

Interactive comment on “Copulas for hydroclimatic applications – A practical note on common misconceptions and pitfalls” by Faranak Tootoonchi et al.

Francesco Serinaldi

francesco.serinaldi@ncl.ac.uk

Received and published: 28 September 2020

In agreement with RC1, I also "get the feeling that the authors are not very familiar with the basic statistical copula literature". However, I also get the feeling that the Authors seem to be not even familiar with some basic statistical concepts as well as literature dealing with applications of copulas to hydrological variables. In this respect, it is quite ironic or paradoxical that a paper discussing "misconceptions" endorses and re-proposes "misconceptions"! I also think that this type of papers should be written/supervised by people with more experience in the field; I mean names like Favre, Genest, Salvadori, De Michele, Bardossy, and some others... almost certainly, this is

C1

not a task for people with limited experience, in my opinion.

As I recognize that the above statements can appear harsh, please, let me discuss only few points to support my opinion.

L70: Before quoting a paper, it is better to be sure about its content. For example, the Authors state "At the annual resolution AghaKouchak et al. (2014) found a significant negative correlation between anomalies of temperature and precipitation over California". However, those data show zero Kendall correlation, as can be seen by reading a bit more carefully the cited Serinaldi (2016), who re-analyzed the same data, and showed that the sample size is not enough to make conclusions on the actual dependence structure.

L88: "Since the early 2000's, copula methods have been adopted in hydrological modeling, which was triggered by the study of Salvadori and De Michele (2010)." Another discussant suggested Favre et al. (2004) and Salvadori and De Michele (2004) as the first applications of copulas in hydrological domain. Well, the first paper applying copulas in the hydrological context is De Michele and Salvadori (2003), whose 15th anniversary was also celebrated by a special issue in Water journal (https://www.mdpi.com/journal/water/special_issues/copulas_hydrology). Knowing history can help.

L103: "only a few papers also relate to practical challenges encountered in hydroclimatic research"... Often, my papers have been criticized as they are too critical when discussing the problems/challenges of performing inference (including copula inference) on hydroclimatic data; so, it is quite funny to discover that people have criticized me for nothing.

L109: "Accordingly, this paper aims at filling this gap and serves as an overview of the state of the art of using copulas in hydroclimatology for practitioners interested in adopting this method for their research." As discussed below, this paper is far from filling any gap, and surely does not report the state of the art in any respect.

C2

L140: "The result would be the joint CDF of uniformly distributed marginal"?? Perhaps, "the joint CDF of uniformly distributed random variables". Before talking about misconceptions, it can be good to familiarize with the nomenclature and the meaning of the words used.

L145: this simulation procedure yields fitted joint distributions characterized by different (theoretical) Pearson correlations. The correct procedure to show the difference of the tail behavior for identical marginals and Pearson/Kendall correlations is to simulate samples from distributions with specified copula, marginals and Pearson/Kendall correlations. However, since the Pearson correlation depends on both copula and marginals (as stressed by another discussant), and generating samples with the same (theoretical) Pearson correlation can be a bit tricky, my feeling is that the Authors opted for this shortcut, which is however theoretically incorrect and numerically imprecise.

Section 2.3: The Authors merge the expression of the population Pearson correlation with the sampling estimators of the Kendall and Spearman correlations. This generates not only confusion but also a subsequent mistake concerning how to account for zeros and more generally for statistical ties (see below). When presenting dependence measures, it can be better to discuss both theoretical expressions and finite sample estimators, just to avoid "misconceptions".

L168: "Aloui, Ben Aïssa, and Nguyen (2013) distinguished between three different types of Archimedean copulas" I'm sure there are better references for this, and "distinguished" seems to me inappropriate (perhaps "discussed" or "focused on").

L179: "After estimating θ based on Kendall's τ (Table 1), the copulas can be computed using their respective formulas." For a paper that should discuss the "state of the art", only mentioning the moment-like estimation based on Kendall τ , which is not even applicable to multi-parameter bivariate copulas, seems to me insufficient. What about max-likelihood (with all its flavors: exact ML, max pseudo-like, etc.), max-entropy, Bayesian inference, etc.?

