Response to SC1:

S. Sippel et al., 2016

The authors scrutinize a recent study (Donat et al. 2016) that reported increasing trends in precipitation extremes and annual totals in the world's dry regions, as defined by precipitation amounts. The authors (1) suggest that the results of the scrutinised study were biased owing to choices of the reference period, and (2) discuss that the findings depend on how 'dry' regions are defined.

We thank the authors for pointing out the statistical issue related to the reference period which is now addressed in a Corrigendum (submitted to Nature Climate Change on 12th September 2016). Importantly, this statistical issue does not affect the major conclusions of the scrutinised study, a point that should be made clearly in the current manuscript. However, the remainder of this manuscript, in particular the discussion related to the definition of dry regions, is biased, inconsistent, ambiguous (misleading), and incomplete as outlined below. Therefore the manuscript needs to be carefully revised before publication.

We thank M. Donat for the partly critical and partly positive comments on our manuscript. We appreciate that pointing out the statistical issues and sensitivities related to the data-analytical tools applied in the original study are welcomed. Please note that partly in response to reviewer comments (who stressed that the normalization-methodology is indeed relatively common), we include some analytical approximation (see pdf-file) that allows to derive analytical estimates of the biases as a function of reference period length and the parameters of the distribution. We hope these estimates are considered as useful, and we intend to include this material in a revised manuscript as Supplement or Appendix.

We also acknowledge the critical comments regarding the definition of dryness, and have carefully addressed the comments raised.

Biased: The current manuscript claims that the only valid definitions of wet and dry regions are those based on surface water availability, referring to what is 'commonly understood' or 'conventional'. However, in everyday language it is common to use 'wet' or 'dry' to refer to high or low precipitation for both regions and times of year. Furthermore, in the scientific literature there are numerous related studies that have defined wet and dry solely based on meteorological parameters such as precipitation (e.g. Allan et al.,2010; Sun et al., 2012; Liu and Allan, 2013), and these are ignored in the current discussion and should be included in a revised manuscript. The current manuscript, therefore, appears biased in that it is largely based on a claim that only a particular definition of dryness is valid, when several other definitions are in common use.

The purpose of the section on the definition of dryness was not to claim superiority of any particular dryness definition. In contrast, the main idea behind this exercise was to test how sensitive trend slopes or period changes are to alternative definitions of dryness, given that early climatological research had used the word "dry" in terms of water availability rather than precipitation totals alone (see our manuscript). However, we do not claim that this is superior. We have made this point clearer in the manuscript, and we also acknowledge that the three papers cited above use a dryness definition based on precipitation totals (annual or monthly climatology), similarly as Donat et al (2016) do in their original study for annual precipitation totals (PRCPTOT).

However, for annual-maximum daily precipitation (Rx1d), we believe it is crucially important to consider this additional point: In contrast to the studies cited above, in the original NCLIM study by Donat et al, the definition of "dryness" has been made based on the annual-maximum daily precipitation amount (i.e., Rx1d). This means that any region with a relatively modest 1-day extreme rainfall would be considered as "dry". This is in contrast to the three papers cited above, because Rx1d is not necessarily strongly related to precipitation totals. For example, the potential for very strong convective rainfall in high Northern latitudes (e.g. Scandinavia, Siberia) might be limited, therefore resulting in moderate annual-maximum daily precipitation, while the region could still be "wet" throughout the year (either in terms of precipitation totals, or in terms of water availability, or both). To illustrate this example, please see the plot below computed from the original HadEX2 data (1951-2010 means, using the 90% threshold for NA-removal):

While there is clearly a relationship between PRCPTOT and Rx1d, we find that only 22% of the "dry" grid cells according to the maximum-annual precipitation definition are also "dry" given annual precipitation totals (see plot below). Hence, while we do understand the notion of exploring (spatial) extremes in the "HadEX2 data space", it becomes an issue of semantics here: We argue that regions with low annual-maximum precipitation should simply be called for example "regions with low maximum precipitation" rather than "dry" as this might lead to confusion (e.g. if cited in IPCC reports, reported by the media, etc.). Similarly, if compared with aridity, the difference between a definition based on precipitation totals, rainfall extremes and aridity becomes very clear (see figures below).

