, 1995.

Faraway, J., Marsaglia, G., Marsaglia, J., and Baddeley, A.: goftest: Classical Goodness-of-Fit Tests for Univariate Distributions, http://CRAN.R-project.org/package=goftest, r package version 1.0-3, 2015.

Fellows, I.: wordcloud: Word Clouds, http://CRAN.R-project.org/package=wordcloud, r package version 2.5, 2014.

Hydrology and
Earth System
Sciences

Open Access

EGU

Discussions

Fisher, N. I. and Switzer, P.: Chi-Plots for Assessing Dependence, Biometrika, 72, 253–265, doi:10.2307/2336078, http://www.jstor.org/stable/2336078, 1985.

Fisher, N. I. and Switzer, P.: Graphical Assessment of Dependence: Is a Picture Worth 100 Tests?, The American Statistician, 55, 233–239, doi:10.2307/2685807, http://www.jstor.org/stable/2685807, 2001.

Frahm, G., Junker, M., and Schmidt, R.: Estimating the tail-dependence coefficient: Properties and pitfalls, Insurance: Mathematics and Economics, 37, 80–100, doi:http://dx.doi.org/10.1016/j.insmatheco.2005.05.008, http://www.sciencedirect.com/science/article/pii/S016766870500065X, 2005.

Gambarelli, P., Moretti, G., and Orlandini, S.: Sviluppo di un Modello Matematico del Funzionamento Idraulico della Cassa di Espansione sul Fiume Panaro, Thesis, 2009.

Ganguli, P. and Reddy, M. J.: Probabilistic assessment of flood risks using trivariate copulas, Theoretical and Applied Climatology, 111, 341–360, doi:10.1007/s00704-012-0664-4, http://dx.doi.org/10.1007/s00704-012-0664-4, 2013.

Genest, C. and Boies, J.-C.: Detecting Dependence With Kendall Plots, The American Statistician, 57, 275–284, doi:10.1198/0003130032431, http://dx.doi.org/10.1198/0003130032431, 2003.

Genest, C. and Rivest, L.-P.: Statistical Inference Procedures for Bivariate Archimedean Copulas, Journal of the American Statistical Association, 88, 1034–1043, doi:10.2307/2290796, http://www.jstor.org/stable/2290796, 1993.

Gräler, B., van den Berg, M. J., Vandenberghe, S., Petroselli, A., Grimaldi, S., De Baets, B., and Verhoest, N. E. C.: Multivariate return periods in hydrology: a critical and practical review focusing on synthetic design hydrograph estimation, Hydrol. Earth Syst. Sci., 17, 1281–1296, doi:10.5194/hess-17-1281-2013, http://www.hydrol-earth-syst-sci.net/17/1281/2013/, 2013.

Kojadinovic, I. and Yan, J.: Modeling Multivariate Distributions with Continuous Margins Using the copula R Package, Journal of Statistical Software, 34, 20, doi:10.18637/jss.v034.i09, https://www.jstatsoft.org/index.php/jss/article/view/v034i09, 2010.

Laio, F., Di Baldassarre, G., and Montanari, A.: Model selection techniques for the frequency analysis of hydrological extremes, Water Resources Research, 45, n/a–n/a, doi:10.1029/2007WR006666, http://dx.doi.org/10.1029/2007WR006666, 2009.

Nelsen, R. B.: An introduction to copulas, vol. 139, Springer Science Business Media, 2013.

Pinya, M. A. S., Madsen, H., and Rosbjerg, D.: Assessment of the risk of inland flooding in a tidal sluice regulated catchment using multi-variate statistical techniques, Physics and Chemistry of the Earth, Parts A/B/C, 34, 662–669, doi:http://dx.doi.org/10.1016/j.pce.2009.03.001, http://www.sciencedirect.com/science/article/pii/S1474706509000345, 2009.

R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org/, 2015.

Requena, A. I., Mediero, L., and Garrote, L.: A bivariate return period based on copulas for hydrologic dam design: accounting for reservoir routing in risk estimation, Hydrol. Earth Syst. Sci., 17, 3023–3038, doi:10.5194/hess-17-3023-2013, http://www.hydrol-earth-syst-sci.net/17/3023/2013/, 2013.

Salvadori, G., De Michele, C., Kottegoda, N., and Rosso, R.: Extremes in Nature: An Approach Using Copulas, Water Science and Technology Library, Springer Netherlands, 1 edn., doi:10.1007/1-4020-4415-1, 2007.

Salvadori, G., De Michele, C., and Durante, F.: On the return period and design in a multivariate framework, Hydrol. Earth Syst. Sci., 15, 3293–3305, doi:10.5194/hess-15-3293-2011, http://www.hydrol-earth-syst-sci.net/15/3293/2011/, 2011.

Salvadori, G., Durante, F., and De Michele, C.: Multivariate return period calculation via survival functions, Water Resources Research, 49, 2308–2311, doi:10.1002/wrcr.20204, http://dx.doi.org/10.1002/wrcr.20204, 2013.

