Dear Dr Serinaldi and anonymous reviewer,

At first, we would like to thank you both for your time and effort devoted to comment on our paper and your reading suggestions.

<u>Comment on uncertainty</u>

Dr Serinaldi, one of your main objections was the lack of the communication of the sampling uncertainty that is inherent in the flood frequency analysis.

As you have already mentioned in your review and in your papers (Serinaldi, 2013; 2016) this part is, almost always, overlooked in a multivariate framework, even though it is an integral component in a univariate framework. In your paper (Serinaldi, 2016) you have suggested a Monte Carlo procedure that accounts for the sampling uncertainty of the distribution of moisture conditions. In that specific case the expression is analytically defined. Unfortunately, in a more complex process (i.e. routing) the computational cost can be prohibitively high, given the limited resources. We did carry out a relatively small uncertainty analysis (generation of 1000 frequency curves of n=3500 sample size) compared to the ones in the literature (Dung et al, 2015; Serinaldi, 2013; 2016). The simulation took approximately 45 hours by building a cluster, adding up to 22 cores that worked in parallel.

The 95% confidence interval of the parameters- calculated by maximum log-likelihood estimation method- obtained from parametric bootstrapping (10000 samples) for each distribution is shown in Table 1.

Table 1. 95% confidence intervals of the parameters of the inverse Gaussian for peak, Rayleigh for volume and Gumbel copula for peak-volume pairs

	Mean	shape	scale	copula parameter
Inverse Gaussian	[367.88,445.85]	[1646.36,3184.86]	-	-
Rayleigh	-	-	[3.43,4.31] x10^7	-
Gumbel copula	-	-	-	[1.79,3.01]

A brief summary of the 1000 Monte Carlo simulations fo the MWL is presented below (Table 2 and Fig.1). The difference in MWL for the 75% interval is increasing with the return period and for T=10, 20 and 50 years it is 0.81, 0.91 and 1.06 m, accordingly. For the 95% interval the difference for 20 and 50 years is approximately the same (2.28 m), which is most likely due to the small number of simulations. It could also be an effect of the filtering capacity of the dam, as the density distribution of the attenuated peaks tends to be narrower than the one of the natural peaks.

Additionally, the range of MWL's with the return period of 20 years overlaps the one of 10 years and similarly the range of 50 overlaps the one of 20 years.

The sampling uncertainty is computed assuming the knowledge of the parent multivariate and univariate distribution. The uncertainty range is bound to get larger when including the parameter uncertainty of other distributions.

Table 2. Confidence intervals of maximum water level (5, 25, 50, 75, 95%) for return periods of 10, 20 and 50 years

	5%	25%	50%	75%	95%
T=10	37.54	38.07	38.50	38.88	39.42
T=20	38.55	39.18	39.65	40.09	40.84
T=50	39.70	40.46	41.07	41.52	41.98

