REVIEW REPORT

Journal: Hydrol. Earth Syst. Sci.
Paper: hess-2016-519
Title: Defining flood risk in a multivariate framework: Application on the Panaro watershed
Author(s): Eleni Maria Michailidi and Baldassare Bacchi

GENERAL COMMENTS.
In my opinion, the present level of the paper is not adequate for the standards of *Hydrol. Earth Syst. Sci.*: apparently, it is not innovative, and several fundamental issues need to be fixed before acceptance. In addition, English must be improved: for instance, in many sentences, the tenses are inconsistent and "at random", and several typos are present as well throughout the paper.

The objections given below should be read as a positive criticism: these suggestions are constructive, they indicate how to fix several problems, and may greatly improve the paper. I think that a MAJOR REVISION is necessary. Overall recommendations are given at the end of this review, as well as some relevant references.

SPECIFIC COMMENTS.

Page(s) 1, Abstract.

Referee. The abstract is somewhat confused, and does not focus on the problem under investigation: practically, it is non-informative, as instead it should be. The Authors should stress the key point of their work: in a nutshell, the assessment of the impact of the shape of the hydrograph on the MWL via a Monte Carlo structural approach.

Page(s) 3, Line(s) 16–17.

Authors. 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…

Referee. If this is the "main objective of this study", it has already been dealt with in literature, including some of the papers cited by the Authors. In particular, a general and consistent mathematical framework, based on Hazard Scenarios, for dealing with a structural approach (such as the one of interest in this work) has recently been outlined in Salvadori et al. (2016).

Page(s) 4, Line(s) 17–ff.

Authors. The marginal distributions of flood peak and volume (after the baseflow removal) were selected…

Referee. What procedure was used for the baseflow removal? It can make a difference. In addition, Eq.s (1)–(2) can be discarded, being well known. Maybe, the use of the Corrected AIC as a selection criterion could be a better choice, in order to penalize possible over-parametrization/fitting.

Page(s) 4, Line(s) 25–26.

Authors. Lastly, the Kolmogorov-Smirnov, the Anderson-Darling and the Cramér-von Mises tests were carried out to test the goodness-of-fit.

Referee. Here the procedural order should be changed: it makes sense to select a distribution *only* among the admissible ones, i.e. those who pass a GoF test. Therefore, first carry out suitable GoF tests, and then apply a selection criterion *only* to the admissible laws.

Page(s) 5, Line(s) 16–ff.

Authors. To graphically assess independence chi-plots and Kendall plots were generated…

Referee. There is no need to "graphically assess" the dependence: objective statistical tests suffice. These two paragraphs, and the corresponding analyses, can be discarded.

Page(s) 6, Line(s) 16–ff.

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

Referee. A better way to make a reference to the seminal literature is as follows: "Copulas provide a powerful tool for modeling 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 De Michele (2007)."

Page(s) 6, Line(s) 6–7.

Authors. … investigated by Genest and Rivest (1993); Wang and Wells (2000).

Referee. Here the correct reference is to the GoF algorithms provided by the R package "copula" (Hofert et al., 2016), which involves the procedures outlined in Genest et al. (2009).

Page(s) 6, Line(s) 8–ff.

Authors. Since we are dealing with extremes we also assessed the tail dependence of the observations…

Referee. Here it is not clear what the Authors did. Either they estimated the TDC using the available data, or they estimated it via a functional expression involving the fitted copula: this should be made clear. In all cases, given the sample size involved, the estimate of the TDC may be affected by large uncertainties. Later, the TDC is used by the Authors to get some "suggestions" concerning the copula selection, but I suspect that this may be questionable (and, indeed, also the Authors claim that the TDC "can only be used qualitatively"). I would suggest to discard this issue: practically, it is useless, it looks like an academic exercise.

Page(s) 6, Line(s) 15–ff.

Authors. 2.5 Defining the return period…

Referee. Section 2.5 can be discarded: it is well known material, it is enough to put seminal references. Note that the cited structural approach by Volpi and Fiori (2014) has recently been generalized and improved in Salvadori et al. (2015, 2016), by relaxing some of the assumptions needed for a consistent Monte Carlo approach.

As a matter of fact, there is no need to discuss (multivariate) Return Periods in this work: the Structural Approach sought by the Authors is an univariate one, it only involves a single target (structural) variable. In turn, the (multivariate) Return Periods of the loads possibly affecting the structure considered are of no interest here.

