REVIEW REPORT

Journal: Hydrology and Earth System Sciences
Paper: hess-2018-159
Title: Modeling the spatial dependence of floods using the Fisher copula
Author(s): Manuela I. Brunner, Reinhard Furrer, and Anne-Catherine Favre

GENERAL COMMENTS.

This is a nice paper: it is clear, well written, it deals with a problem of interest for the readers, and introduces several elements of novelty, which are well combined together in an appropriate way. Therefore, I may anticipate that I am in favour of having this work published. However, a few critical issues must be fixed before acceptance. Below, please find some indications: the objections should be read in a constructive way, since they may help the Authors improve the paper.

As a final note, for the benefit of unskilled readers and practitioners, the Authors should provide some basic and thorough references involving seminal books, papers, and guidelines about copula modeling, like writing:

"For a theoretical introduction to copulas, see Nelsen (2006); Joe (2014); Durante and Sempi (2015); for a practical/engineering approach, see Genest and Favre (2007); Salvadori and De Michele (2007); Salvadori et al. (2007). In particular, elementary Guidelines for using copulas are illustrated in Favre et al. (2004); Salvadori and De Michele (2004); Salvadori et al. (2014) (and references therein) for multivariate frequency analysis and design, and in Salvadori et al. (2015, 2016) for a multivariate structural approach."

SPECIFIC COMMENTS.

Page(s) 1, Line(s) 15–16.

Author(s). The Fisher copula is therefore a suitable model for the stochastic simulation of flood event sets at multiple gauged and ungauged locations.

Referee. Such a claim is too strong and general, since it is based on a single case study: please make it weaker.

Page(s) 5, Line(s) 5–6.

Author(s). First, flood events were identified at a local scale for each individual station using a peak-over-threshold approach with the 0.9975 quantile as a threshold.

Referee. Is this the 99.75% quantile of the empirical distribution of the hourly data collected at each station? Did I get it right?

Page(s) 5, Line(s) 9–10.

Author(s). This procedure allowed for the composition of an event set with events during which at least one station exceeded its 0.9975 quantile.

Referee. If I understand it correctly, the Authors use a multivariate "OR" Hazard Scenario approach, as thoroughly conceptually defined and discussed in Salvadori et al. (2016): please make the point clear.

Page(s) 5, Line(s) 15–16.

Author(s). Criterion one was a low variability of the ranks of one event across different stations (standard deviation of ranks ¡ 50)...

Referee. The explanation is somewhat obscure, and maybe I did not get it right. Why using the standard deviation of the ranks and not of the observations? The ranks are the same for all stations (i.e., integers from 1 to N), and cannot express the actual intensity/magnitude of the phenomenon. To be clear: the observations may have the same ranks at two different stations, but the discharges measured at one station may be, say, 10 times larger than the ones observed at another station. Please make the point clear, and similarly for Criterion 2.

Page(s) 6, Figure(s) 2.

In the caption of Figure 2, the Authors may add that patterns of positive association are clearly visible in all cases.

Page(s) 6, Line(s) 9–10.

Author(s). It shows that there is a dependence...

Referee. The Authors may add that, in particular, the variables are in general positively associated.

Page(s) 6, Line(s) 11–13.

Author(s). Both upper and lower tail dependence were present in the data according to the estimator of Schmidt and Stadtmuller (2006) which needs to be used with care since it provides unreliable estimates for small sample sizes (Serinaldi, 2015).

Referee. The Authors correctly warn the reader about the problems concerning the estimate of the Tail Dependence coefficients. However, later (e.g., at page 19), they write sentences like "this model fits the tail dependence better than the other model". I would suggest to check the manuscript, and change (if not discard) claims like the one mentioned above. In fact, given the uncertainty of the estimates of the Tail Dependence coefficients (if not "randomness" of the estimate, as in the numerical experiments under controlled conditions I personally carried out using the same estimators suggested by the Authors), I think it could be dangerous to use, and to rely on, the notion of Tail Dependence.

Page(s) 7, Line(s) 4–5.

Author(s). We used river distance as a distance measure since it has a hydrological meaning.

Referee. This is an interesting point, and may provide a valuable solution: I like it!

Page(s) 7, Line(s) 14–15.

Author(s). The generalized extreme value distribution (GEV) (Coles, 2001) was not rejected for both types of events according to the Kolmogorov-Smirnov test statistic (level $\alpha = 0.05$).

Referee. This is a critical statistical point: how was the p-value computed? In fact, as is well known (e.g., simply read the help of Matlab), the KS test requires that the theoretical distribution be known a priori, it cannot be the fitted one. In the latter case, suitable (but simple) Monte Carlo techniques can be used to estimate an approximate p-value. Please clarify the issue.

Page(s) 7, Line(s) 24–15.

Why the Y-coordinate is not present in the case of the Location parameter?

Page(s) 7, Line(s) 29.

Author(s). They resulted in absolute prediction errors over the ten stations of 0.11, 0.21, and 0.15 respectively.

Referee. This result is somewhat difficult to interpret from a practical point of view: please provide some explanation.

Page(s) 9, Line(s) 7–ff.

Author(s). ... the d-dimensional Fisher copula can be expressed by...