Salvadori, G., Durante, F., De Michele, C., Bernardi, M., and Petrella, L.: A multivariate copula-based framework for dealing with hazard scenarios and failure probabilities, Water Resources Research, 52, 3701–3721, doi:10.1002/2015WR017225, http://dx.doi.org/10.1002/2015WR017225, 2016.

5 Schmidt, R. and Stadtmüller, U.: Non-parametric Estimation of Tail Dependence, Scandinavian Journal of Statistics, 33, 307–335, doi:10.1111/j.1467-9469.2005.00483.x, http://dx.doi.org/10.1111/j.1467-9469.2005.00483.x, 2006.

Schwarz, G.: Estimating the Dimension of a Model, The Annals of Statistics, 6, 461–464, http://www.jstor.org/stable/2958889, 1978.

Serinaldi, F.: Dismissing return periods!, Stochastic Environmental Research and Risk Assessment, 29, 1179–1189, doi:10.1007/s00477-014-0916-1, http://dx.doi.org/10.1007/s00477-014-0916-1, 2014.

10 Serinaldi, F., Bárdossy, A., and Kilsby, C. G.: Upper tail dependence in rainfall extremes: would we know it if we saw it?, Stochastic Environmental Research and Risk Assessment, 29, 1211–1233, doi:10.1007/s00477-014-0946-8, http://dx.doi.org/10.1007/s00477-014-0946-8, 2015.

Singh, V. P.: Effect of spatial and temporal variability in rainfall and watershed characteristics on stream flow hydrograph, Hydrological Processes, 11, 1649–1669, doi:10.1002/(SICI)1099-1085(19971015)11:12<1649::AID-HYP495>3.0.CO;2-1, http://dx.doi.org/10.1002/(SICI)1099-1085(19971015)11:12<1649::AID-HYP495>3.0.CO;2-1, 1997.

15 Singh, V. P. and Zhang, L.: IDF Curves Using the Frank Archimedean Copula, Journal of Hydrologic Engineering, 12, 651–662, doi:doi:10.1061/(ASCE)1084-0699(2007)12:6(651), http://ascelibrary.org/doi/abs/10.1061/%28ASCE%291084-0699%282007%2912%3A6%28651%29, 2007.

Sklar, M.: Fonctions de répartition à n dimensions et leurs marges, Université Paris 8, 1959.

20 Smyth, G., Hu, Y., Dunn, P., Phipson, B., and Chen, Y.: statmod: Statistical Modeling, http://CRAN.R-project.org/package=statmod, r package version 1.4.21, 2015.

Volpi, E. and Fiori, A.: Hydraulic structures subject to bivariate hydrological loads: Return period, design, and risk assessment, Water Resources Research, 50, 885–897, doi:10.1002/2013WR014214, http://dx.doi.org/10.1002/2013WR014214, 2014.

Walker, A.: openxlsx: Read, Write and Edit XLSX Files, http://CRAN.R-project.org/package=openxlsx, r package version 2.4.0, 2015.

25 Wang, W. and Wells, M. T.: Model Selection and Semiparametric Inference for Bivariate Failure-Time Data, Journal of the American Statistical Association, 95, 62–72, doi:10.2307/2669523, http://www.jstor.org/stable/2669523, 2000.

Yan, J.: Enjoy the Joy of Copulas: With a Package copula, 2007, 21, 21, doi:10.18637/jss.v021.i04, https://www.jstatsoft.org/index.php/jss/article/view/v021i04, 2007.

Yee, T. W.: The VGAM Package for Categorical Data Analysis, 2010, 32, 34, doi:10.18637/jss.v032.i10, https://www.jstatsoft.org/index.php/jss/article/view/v032i10, 2010.

