Comments addressed to the first reviewer (Dr F. Serinaldi)

Abstract: The abstract was completely rewritten as the aim of the research was focused on the uncertainty, upon the two reviewers' suggestions.

Section 1.1: The Literature Review (now titled Introduction) was completely rewritten and uninformative references were avoided. Only relevant studies and our eventual contribution were mentioned.

Section 1.2: We have followed the reviewer's suggestion.

P4L11: The numerical effect of the correlation that you mention also in your paper (Serinaldi and Kilsby, 2013) can be a result of the high threshold (80^{th}, 90^{th}, 95^{th} percentile) on the discharge measurements. Here, the threshold was basically equal to each episode's baseflow. The dependence structure of peak and volume does not seem to change (Figure 1), but for a few values, when only the "net" volume is considered. However, given the limited data set we understand that such correlation is difficult to observe.

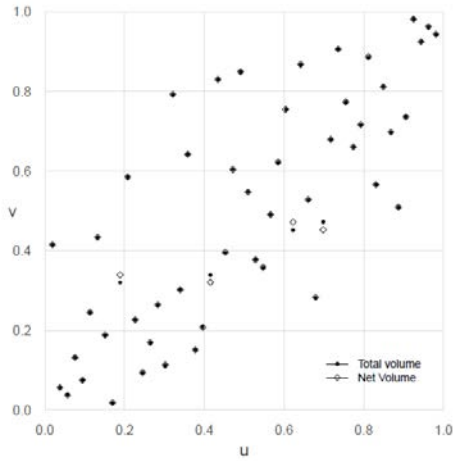

Figure 1. Comparison of peak and volume pairs in the case of total and "net" volume

Section 2: The scope of this paper has changed, focusing on the uncertainty of the structure-based return period of the maximum water level reached in a reservoir. A Bayesian framework was also implemented to comprehend how the levels and the associated uncertainty is affected after the introduction of additional information.

Section 2.1: AIC and BIC formulas were removed and the relevant literature was referenced. The use of the indices was also mentioned.

Section 2.2: Chi- and K-plots were removed.

P4L11 and Section 2.3: Firstly, a preliminary analysis of statistical dependence between peak and volume (Table 5 of submitted manuscript) demonstrated that the hypothesis of zero correlation is rejected. The significance of zero correlation is numerically lower than 10^{-9}. This result of course does not imply definitively the tail dependence condition but when extraordinary peaks occur, extraordinary volume are expected, producing an extraordinary event (Bacchi and Maione, 1984)

Many significant events, at least in Italy, occur when a frontal perturbation generated by the cold high masses coming from the North Atlantic Ocean or the Arctic Ocean, meets Mediterranean southward warm fronts. Depending on the persistence of the south and north current, the generated front begins to develop covering a large area (e.g. 10^4 km^2). Inside this warm front, the energy content is very high. This causes local convective phenomena enhanced by orographic effects. So, thunderstorms can appear locally producing rainfall whose values can surpass one third of the mean annual in 24-30 hours. In the vicinity of the local thunderstorms the rainfall is moderately high producing large soil saturation and increasing, significantly, the contribution to the groundwater. This kind of rainfall events produce not only maximum observed peaks of flood in many rivers

of low (<100 km^2) and medium size (<2000 km^2) but also the largest observed volumes associated with the persistence of the global event. This is the case, for example, for the flood in Florence and Triveneto on 4th November 1966, in Valtellina on 18-25th July 1987, along Tanaro on 5-6th November 1994, along Po in Piedmont on 17-21st October 2000, etc.

We understand, however, that such tail dependence is difficult to detect in a small sample size and that such coefficients can be defective. Thus, they were removed.

Section 2.5: We have removed the part of the various multivariate return periods.

Section 4.1 and 4.2: See response in Section "Comment on the marginal distribution inference" during the discussion phase. Model inference was not the scope of this paper, so, as we have made clear in the reviewed manuscript, there is no clear way to see each model's superiority in a limited data set. We based our choice on the models' parsimony.

Section 4.3: We have changed the title appropriately and the communication of the associated uncertainty has an integral role in the reviewed revision.

Comments addressed to the second reviewer

Page(s) 1, Abstract: See response in the previous section.

Page(s) 3, Line(s) 16-17: The scope of the paper was changed, as mentioned before.

Page(s) 4, Line(s) 17-ff: The baseflow was removed from the analysed hydrographs in an attempt to render, as much as possible, independent the peaks from the previous rainfall events. The baseflow removal is reported also in other works (Apel et al., 2004; Aronica et al., 2012). The abrupt change in the discharge marks the start of the direct runoff. The end of the direct runoff- or at least the runoff caused by the same rainfall event- is considered when the discharge falls below a certain empirical threshold and, at the same time, the gradient of the recession limb becomes steadily small. We considered that baseflow was changing linearly between the start and the end of the direct runoff. Each hydrograph was then visually inspected and heuristic corrections were made when necessary.

The other comments are accepted.

Page(s) 4, Line(s) 25-26: We agree with the reviewer.

Page(s) 5, Line(s) 16-ff: We agree with the reviewer but we would like to add that statistical tests were derived as a more satisfactory method after graphical control.

Page(s) 6, Line(s) 16-ff: We have made the appropriate change.

Page(s) 6, Line(s) 6-7: We accept the suggestion.

Page(s) 6, Line(s) 8-ff: We agree on the existence of uncertainty concerning the TDC but as described before, the hypothesis of the positive TDC is "supported" by physical extreme flood processes.

Page(s) 6, Line(s) 15-ff: We have made the correction.

Page(s) 8, Line(s) 15-ff: Bootstrapped values were considered.

Page(s) 8, Line(s) 18-19 & 29-31: See response to the first reviewer.

Page(s) 9, Line(s) 6-8: The mentioned figures were discarded.

Page(s) 9, Line(s) 9-ff: Regarding the copula information criteria (Grønneberg and Hjort, 2014), we have conducted a Monte Carlo simulation for our sample size (52), the sample's Kendall tau and three one-parameter copulas that gave the smallest AIC, namely Gaussian, Gumbel and Frank. In 76% of the generated samples from a Gaussian copula the AIC was able to "identify" the copula in comparison with 56% of the Copula

information Criteria. For samples drawn from the Gumbel copula the percentages for AIC and CIC were 83% and 73%, accordingly. Of course, these percentages are very likely to change as more copulas are added into the selection and they are dependent on the sample size, as well as Kendall's tau. The difficulty in understanding the systematic difference in performance of the two criteria is supported by Jordanger and Tjøstheim (2014).

Page(s) 9, Line(s) 11: The sample size is small, statistically speaking, so the appropriate change will be made. However, we would like to add that it is considered at least of medium size, hydrologically speaking.

On the other hand, if we consider a quite long hydrological series of 70-100 years of flows, frequently it is not possible to state if we are dealing with a sample extracted from the same population (hypothesis of stationarity of the process) or if the series is composed by the union of two or three different populations. As an example, the dams that were built in Sicily in the period of 1970 up to 2000 only during the years 2005 up to 2016 have reached the maximum regulation level. During the building period, a vivid discussion began between politicians and engineers that regarded the possible overestimation of the dam's size. This discussion is now closed, since the annual rainfall is not similar to that of the period from 1945 to 1970 that was used for storage design.

Page(s) 9, Line(s) 21-22: We agree on the comment.

Page(s) 9, Line(s) 27-ff: Upon the reviewers' suggestions the objective of the paper has completely changed, focusing on the uncertainty, which is not dealt in De Michele et al. (2005).

Page(s) 10, Line(s) 5-8: This sentence was removed.

Page(s) 10, Line(s) 25-26: Any references to multivariate return periods were removed.

References

Apel, H., Thieken, A.H., Merz, B., Blöschl, G., 2004. Flood risk assessment and associated uncertainty. Natural Hazards and Earth System Sciences, 4, 295-308.

Aronica, G. T., Candela, A., Fabio, P., and Santoro, M., 2012. 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.

Bacchi, B., Maione, U., 1984. Influenza di un volume di laminazione sulla distribuzione di probabilità dei colmi di piena. Energia Elettrica, 10, 399-413.

De Michele, C., Salvadori, G., Canossi, M., Petaccia, A., Rosso, R., 2005. Bivariate statistical approach to check adequacy of dam spillway. Journal of Hydrologic Engineering, 10(1), 50-57.

Grønneberg, S., Hjort, N. L., 2014. The copula information criteria. Scandinavian Journal of Statistics, 41 (2), 436–459.