Referee. In the definition of the Fisher copula, it is not clear what the ϵ's are. The mathematical notation used is confusing (as well as in the cited original paper by Favre et al., 2018). Are these variables/parameters continuous or discrete, as it seems (the braces notation $\{-1, +1\}^d$ does not help)? It is not clear whether they just take the values -1 and $+1$, or all values in the subset (open? closed?) $(-1, +1)$. Please clarify the issue. Furthermore, considering a practical perspective, what is the "role/contribution" of the ϵ's to the dependence structure? How do they affect the copula? Sorry, I am puzzled: a better explanation would help the reader.

Page(s) 10, Line(s) 2–3.

Author(s). Max-stable processes assume asymptotic dependence (i.e. dependence will not disappear at very large distances)...

Referee. This claim is not clear: asymptotic dependence looks like a mathematical (analytical) property: what is the notion of "distance" mentioned by the Authors? Is it a physical/geographical distance? Please make things clear.

Page(s) 10, Line(s) 12.

Author(s). To get values on the original scale, these values had to be back-transformed.

Referee. Practically, the Authors used a Probability Integral Transform procedure, isn't it?

Page(s) 10, Line(s) 15.

Author(s). The most suitable model was defined as the one that best reproduced the observed data.

Referee. This sentence is "void", since no criteria are specified to define the notion of "best reproduced". Please fix this claim.

Page(s) 10, Line(s) 16.

Author(s). ... no quantitative goodness of fit test is available to help in selecting the best model...

Referee. This sentence is (statistically) questionable: a GoF test can only be used to reject a distribution, surely it CANNOT / MUST NOT be used to select a distribution. Please fix this claim.

Page(s) 11, Line(s) 10–ff.

Usually, in hydrology, an elementary way to test a procedure for ungauged sites is to check it by using all gauged locations except one, whose records are known (but are supposed to be unknown). It is not clear whether the Authors followed this approach. If not, first the Authors should use this protocol, and exclude, one at a time, each one of the known stations, and check how, and to what extent, the model they propose is able to reproduce the known (but discarded) data.

Page(s) 13, Line(s) 1–2.

Author(s). The samples generated using the dependence models and back transformed using the regionalized GEV parameters showed a very similar picture to these regionalized marginal distributions.

Referee. Is it possible to provide a table of p-values? Visual statistics is intuitive, but some more "objective" results would be better.

Page(s) 13, Line(s) 4.

Author(s). Figure 6 shows the F-madogram of the observations and the different dependence models.

Referee. Figure 6 is confusing. I would suggest to use a 4x2 frame, and show all the plots of individual comparisons: this would make graphically clear the ability of each model to fit (or not) the data.

Page(s) 17, Figure(s) 8.

To the best of my knowledge, the Tail Dependence coefficient ranges from 0 to 1: why the colorbar ranges from -1 to $+1$? Just a software feature?

Page(s) 17, Line(s) 13–14.

Author(s). Similarly, the Gumbel copula was not able to model the dependence structure in the data despite its asymmetry.

Referee. The Authors should make the claim more precise. The Gumbel copula is symmetric, being Archimedean: the asymmetry concerns the tail dependence (only upper, not lower). Please fix the sentence.

Page(s) 19, Line(s) 15–16.

Author(s). Currently, to our knowledge, no copula model in more than three dimensions is available which models asymmetric lower and upper tail dependence.

Referee. Intuitively, this should be possible by using the Khoudraji-Liebscher copulas introduced by Durante and Salvadori (2010), but I did not check it.

Page(s) 20, Line(s) 5–6.

Author(s). However, more sophisticated regionalization techniques such as…

Referee. The Authors may also consider mentioning recent regionalization approaches entirely based on copulas, such as the ones outlined in Grimaldi et al. (2016) and Pappadà et al. (2018).

References

Durante, F., Salvadori, G., 2010. On the construction of multivariate extreme value models via copulas. Environmetrics 21, 143–161.

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

Favre, A.-C., El Adlouni, S., Perreault, L., Thiémonge, N., Bobée, B., 2004. Multivariate hydrological frequency analysis using copulas. Water Resources Research 40 (1).

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.

Grimaldi, S., Petroselli, A., Salvadori, G., De Michele, C., 2016. Catchment compatibility via copulas: A non-parametric study of the dependence structures of hydrological responses. Advances in Water Resources 90, 116 – 133, doi: 10.1016/j.advwatres.2016.02.003.

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.

Pappadà, R., Durante, F., Salvadori, G., De Michele, C., 2018. Clustering of concurrent flood risks via hazard scenarios. Spatial Statistics 23, 124–142, doi: 10.1016/j.spasta.2017.12.002.

Salvadori, G., De Michele, C., 2004. Frequency analysis via copulas: theoretical aspects and applications to hydrological events. Water Resour. Res. 40, W12511, doi: 10.1029/2004WR003133.

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 52 (5), 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.

Salvadori, G., Tomasicchio, G. R., D'Alessandro, F., 2014. Practical guidelines for multivariate analysis and design in coastal and off-shore engineering. Coastal Engineering 88, 1–14, doi: 10.1016/j.coastaleng.2014.01.011.