30 Zhang, L. and Singh, V. P.: Bivariate rainfall frequency distributions using Archimedean copulas, Journal of Hydrology, 332, 93–109, doi:10.1016/j.jhydrol.2006.06.033, 2007.
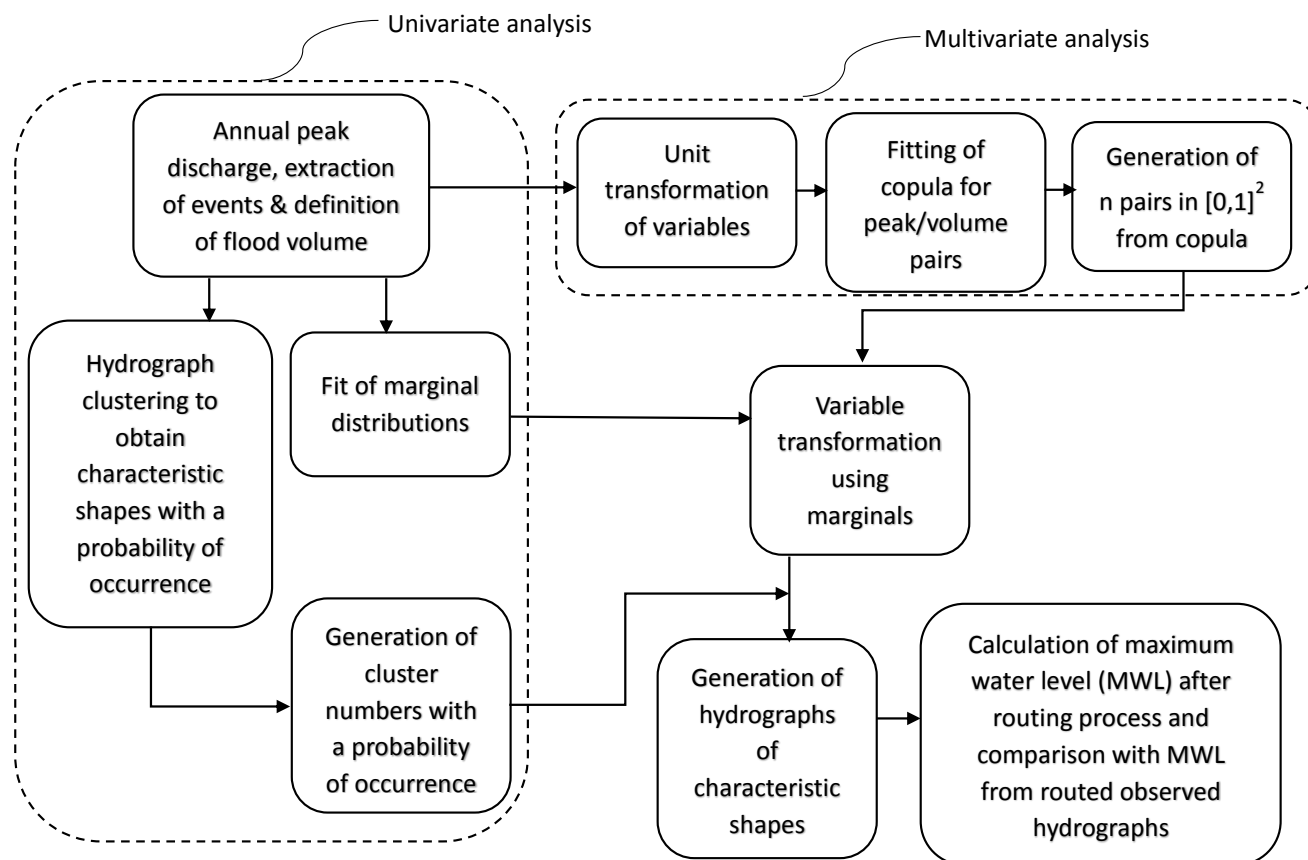
Figure 1. Flowchart of the methodology

Table 1. Main hydrological characteristics of the Panaro watershed (area A, main stream length L, minimum elevation H_{min}, elevation drop ΔH, time of concentration t_c)

A [km^2]	876
L [km]	124.4
H_{min} [m]	32
ΔH [m]	1449.9
t_c [h]	10.0

Hydrology and
Earth System
Sciences
Discussions

Open Access
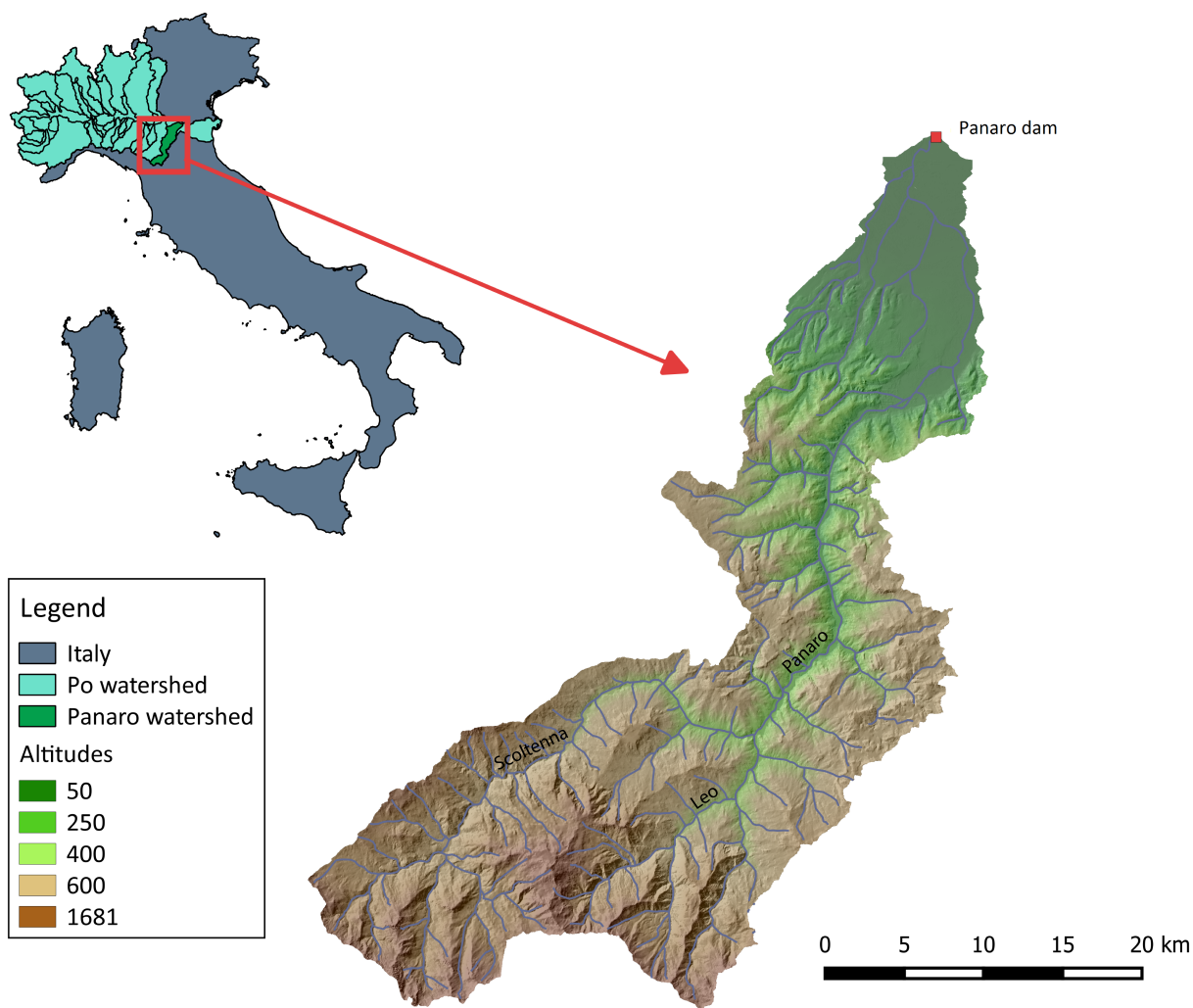
Figure 2. Panaro study watershed

