REVIEW REPORT

Journal: Natural Hazards and Earth System Sciences (NHESS)
Paper: nhess-2017-389
Title: A hazard model of subfreezing temperatures in the United Kingdom using vine copulas
Author(s): Symeon Koumoutsaris

GENERAL COMMENTS.

The paper is an interesting one, and outlines an original multivariate investigation concerning subfreezing temperatures. The comments posted by Referee 1 already provide an excellent, detailed review, with which I (almost) fully agree. Below, please find further notes: my objections should be read as constructive advices. Some relevant bibliography is reported at the end of this review.

1. I noticed that there is some confusion between the notions of probability distribution function and probability density function (e.g., Page 10, Lines 5–7: "The uncertainty intervals in the historical data are computed as the 5th and 95th quantile of the probability density function (Folland and Anderson, 2002)"). The probability distribution function is the integral of the probability density function (if it exists). The quantiles are the inverses of the probability distribution function (a non-decreasing one), not of the density function (which may not even be monotone). The Author must check the paper and fix all the points where such a confusion arises, otherwise the paper is not correct from a probabilistic point of view.

2. I was puzzled by the comment of Referee 1 concerning the sample size, and I ask the Author to clarify the issue: here, 170 variables are at play, each observed 51 times. To the best of my understanding, the idea revolving around Vine copulas is that any multivariate density can be decomposed into a (suitable, maybe not unique) product of univariate marginal densities and bivariate copula densities: in turn, only univariate and bivariate fits should be needed, isn't it? Thus, apparently, the fitting problem may not be so severe.

   Clearly, trying and fitting the upper tail of a GEV law using only 51 observations may be difficult (although the TWMLE escamotage is used), but it may not be impossible. Similarly, trying and fitting a bivariate copula using only 51 pairs may not be advisable, but it is not uncommon in practical applications. Overall, should my interpretation be correct, the game played by the Author may not be a "Mission Impossible", rather an "Uncertain Mission"...

   Thus, I kindly ask the Author to clearly explain the situation, and to provide estimates of the uncertainties as explained below.

3. I definitely agree with the comment of Referee 1 concerning the procedure to estimate the uncertainties (Page 9, Lines 23–ff.). As a rule of thumb, 1000 independent repetitions of the 10,000-years Monte Carlo simulations are usually suggested in literature, in order to provide "reasonable(?)" estimates of the confidence intervals of interest (clearly, it may be adjusted depending on the computational burden).

4. My main "perplexity" concerns a methodological issue. In this work, I can see the Mathematics/Statistics, but I do not see the Physics, which, instead, should be the starting point. To be clear, and to the best of my knowledge, the procedure used to construct the 170-dimensional copula finds its justification in an aggregation/clustering algorithm based solely on statistical considerations

(Page 9, Lines 13–14: "The method follows an automatic strategy of jointly searching for an appropriate R-Vine tree structure"). If I remember it correctly, the algorithm is based on the Kendall τ and/or on the Kendall Distribution Function $\mathbf{K}$, and/or, in general, on the strength of the statistical association between the variables at play. While interesting and meaningful from a mathematical point of view, such a procedure may eventually (statistically) associate grid cells having little, or negligible, physical link (for instance, could this be the case of the grid cells corresponding to Edinburgh and London, quite far apart from a spatial and a climatic point of view?)

In other words, important information like, e.g., the latitude (corresponding to different climatic regions) may not be considered/used by the statistical procedure adopted for constructing the overall copula. The Author is kindly asked to discuss the issue, and to provide suitable justifications. Is it possible to modify the construction of the 170-dimensional copula in order to take into account the physics of the phenomenon?

5. The Author has modeled the historical data, but, should the climate be really changing, then (at least from an Insurance point of view) the Author should account for it in his model, e.g. by introducing (in the long term simulations) suitable temporal patterns in the GEV/copula parameters according to available projections of the future climate (like, e.g., in IPCC scenarios). A comment is required on this issue.

6. In Section 3.1 "Results and discussion", the Author mentions the actual debate about climate changes (already commented by Referee 1). I would suggest to take a look at a recent paper by Vezzoli et al. (2017), where the traditional validation criteria of climate models are discussed, and an advanced/thorough distributional perspective is outlined: it may partially explain why several crucial hypotheses are "still largely under debate" (as claimed by the Author and Referee 1), and may partially account for the general inability to draw up clear settlements.

SPECIFIC COMMENTS.

Page(s) 2, Line(s) 23–ff.

For the benefit of unskilled readers and practitioners, here the Author should provide general references involving seminal books, papers, and guidelines concerning copulas, like writing: "For a theoretical introduction to copulas, see Nelsen (2006); Joe (2014); Durante and Sempi (2015); for a practical/engineering approach and guidelines, see Genest and Favre (2007); Salvadori and De Michele (2007); Salvadori et al. (2007, 2014, 2015)". Instead, citations concerning Vine copulas, being more specific and related to the modeling outlined in this work, may be postponed later.

Page(s) 9, Line(s) 20–22.

Author. "Goodness-of-fit is performed for the final selected R-Vine Model (RVM) based on the RVineGofTest algorithm of the same R package (Schepsmeier, 2013). The Cramer von Mises test, which compares the empirical copula with the RVM, has a value of 0.019 and a p.value = 1, which indicates that the fitted RVM cannot be rejected at a 5% significance level."

Referee. I am puzzled by such a large p-Value: in my opinion, it may entail a large probability of Type II error, i.e. accepting a False Null Assumption (this a typical performance of Cramer-von-Mises and similar tests, when the sample size is insufficient). The Author is kindly asked to discuss the issue, and to provide suitable justifications.

References

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.

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., 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.

Vezzoli, R., Salvadori, G., De Michele, C., 2017. A distributional multivariate approach for assessing performance of climate-hydrology models. Scientific Reports 7:12071, doi: 10.1038/s41598-017-12343-1. URL www.nature.com/scientificreports