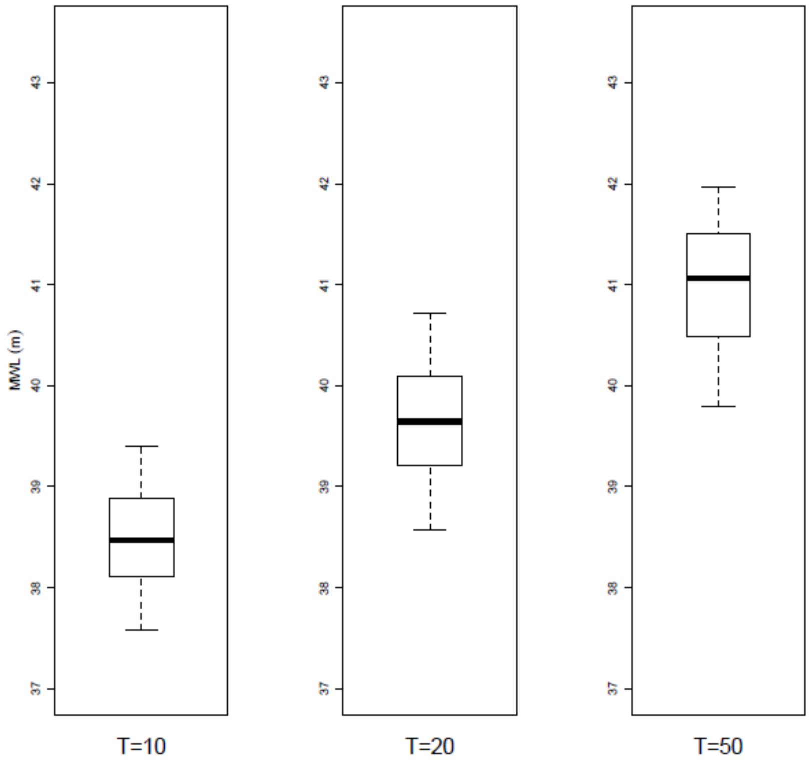

Figure 1. Box plot of confidence intervals of maximum water level (5, 25, 50, 75, 95%)

<u>Comment on the marginal distribution inference</u>

Regarding the choice of the marginal distributions, it is true that the differences between the AIC and BIC values were small (Table 3). We preferred a distribution that gave the smallest BIC (or better, upon your suggestion, the more comparable BIC weights) and that was more parsimonious (e.g. two parameters instead of three of the GEV), thus reducing the additional statistical uncertainty introduced by an extra parameter, following the logic of Occam’s razor. Also, we have graphically assessed the fit with Q-Q, P-P and CDF plots (Fig. 2), and have used extra information on the annual peaks that corroborated the choice.

Table 3. Corrected AIC, BIC, Akaike and BIC weight values for the discharge peak

	AICc	BIC	Akaike weight	BIC weight
Inverse Gaussian	674.89	678.55	0.4097	0.4901
Lognormal	675.49	679.15	0.3031	0.3627
GEV	675.60	680.95	0.2872	0.1472

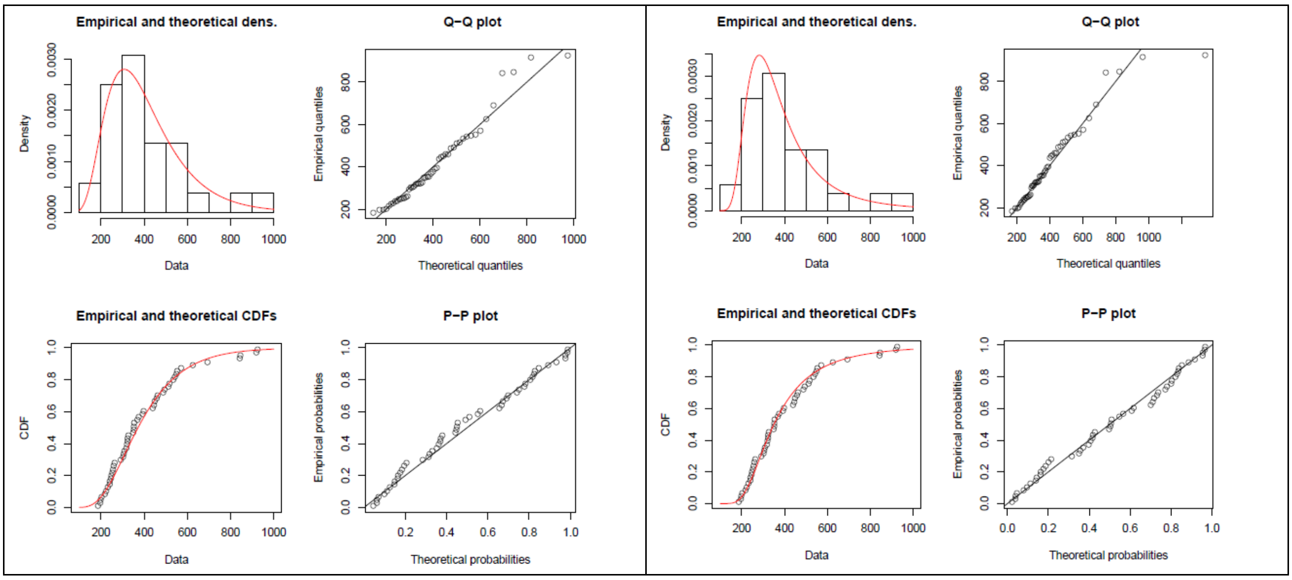

Figure 2. Graphical comparison of Inverse Gaussian (left) and GEV (right) distributions