Table 2. Data series summary

Series beginning	1923
Series end	1983
Available years	52
Sampling interval [h]	1.0
Q_{max} [m^3/s]	925
Q_{ave} [m^3/s]	17.6
Q_{min} [m^3/s]	0.0
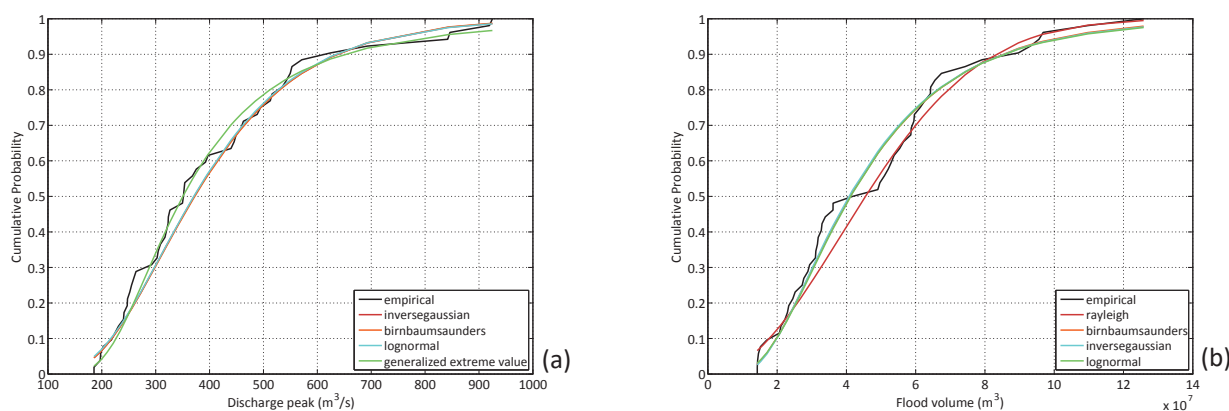
Figure 3. Distribution fittings for flood peak (a) and volume (b)

Table 3. Distribution parameters for the marginals

Variable	scale	shape
Q (Inverse Gaussian)	405.60	2179.78
V (Rayleigh)	38657781.47	-

Table 4. Statistics and p-values from goodness-of fit tests for the marginals

Variable	Kolmogorov-Smirnov	Anderson-Darling	Cramér-von Mises
Flood peak	0.0842 (p =0.824)	0.454 (p =0.794)	0.0632 (p =0.796)
Volume	0.126 (p =0.347)	0.668 (p =0.585)	0.112 (p =0.531)
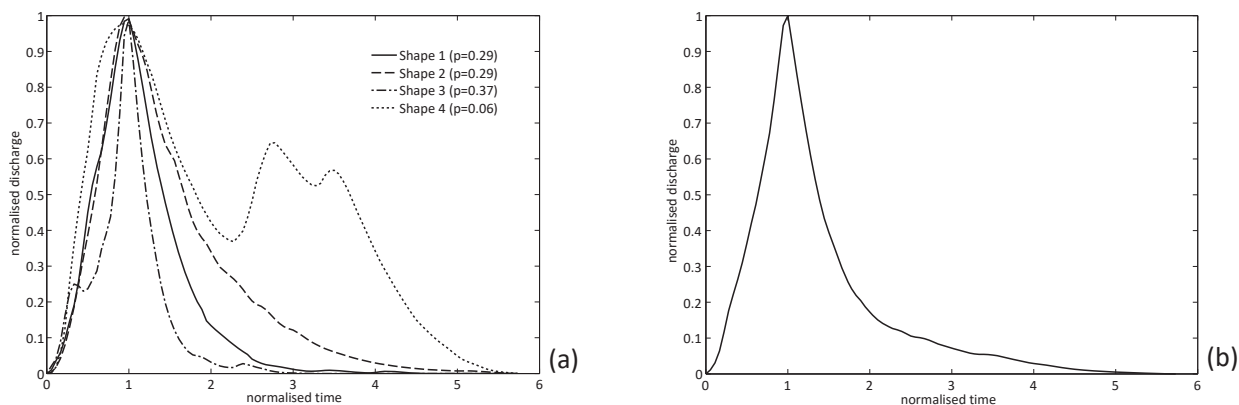
Figure 4. Normalised clustered hydrographs and their probability of occurrence; 4 clusters (a) & 1 cluster (b)