In summary, we have changed the manuscript such that it becomes clear that we simply explore alternative definitions

of dryness, we do not claim superiority or call it "common understanding", etc., and we now also state that definitions based on precipitation totals had been used previously in the literature.

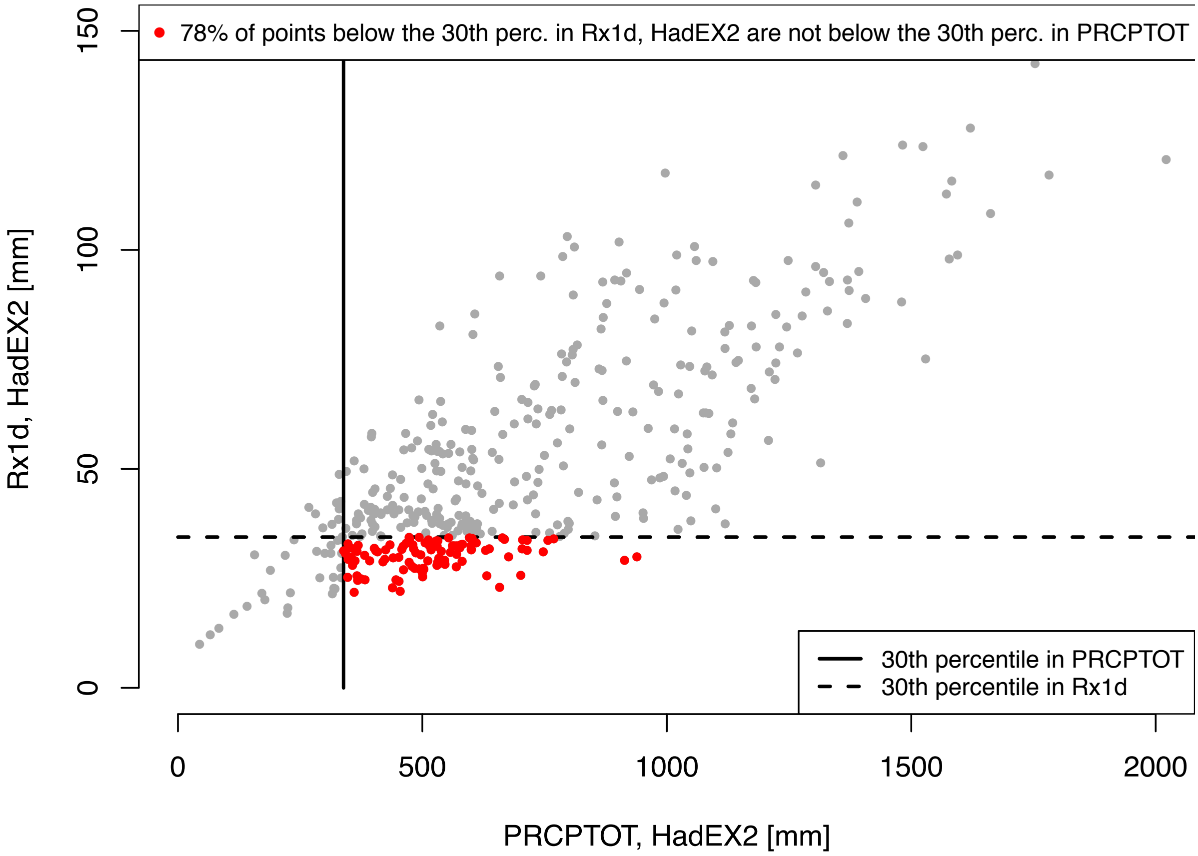

Figure 1: Relationship between PRCPTOT and Rx1d in HadEX2

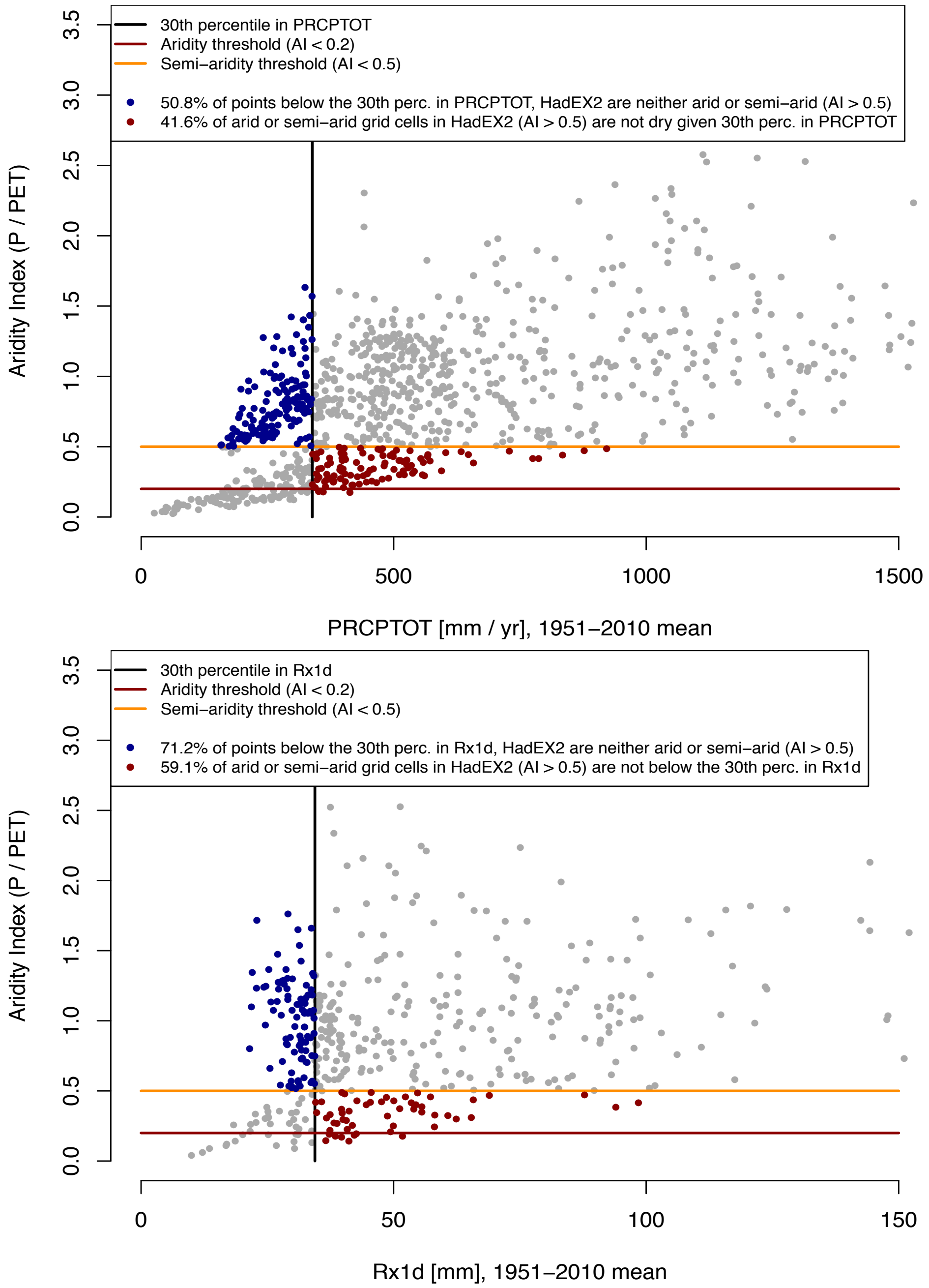

Figure 2: Relationship between a) PRCPTOT and Aridity, and b) Rx1d and Aridity in the HadEX2-GHCNDEX merged dataset. Potential evapotranspiration is taken from the CRU-TS3.23 dataset (Harris et al., 2014).