C3

Sec. 2.3.2: please use Φ_R to denote an elliptical multivariate distribution (as per Fig. 3).

L203: "Embrechts, Mcneil, and Strauman (1999) estimated the Elliptical copula parameter from the rank-based Kendall's τ or Spearman's S ." this is valid only for elliptical copulas, such as Gaussian, that are only characterized by pairwise correlation matrices. For families such as the Student copula, we need to estimate additional parameters (e.g. the degrees of freedom), and therefore other estimators are required (if we want to report "the state of the art").

L237: "The null hypothesis of the test is the acceptance of the parametric copula. For a particular copula, the p-value is sufficient to determine the acceptance or rejection of the null hypothesis with the significance level of α , but in a group of different acceptable copulas, the best alternative is the one with the smallest S_n or D_n and the greatest p-value (Madadgar and Moradkhani, 2014)." This sentence is incorrect. Leaving aside the meaning (or lack of meaning) of the hybrid of Fisher-Neyman-Pearson hypothesis testing framework (see Wasserstein et al. (2019) and references therein along with ASA recommendations), goodness-of-fit tests are never confirmatory; they can only conclude that "the null cannot be rejected", meaning that there is not enough information to exclude the null hypothesis. The last part of the of Authors' statement describes another widespread misconception. Indeed, test statistics and p-values cannot be used to rank the models. Why? Because the model parameters are estimated on the data, and therefore the KS and CvM tests are no longer distribution free, and then those p-values correspond to different quantiles of different distributions of the test statistics. This is the reason why the null distribution of GRB goodness-of-fit test is computed by MC simulation, which must be performed every time we consider a different data set. Instead of reporting suggestions described by hydrologists "playing" with statistics, I suggest reading statistically written by statisticians such as D'Agostino and Stephens (1986)... or hydrologically oriented but theoretically grounded papers such as Laio (2004), for a gentle discussion.

C4

L242: Using an ensemble of performance measures makes sense if these measures highlight different "fitting" aspects. However, these indexes are often selected quite randomly, as in this case, overlooking their redundancy; indeed, Nash-Sutcliffe is nothing but a similarity measure corresponding to the (R)MSE (see e.g. Hyndman and Koehler (2006), Jachner et al. (2007), Dawson et al. (2007), and Reusser et al. (2009) for a discussion on a more appropriate use of performance metrics). Moreover, selecting the model with the lowest/better metrics is also questionable especially in the case of (usually) small samples, as these metrics are affected by uncertainty, and the model rank can change by changing the sample. This problem is well known in the field of "information criteria" (AIC, BIC, etc.), where the selection either relies on the significance of the differences between two models (in terms of model evidence) or it is somewhat avoided by using model averaging.

Sec.3 and Fig. 8: In my opinion, most of the supposed pitfalls and misconceptions listed in Sec. 3 and summarized in the flow chart in Fig. 8 result from some Authors' misconceptions or rather superficial approach to the topic. Firstly, copulas are general models that can be used for data at any spatio-temporal scale. Even if data at some scales can generally be more or less serially correlated, for instance, this does not prevent the use of copulas, taking for granted that serial correlation should be accounted for in some way. However, the problem should be considered case-by-case rather than ascribed to spatio-temporal scales. Moreover, serial correlation or other properties depend on the variables at hand. Furthermore, checking for the significance of the cross-correlation of two (or more) variables is a false problem. Indeed, we use copulas to build joint distributions, thus meaning that we need the joint distributions, independently of the correlation value. When the correlation is close to zero, this simply means that the product copula is a feasible option, and this copula is often a special case of other copulas. As mentioned by another discussant, it seems to me that the Authors confuse correlation and dependence structure. Zero correlation does not mean that the joint probability is zero or does not exist, or it is not of interest; it means that the joint probability can reasonably be described by the product of marginals or, the same,

C5

by the product copula. In this respect, the sentence "...there are months in which the correlation is not significant at the 5% level, thus cannot be considered as reliable correlation between variables." makes little sense. Zero correlation is what is, it is not less/more "reliable" than 0.2, 0.5, or 0.9. Reliability depends at most on the sample size used to estimate the correlation values. We can use copulas for every month in Fig. 5 if we need the joint distribution of every month. Selection of June and July is artificial, not required, and not justified neither empirically nor theoretically.