Table 5. Coefficients of dependence and their p values

	Flood peak/Volume
Kendall	$0.58\,(p = 1.4 \times 10^{-9})$
Spearman	$0.77\,(p = 2.2 \times 10^{-16})$
Pearson	$0.81\,(p = 2.6 \times 10^{-13})$

Table 6. Gumbel copula, its parameter space and upper tail dependence coefficient

	$C_\theta(u,v)$	θ space	$\lambda_U^C(\theta)$
Gumbel	$exp[-((-log(u))^\theta + (-log(v))^\theta)^{1/\theta}]$	$[1,\infty]$	$2 - 2^{1/\theta}$

Table 7. Statistics and p values from goodness-of fit tests for the selected copula

	Copula parameter	$\lambda^C(\theta = 2.27)$	Kolmogorov-Smirnov	Cramér-von Mises
Q/V (Gumbel)	2.27	0.643	$0.744\,(p = 0.219)$	$0.078\,(p = 0.327)$

Hydrology and
Earth System
Sciences
Open Access
Discussions
EGU
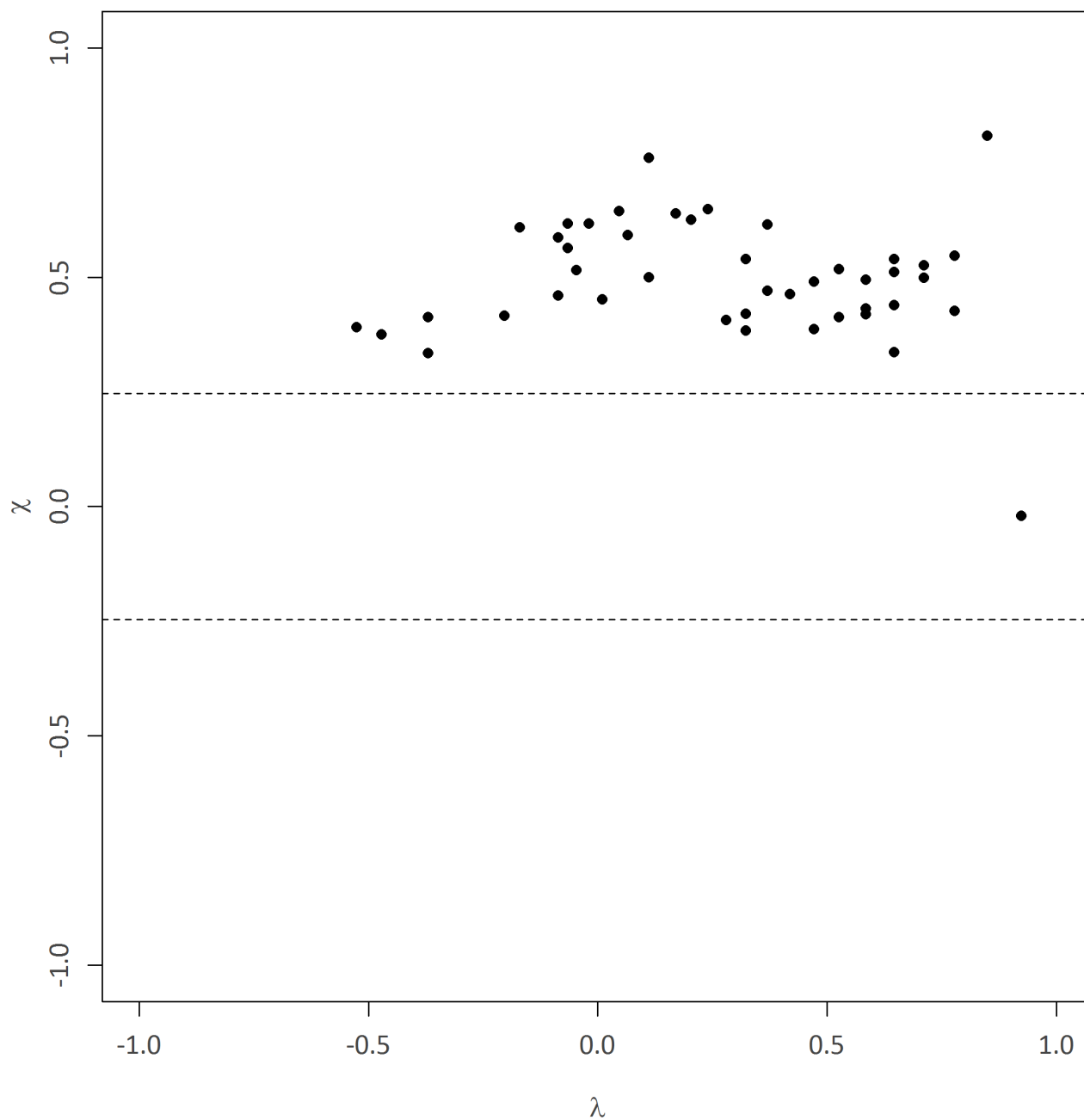
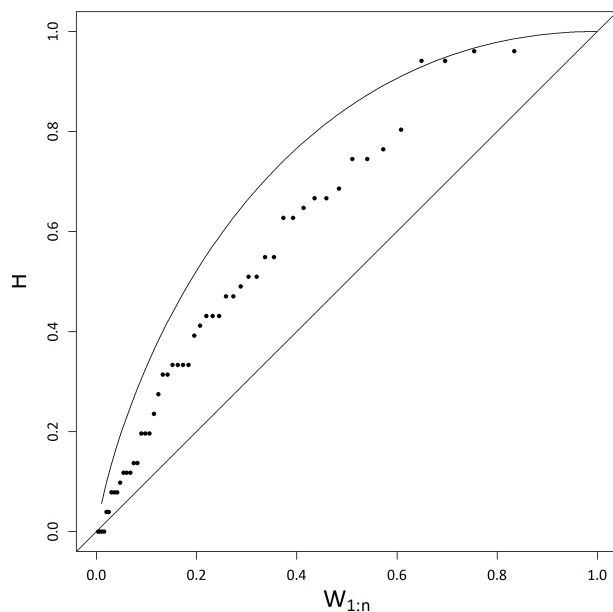**Figure 5.** Chi-plot of peak/volume pairs