When calculating the parameters of the GEV for the peak using the L-moments method, the fit seemed visually better (Fig.3), but the slightly better fit of the Inverse Gaussian and, most importantly, the attractiveness of having two parameters instead of three justified our choice.

Unfortunately, as it is already known and accepted, even under complete certainty of the parent distribution the estimated parameter can vary substantially (Table 1).

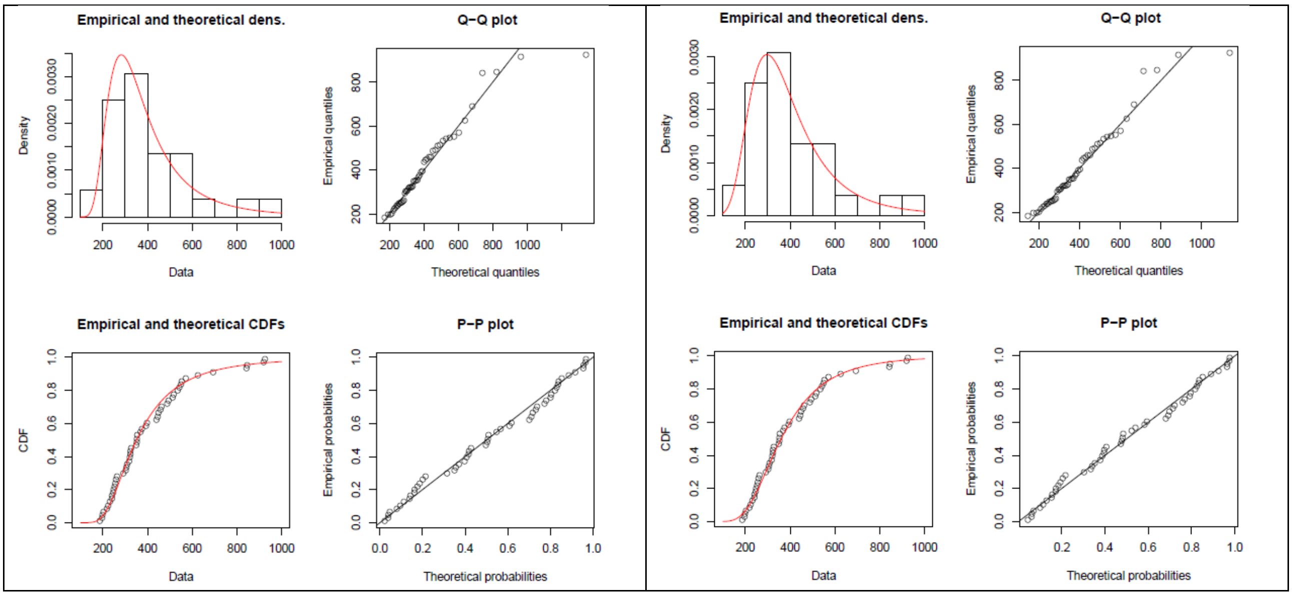

Figure 3. Graphical comparison of GEV distributions with parameters estimated using maximum likelihood (left) and L-moments (right)

<u>Comment on the copula model inference</u>

Of course, when choosing a copula (multivariate distribution) the problem becomes more difficult since more data are needed. The Gumbel and Gaussian copulas, which have the same number of parameters, gave the smallest AIC values, corrected AIC for sample size and Akaike weights of 0.255 and 0.463, accordingly. The final choice was based on previous published research and the obvious preference towards the Gumbel (mentioned in the manuscript's literature review), including conference proceedings by Balistrocchi et al.