Jordanger, L.A., Tjøstheim, D., 2014. Model selection of copulas: AIC versus a cross validation copula information criterion. Statistics & Probability Letters, 92, 249-255.

Serinaldi, F., Kilsby C. G., 2013. The intrinsic dependence structure of peak, volume, duration, and average intensity of hyetographs and hydrographs, Water Resources Research, 49, 3423–3442.

~~Defining flood risk in a multivariate framework: Application on the Panaro watershed~~ Dealing with uncertainty in the probability of overtopping of a flood mitigation dam

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 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 hydrological 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 hydrograph 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.~~

In recent years, copula multivariate functions were used to model, probabilistically, the most important variables of flood events: discharge peak, flood volume and duration. However, in most of the cases the sampling uncertainty, from which small-sized samples suffer, is neglected. In this paper, considering as a case study a real reservoir controlled by a dam, we apply a structure-based approach to estimate the probability of reaching specific reservoir levels, taking into account the key components of an event (flood peak, volume, hydrograph shape) and of the reservoir (rating curve, volume-water depth relation). Additionally, we improve information about the peaks from historical data and reports through a Bayesian framework, allowing the incorporation of supplementary knowledge from different sources and its associated error. As it is seen here, the extra information can result in a very different inferred parameter set and consequently this is reflected as a strong variability of the reservoir level, associated with a given return period. Most importantly, the sampling uncertainty is accounted for in both cases (single site and multi-site with historical information) and Monte Carlo confidence intervals for the maximum water level

are calculated. It is shown that water levels of specific return periods in a lot of cases overlap, thus making risk assessment without providing confidence intervals deceiving.

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. 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) 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 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.~~

~~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 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 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 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. Defining the return period)~~

~~De Michele et al. (2005) were the first (to our knowledge) to check the adequacy of a dam's spillway under a bivariate 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 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.~~

In the relatively recent literature there is a wide application of the copula functions to model the natural variability of hydrometeorological variables, ranging from rainfall (De Michele and Salvadori, 2003; Zhang and Singh, 2007; Balistrocchi and Bacchi, 2011; Singh and Zhang, 2007; Ariff et al., 2012) to floods (Aronica et al., 2012; Balistrocchi et al., 2014; Candela et al., 2014; Domeneghetti et al., 2013; Gräler et al., 2013; Ganguli and Reddy, 2013)

An important application of this multivariate analysis is the determination of the risk of failure of a hydraulic structure. De Michele et al. (2005) were the first to check the adequacy of a dam's spillway under a bivariate hydrological load, followed by Requena et al. (2013), while Volpi and Fiori (2014) formalised the idea 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 by fixing a "structure-based" return period.

In the same conceptual framework, Serinaldi (2016) suggested that the choice between a univariate and multivariate risk assessment should not be based on whether one or the other over/under estimate the risk but rather on the operational criteria of the problem, or simpler on what is the mechanism of failure.

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 majority of the studies, the communication of the sampling uncertainty- an integral component in a univariate framework- is overlooked in a multivariate case. Serinaldi (2013) studied the effect of sample size on the confidence bands of the probability of exceedance curves of a joint peak-volume event and showed that in small and medium sample sizes these curves largely overlap. Similarly, Zhang et al. (2015) implemented a Bayesian inference approach to account for the uncertainty of parameter estimation and for the occurrence of a drought, coming to the conclusion that the 95 % confidence interval of a 20-year event can include the expected values of 10 up to 50-year events

In order to account for the sampling uncertainty of multivariate cases, where a variable of interest can be expressed as a function of one or more variables, Serinaldi (2016) proposed a Monte Carlo procedure . He also underlined the importance of including confidence intervals when providing point estimates of a variable of interest, which is even more necessary in the multivariate frequency analysis, where the unknown dependence structure contributes to the uncertainty.

The sampling uncertainty in a joint peak-volume event was quantified by Dung et al. (2015) who used two bootstrapping methods- one developed by the authors and the second by Serinaldi (2013)- and concluded, as the previous, that the model selection and parameter estimation methods is of minor significance in uncertainty estimation in respect with sampling uncertainty, even in relatively large sample sizes. They suggested that efforts should be focused on the expansion of the data set in order to achieve a reduction of uncertainty.

The data expansion can be temporal, spatial and causal (Merz and Blöschl, 2008) , thus enriching the available evidence with information from neighbouring basins, previous periods and by comprehension of the flood-generating mechanisms. In the past, many researchers (Parent and Bernier, 2003; Reis Jr and Stedinger, 2005; Gaume et al., 2010; Halbert et al., 2016; Parkes and Demeritt, 2016; Viglione et al., 2013) have dealt with the extension of the available data using information on paleo-floods, historical flood reports, marks of the river stage during important flood events, expert judgement etc. with the aim of reducing the range of uncertainty bands or simply to reach a more realistic design value. Bayesian inference allows the integration of information from different sources and their associated uncertainty and errors and provides a mean of conveying hydrological reasoning in a mathematical context.

In this research we validate a methodology of flood risk assessment in a real case-study, where risk is expressed in terms of probability of exceeding a given reservoir level in an on-line flood mitigation dam. We consider this level as a function of flood peak, volume and hydrograph shape and, consequently, multivariate modelling is implemented with the use of copulas. The characteristics of the reservoir- also a function of the level- are synthesised in the rating curve and the volume-level curve. The main scope is to integrate the associated sampling uncertainty and to build confidence intervals for each water level through

Monte Carlo simulations. Furthermore, we incorporate additional information on the peaks in a Bayesian framework and we examine its effect on the distributions, their confidence intervals, as well as the ones of the reservoir level frequency curve.

1.1 ~~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 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. Defining the return period). 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 Case study and data

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~~

upstream of the Panaro dam (Fig. ~~2~~1) ~~, that~~ and 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).

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 from the Bomporto station located downstream, near the current location of the dam. The hydrological characteristics of the study basin ~~as well as a summary of the available data~~ are briefly presented in Table 1~~and 2~~.

Additional data included the annual peaks of the missing years from the same station (Servizio Idrografico Italiano, 1939, 1953) and recent annual peaks from upstream stations, after consulting the annual hydrological reports of ARPA- Emilia Romagna published in its website (www.arpae.it); in specific for 2003 from Pievepelago, for 2004, 2005 and 2015 from Ponte Samone, and for 2006 to 2014 from Spilamberto.

In a report about natural disaster risks in the city of Modena (Nora and Ghinoi, 2009) it is also stated that the most disastrous flood events of the 20th century happened during the last 40 years (1966, 1969, 1972 and 1973). In November 1966 the flooded area from the Panaro covered 9400 ha, in September 1972 2540 ha and in September 1973 6000 ha (Nora and Ghinoi, 2009).

3 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~~

$$AIC = -2\log(L) + 2k$$

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

~~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.~~

2.1 Data regionalisation

In order to rescale the flood information from subcatchments and from the downstream station, depending on the area, the following scale function was used:

$$Q(A_1) = Q(A_2)(A_1/A_2)^m \tag{1}$$

where $Q(A_1)$ is the rescaled discharge, $Q(A_2)$ a known discharge and m a regional scale exponent.

De Michele and Rosso (2002) clustered basins with similar flood generation mechanism and checked the homogeneity of the grouped regions. The study area was located in North-western Italy and included the Panaro watershed. The proposed scale exponent m for this region is 0.772 with a standard deviation of 0.072. In our case the rescaling regarded different locations of the same basin, although in theory neighbouring basins could have been used (e.g Secchia), but they did not add additional information here.

2.2 Incorporating additional data

Thomas Bayes' theorem expresses how an individual's degree of belief can change after the presence of new evidence. Bayes' theorem can be formulated as:

$$p(\theta \mid D) = \frac{l(D \mid \theta)\,\pi(\theta)}{\int l(D \mid \theta)\pi(\theta)d\theta} \propto l(D \mid \theta)\,\pi(\theta) \tag{2}$$

where $\pi(\theta)$ is the prior density distribution of the parameters θ, $p(\theta \mid D)$ is the posterior distribution after the introduction of the observed information D and $l(D \mid \theta)$ is the likelihood of the data. The denominator serves only as a normalisation constant to ensure unity of the area under $p(\theta \mid D)$, so the equality sign can be substituted with the proportionality sign. This integral cannot be solved analytically, so for its computation Monte Carlo Markov Chain algorithms are employed. In each Markov

chain the aim is the maximisation of the logarithm of the unnormalised joint posterior distribution starting from an initial value and proceeding iteratively in order to arrive at each target distribution (Statisticat, LLC, 2016).