Sec. 3.3: Another discussants already made comments on the stationarity issue. I would only like to stress that the sentence "when the correlation is highly sensitive to the selected time period, it is an indication of a non-stationary behaviour" makes little sense. Non-stationarity requires a (known) law of evolution, while sub-sample fluctuations do not indicate any non-stationarity per se. The Vattholma example and corresponding numerical experiment and interpretation is also meaningless. Indeed, that bootstrap experiment does not (and cannot) reveal non-stationarity; it simply shows the sampling variability for 10-year samples under stationarity! Why? Because the bootstrap experiment (selection of ten randomly chosen years) is designed to destroy whatever supposed time evolution. Fig. 5 only shows the seasonality of the correlation and its sampling uncertainty for 10-year spanning samples.

Sec. 3.4 has missed the key works of Andrew Patton providing the theoretical basis and conditions required to apply the so-called conditional or dynamic copulas, i.e. the models applied in the cited references. The sentence "In contrast, other papers argued that copulas are still applicable in case of only low-degree (removable) auto-correlation in the time series" denotes once again some lack of familiarity with the existing literature. Indeed, these models are widely applied in econometrics, involving complex and often strongly persistent processes (see also Serinaldi and Kilsby (2017) for an example on hydrological data). Relying exclusively on hydrological literature to get information about statistical tools is never a good idea; based on my experience, hydrology is one of the disciplines with the most superficial (amateur) use of statistics.

C6

Sec. 3.5: "Correlation for data with the same rank (dry days)". Data with the same rank are called statistical ties, and unlike stated in this section they are not a big problem and do not require any rainfall threshold, post-processing or such. What is needed is only a decent literature review revealing that (i) estimators of e.g. Kendall correlation accounting for ties and zero-inflation already exist and do not require any pre/post processing (check Kendall, Gibbons and others' works), and (ii) joint distributions of two variables with discrete-continuous and continuous marginals, respectively, are special cases of the models described for instance by Shimizu (1993) and Herr and Krzysztofowicz (2005), and formalized in terms of copulas by Serinaldi (2009). As mentioned above, distinguishing between population and sampling version of a given statistic can help avoiding mistakes.

L379: "Since a significant correlation is the basis for applying copulas". I do not think so: the need for joint distributions with specified dependence structure and marginals is the basis for applying copula. If such a copula is the product copula in some cases, it does not matter very much.

L389: "We here demonstrate how such performance measures can be applied practically"; as mentioned above, Sec. 3.6 demonstrates the opposite, i.e. how not to apply these indexes.

- S_n or D_n are associated to goodness-of-fit tests that have only dichotomous outcome (rejection or not rejection);
- they (and p-values) cannot be used to rank the models (as explained above);
- the range of the four indexes are different and there is no empirical or theoretical support for the statement "Here, evaluation of copulas by E_{RMS} or E_{NS} did not reveal the weak performance of Clayton copula".

Based on Table 2 and considering the similarity of Gaussian and Student copulas (as p-values are not reported), the interpretation should be as follows:

C7

- CvM and KS tests probably say that those two models cannot be rejected, while Clayton can be (p-values should be shown to confirm this);
- E_{RMS} or E_{NS} do not allow any conclusion without complementing them with uncertainty assessment, which quantify the sampling fluctuations of these metrics (and the significance of their differences).

In my opinion, this paper does not provide "an overview of the state of the art of using copulas in hydroclimatology", but something like the opposite. Based on a very superficial literature review (which neglects theoretical literature), and an apparent lack of familiarity with the topic, this manuscript is also quite superficial itself, iterates some misuses of statistical tool (which are widespread in the hydrological literature), and does not provide a good service to a community that already suffers from confusion when coming to applied statistics.

