Response to RC2:

S. Sippel et al., 2016

Review of Sippel et al., 'Have precipitation extremes and annual totals been increasing in the world's dry regions over the last 60 years?'

This paper (which can effectively be considered as a comment on the Donat et al (2016) paper) raises two issues with the Donat et al (2016) (hereafter D2016) paper – the way in which spatial averaging has been used and the way in which dry regions have been defined.

Both of these are legitimate concerns. However, in my view both this paper and D2016 miss what I think is the main point with respect to the definition of dry regions – namely, that most of the world's driest regions (in particular, almost all of the Sahara and the Middle East) are excluded because of a lack of data. (Similarly, many of the world's wetter regions in South America, equatorial Africa and southeast Asia are also excluded). Any definition, whether it is the one used in D2016 or in this paper, is likely to give an unrepresentative sample of the world's dry regions given that the data availability is largely confined to North America, Eurasia and parts of Australia. Put another way, the HadEX2 data set in its current form is not capable of providing a fully representative sample of the world's dry regions, which is particularly important given that there is no reason why we would expect tropical arid and semi-arid zones (e.g. the Sahel), subtropical deserts (e.g. southwest US) and high-latitude low-precipitation regions to have similar long-term trends. A casual reader encountering either this paper or D2016 would expect the papers to be covering a very different range of areas to that which they actually do.

We thank the reviewer for highlighting this issue. The reviewer is of course correct (and the other reviewers who have pointed this out), data scarcity is a large caveat in analyses about precipitation characteristics in "the world's dry regions".

However, we believe that scarcity of data alone should not prohibit scientific analyses on precipitation characteristics such as the one by Donat et al. (or ours) in those dry regions where data are available. Therefore, in the revised manuscript we will make this point more clear in the text, and we particularly highlight that "dry regions of the world" actually implies "dry regions of the world where data are available". Moreover, to make this point clear to casual readers also, we will include Appendix A (Figure 0, the maps of the dry region definitions) in the main manuscript (as Reviewer #1 also recommended).

(I would view both this paper's method and the D2016 method as being reasonable ways of defining dry regions – the issue is that neither is representative given the gaps in the data set).

Again, we agree with the Reviewer: Our manuscript was not intended to reject the definition used by Donat et al (please see also our reply to the short comment by M. Donat). We simply intended to illustrate that for the overall change in dry region extreme precipitation, it does make indeed a relatively large difference (i.e. there is a large sensitivity to how one defines a dry region) if one analyses extreme precipitation (Rx1d) trends in "regions of light extreme rainfall" (that could be humid throughout the year, cf. our response to the short comment by M. Donat), or whether one studies trends in "dry" (arid) regions.

Averaging precipitation indices is another challenge – whilst the averaging period (as mentioned in this paper) is one issue, another is whether it is appropriate to average values from a distribution which is bounded below by zero and highly non-Gaussian. If one averages absolute values, area averages are likely to be dominated by the wetter areas; if one averages normalised values, there will be much more volatility in the driest areas. (Somewhat ironically, the fact that the HadEX2 data set excludes most of the world's really dry areas averted a bigger problem here – in climates where mean annual values are, say, below 10 mm, annual totals in excess of 1000% are plausible, which would completely overwhelm less variable climates in a spatial average).

The Reviewer raises an interesting and highly relevant point - and we have investigated in detail as to whether the methodological issues due to reference period normalization and subsequent spatial averaging would have been worse in arid regions where there is currently no data available.

First, we would like to point out that we agree with the Reviewer: Averaging absolute values would lead to wetter areas dominating the response, and that is why a normalization of some sort is required. Second, random variation in dry regions and the normalization procedure with subsequent spatial averaging (as illustrated by the Reviewer in the 1000% example) is precisely the root cause of the biases identified in our paper. These biases are related to the length of the reference period, but also to the ratio of mean:standard deviation (or location:scale in a GEV distribution) in the Rx1d time series (as implied by the Reviewer; and we will make this point clearer in the revised manuscript). Hence, it can be seen analytically that 1) these biases are systematically positive outside the reference period, and 2) the biases scale with the change (i.e., trend) in a multiplicative way (please see the attached pdf-file, we intend to include this material in the Appendix of a revised manuscript); and it is also

possible to derive a (first-order) analytical approximation of the expectation value of the biases as a function of the ratio mean:sd, and the length of the reference period. The analytical approximation allows to derive some estimates of the magnitude of the biases if a grid-cell scale normalization is followed by spatial averaging. We believe that this is useful to have because the applied methodology is quite common and not specific to the Donat et al. study - for instance similar data processing methodologies based on fixed reference periods are used to bias-correct relative precipitation anomalies from climate models to some observational datasets (Hempel et al. 2013, ESD, doi:10.5194/esd-4-219-2013), or observational datasets are derived based on station anomalies from a fixed baseline (Harris et al., 2014, doi/10.1002/joc.3711). However, we also note that in many real-world cases, the introduced biases will be small because the ratio of mean to standard deviation be high (e.g. in humid regions), but nonetheless it is important to note that this type of bias exists.

However, having the analytical approximation at hand we can assess how the expectation values of the biases would differ between regions, in particular whether the normalization-induced bias would become worse if more stations from data-scarce arid regions would be available. To do so, we downloaded all available arid-region station data from the GHCNDEX database (http://www.climdex.org/sewocs.html), and disregarded all stations with less than 30 years of data. Subsequently, we stratify these stations according to whether they lie in regions with or without data availability in the HadEX2 dataset, and calculate the long-term mean and standard deviation of each individual station.

