Review of the manuscript “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, Eric Gilleland, and Andrew W. Wood.

General comment

The authors introduce a multi-site multi-variable weather generator (*PRSim.weather*), which allows for simulating Temperature and Precipitation over the US during 100*28 years. While the weather generator has some limitations (that the authors discuss), the output is overall satisfying. The simulated data allows for analysing both (1) events that tend to be characterised by hot and dry conditions and (2) the spatial extents of these events. The authors illustrate and discuss the characteristics of these events across the US. Some improvements is needed, especially in the presentation of some methodological aspect (selection of concurrent hot and dry events and method for analysis in Figure 10). Generally, the paper is definitely well structured and I found it interesting.

I recommend the authors to consider my specific comments below. Those marked with *** are the less technical.

Specific comments

L9 **meteorological** drought indicators

L25. Could you mention, very *briefly* as it is an introduction, what were the causes for the changes in hot and dry events in these studies, e.g., temperature/precipitation trends?

L30 I suggest re-shaping the sentence slightly. That is, including the words “local” and “regional” (or “aggregated over a region”). The local impact depends on frequency and duration. The aggregated regional impacts depend, in addition, also on the extent.

*** L40. I think that the terminology could be improved, not only here, despite it is not wrong as there is no full agreement on this matter in the community. For example, here: “(2) spatial extents of compound events”.

- You use “compound event” to refer to concurrent events or multivariate events (such as hot-dry events), which is a type of compound event.

- Note that the spatial characteristics of an event, make the event compound on its own (Zscheischler et al., 2020).

Therefore, the considered events are compound for two reasons, however you refer to the multivariate characteristic as a compound element, but you do not do the same for the spatial part. Of course, you cannot say twice “compound”, but why is one privileged? Talking of “spatial extents of concurrent hot and dry events” may make things better in the paper. This would lead to reshaping a bit, for example, lines 40-45.

L40. Spatial patterns and spatial extents. Please, make the difference clear. I know what you mean, but I suspect that it will not be obvious to everyone.

L45. This statement is interesting. We have recently worked on the topic and shown that it is very

difficult to study seasonal precipitation extreme extents without large ensemble simulations (discussed at the end of the "Present-day spatial scale extremes" section):

Bevacqua, E., Shepherd, T.G., Watson, P.A.G., Sparrow, S., Wallom, D., and Mitchell, D. (2020). "Larger spatial footprint of wintertime total precipitation extremes in a warmer climate". Submitted. Preprint's DOI: 10.1002/essoar.10505310.1

*** L45 You write: "Studying such spatial compound events is challenging because they are rare in observational records (Zscheischler et al., 2018). This challenge can be tackled by developing stochastic simulation approaches to generate large data sets with similar statistical properties as the observations "

The weather generator is calibrated on and learn from the limited observation (or available data). So, does using a weather generator address completely the challenge of limited data? I suggest discussing this, especially the limitations, for a non-expert reader.

L76 add "daily" to "time series".

L83 Also in the procedure. You simulate, in the end, daily time series of P and T. Could you state this explicitly somewhere, maybe simply adding a "daily" somewhere?

Caption Fig 1. Add "daily" and "monthly" where required. E.g., in Step (2), I suggest moving the "monthly": "fit SEP distribution to T and E-GP distribution to P **monthly** time series of all sites"

L119 Adding "aggregated" somewhere may help to make very clear that you will pull together all the weather generator output in a unique aggregated time series of 2800 years (one may in principle repeat the analysis on the 100 weather generator output and get, e.g., a mean).

Fig 2. What time scale are you using here for computing the indices? Please, specify.

*** L133 "in events where both STI and -SPI are jointly exceeded." Not clear what is "jointly exceeded", though this is described rigorously later. At this point, I tended to expect a method that would catch events where STI and -SPI high values are jointly exceeded (e.g., concurrent values above the 99.5th percentile). In fact, the authors also refer to "The highest probability of concurrent hot-dry events" at line 172 and later in the paper, when discussing the results based on the copula-related metric. Is there any particular reason for opting for this particular copula-based threshold criterion?