Based on my experience, most of the misconceptions concerning copulas, and more generally applied statistics, are related to hydrologists' statistical background, which is on average much more limited than that of people working on other fields, such as economics, biology, medicine, etc., where statistical analysis is routinely performed by (or with the help of) professional statisticians, or people with much more solid statistical background. I also think that the sampling and model uncertainty mentioned by the Authors in their conclusions is one of the most important aspects to draw meaningful conclusions. In most of the submitted or published literature on these topics, data are not enough to draw any definite conclusion, and statistics is somewhat (ab)used to overcome this lack of information, with the incorrect hope that it can give that certainty (or rigor) that the data and meta-data cannot provide. Concerning the uncertainty in copula inference it could be fair mentioning the contributions of Serinaldi (2013), Dung et al. (2015), and Zhang et al. (2015). Once uncertainty is accounted for, copula inference often reveals that the discrimination among different models and/or preliminary assumptions is very difficult, if not impossible without additional information.

C8

My final (not requested) suggestion for the Authors is to reconsider their approach to this topic, going beyond the concepts that can be deduced from few tens of hydrological papers, which may be not even the best ones in terms of quality. As mentioned above, such a kind of papers require a wider/deeper familiarity with theory, applications, literature (from different fields), etc.

Sincerely

Francesco Serinaldi

References

D'Agostino RB, Stephens MA (1986). Goodness-of-fit-techniques (Vol. 68). CRC press
Dawson, C. W., Abrahart, R. J., See, L. M. (2007). HydroTest: a web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. *Environmental Modelling Software*, 22(7), 1034-1052

De Michele, C., and G. Salvadori (2003) A Generalized Pareto intensity-duration model of storm rainfall exploiting 2-Copulas, *J. Geophys. Res.*, 108(D2), 4067, doi:10.1029/2002JD002534

Dung NV, 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

Herr HD, Krzysztofowicz R (2005) Generic probability distribution of rainfall in space: the bivariate model. *J Hydrol* 306:234–263

Hyndman, R. J., Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4), 679-688

Jachner, S., Van den Boogaart, G., Petzoldt, T. (2007). Statistical methods for the qualitative assessment of dynamic models with time delay (R Package qualV). *Journal C9*

of *Statistical Software*, 22(8), 1-30

Laio, F. (2004), Cramer-von Mises and Anderson-Darling goodness of fit tests for extreme value distributions with unknown parameters, *Water Resour. Res.*, 40, W09308, doi:10.1029/2004WR003204.

Patton A.J. (2009) Copula-Based Models for Financial Time Series. In: Mikosch T., Kreiß JP., Davis R., Andersen T. (eds) *Handbook of Financial Time Series*. Springer, Berlin, Heidelberg

Reusser, D. E., Blume, T., Schaepli, B., Zehe, E. (2009). Analysing the temporal dynamics of model performance for hydrological models. *Hydrology and earth system sciences*, 13(7), 999-1018

Serinaldi F. (2009) Copula-based mixed models for bivariate rainfall data: an empirical study in regression perspective. *Stochastic Environmental Research and Risk Assessment*, 23(5), 677–693.

Serinaldi F (2013) An uncertain journey around the tails of multivariate hydrological distributions. *Water Resour Res* 49(10):6527–6547

Serinaldi F, Kilsby CG (2017) A blueprint for full collective flood risk estimation: demonstration for European river flooding. *Risk Analysis*, 37(10), 1958-1976

Shimizu K (1993) A bivariate mixed lognormal distribution with an analysis of rainfall data. *J Appl Meteorol* 32:161–171

Zhang Q, Xiao M, Singh VP (2015) Uncertainty evaluation of copula analysis of hydrological droughts in the East River basin, China. *Global Planet Change* 129:1–9

Wasserstein RL, Schirm AL, Lazar NA (2019) Moving to a World Beyond “ $p < 0.05$ ”, *The American Statistician*, 73:sup1, 1-19, DOI: 10.1080/00031305.2019.1583913

Interactive comment on *Hydrol. Earth Syst. Sci. Discuss.*, <https://doi.org/10.5194/hess-2020->

C11