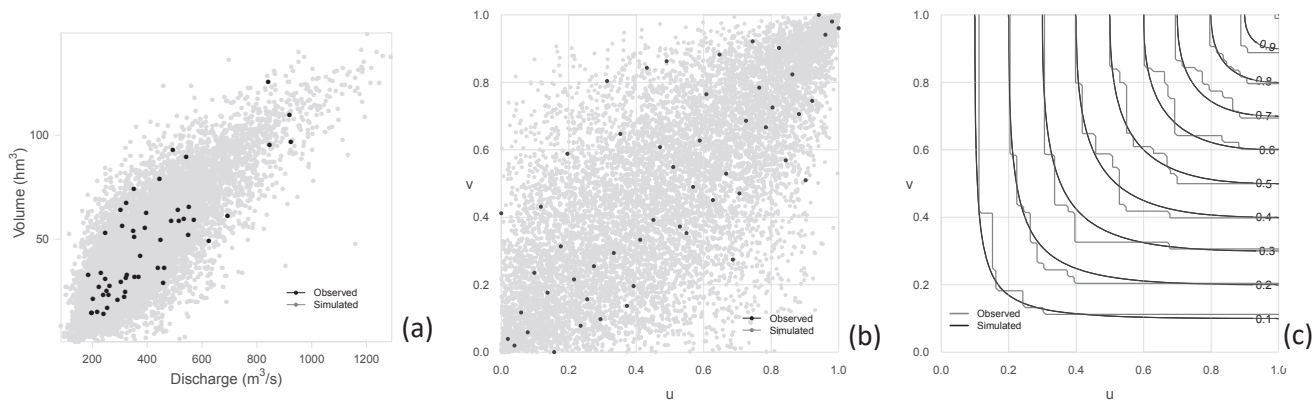**Figure 6.** K-plot of peak/volume pairs