(2014), that fitted the Gumbel on peaks obtained from a Peak-Over-Threshold method on the same discharge time series. Additionally, 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.

The difference of the two level frequency curves (Figure 4) derived from the Gumbel and the Gaussian copula is minimal. However, the uncertainty bands of the two curves would certainly overlap.

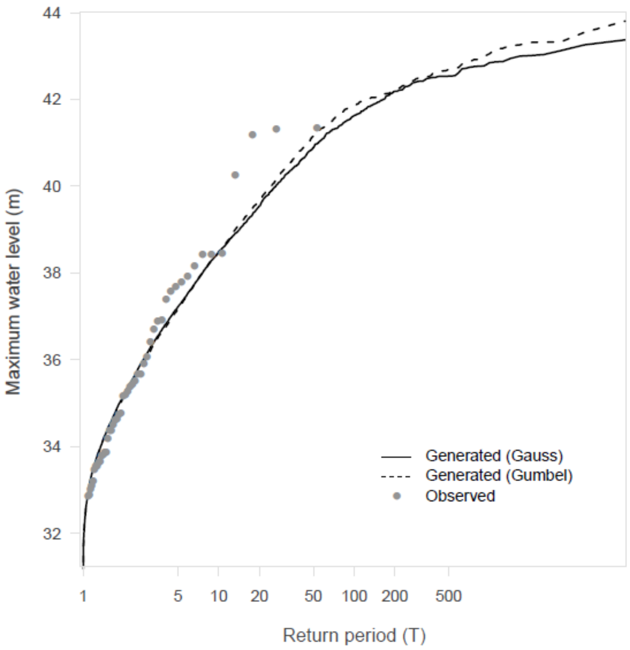

Figure 4. Comparison of MWL curves when considering the Gauss and the Gumbel copula

Comment on the copula information criteria

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).

Comment on the numerical effect of the correlation

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. In our case, the dependence structure of peak and volume does not seem to change (Figure 5), but for a few values, when only the "net" volume is considered.

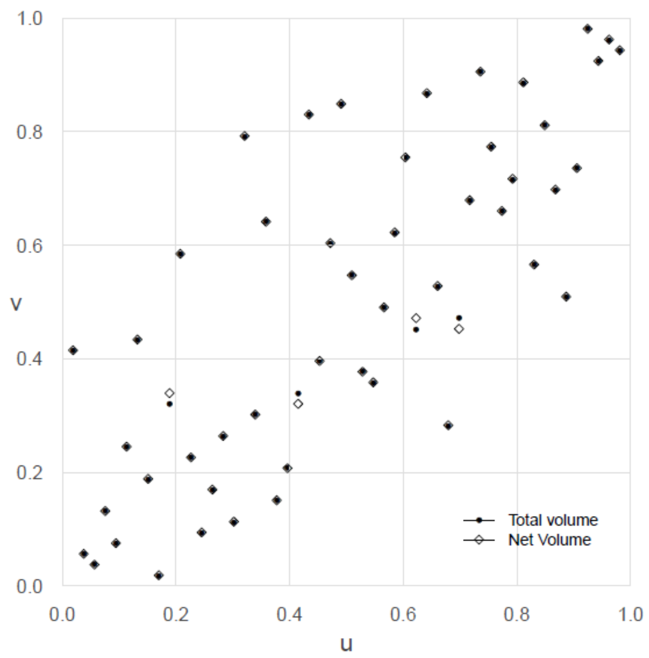

Figure 5. Comparison of peak and volume pairs in the case of total and "net" volume

<u>Specific comments addressed to the first reviewer (Dr F. Serinaldi)</u>

Abstract: We intend to rewrite completely the abstract, better focusing on the objective of the paper and the results.

Section 1.1: We agree with the different approach in unfolding the literature review. The review will be inserted in the context of illustrating the assessment of the problem. Section 1.1 will be merged in Section 1.2

Section 1.2: The first part of Section 1.2, as suggested, will be moved in Section 2 and will be more thoroughly developed.

