## **Reviewer 1**

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.

**Reply:** We thank the reviewer for their detailed and thoughtful comments, which particularly helped to improve the presentation of some methodological aspects. Please find our responses to the individual comments below.

## **Specific comments**

## L9 meteorological drought indicators

**Reply:** We specified that we are referring to 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? **Reply:** Most of the studies that look at the drivers of changes indicate that increasing temperatures can at least partly explain the changes in hot-dry events. We specify that 'Substantial increases in the number of concurrent droughts and heat waves over the last few decades that are partly explained by increasing temperatures have been reported not just for the US...'

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.

**Reply:** We integrated the terms local and regional by writing: 'While frequency of occurrence is an important factor determining local and regional impacts, the severity of impacts related to compound events likely also depends on their spatial extent, i.e. how large the affected region is, and their time scale, i.e. whether they just last weeks or extend over a longer period of time.'

\*\*\* 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.

**Reply:** Thank you for this note. We are aware that the spatial characteristics of an event make it compound on its own. However, the word compound itself is unspecific as it does not tell us anything about where the 'compoundedness' comes from. In our attempt to specify the nature of compoundedness, we came up with the term 'spatial compound'. We agree though that this term is still not specific enough. We therefore either replaced the term 'spatial compound' by the more specific term 'spatial multivariate' or by the more general term 'compound'. Modifications were applied throughout the document.

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.

**Reply:** We added the following specifications: '(1) spatial patterns of compound event affectedness (i.e. where in the US hot-dry events are most frequent), (2) spatial extents of compound events (i.e. how large compound events are).'

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* 

**Reply:** We agree that using large ensemble simulations would be an alternative to using stochastic models. We therefore slightly adjusted the sentence to: 'This challenge can for example be tackled...'. In the discussion section, we add that: 'If physical consistency is a requirement for a specific application, stochastic approaches may be combined with physical approaches as e.g. in the weather generator AWE-GEN-2 by Peleg et al (2017) or one may rely on large climate ensemble simulation approaches (Deser et al. 2020; Bevaqua et al, 2020).

\*\*\* 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.

**Reply:** Yes, as every other calibrated/fitted model, the weather generator is based on the available observations. We specified in the Methods section that: 'The simulation of yet unobserved magnitudes becomes possible thanks to the use of parametric distributions for T and P in Step 2.' We added the following point to the discussion: 'Please note that even though the model generates

yet unobserved observations, the simulations are not independent of the limited sample size used to fit the model because the model is data-driven as any other calibrated/fitted model.'

L76 add "daily" to "time series". **Reply:** *We added 'daily'.* 

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? **Reply:** *We specified that PRSim.wave simulates compound hot-dry events 'at a daily scale'.* 

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" **Reply:** We added 'daily' to clarify that both input and output are at a daily scale and we clarified that the SEP and E-GP distributions were fitted 'at a monthly scale'.

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). **Reply:** Thank you for this suggestion. We specified that: '…increase the sample size available for the assessment of compound hot-dry events by pooling the different model runs.'

Fig 2. What time scale are you using here for computing the indices? Please, specify. **Reply:** *We specified that Figure 2 shows 'monthly' time series.* 

\*\*\* 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"?

**Reply:** Thank you for highlighting that the definition of compound hot-dry events needed further explanation. We applied the threshold to the bivariate distribution of T and P instead of their marginal distributions, which is one potential way of identifying bivariate extremes. A joint definition where both T and P have to exceed a marginal threshold will extract only events in the upper right corner of the distribution. Using a threshold on the bivariate distribution also includes these upper right corner events and adds a few events which are also critical but only extreme with

respect to one of the margins. A nice illustration of the different probability spaces we are talking about is given in Figure 1 of Serinaldi et al. (2015)

(https://link.springer.com/article/10.1007/s00477-014-0916-1). We agree that talking about 'joint exceedances' and about 'concurrence' is confusing in this context and therefore replaced the terms by more appropriate ones (concurrence = compound). We also added the following specification to the methods section: 'This copula-based threshold procedure slightly differs from an approach where both margins (-SPI and STI) have to jointly exceed a threshold in order for an event to be defined as a compound event. The bivariate threshold procedure includes a slightly bigger event space, which besides the jointly marginally extreme events also includes those events that are extreme in terms of the bivariate distribution but not necessarily in terms of both margins.'

