Earth Syst. Dynam. Discuss., referee comment RC2
https://doi.org/10.5194/esd-2021-5-RC2, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on esd-2021-5

Anonymous Referee #2

Referee comment on "Space–time dependence of compound hot–dry events in the United States: assessment using a multi-site multi-variable weather generator" by Manuela I. Brunner et al., Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2021-5-RC2, 2021

Summary:

In this paper, the authors introduce a multi-site multi-variable stochastic weather generator called "PRSim.weather" to assess the (joint) occurrence probabilities, severity, and spatial patterns of compound hot-dry events in the US at various time scales (1 week, 1 month, 3 months, 6 months, 1 year). The proposed weather generator is a simple extension of a previously published version for a single variable, and they here make some necessary adjustments for its application to study high temperatures / low precipitation. The authors conclude that their model correctly replicates the distribution and dependencies in observed data, and their analysis further reveals that

(1) Northwestern/Southeastern US are more likely to experience hot-dry events

(2) the time scale influences the size of compound hot-dry events (i.e., shorter time scales imply larger spatial extents of joint extreme events)

(3) temperature mostly determines compound events for short time scales, while precipitation is the key factor for longer time scales.

Assessment:

Overall I like the paper and the data analysis. The topic tackled by the authors, namely to understand the spatio-temporal distribution of compound extreme weather events, is difficult and timely. The paper is well written, is relatively concise and the authors precisely detail the findings of their analysis. The proposed approach (PRSim.weather) has, however, some limitations that the authors should, I think, better acknowledge and discuss more openly. I discuss some of those in my comments below. Another point to mention is that although the data analysis and the findings are well supported and of practical interest, the methodological novelty is rather limited, since the proposed method is a simple adjustment of an already published approach.

Comments:

1. I found the 5 steps of the method on page 4 (lines 90-109) difficult to understand. For example,

- how do you "fit monthly distributions to T and P"? do you first a distribution to the data within each month separately assuming that they are iid during that month?

- what do you mean by "we combine the E-GPD with as many zero-values as in the observations"? Do you mean that you don't simulate zero observations, but keep them fixed like in the data? If so, is this not "cheating" (i.e., over-fitting)? and do you keep the zeros at the same time points?

- how do you apply the continuous wavelet transform? and how to interpret the amplitude and phase signals?

- in point 4., what do you generate a random time series for both T and P? Or just one time series?

- in point 5., how to you do the "rank-transform" exactly? Do you mean that you apply the probability integral transform?

Bottomline: I think it is needed to clarify the methodology. It seems necessary to me to provide further mathematical equations to clarify each point and to illustrate the wavelet transform with a simple example in order to faciliate interpretation.

2. The methodology seems to have certain limitations that may be concerning:

- The authors mention that the same random phases are used at all sites and for both variables. Is this not too restrictive, and will this not create too strong spatial or cross-dependencies?

- In point 4., a time series of one site is chosen at random. Are all sites "exhangeable"? What is the implication of this approach?

- Again in point 4., a random time series is generated by bootstrap by resampling years with replacement. This implies that years are exchangeable and therefore that any time trend is ignored. Is this not a major issue for temperatures (and perhaps also precipitation)? If so, this should be further acknowledged and discussed.

- Using a bootstrap-based approach implies implicitely that simulated events will NEVER be more extreme than what has been observed in the data. This is a major limitation since the goal here is to enrich the dataset with more simulations of compound extreme events.

- Estimating a copula using the empirical copula (based on ranks) implicitely implies that the data are stationary over time, thus without time trend (or seasonality) again. Is this a reasonable assumption here?

3. L129, p5, "site-specific Gamma distribution": should this not be the E-GPD distribution as specified in the methods section (point 2.)?

4. p6, top: further details on copulas are required to introduce the notation properly...

- What is a copula => Joint distribution with uniform Unif(0,1) margins

- What is C(u,v)? => the copula of T and -P

- What are the ranks R_i and S_i? => ranks of T or P values across the time series

- In Figure 2, what does "Empirical copula" mean? => the values of C_n(R_i/(n+1),S_i/(n+1)), i.e., the empirical copula evaluated at the observed uniform values.

5. In Figure 3, the results are almost too good to be true in my opinion. Does this not hide some issues of overfitting? Again, how do you simulate the zeros in precipitation for example?

6. When the goal is to simulate many more compound events, it is crucial to check if the marginal and joint tails are captured correctly. For marginal tails, I would suggest to consider comparing long-term return levels of simulated vs observed data (on a scale that zoom into the tail rather than the bulk). For joint tails, a possibility is to look at the tail correlation coefficient (\lambda(u) = P(U1>u | U2>u)) for increasing thresholds u=0.8,0.9,0.95,0.98,0.99,0.995,0.999, say. Such diagnostics would complement the results in Figure 3.

7. In Figure 4, the simulated fields appear smoother than observations. Why is that the case?

8. In Figure 5, it seems like the spatial extent of very extreme events is largely overestimated. Is this because a single random phase is chosen across sites? Or is this a false impression due to the fact that there are less extreme events available in

observations than simulations?

9. In Figure 6, simulations severley underestimate the joint probability of concurrent events for severe and extreme events... and also for moderate events in the Southeastern part of the US... Is this due to using the empirical copula approach? What is the cause of this and how to remediate this (fairly severe) issue?

10. Figure 8 plots the "median spatial extent of concurrent events affecting grid cell". How was that calculated? I don't think it is clearly explained in the text...

11. Figure 10 reports the values of Kendall's tau between T and the bivariate empirical copula, as well as between P and the bivariate empirical copula. However, given that the empirical copula is itself calculated from T and P, I'm not convinced that such "correlation" values make sense... Wouldn't it make more sense to report the actual ranks R_i/(n+1) and S_i/(n+1), which already give the importance of T and P in the calculation of the empirical copula?