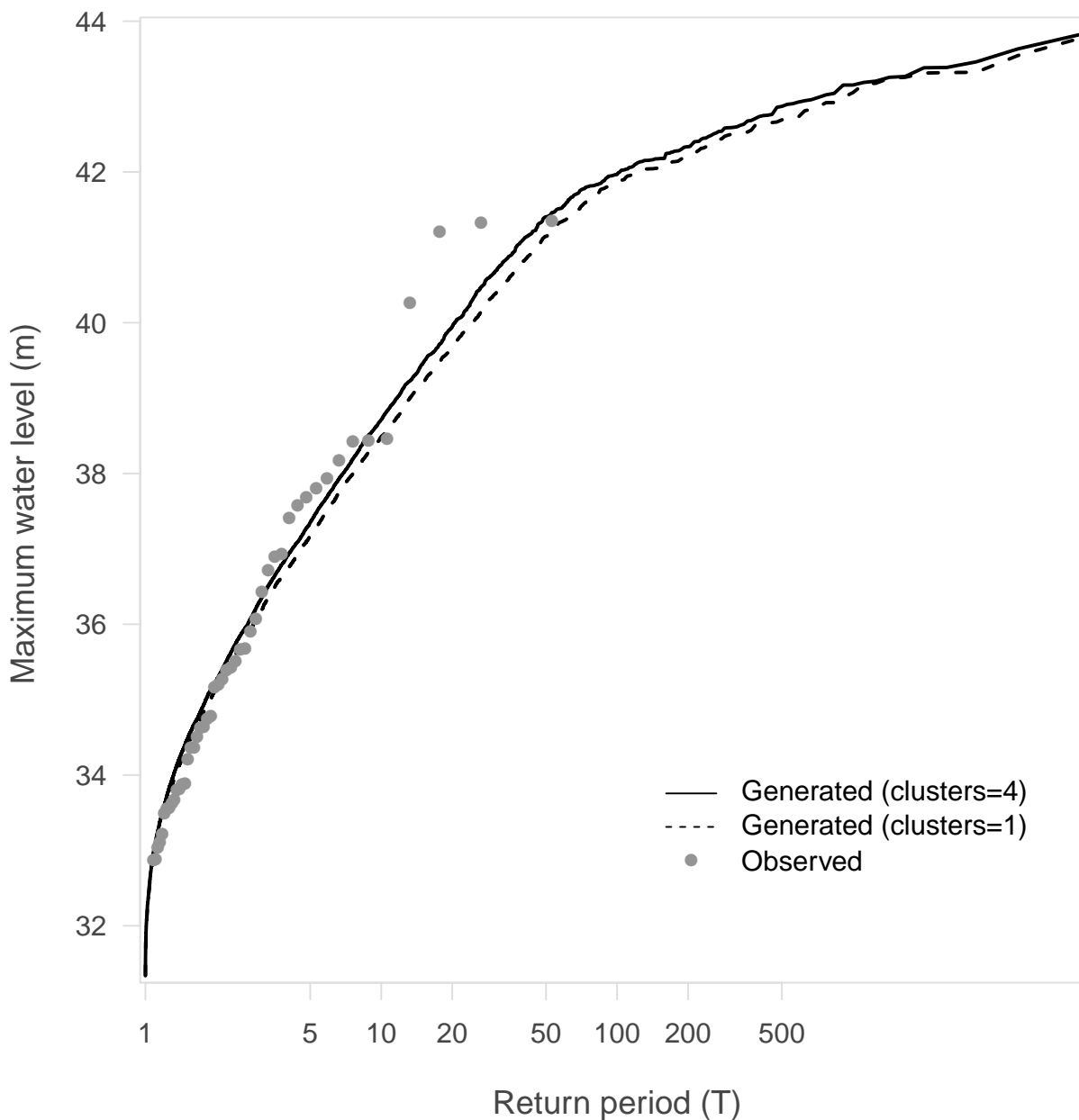**Figure 7.** Scatter (a & b) & contour plot (c) of the theoretical and observed values

Figure 8. Frequency curve of MWL of synthetic and observed hydrographs for four clusters and one cluster
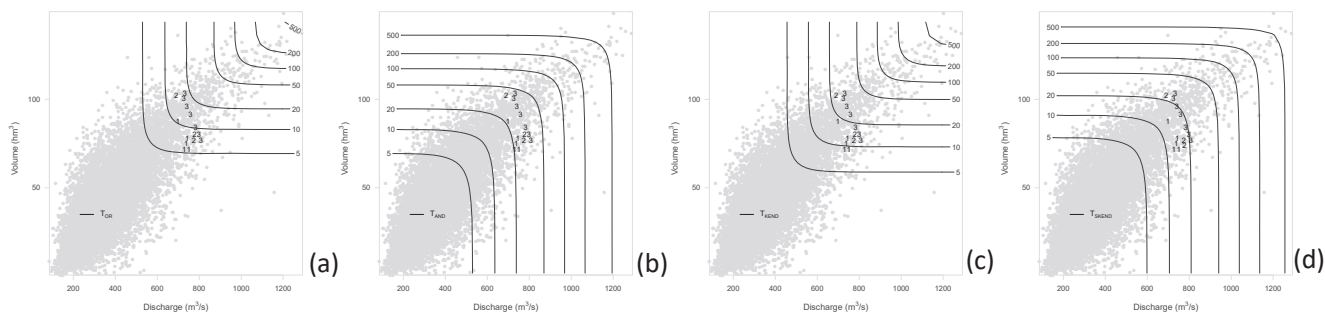
Figure 9. Comparison of joint RP curves (OR, AND, KEND, SKEND) with events (assigned to a hydrograph shape) with RP T=20 years
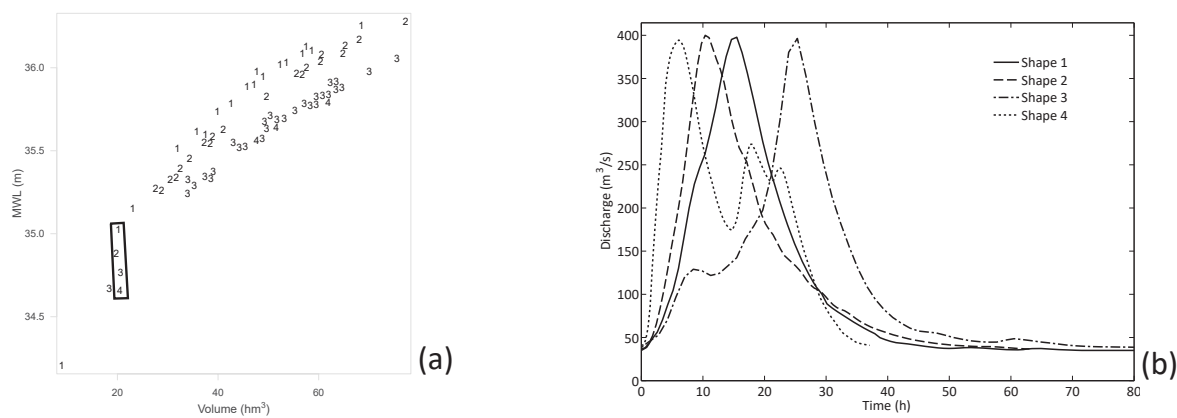