In a Bayesian framework the model's parameters are handled as stochastic variables in order to incorporate the uncertainty of their values (Ouarda and El-Adlouni, 2011). In the present case the model's parameters were the parameter of the peak marginal distribution and the scale exponent. We used non-informative prior distribution for the marginal and a normal distribution prior for the exponent $m \backsim N(0.772, 0.072)$. We integrated a perception threshold X_P- a value which only in k number of years in a historic period of h years was exceeded; here it is set as at about 1000 m^3/s, a value thought to be exceeded only once in a historic period of 117 years since flood reports indicate that during the early years of 1900s, when systematic records were non-existent, the flood events were of less significance in comparison with the events occurring at the $70s$. Additionally we have introduced the uncertainty of the scale exponent m. Therefore, the likelihood function of the data is set as:

$$l(D \mid \theta) = \begin{bmatrix} h \\ k \end{bmatrix} F_X(X_P)^{(h-k)} \prod_{i=1}^{s} [\prod_{j=1}^{n_i} f_x(y_{ij}(A/A_i)^m)] \tag{3}$$

where $\begin{bmatrix} h \\ k \end{bmatrix} = \frac{h!}{k![(h-k)]}$ is the binomial coefficient, s is the number of different sites of the recorded flood peaks, n_i is the number of recorded peaks for each site, y_{ij} are the annual peaks from the different sites.

The Bayesian inference was conducted in R with the package LaplacesDemon (Statisticat, LLC, 2016) and the MCMC algorithm utilised was the Componentwise Hit-And-Run Metropolis. The logarithm of the posterior distribution to be maximised is the sum of the logarithm of the likelihood and the logarithm of the priors:

$$log(p(\mu, \sigma, m \mid D) = log(l(D \mid \mu, \sigma, m) + log(\pi(m)) \tag{4}$$

where μ and σ are the mean and shape parameters of the peak distribution.

2.3 Copula and marginal inference

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, $i = 1, ..., d$ are the marginals of F and C is the copula function.

Copulas provide a powerful tool for the statistical modelling of multivariate data: for a theoretical introduction see Nelsen (2006); Joe (2014); Durante and Sempi (2015), for a practical engineering approach see Genest and Favre (2007); Salvadori et al. (2007); Salvadori and DeMichele (2007).

The application of copula functions has facilitated overcoming some 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).

The degree of relation 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. For the observed discharge/volume pairs these were equal to 0.58, 0.77 and 0.81, accordingly and indicate strong dependence.

In the absence of a long sample, the copulas that fit the data can be numerous and goodness-of-fit tests cannot distinguish between them (Serinaldi, 2013). Since inferring the "correct" copula model is not the aim of this research and since this endeavour at this point can be futile, given the available data set, the final choice was based partly on the preference of previous published research towards the Gumbel, including conference proceedings by Balistrocchi et al. (2014) , who fitted the Gumbel on peaks obtained from a Peak-Over-Threshold method on the same discharge time series. In the present case both the Gaussian and the Gumbel-Hougaard one-parameter copula passed the goodness-of-fit tests (Cramer-von Mises and Kolmogorov-Smirnov) and demonstrated the smallest Akaike weights- or else the probability that the chosen model is the most apt among the tested ones (Burnham and Anderson, 2004). However, we thought that if tail-dependence exists, Gumbel would be more appropriate (belonging to the extreme value copula family), as the Gaussian has no tail dependence. We recall the Gumbel-Hougaard copula as:

$$C(u,v) = \exp[-((-\log(u))^{\theta} + (-\log(v))^{\theta})^{1/\theta} \tag{6}$$

where u, v are the pseudo-observations and θ the copula parameter.

The existence of tail dependence between peak and volume was also implied by some historical evidence. Many significant events in Italy occur when a frontal perturbation, generated by the cold high masses coming from the North Atlantic Ocean or the Arctic Ocean, meets Mediterranean southward warm fronts. Depending on the persistence of the south and north current, the generated front begins to develop covering a large area (e.g. 10^4 km^2). Inside this warm front, the energy content is very high. This causes local convective phenomena enhanced by orographic effects. So, thunderstorms can appear locally producing rainfall whose values can surpass one third of the mean annual in 24-30 hours. In the vicinity of local thunderstorms the rainfall is moderately high producing large soil saturation and increasing, significantly, the contribution to the groundwater. This kind of rainfall events produces not only maximum observed peaks of flood in many rivers of small (<100 km^2) and medium size (<2000 km^2) but also the largest observed volumes associated with the persistence of the global event. This is the case, for example, for the flood in Florence and Triveneto on 4th November 1966, in Valtellina on 18-25th July 1987, along Tanaro on 5-6th November 1994, along Po in Piedmont on 17-21st October 2000, etc.

Unfortunately, tail dependence estimators such as the ones of Frahm et al. (2005) and Schmidt and Stadtmüller (2006) can be biased and susceptible to high uncertainty even in large sample sizes (Serinaldi et al., 2015), thus their use in this case is discouraged.

Regarding the choice of the marginal distributions, we preferred distributions that were more parsimonious, thus reducing the additional statistical uncertainty introduced by an extra parameter, following the logic of Occam's razor, and that provided a nice visual fit. The differences between the corrected AIC, BIC and Akaike weighted values were not sufficient to make a safe distinction between the models. The peaks were modelled with the Inverse Gaussian distribution (two parameters instead

of three of the GEV) and their corresponding volumes with the one-parameter Rayleigh. It is, however, imperative to note that there was no clear indication of overall performance superiority of the chosen distributions.

The parameters of the inferred distributions (copula and marginals) are presented in Table 2.

2.4 Hydrograph ~~clustering~~ selection

The shape of the "design hydrograph "is often considered 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 were determined from the annual maxima flood events extracted from the available 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

$$Q_{norm} = (Q - q_{base})/(Q_{max} - q_{base}) \tag{7}$$

$$t_{norm} = t/t_{Qmax} \tag{8}$$

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).

2.5 ~~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 (Fischer and Schwarz, 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).

2.5 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) :

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

for all $\mathbf{x} \subset R^d$, 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 River (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

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

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 :

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

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:

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

Similarly, the AND return period is defined as:

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

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:

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

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.

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

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.

2.5 Structure-based risk analysis

Volpi and Fiori (2014) associated the RP return period 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 return period of a value z takes the formwould be expressed as :

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

where μ_T is the mean interarrival time between two consecutive occurrences of z (in our case $\mu_T = 1$ year), F_Z is the probability distribution function of the derived variable Z which in this case is the reservoir level and $\mathbf{X} \equiv (Q, V, shape, ...)$ ~~In this case~~ Here, the structure function is very complex , since the reservoir level is a function of the spillway's rating curve and the flood's natural variables, and therefore the whole analysis must be based on Monte Carlo simulations. This adds to the computational burden, specifically when dealing with the quantification of the uncertainty.

2.6 Uncertainty estimation

In order to account for sampling uncertainty and to estimate the confidence intervals the following Monte Carlo procedure was implemented, originally proposed by Serinaldi (2016).

1. Estimate the parameter $\hat{\theta}$ of the copula for the observed sample as well as the parameter of the flood volume distribution.
2. Simulate B bivariate samples of size n (equal to the number of years of the observed sample) using the estimated copula parameter and then transform into volume using the estimated parameter of the marginal.
3. Calculate the copula parameter $\hat{\theta}$ and the volume marginal parameter for each sample with the same estimation method used for the observed sample.
4. Simulate B bivariate samples of size M with the copula parameter $\hat{\theta}$ estimated in the previous step.
5. Transform the samples from the unit interval to discharge and volume using the estimated marginal parameters. Generate $B \times M$ hydrographs with an assigned peak, volume and shape and route them to calculate the reservoir level and the frequency curves of all B samples.
6. Build the confidence intervals of the reservoir level frequency curves.

In the present research B was set equal to 10000 and M equal to 1000.

The confidence intervals of the peaks marginal distribution parameters have been estimated in a Bayesian framework, as stated previously, in order to incorporate the additional knowledge and to account for the scaling uncertainty.

The parameter uncertainty of additional distributions that fit the data could be introduced in the procedure, leading in larger confidence intervals. However, in this case, only the parameter uncertainty of the inferred models was of interest.