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. "

**Reply:** Thank you for this rephrasing suggestion, which we integrated as: 'To assess the spatial extent of compound events at different time scales, we define the spatial extent of the compound event as the percentage of grid cells affected by the compound event at any given time scale.'

\*\*\* 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.

**Reply:** Thank you for pointing out the need for clarification. We specified that: 'we compute Kendall's correlation between the median bivariate distribution (empirical copula) and the median standardized indices STI and SPI over all simulation runs at different time scales.' We further clarify that 'This correlation analysis is performed for nine hydro-climatic regions in the United States to quantify the regional spread in the role of STI and SPI for compound event development, i.e., correlation is computed between median bivariate distributions and median STI or SPI at different grid cells within a region.'

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.

**Reply:** The term 'extreme droughts' was indeed confusing. We replaced it with 'compound event' to use consistent terminology.

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.

**Reply:** We specified that the different lines in a-d refer to the different simulation runs. Adding confidence intervals to the observations would be possible by applying some bootstrap procedure. However, we think that the observed and simulated acfs appear to be similar enough not necessarily requiring this additional uncertainty information. Panels e and f are displayed on the same scale, which is indeed important to facilitate comparison. However, this might not have been sufficiently clear because there are two separate y-axes, one for T and one for P. We therefore specified in the caption that the left y-axis refers to precipitation and the right axis to temperature. We here present a model evaluation for the local T and P characteristics for one single grid cell, which was not sufficiently clear in the previous version of the manuscript. We specified that local model performance was assessed on an example grid cell and highlight that 'The above-described model evaluation can be generalized to other grid cells in the data set.' We also slightly adjusted the last sentence of the paragraph to highlight the main benefit of the stochastic simulation approach, i.e. increasing the available sample size of compound events: '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 and enables increasing the sample size of compound events.'

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). **Reply:** We added an 'and' to separate the descriptions of subpanels (a) and (b). We also added a short description of a variogram including a suitable reference: 'which describe the degree of spatial dependence of a field (Cressie, 1993).'

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).

**Reply:** We split up the sentence into two sentences with a second sentence stating: 'These spatial samples enable comparing observed and simulated STI and SPI patterns for different levels of extremeness'. Yes, the indices are computed for observations and simulations independently.

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).

**Reply:** We specified that: 'The spatial STI and SPI patterns are reflected in the spatial distribution of the probability of compound hot-dry events, which is also realistically represented but slightly underestimated by PRSim.weather.'

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.

**Reply:** Yes, we definitely agree. We slightly expanded the sentence by writing: 'However, the probability of compound events decreases with increasing time scale, as can be expected due to the aggregation over increasingly longer periods of multiple weather events that may not all favor instantaneous compound hot and dry conditions, and event extremeness.'

L196, do you mean "importance of T and P"? **Reply:** *Yes, we changed the sentence to 'importance of T (STI) and P (SPI)'.* 

\*\*\* L196-200 is not clear, see my comment above on the methodology. Please, improve this. **Reply:** In addition to the changes already applied to the methods section, we provided more guidance for reading Figure 10 in the results section: 'T is a particularly important driver at short time scales as indicated by the high correlation between median STI and median bivariate distribution of grid cells within a specific hydro-climatic region' and adjust the Figure caption: 'Importance of T and P as drivers of compound events across time scales and extremeness levels. Correlation of median bivariate distribution (empirical copula) with (a) STI and (b) SPI across simulation runs between grid cells in nine hydro-climatic (Bukovsky) regions (spread of boxplot) per time scale (color) and level of extremeness (hue).'

\*\*\* 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.

**Reply:** It is true that spatial patterns of compound hot-dry events could also have been studied using observations only. However, studying rare spatial events with large extents would have been difficult/impossible. We added the following point to the discussion section: 'It [the stochastic model] enables studying rare spatial multivariate events, which would not be possible using observations only.'

L219, is this reasoning apply also for the yearly time scale? **Reply:** *No. Spatial variations disappear at annual time scales. We specified that this paragraph refers to the findings for 'sub-annual time scales'.* 

L228, consider adding some references. **Reply:** *We added a suitable reference to the statement.*