Page(s) 8, Line(s) 15–ff.

Authors. The p-values were much greater than . . .

Referee. How is calculated the p-value here? Is it the one computed by R/Matlab, simply comparing the available data and the fitted distribution? If yes, then (as also explained in the Matlab help) such a p-value might not be a valuable one: in this case, a bootstrap p-value should be calculated (it only takes a few lines of code in R/Matlab).

Page(s) 8, Line(s) 18–19.

Authors. 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.

Referee. Statistically speaking, this choice is definitely questionable: the Authors should appropriately justify it. With a sample of size 52, almost everything (or nothing) can be said. Did the Authors try other estimation techniques (beyond ML), such as L-moments of PWM's, which in literature are reputed to provide "better" estimates of the parameters in the GEV case (especially the shape one)?

Page(s) 8, Line(s) 29–31.

Authors. 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).

Referee. Same comments as in the previous two items above.

Page(s) 9, Line(s) 6–8.

Authors. 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.

Referee. Table 5 suffices: Figures 5 and 6 can be discarded, since they add no useful information.

Page(s) 9, Line(s) 9–ff.

Authors. Among all the copula distributions that were tested. . .

Referee. The use of AIC/BIC for selecting copulas (or multivariate distributions) has recently been questioned by Grønneberg and Hjort (2014). For this reason, the cross-validation procedure provided by the function `xvCopula` (available in the new R package "copula"), which computes the leave-one-out cross-validation criterion (or a k-fold version of it) for the hypothesized parametric copula family, should be used here.

Page(s) 9, Line(s) 11.

Authors. This similarity is a consequence of the medium sample size (set of 52 values).

Referee. Right, there might be a model "identifiability" problem here, due to the small sample size.

Page(s) 9, Line(s) 21–22.

Authors. 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.

Referee. In turn, why do not the Authors just forget about the TDC? Practically, it makes no difference, it is useless for their purposes. Statistically speaking, it is generally a dream trying to carry out an asymptotic analysis with a sample of size 52 (and, by definition, the TDC's are asymptotic limits).

Page(s) 9, Line(s) 27–ff.

Authors. Following the generation of 10000 pairs of flood peak and volume values...

Referee. Here, the Authors should stress and emphasize the differences (and the possible improvements) of their work with respect to (and over) the seminal paper by De Michele et al. (2005).

Page(s) 10, Line(s) 5?–8?.

Authors. 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.

Referee. This paragraph, as well as Figure 9, is not clear to me. What does it mean "This variability prohibits the clear definition of a region where all multivariate events produce risk lower than an assigned value"? What is the notion of "risk" the Authors are considering? Please provide some thorough explanation.

Page(s) 10, Line(s) 25?–26?.

Authors. In order to highlight the importance of choosing the appropriate return period expression we have included the comparison between the various definitions.

Referee. To the best of my understanding (and according to Serinaldi (2015); Salvadori et al. (2016)), the Return Period should be chosen *a priori*, considering the physics of the phenomenon and/or the response of the structure, i.e. according to the problem at hand. It is not clear why such a "comparison" is presented. Please provide some clear explanation.

In all cases, as stressed above, here multivariate Return Periods are not needed: to all appearances, they will disappear in the revised version of the paper. In turn, just discard this sentence and the related comparisons.

Overall recommendations. In general, I think that the paper should be "oriented" in a different direction, and substantial changes should be introduced.

Essentially, the ultimate target of the Authors is a Structural Approach, similar to the one worked out in the seminal paper by De Michele et al. (2005). Useful indications and guidelines for carrying out a multivariate Structural Approach are given in Volpi and Fiori (2014); Salvadori et al. (2015, 2016).

In particular, since the notion of Return Period has recently been "dismissed" by Serinaldi (2015), I would suggest to use instead the concept of Failure Probability, and work out analyses similar to the ones outlined in Salvadori et al. (2016). In this latter paper, a target variable W is deterministically linked, via a structure function Ψ, to a pair of random environmental loads H and D, yielding a functional form like $W = \Psi(H, D)$. Apparently, this is exactly the same situation the Authors have to assess. Given the fact that univariate/multivariate fits are already available, it will take a short time to adapt the software code in order to compute the new quantities of interest, and draw the corresponding plots.