3 Results and discussion

Initially we have clustered the hydrograph shapes into four characteristic groups. After simulating 10000 peak-volume pairs from the inferred distributions, we assigned to each one a specific hydrograph shape (respecting their frequency of occurrence). Then, we denormalised and routed the hydrographs; we repeated the same procedure but after clustering into only one group, thus considering a global "mean" hydrograph. The characteristic shapes are depicted in Fig. 2a, along with the "mean" shape (Fig. 2b). The level frequency curve showed small differences between the two cases (Fig. 3) and as we shall see later this

difference is negligible in comparison with the estimated uncertainty. So, we have proceeded with the uncertainty assessment considering one "mean" hydrograph.

We have implemented the Bayesian framework on the peaks extracted from the systematic discharge series recorded at Bomporto station, adding also the uncertainty of the scaling exponent of the regionalisation relation. In the second scenario, we also included recently recorded annual peaks from other hydrometric stations of the same basin, mentioned previously, as well as information from flood reports, while integrating the uncertainty of the scaling exponent. As it can be seen in Fig. 4, when ignoring the additional information, the estimate of discharge peak is lower for a given return period. We focus our attention on relatively small return periods since any extrapolation beyond the available time period is subject to great uncertainty. Indicatively, a peak with a univariate return period of 50 years can increase by 18%, exceeding the confidence intervals of the fitted distribution. This occurs because during the last 10 years crucial flood events appeared in the area. In Table 2 we note that in the second case the distribution's mean is bigger and the shape parameter is smaller with a significantly lower standard deviation.

The 95 % confidence interval of both of the peak distributions can be wide (Fig. 4), e.g. for a univariate return period of 50 years it can span from 665 to 941 m^3/s (29 % difference) and from 837 to 1092 m^3/s (23 % difference) for the first and second case.

Similarly, in the case of the flood volume the 95 % confidence interval for a univariate return period of 50 years can span from 95.5 to 120 hm^3 (20 % difference) (Fig. 5).

The confidence intervals of the parameters of the inferred distributions are presented in Table 2.

The results of the increased peaks are reflected also on the frequency curve of the maximum water level (MWL). The return period here corresponds to a water level, so it is considered structure-based, since the level is a function of the structural and operational characteristics of the dam, among others. As it can be seen (Fig. 6), the MWL is significantly lower in the case of no extra information, especially for greater return periods. For a return period of 50 years the average MWL can differ 1.2 m- a considerable magnitude in terms of volume and when considering that the safety margin above the spillway's crest is in some cases 1 m

In Fig. 7 highest density regions depict a sort of confidence intervals of the MWL for specific return periods for each case (single site information & multi-site with historical information). These regions can be defined as the smallest areas in the sample space with a certain probability coverage and they have the advantage of displaying multimodal distributions, thus they may consist of disjoint subsets (Hyndman, 1996) . They are particularly suitable in multivariate cases or for asymmetrical distributions. In the HDR boxplots the mode (horizontal line) substitutes the median and the darker region corresponds to a probability coverage of 50 %, the lighter to a coverage of 95 % and the points outside to the data beyond the 95 % probability.

The span of the highest density regions slightly decreases as more information is introduced. However, this decrease in uncertainty seems small and we cannot conclude whereas the extra information has contributed to a systematic uncertainty reduction. An increase in the simulation size could lead to a slightly different picture, however the added computational burden is prohibitive and, anyhow, as Salvadori et al. (2015) stated, a clear rule-of-thumb regarding the simulation size does not exist. In any case the size is considered large enough for some safe conclusions, especially for the smaller return periods

Within a certain return period the parameter uncertainty can lead to substantial MWL variations, e.g. for 20 years (in the case of extra information), the span of the MWL with a density of 50 % and 95 % is of 0.8 and 2.3 m, accordingly, which correspond to huge volume differences. These can have devastating effects not only in the case of overtopping but also because the remaining water can cause bank failure due to piping. These spans could increase for larger return periods, where the uncertainty is bound to get vaster.

As the results of previous studies suggested (Serinaldi, 2013; Dung et al., 2015; Zhang et al., 2015; Serinaldi, 2016), the regions of the return periods can overlap; in this case the 95 % confidence interval of an event of 20 years can include, marginally, the expected values of events of 10 up to 50 years (e.g. single-site scenario). For the multi-site with extra information scenario the overlapping region is smaller.

In Fig. 8 the highest density regions (50, 75 & 95 %) are depicted in a two-dimensional plane (discharge-volume) that correspond to a return period of 50 years. For events that result in a MWL with a specific return period, the variation of discharge and volume can be huge even when looking in the smaller density regions (e.g. 50 %). For example, in the 95% region the discharge can assume values with a univariate return period from 1 to 50 years (for the multi-site scenario), which is a strong indicator of the non-linearity of the problem. The same applies for the flood volume.

3.1 ~~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.

3.1 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. Literature review) 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.

3.2 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 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, under these conditions, the differences in the recession limbs of the typical shapes will be annulled. This is visible in Fig. 10a, 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.

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 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.

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).

4 Conclusions

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 consideration 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 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.

This analysis focuses on the uncertainty introduced when calculating the probability of exceeding specific water levels in a flood control reservoir, which is a result of the parameter uncertainty of the marginals of the hydrological variables, as well as the copula multivariate function, due to the small size that characterises in most cases a hydrological sample. Therefore, we attempted to quantify this uncertainty, without aiming our attention to copula/marginals inference. Instead, we studied the effect of additional flood information not only on the distribution parameters but also on the uncertainty range in a Bayesian framework that among else permits the consideration of errors from different sources.

The extra flood data that included additional peaks from different hydrometric stations leaded to a peak distribution with bigger mean and smaller shape parameters and thus to elevated peaks, since it includes flood events of recent years that exceed in magnitude the events of the historical data series. Consequently, including the additional information translates into a general bigger estimate of the peaks, which is also reflected on the MWL's, as the peak is a driving factor of the routing process.

The uncertainty range of discharge and volume is considerable and affects, along with the copula parameter, the MWL. The variations in the MWL for the same structure-based return period correspond to significant variation in the stored water volume. Most importantly, the return period of a specific water level cannot be determined with certainty because the return periods of the events overlap. Naturally, the range of discharge and volume values for a given structure-based return period is very ample due to the wide range of the parameters of the inferred distributions.

A clear observation of whether uncertainty is systematically reduced with the introduction of additional information cannot be made here. Nonetheless, a Bayesian framework allows a certain degree of transparency (Parkes and Demeritt, 2016) . Incorporating knowledge about water levels during historic events, e.g. at the locality of Navicello for the 1783 and 1842

flood, could result in a more significant change in the uncertainty range, as past researches have shown. But, one must consider also the great amount of error involved in these data in the Bayesian framework.

As a general remark one can deduce that the process of risk estimation is inherently crippled by uncertainty that can be quantified or at least approximated. Any attempt to obscure this uncertainty could create a false notion about its existence in a multivariate problem with eventual implications in dam safety.

Data availability

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

The authors declare that they have no conflict of interest.

Acknowledgements. We would like to thank Stefano Orlandini for providing us with all the information regarding the Panaro dam.

For the routing of the hydrographs we used the model developed by Gambarelli et al. (2009). The analysis was implemented in R Statistical Computing Software (R Core Team, 2015) with the use of the following contributed packages: ~~akima (?),~~ CDVine (Brechmann and Schepsmeier, 2013), Cairo (Urbanek and Horner, 2015), copula (Kojadinovic and Yan, 2010; Yan, 2007), fitdistrplus (Delignette-Muller and Dutang, 2015), goftest (Faraway et al., 2015), ~~openxlsx (Walker, 2015),~~ hdrcde (Rob J Hyndman with contributions from Jochen Einbeck and Matt Wand, 2013), ks (Duong, 2016), LaplacesDemon (Statisticat, LLC, 2016), lmomco (Asquith, 2015), statmod (Smyth et al., 2015), VGAM (Yee, 2010), wordcloud (Fellows, 2014).

References

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, 2012.

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, 2012.

Asquith, W.: lmomco—L-moments, censored L-moments, trimmed L-moments, L-comoments, and many distributions, http://www.cran.r-project.org/package=lmomco, r package version 2.1.4, 2015.

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 (in Italian), 2006.

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, 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 Software, 52, 27, doi:10.18637/jss.v052.i03, 2013.

Burnham, K. P. and Anderson, D. R.: Multimodel inference understanding AIC and BIC in model selection, Sociological methods research, 33.2, 261–304, 2004.