The sample mean of the stations are indeed lower in data-scarce regions (Fig. 1a), i.e. arid regions without data in HadEX2 (and those stations in arid regions without data that lie in Africa only, Fig. 1a) tend to systematically receive less extreme rainfall (Rx1d). However, we notice that the ratio of mean:sd parameters is approximately similar in these regions (which would indicate a similar magnitude of the biases). An approximation of the expected bias in data-scarce regions (Fig. 3d) indicates that the expected bias would be slightly higher (+0.67\% vs. +0.71\%, +0.85% in stations that lie in Africa).
Hence, we conclude that the reviewer is correct, the systematic biases would be slightly larger in data-scarce arid regions if data would be available there, but it would not be a completely different story. We do not intend to include this analysis in the revised manuscript (because it might lead a bit too far away from the actual story thereby compromising readability and clarity), but we will include a comment in the discussion of the normalization-induced biases (p.3, l. 1-14) to the fact that these biases will be higher if the mean:sd ratio is lower (as seen from the analytical approximation),

which one might expect in very dry, currently data-scarce regions.

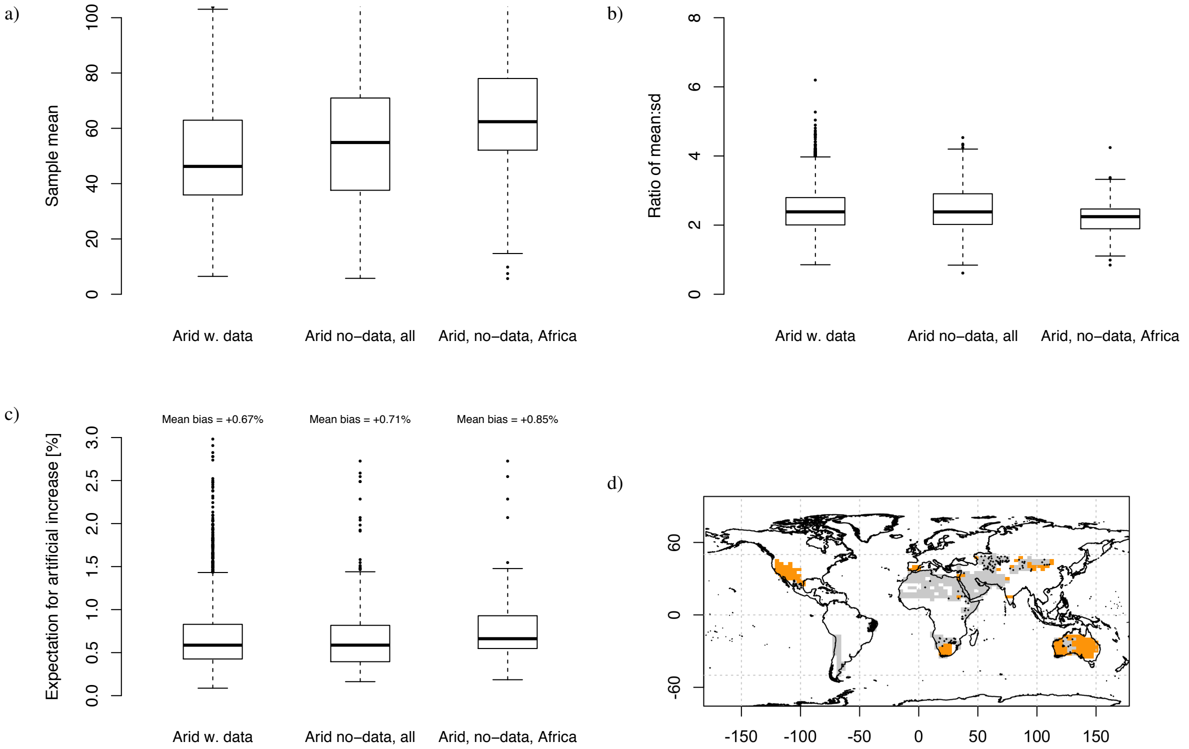

Figure 1: **(a) Sample mean, (b) ratio of mean:sd at individual stations of the GHCNDEX (stations-based) dataset in arid regions ("Arid w. data": Stations in arid regions with data in HadEX2; "Arid no-data, all": Stations in arid regions without data in HadEX2; "Arid no-data, Africa": Stations in arid region in Africa without data in HadEX2). c) Expectation for artificial increase according to Eq. A8 (with only the first term of the Taylor series) in attached pdf-file. d) Map of arid regions of the world (Greve et al. 2014): Orange regions indicate data availability in the gridded HadEX2 and GHCNDEX datasets, whereas grey areas indicate data gaps. Individual stations in grey areas are denoted by black dots.**

In my view, it would be better not to try to do spatial averages at all, and instead report using indicators such as the % of gridpoints which show significant positive/negative trends. That said, if you are going to average precipitation indices, then this paper has identified a genuine issue with the D2016 methodology.

We agree with the reviewer that in general reporting station-based indicators would be at least equally important. However, in order to compare quantities to climate models, or to study globally aggregated quantities relative to a common baseline (for which there seems to be a demand, see e.g. Hansen et al. 2012, Huntingford et al. 2013, Seneviratne et al. 2014, or similar papers that outline globally aggregated temperature characteristics), we believe that this approach can be indeed useful, too.

In summary – I think this paper accurately documents valid issues with the D2016 paper, and as such I think it is appropriate for publication, but I also think it would be improved if it engages substantially more with the issues identified above.

We thank the Reviewer for the positive evaluation and hope that we have addressed your concerns properly.