An important point that emerges from this discussion is that it is desirable to specify more clearly which type of definition of dry and wet is being used in studies of climate change. Indeed this is something the current manuscript could do better; see 'ambiguous' section below. We suggest to the authors that they make the conclusion of their paper and abstract a call for more specificity in the use of 'dry' and 'wet' in climate-change studies. For example, one could refer to 'meteorological' or 'hydrological' wet and dry regions, by analogy with the standard definitions of 'meteorological' or 'hydrological' drought. This would be of greater value than arguing that only one type of definition is valid.

Thanks for this important point, we agree. We have shaped our manuscript more in this direction and distinguish between precipitation totals, regions of moderate annual-maximum precipitation (for Rx1d) and aridity-based definitions.

Inconsistent: The analysis in Section 3 is likely affected by the same "regression to the mean" bias discussed in Section 2, because the dry-regions masks that include water demand were not defined over the entire study period 1951-2010.

We do not think this is the case: In Section 3, we use two different dry-regions masks, one based on Köppen-Geiger, and another one based on Greve et al (2014). The former had been derived in 1900 without any data from 1951-2010, so it cannot be affected. The latter dry region mask has been derived from the 1948-1968 period, but based on a very large number of dataset combinations (77) at the gridbox level and based on the aridity index (i.e., for two different variables: Potential evapotranspiration, Ep and precipitation, P). Therefore, estimates of the annual-maximum precipitation (Rx1d) should be virtually independent, as this variable is not (at all) related to Ep, and only weakly related to P. Similarly, the "regression to the mean" problem should be virtually eliminated also for precipitation totals due to the different combinations of datasets that inhibit random variation within one dataset, and because the Greve et al dry-region mask "is in good agreement with the commonly used standard climate-classifications of Köppen-Geiger" (Greve et al, 2014, NGEO). Any remaining "regression to the mean" issue in PRCPTOT would lead to a positive bias in the trend slopes and period changes relative to the Köppen-Geiger mask, but in fact the trend slopes and period increments obtained with the Greve-mask are smaller than those obtained with the Köppen-Geiger mask (see Table 2 in the manuscript). Hence, we conclude that "regression to the mean" is not an issue in our manuscript.

Ambiguous: The current text uses 'dry' for different concepts, and this is likely to confuse readers. To avoid confusion, the authors should specify whether they are talking about 'low-precipitation' or 'arid'/'water-limited' regions. This is particularly problematic e.g. in the Abstract lines 3-5 where dry is defined in terms of water availability but then immediately used to refer to the scrutinised study in which dry means low precipitation. Similarly in the introduction it needs to be specified which concepts of 'dry' the authors refer to in each case.

Thanks for this point. As indicated above, we believe this is a very good idea and we have incorporated it in the manuscript.

Incomplete: The main reason why Sippel et al. don't find a (statistically significant) increase in Rx1day in arid regions seems to be related to scarcity of data. It is unfortunate reality that arid regions are insufficiently covered by observations. Aggregating only over a few grid cells results in relatively noisy time series, so that – despite a positive trend slope – the p-value of the applied trend test is too high to reject the null hypothesis of 'no change'. A relatively easy attempt to optimise spatial coverage by merging the two existing datasets HadEX2 (Donat et al., 2013a) and GHCNDEX (Donat et al., 2013b) gives a few additional grid cells with data in arid regions. Aggregating over this just slightly improved coverage results in a more robust trend estimate in observations and in the CMIP5 ensemble mean (Figure 1). This suggests that a major uncertainty when analysing precipitation changes in arid regions comes from the limited availability of observations. Also, if using the complete coverage as provided e.g. by the ensemble of CMIP5 models as used in Donat et al. (2016), the authors would find statistically significant increases in ensemble mean over the arid regions (not shown). Therefore we assume that the main reason why Sippel et al. conclude there is 'no significant increase in heavy precipitation' in arid regions is related to the scarcity of observations.