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, 2014.

De Michele, C. and Rosso, R.: A multi-level approach to flood frequency regionalisation, Hydrol. Earth Syst. Sci., 6, 185–194, doi:10.5194/hess-6-185-2002, 2002.

De Michele, C. and Salvadori, G.: A Generalized Pareto intensity-duration model of storm rainfall exploiting 2-Copulas, Journal of Geophysical Research: Atmospheres, 108, n/a–n/a, doi: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), 2005.

Delignette-Muller, M. L. and Dutang, C.: fitdistrplus: An R Package for Fitting Distributions, 2015, 64, 34, doi:10.18637/jss.v064.i04, 2015.

Domeneghetti, A., Vorogushyn, S., Castellarin, A., Merz, B., and Brath, A.: Probabilistic flood hazard mapping: effects of uncertain boundary conditions, Hydrol. Earth Syst. Sci., 17, 3127–3140, doi:10.5194/hess-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, 2015.

Duong, T.: ks: Kernel Smoothing, http://CRAN.R-project.org/package=ks, r package version 1.10.4, 2016.

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.

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, 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, 2013.

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, 2013.

Hyndman, R. J.: Computing and Graphing Highest Density Regions, The American Statistician, 50, 120–126, doi:10.2307/2684423, 1996.

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, 2010.

Merz, R. and Blöschl, G.: Flood frequency hydrology: 1. Temporal, spatial, and causal expansion of information, Water Resources Research, 44, n/a–n/a, doi:10.1029/2007WR006744, 2008.

Nora, E. and Ghinoi, A.: Alluvioni e terremoti: Principali rischi naturali di Modena nel Novecento, http://www.comune.modena.it/lecittasostenibili/documenti-cittasostenibili/annale-900-citt-ambiente/alluvioni-e-terremoti, 2009.

Ouarda, T. B. M. J. and El-Adlouni, S.: Bayesian Nonstationary Frequency Analysis of Hydrological Variables1, JAWRA Journal of the American Water Resources Association, 47, 496–505, doi:10.1111/j.1752-1688.2011.00544.x, 2011.

Parkes, B. and Demeritt, D.: Defining the hundred year flood: A Bayesian approach for using historic data to reduce uncertainty in flood frequency estimates, Journal of Hydrology, 540, 1189–1208, doi:http://dx.doi.org/10.1016/j.jhydrol.2016.07.025, 2016.

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, 2013.

Rob J Hyndman with contributions from Jochen Einbeck and Matt Wand: hdrcde: Highest density regions and conditional density estimation, http://CRAN.R-project.org/package=hdrcde, r package version 3.1, 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., Durante, F., Tomasicchio, G. R., and D'Alessandro, F.: Practical guidelines for the multivariate assessment of the structural risk in coastal and off-shore engineering, Coastal Engineering, 95, 77–83, doi:http://dx.doi.org/10.1016/j.coastaleng.2014.09.007, 2015.

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, 2006.

Serinaldi, F.: An uncertain journey around the tails of multivariate hydrological distributions, Water Resources Research, 49, 6527–6547, doi:10.1002/wrcr.20531, 2013.

Serinaldi, F.: Can we tell more than we can know? The limits of bivariate drought analyses in the United States, Stochastic Environmental Research and Risk Assessment, 30, 1691–1704, doi:10.1007/s00477-015-1124-3, 2016.

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, 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, 1997.

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), 2007.

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

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.

Statisticat, LLC: LaplacesDemon: Complete Environment for Bayesian Inference, https://cran.r-project.org/web/packages/LaplacesDemon/, r package version 16.0.1, 2016.

Urbanek, S. and Horner, J.: Cairo: R graphics device using cairo graphics library for creating high-quality bitmap (PNG, JPEG, TIFF), vector (PDF, SVG, PostScript) and display (X11 and Win32) output, http://CRAN.R-project.org/package=Cairo, r package version 1.5-9, 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, 2014.

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

Yan, J.: Enjoy the Joy of Copulas: With a Package copula, 2007, 21, 21, doi:10.18637/jss.v021.i04, 2007.

Yee, T. W.: The VGAM Package for Categorical Data Analysis, 2010, 32, 34, doi:10.18637/jss.v032.i10, 2010.

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.

Zhang, Q., Xiao, M., and Singh, V. P.: Uncertainty evaluation of copula analysis of hydrological droughts in the East River basin, China, Global and Planetary Change, 129, 1–9, doi:http://dx.doi.org/10.1016/j.gloplacha.2015.03.001, 2015.

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

~~Flowchart of the methodology~~

~~Data series summary~~

~~Distribution fittings for flood peak (a) and volume (b)~~

~~Distribution parameters for the marginals~~

~~Statistics and p-values from goodness-if-fit tests for the marginals~~

~~Coefficients of dependence and their p values~~

~~Gumbel copula, its parameter space and upper tail dependence coefficient~~

~~Statistics and p-values from goodness-if-fit tests for the selected copula~~

Table 2. Estimated parameters of the inferred distributions and their confidence interval (95%).

	Parameter name	Estimated parameter	Confidence interval (95%)	Standard deviation
Gumbel-Hougaard	θ	2.27	[1.79,3.01]	0.38
Inverse Gaussian for peak (single site)	μ	364.12	[323.64,416.40]	23.67
	λ	1957.16	[1288.05,2820.61]	380.17
Inverse Gaussian for peak (multi-site &	μ	398.61	[358.42,442.54]	21.89
historical information)	λ	1533.41	[1108.84,2058.99]	256.37
Rayleigh for volume	σ	386.58×10^5	$[343.18,430.53]\times10^5$	26.54×10^5

~~Chi-plot of peak/volume pairs~~

~~K-plot of peak/volume pairs~~

~~Scatter (a & b) & contour plot (c) of the theoretical and observed values~~

~~Comparison of joint RP curves (OR, AND, KEND, SKEND) with events (assigned to a hydrograph shape) with RP T=20 years~~

~~(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~~

~~Risk region of events with RP of T=10, 20 and 50 years derived from one hydrograph group~~