In particular, Item 5 "Structural Scenario" in Section 3 "Hazard Scnearios" in Salvadori et al. (2016) well describes the procedures/algorithms to be used, including a Monte Carlo approach. Later, in Section 4 "The Failure Probability Approach", Item 5 "Structural Scenario" provides a formula of interest. Finally, the structural case study presented in Section 5.2 clearly illustrates the steps needed to carry out a sensible structural approach. Should the structural function used by the Authors be a tricky one (such as the one in equation (48) in Salvadori et al. (2016)), the analyses, as well the estimation of the uncertainties, can be worked out via a Monte Carlo procedure, as illustrated in the cited paper, and presented in the corresponding Figures (where confidence bands for the Failure Probabilities of interest, as well as box-plots for the design variable, are shown). I do not think that the structural case study investigated by the Authors is more difficult/tricky than the one assessed in Salvadori et al. (2016).

A final note. While preparing this review, I had the chance to read the comments of Prof. Francesco Serinaldi (the other referee). In my opinion, he made an excellent technical review: on the one hand, he indicated several critical issues; on the other hand, he provided valuable suggestions and solutions to make the paper publishable. My personal recommendations are as follows.

1. The "tones" of the report of Prof. Serinaldi are strong, but this is a feature of his character (he usually writes reviews like the one received by the Authors): the Authors should not concentrate on the "folkloristic" way used by Prof. Serinaldi to express his opinions, rather they should take into serious account the (correct) technical objections he made.

2. My review, as well as the one of Prof. Serinaldi, clearly indicate how to fix the paper: in particular, practically, Prof. Serinaldi re-wrote several parts of the work, and provided valuable suggestions about how to proceed. I personally agree with his technical indications, and I tried to provide further useful procedural guidelines.

I think that ignoring the above recommendations will result in an inevitable REJECTION of the paper. The Authors should eventually acknowledge helpful comments and suggestions by Prof. Serinaldi (and an anonymous reviewer as well).

References

De Michele, C., Salvadori, G., Canossi, M., Petaccia, A., Rosso, R., 2005. Bivariate statistical approach to check adequacy of dam spillway. J. Hydrol. Eng. 10 (1), 50–57, doi: 10.1061/(ASCE)1084-0699(2005)10:1(50).

Durante, F., Sempi, C., 2015. Principles of copula theory. CRC/Chapman & Hall, Boca Raton, FL.

Genest, C., Favre, A., 2007. Everything you always wanted to know about copula modeling but were afraid to ask. Journal of Hydrologic Engineering 12 (4), 347–368.

Genest, C., Rémillard, B., Beaudoin, D., 2009. Goodness-of-fit tests for copulas: A review and a power study. Insurance: Mathematics and Economics 44, 199–213.

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

Hofert, M., Kojadinovic, I., Maechler, M., Yan, J., 2016. copula: Multivariate Dependence with Copulas. R package version 0.999-15 Edition.
URL http://CRAN.R-project.org/package=copula

Joe, H., 2014. Dependence Modeling with Copulas. CRC Monographs on Statistics & Applied Probability. Chapman & Hall, London.

Nelsen, R., 2006. An introduction to copulas, 2nd Edition. Springer-Verlag, New York.

Salvadori, G., De Michele, C., 2007. On the use of copulas in hydrology: theory and practice. J. Hydrol. Eng. 12 (4), 369–380, (Special Issue: Copulas in Hydrology; doi: 10.1061/(ASCE)1084-0699(2007)12:4(369)).

Salvadori, G., De Michele, C., Kottegoda, N., Rosso, R., 2007. Extremes in Nature. An approach using Copulas. Vol. 56 of Water Science and Technology Library Series. Springer, Dordrecht, ISBN: 978-1-4020-4415-1.

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 Resources Research 53, 3701–3721, doi: 10.1002/2015WR017225.

Salvadori, G., Durante, F., Tomasicchio, G. R., D'Alessandro, F., 2015. Practical guidelines for the multivariate assessment of the structural risk in coastal and off-shore engineering. Coastal Engineering 95, 77–83, doi: 10.1016/j.coastaleng.2014.09.007.

Serinaldi, F., 2015. Dismissing return periods! Stochastic Environmental Research and Risk Assessment 29 (4), 1179–1189, doi: 10.1007/s00477-014-0916-1.

Volpi, E., Fiori, A., 2014. Hydraulic structures subject to bivariate hydrological loads: Return period, design, and risk assessment. Water Resour. Res. 50 (2), 885–897.