Section 2.1: See response in Section "Comment on the marginal distribution inference".

Section 2.2: See response in Section "Comment on the numerical effect of the correlation".

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, of 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 ask, then, what is the illogical aspect in the application of the existence of tail dependence, despite a few tens of observations?

Finally, we intend to collect more data for the demonstration of this “intuitive” hypothesis analysing floods with peaks lower than the annual maximum.

Section 2.5: We accept the comment.

Section 4.1 and 4.2: See response in Section “Comment on the marginal distribution inference”.

Section 4.3: We intend to change the title appropriately and the communication of the associated uncertainty will have an integral role in the future revision.

<u>Specific comments addressed to the second reviewer</u>

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

Page(s) 3, Line(s) 16-17: We generally agree with the referee that the risk of failure in a given life time period T is defined by eq. (46) in Salvadori (2016). However, the argument of our paper deals with the division of the variable space in two regions; one that corresponds to the variables that produce an output that surpasses a certain threshold of the derived variable and the other that is below that threshold.

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 accept the suggestion.

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 accept the suggestion as suggested.

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

Page(s) 8, Line(s) 18-19: In general, we agree on the comment but in this case the choice was based looking at the distributions plots, the AIC values and the number of parameters of the tested distributions (see Section “Comment on the marginal distribution inference”).

Page(s) 8, Line(s) 29-31: Same comment as previous.

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

Page(s) 9, Line(s) 9-ff: See Section “Comment on the copula information criteria”.

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: In the seminal paper of De Michele et al. (2005) the hydrographs are generated through the Nash model. This adopted approach is based on the clustering of the historical hydrographs into groups in order to preserve some of the peculiarities of the flood wave that, of course, depend on the temporal evolution of the rainfall events.

Page(s) 10, Line(s) 5-8: We will rewrite the phrase in a clearer manner.

Page(s) 10, Line(s) 25-26: We agree with the referee. In particular, we support that return period should be applied on single variables. When one considers an aleatory variable ensemble $(X_1, \dots, X_N)$ the concept of the return period can be applied on derived single variables; each of these is a function of the preceding variables e.g. $Y_1 = Y_1(X_1, \dots, X_N)$,…, $Y_K = Y_K(X_1, \dots, X_N)$.

Final notes

Finally, as Dr Serinaldi has guessed correctly, I am a PhD student at the start of my 3rd year. Sometimes we, students, base our choices (e.g. structure of a paper, methodological procedures) on reviewed works of professors, experienced researchers, etc., published in (high-profile) journals. This was also the case. As much as I understand that it is imperative for the reader to process and filter the published information, I do feel that the obligation to maintain a certain quality should be preserved by the opposite part. For that, I thank you once more for your thorough review.

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.

Balistrocchi, M., Ranzi, R., Bacchi, B., 2014. Multivariate statistical analysis of flood variables by copulas: Two italian case studies. 3rd IAHR Europe Congress.

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.

Delignette-Muller, M. L., Dutang, C., 2015. fitdistrplus: An R Package for Fitting Distributions. Journal of Statistical Software, 64(4), 1-34. URL http://www.jstatsoft.org/v64/i04/.

Dung, N.V., Merz, B., Bárdossy, A., Apel, H., 2015. Handling uncertainty in bivariate quantile estimation – an application to flood hazard analysis in the Mekong Delta. J Hydrol, 527, 704–717.

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.

Salvadori, G., Durante F., De Michele C., Bernardi M., Petrella L., 2016. A multivariate copula-based framework for dealing with hazard scenarios and failure probabilities, Water Resour. Res., 52, 3701–3721.

Serinaldi F., 2013. An uncertain journey around the tails of multivariate hydrological distributions. Water Resources Research, 49, 6527-6547.

Serinaldi F., 2016. Can we tell more than we can know? The limits of bivariate drought analysis in the United States. Stochastic Environmental Research and Risk Assessment, 30(6), 1691-1704.

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.