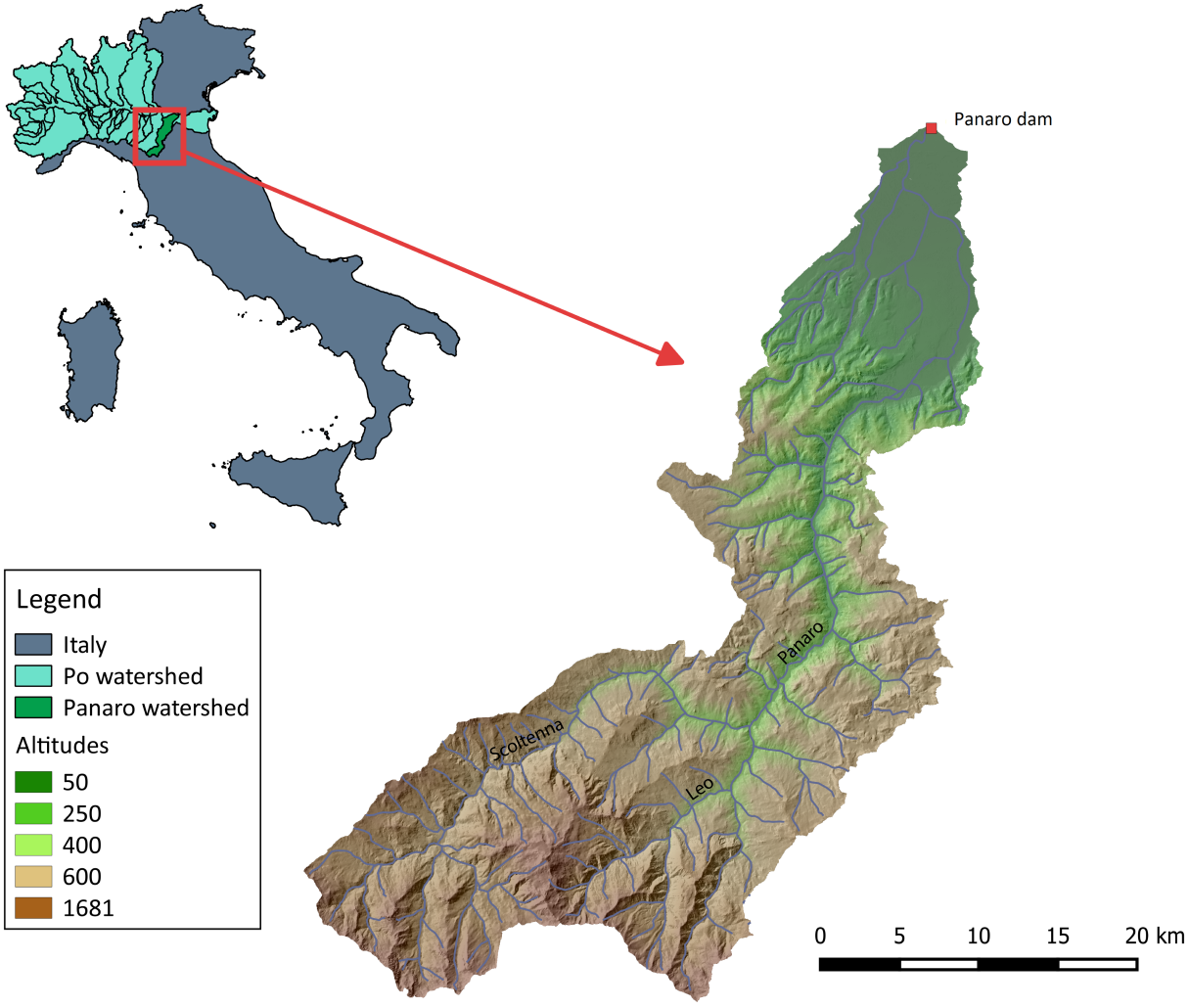

Figure 1. Panaro study watershed

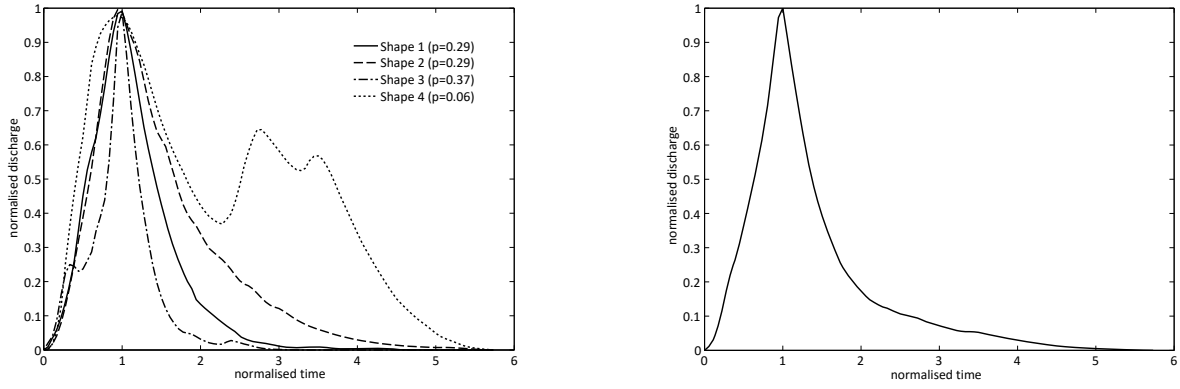

Figure 2. ~~Normalised clustered hydrographs and their probability of occurrence; 4 clusters (a) & 1 cluster (b)~~Characteristic normalised hydrograph shapes (a) with a certain probability of occurrence; (b) "Mean" normalised hydrograph shape.

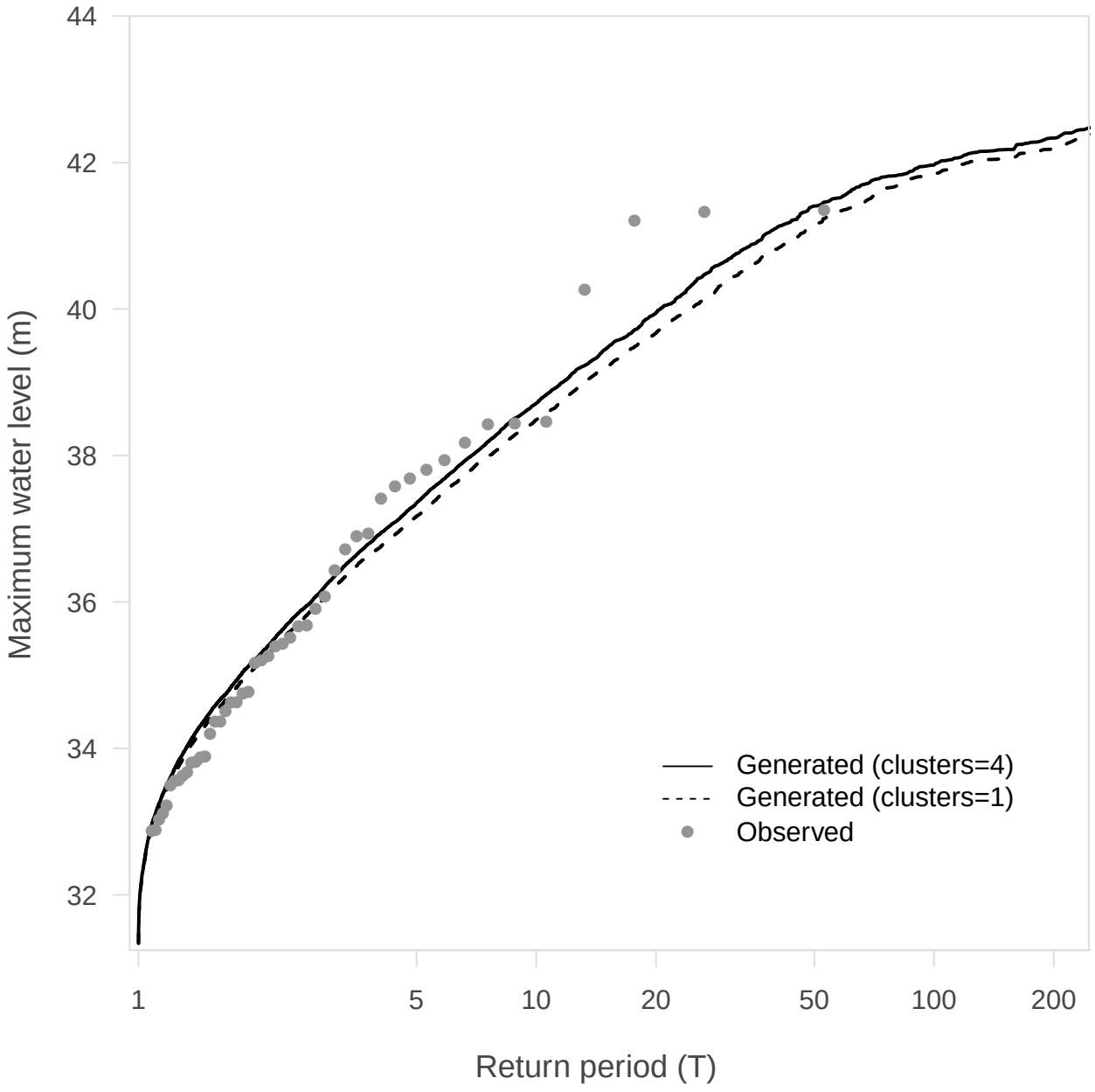

Figure 3. ~~Frequency curve of MWL of synthetic and observed hydrographs for four clusters and one cluster~~ Frequency curves of maximum water level of synthetic hydrographs for four clusters and one cluster and corresponding levels of observed hydrographs.