Selecting (u,v) pairs such that C(u,v)> threshold implies to pick up values of (u,v) which are beyond the "threshold curve" defined by C(u,v)=threshold. Depending on the dependence between -SPI and STI (which depends on the location), the "threshold curve" in the [0,1]x[0,1] space will be different (also the number of selected events will depend on the dependence, which is not something to criticise). Hence, one may wonder whether this leads to comparing events at different locations that are different in nature. Hence, whether using concurrent extreme would not lead a more natural interpretation of the results.

I would appreciate a brief discussion that considers the above, such to provide some insights to the reader.

Hence, in the next, could you find and use a different term than "concurrent hot and dry events"?

L140, do you mean? "For any given time scale, we define the spatial extent of the compound event as the percentage of grid cells affected by the compound event. "

*** L 143-145. This is not fully clear. E.g., “median” among which sample? Therefore I had issues in understanding the results on this topic fully. Please, clarify.

L131-140, Please use the same term when you refer to the same concept to avoid misunderstanding. I got that with “compound hot-dry events”, “extreme droughts”, and “compound events” you are referring to the same thing in these lines.

Figure 3,
- I assume that the different simulated lines correspond to the 100 simulated samples. Please specify in the caption.
- In b and d, precipitation appear to behave a bit differently from observations. However, this may just be a result of higher variability of the precipitation, compared to temperature. Hence, if there were confidence interval around observations, one may find that both T and P behave similarly in term of overlapping the confidence interval. Please, consider the following: Would it be possible to add some confidence interval of the observation estimates? For the autocorrelation function, adding a line highlighting the level of significant correlation may help.
- Panel e-f should have the same axis to facilitate the comparison.

If the above lead to some changes in the interpretation of the graphs/evaluation, then this should be mentioned in the text. However, overall, given that the aim is to discuss the model performance, I do not think that the text should be too much related to the specific performance at an individual grid point. Rather, try to summarise the characteristics of the model at most grid points (as I guess you did already via selecting a representative grid point).

You could add a few words to the last sentence (“The model is considered suitable for the analysis of compound hot-dry events because it has an acceptable performance with respect to all three aspects.”) such to highlight that, despite there are some limitations, your model do at least offer a way to tackle the challenging study of such a compound event.

Figure 4. Missing a full stop before “(b)”. You could add, probably in the caption, a brief reference to the variogram, e.g., that describes the degree of spatial dependence of a field (add reference).

L164. I would divide the first sentence in two sentences. The second sentence should highlight (as you already imply) that the maps allow for evaluating the spatial pattern of the indices, rather the magnitude (I guess that the index is computed on observed and simulated sample independently so they provide information the anomalies relative to the climatology in observations, and simulations, respectively).

L176, Figure 6. The author should mention that the simulations tend to underestimate compound hot and dry events. This seems in line with what discussed at line 158 (on the dependence between P and T).

L179 This is an interesting result.
“However, the probability of concurrent events decreases with increasing time scale, as can be expected due to the increasing aggregation of multiple weather events in longer periods…”

I understand that the point is that multiple weather events, not all of which favouring instantaneous concurrent hot and dry conditions, are pulled together at long time scales. Hence, the overall dependence is influenced by a combination of weather events, some of which causing and others not causing dependence. As a result, the dependence is weakened compared to the short term case where dependence-driving weather system are considered individually. You may explain this more explicitly, if you agree with me.

L196, do you mean “importance of T and P”?
*** L196-200 is not clear, see my comment above on the methodology. Please, improve this.

*** L216, This is a finding that could have been found also based on observations only. I am wondering whether the authors could highlight in the discussion the features that the weather generator (e.g., longer time scale) allowed, in this analysis, to understand better than based on observations only.

L219, is this reasoning apply also for the yearly time scale?

L228, consider adding some references.