

## **Response to RC4:**

### **S. Sippel et al., 2016**

Overall, I am pleased with the topic of the Sippel et al. paper, which is an evaluation and criticism of some of the methods used in the Donat et al. 2016 paper "More extreme precipitation in the world's dry and wet regions". This is the type of check-and-balance that keeps our science robust. Sippel et al. address two main criticisms of the Donat 2016 paper, (1) the introduction of a statistical bias when the rainfall data is normalised, and (2) the introduction of another statistical bias based on the regions that are selected as "dry" and "wet" regions. The overall flow and readability of the paper was dense, but not unfollowable. However, I understood the context of the paper, and the authors' intention, much better after I read the Donat et al. 2016 paper. The authors could use more precise wording to clarify that the methods used were done to recreate the results from Donat et al. 2016.

**We thank the reviewer for the positive evaluation of our manuscript. We agree that the original manuscript was partly very dense, and to address this issue we improve the embedding of Figures and Tables, provide better context, and add a second simple "illustrative example" to illustrate the "regression to the mean issue" a bit better. For more details, please see our reply to Reviewer #1, who raised a similar comment. We will also emphasize in a revised manuscript more precisely that the aim of our paper is to reanalyze the same dataset with a different methodology and choices for dryness definition in order to corroborate and test the sensitivity of the results.**

On the topic of the introduced bias from normalising the data; the process of normalising data is pretty common and ensures that areal averages are not dominated by very wet regions. However, this needs to be done with care. The authors unpack and clearly describe the statistical changes that are introduced from the normalisation process. I liked the illustrative example found on page three in lines six through 11 and the quantification of the bias (%) in appendices A and B provided good support for the argument. (Although it isn't clear why these are included as appendices and not tables in the paper). Furthermore, the authors do well to point out the changes that arise by using different reference periods to deconstruct the data (i.e. Figure 2). I note that it was not really clear from reading the Donat (2016) paper why they used the 1951–1980 period to normalise the data.

**Thanks again for the positive comments. We'll move the Appendices in the main manuscript and hope this will improve readability.**

**We agree with the reviewer that some normalization is often necessary to avoid that wet regions dominate spatial averages. Partly because this methodology is common (as pointed out by the Reviewer), please note that we have tried in addition to derive an analytical understanding / approximation of the biases. This will hopefully allow to estimate whether the systematic biases induced by reference period normalization in any particular case or study are worrisome or whether they are small and can be ignored. Please see the attached pdf-file, we intend to include this material in a revised manuscript as an Appendix or Supplement.**

I don't completely agree with the argument for selecting dry regions. The criteria and thresholds used to define a dry region are very subjective. As Sippel et al. point out, precipitation alone is not enough to determine if a region is wet or dry—e.g. at very high latitudes where even small amounts of rainfall can exceed the potential evapotranspiration. However, the criteria used are dependent on the question to be answered. If the question to be answered is, "How are global precipitation patterns changing?" then an analysis of precipitation alone would be sufficient. If you are trying to address, "Are wet/dry regions getting wetter/drier?" then the hydrology/aridity or climate classification of the region would need to be considered.

**We appreciate your comment: Please note that our intention was not to reject any particular dryness definition, but simply to explore the robustness of the results to this choice. However, we admit that this was not clear enough in the original manuscript, and we'll stress that both definitions can be appropriate depending on which question is being asked. Therefore, in a revised manuscript, we will refer to "regions with moderate extreme precipitation" (for Rx1d), "meteorologically dry regions" (with low precipitation totals), and to dry (arid) regions.**

The authors quantify the "regression to the mean" bias (as shown in appendices A and B) that arise by defining dry areas as the lowest 30%. The authors further demonstrate that by using the Köppen classification and the Greve (2014) definition that the large trends found by Donat et al. are dramatically minimised. I think this argument is a moot point because, as other reviewers have already pointed out, the HadEX2 dataset does not have data over the world's driest regions (e.g. the Sahara, Western Australia) or some of the wettest regions (e.g. the Amazon or the Maritime Continent region).

A global analysis of precipitation extremes or precipitation trends using HadEX2 data would deliver incomplete results.

**We agree that changes in precipitation characteristics as studied in our analyses are not representative or complete**

**given data-scarcity in many of the world's dry regions. However, we also believe that data-scarcity should not prevent scientific analyses being done with the data that is available at present. Therefore, we will emphasize this point clearly in the revised manuscript. In addition, please see our analyses in reply to Reviewer #2, where we have studied some of the characteristics of the data-scarce regions in more detail.**

Specific comments: 1. Page 4, line 12: mentions a two-sided trend test. Is this the same as the Mann-Kendall test used by Donat et al. and mentioned in the caption of figure 3? It is not really clear in the body of the text why or how this test was chosen.

**In the study of Donat et al. a one-sided Mann-Kendall trend test is used. Therefore, in all our figures and tables we report both one-sided ( $H_0$ : No positive trend; value from the Donat et al. study are reproduced), and two-sided ( $H_0$ : No trend) p-values. In a revised manuscript, we will phrase the text more in terms of the reduction of the trend slopes, rather than p-values only, because the latter can be misleading for relatively noisy time series (see e.g. Short comment by Donat et al. and our reply).**

2. Appendix A, Figure 0, caption: check the spelling of Köppen. This figure was hard to understand. After reading the caption a few times I understood that it is basically built as a table with the first (left) column being the PRCPTOT data and the second (right) column being the Rx1D data. It would be nice to have the rows/columns clearly labelled.

**Thanks for this point, we clearly see the need to improve the labels and caption and will do so.**

3. Figure 2: The caption mentions red lines. The lines look orange to me.

**Yes they are (erroneously), and they will be changed. Thanks for reading thoroughly.**

4. Figure 3: I found this figure very difficult to understand. There is a lot of information that is overlaid on other information. The grey text is too light against the white background.

**We will improve readability of Figure 3 by changing the colour and expanding the caption.**

5. Your methods for producing this graph (grey and black lines) are not clear. You mention the grey lines have been

corrected for “statistical artefacts”; I could not find this correction explained anywhere. Which artefacts have you corrected for? Is it the bias from the normalisation? Likewise, the process for producing the black line, or removing the incomplete data, is not explained.

**Yes, both the grey and black lines are produced by normalizing with the period means of the whole period, therefore avoiding the bias. Grey lines are based on the 90% completeness threshold in Donat et al., black lines are based on only 100% complete time series. We will improve this explanation in a revised manuscript.**

6. The label on the first row of graphs mentioned the Köppen–Geiger climate classification, but the caption references Köppen (1900). The Köppen–Geiger classifications were not published until Geiger (1954 and 1961). Kottek et al. 2006, which was mentioned in the text, is of the Köppen–Geiger classifications. Should the caption reference Kottek et al. 2006 rather than Köppen (1900)?

**Thanks for this hint! Yes, it should.**

7. Are graphs 3.e and 3.f from the Greve data, dry+transitional regions? It is not clear from the caption.

**Yes, they are. Thanks for reading thoroughly, we will add this to the caption.**