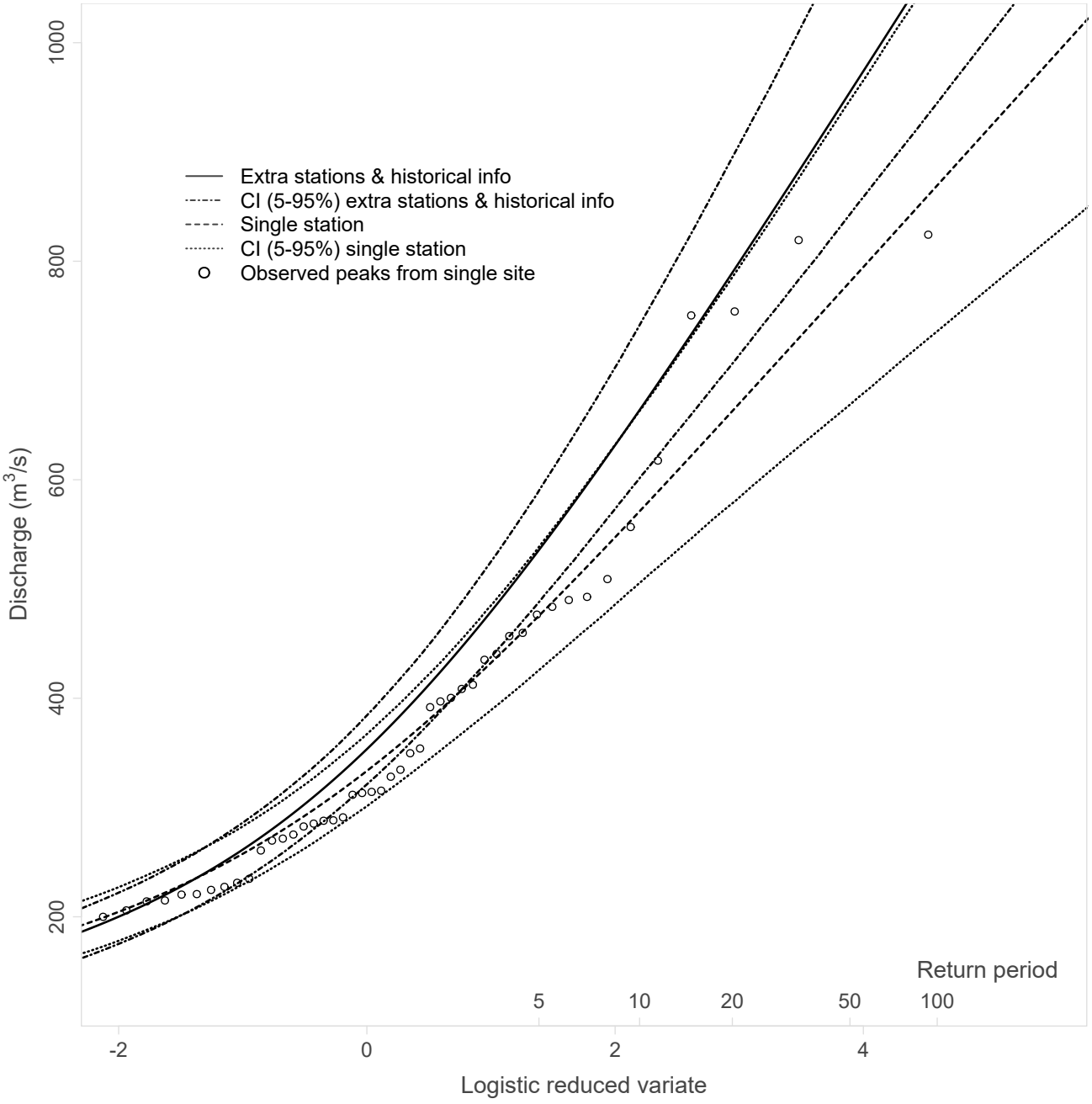

Figure 4. Flood frequency curves with 95 % confidence limits for the single-site data and the multi-site data with extra information case. Observed peaks are also plotted with the Gringorten plotting position.

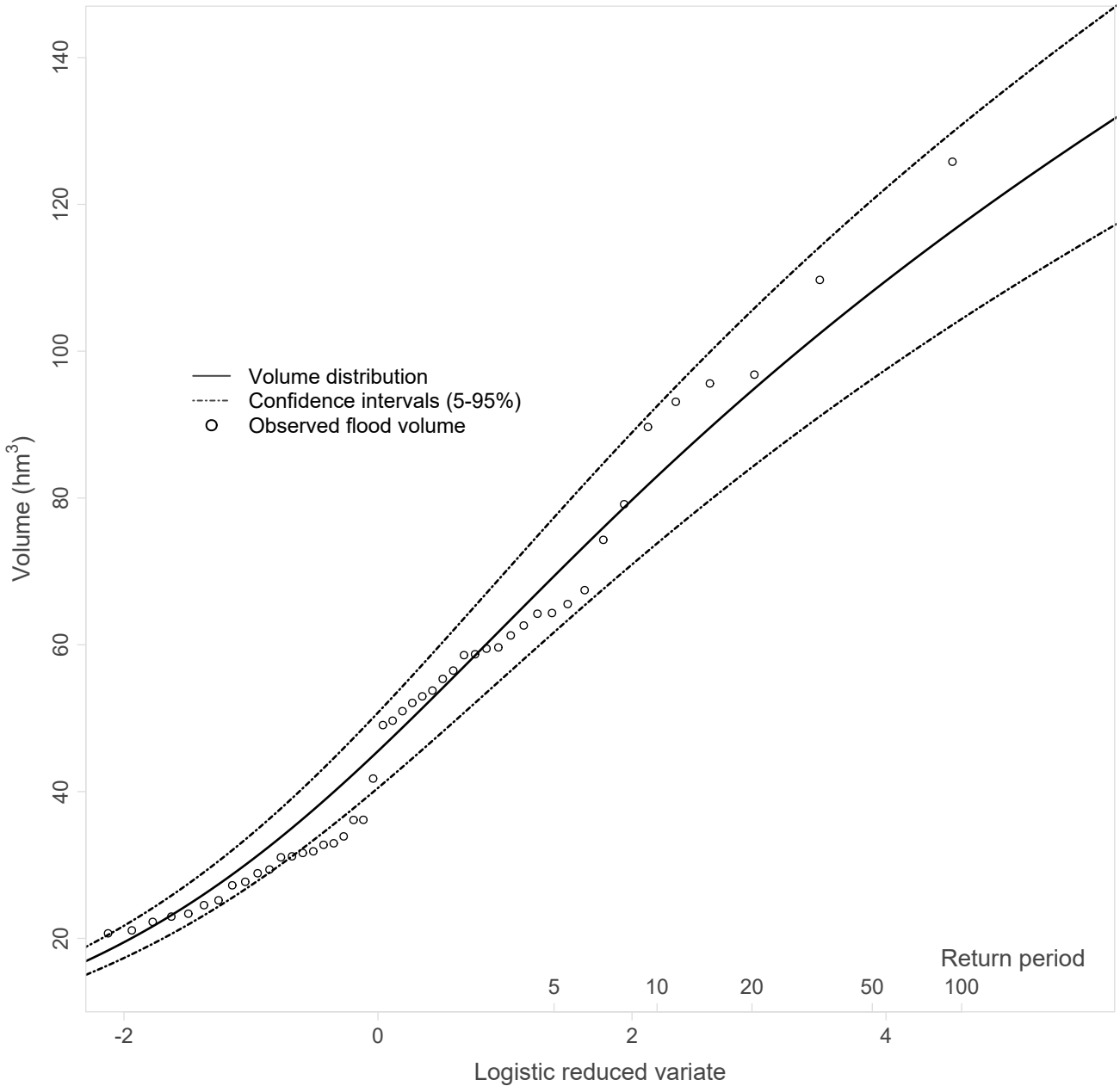

Figure 5. Flood volume curve with 95 % confidence limits. Observed flood volumes are also plotted with the Gringorten plotting position.