Figure 10. (a). Comparison of effect of hydrograph shape on MWL for events with equal peaks; (b). Selected (noted in rectangle) hydrographs with equal peaks and volumes
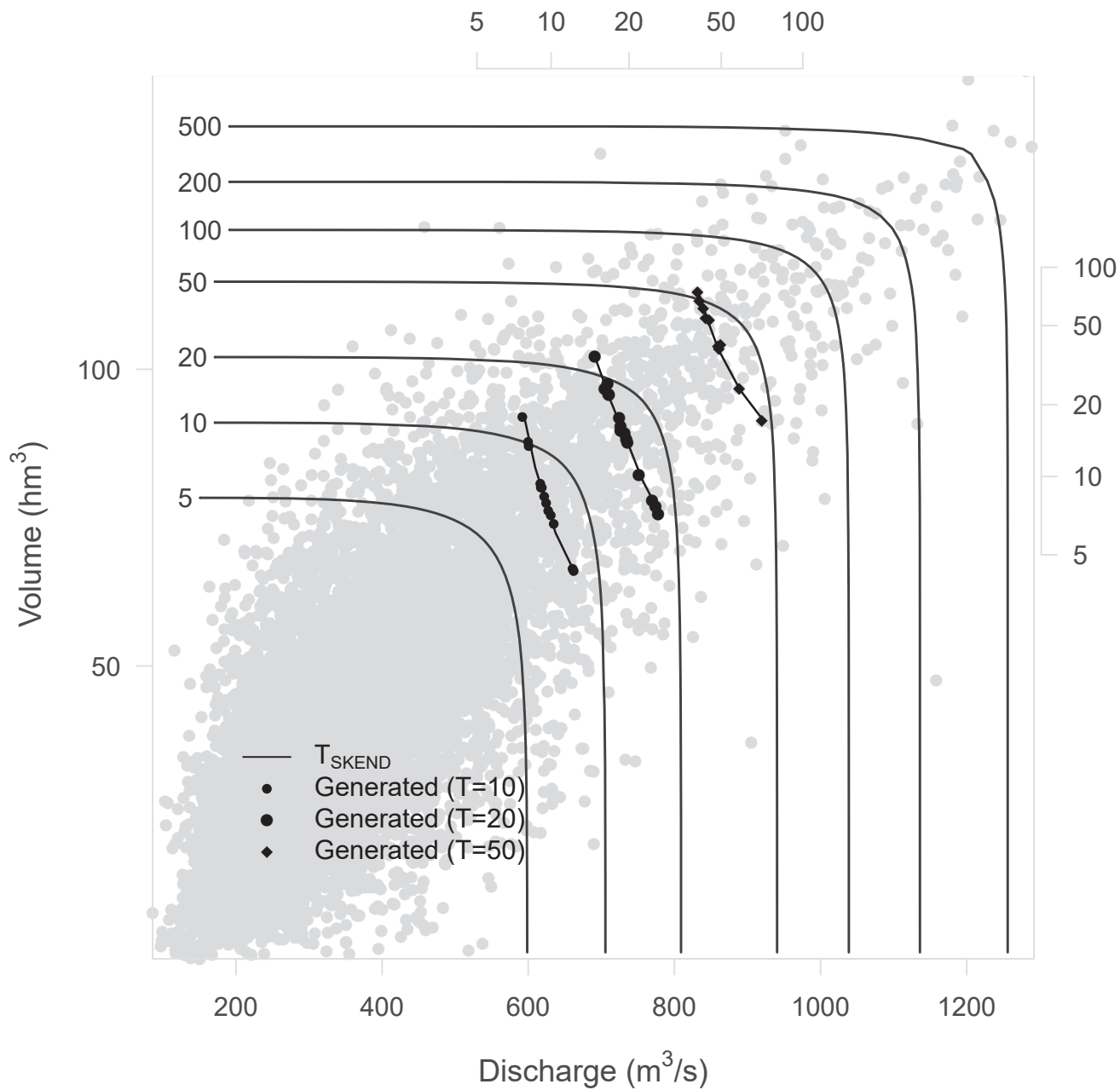
Figure 11. Risk region of events with RP of T=10, 20 and 50 years derived from one hydrograph group