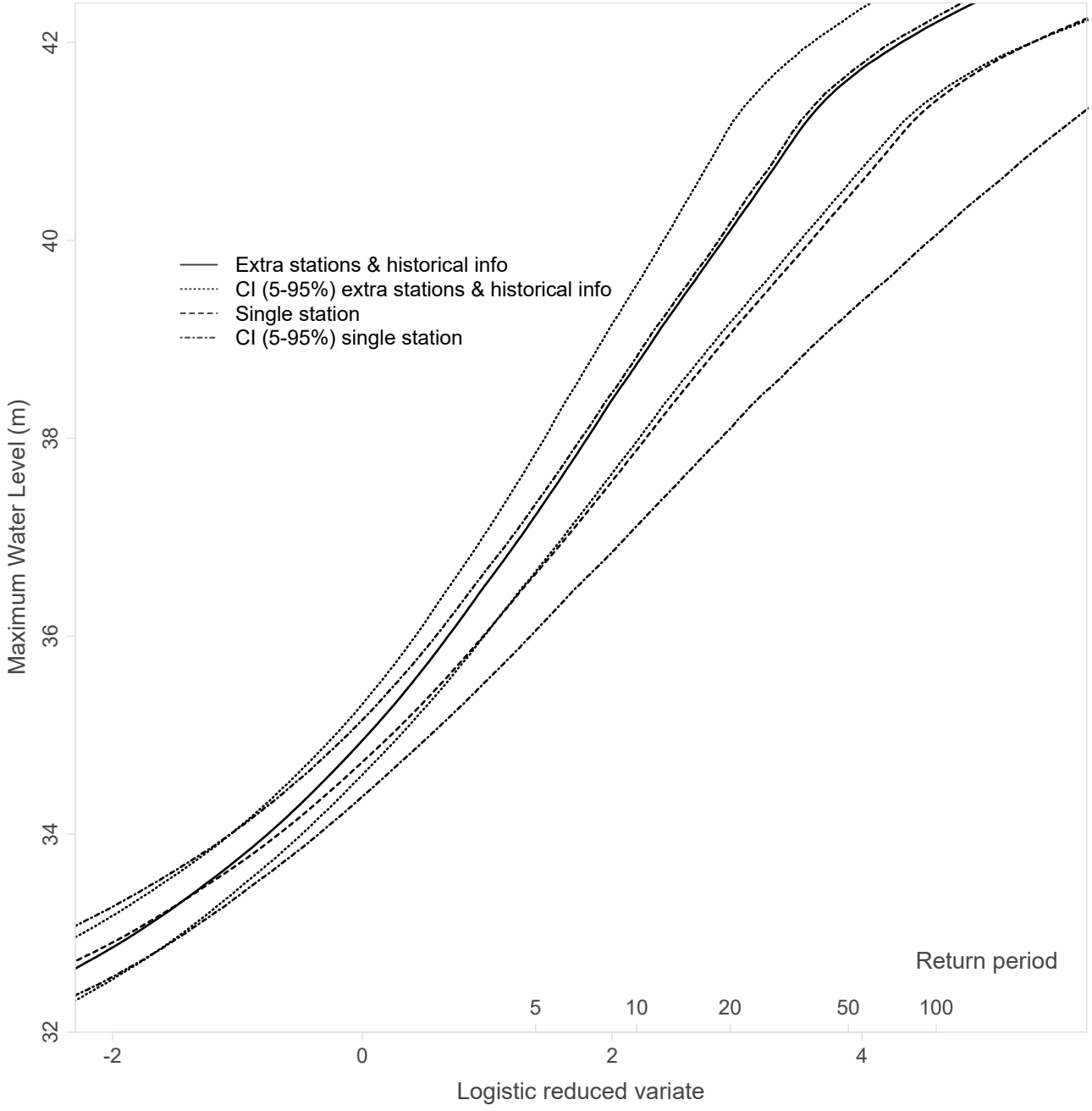

Figure 6. MWL frequency curves with 95 % confidence limits for the single-site data and the multi-site data with extra information case.

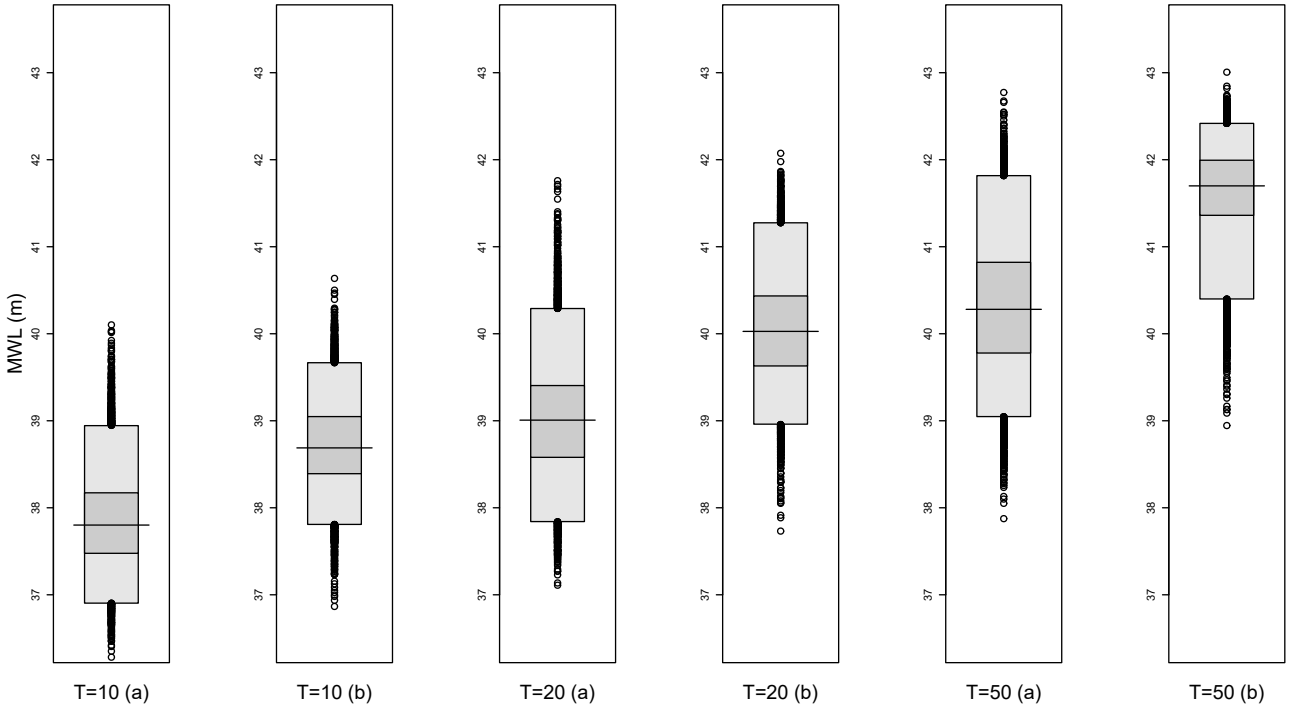

Figure 7. High density region boxplots of MWL for return periods of 10, 20 and 50 years for (a) single-site data and (b) multi-site data with historical information.

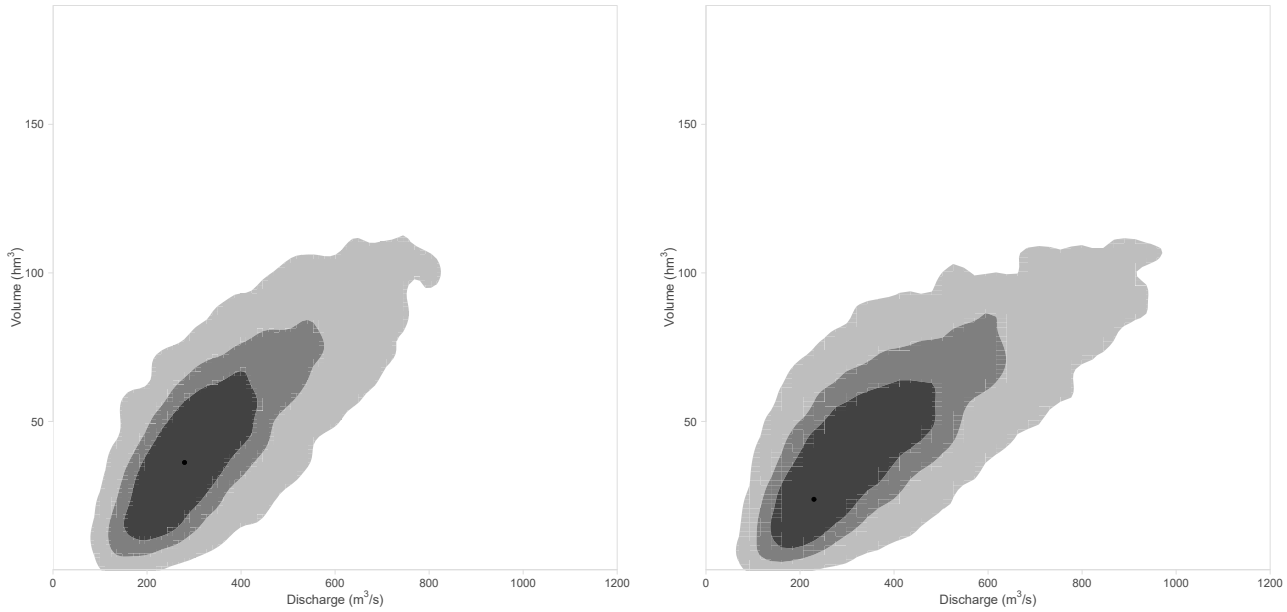

Figure 8. 50, 75 and 95 % of the kernel density areas of MWL's with a return period of 50 years on the discharge-volume plane, for single-site data (a) and multi-site data with historical information (b).