Thanks. We agree that the scarcity of observational coverage and resulting noisy time series can be a major obstacle to detect significant trends. As suggested, we have merged the HadEX2 dataset with the additional GHCNDEX dataset that contains data. This results in a minor upwards change in trend slopes and period increments, and that several (but not all) trend slopes are indeed significantly increasing. Hence, we report these additional results in the revised manuscript. For example, our revised Conlusion reads:

"Monitoring and an accurate quantification of trends in meteorological risks in a rapidly changing Earth system is a prerequisite to informed decision-making in the context of climate change adaptation (IPCC 2014). Therefore, short reference periods that are defined on a subset of the available dataset for normalisation or data pre-processing purposes should be avoided, as this procedure inevitably introduces biases (Zhang et al., 2005; Sippel et al., 2015). In the present study under scrutiny, these statistical effects reduce the reported trend slopes and period changes by up to 40%, but the direction of the overall signal remains unchanged (i.e. increasing trends in Rx1d and PRCPTOT in regions of moderate extreme precipitation and low annual totals, respectively).

Furthermore, the definition of a `dry region' induces considerable uncertainty in quantifying changes in precipitation extremes or totals. If dryness is defined based on water supply and demand (i.e. aridity), we find a systematic and significant reduction of trend slopes and period increments in annual-maximum extreme precipitation and precipitation totals, which yields some significant and some in-significant (depending on precipitation characteristic, pre-processing, and specific dryness definition considered) but exclusively positive trend slopes (Table~\ref{table2} and Table~\ref{table3}). Hence, overall we confirm an indication towards increases in both metrics in the world's dry regions.

However, as a caveat to the present study, it is important to stress that a large part of the world's dry regions, such as large arid and semi-arid regions in Africa, the Arabian peninsula, and partly South America are not covered by monitoring datasets that are available at present. This fact highlights the importance of consistent, long-term monitoring efforts, data quality control, development and maintenance of long-term datasets (Alexander et al., 2006; Donat et al., 2013a,b), and also emphasises that the results reported here should be regarded as indicative only for those arid regions where there is data available at present.

In summary, understanding and disentangling on-going changes in precipitation characteristics in the world's dry regions remains a research priority of high relevance. In this context, our paper demonstrates that 1) data pre-processing methods can introduce substantial bias, and 2) trends and period changes can be sensitive to the specific choice of dryness definition that is used; therefore we urge authors to be considerate and specific regarding both choices and to consider associated uncertainties. "

Specific comments:

Page 2, line 3: 'if there is enough moisture available' – do you mean annual average moisture availability? Or seasonal? Or on the day the rainfall extreme occurs?

By "if there is enough moisture available", we mean enough moisture available for the extreme precipitation to occur, i.e. sensu e.g. Trenberth (2003). However, this sub-sentence is not necessary for the meaning and we have clarified the first sentence.

Page 3, line 24: It would avoid possible confusion to include a clarification at the end of Section 3 that despite having effects on the quantification of trends, these biases do not affect the conclusions in the study under scrutiny. When avoiding the discussed biases, there are still statistically significant increases in Rx1day and PRCPTOT in the dry (i.e. low-precipitation) regions.

This is correct and had not been disputed in the original manuscript. However, to make this point crystal-clear, we have added a clarifying sentence as suggested.

Page 3, lines 26-30: To avoid the impression of bias, it is important to mention other definitions of 'dry' here that are also commonly used in the scientific literature.

Page 3, lines 31-33: Donat et al. provided a number of sensitivity tests, and also analysed Rx1day changes in the dry regions defined based on PRCPTOT (see their Supplementary Information SI4) – in this mask Scandinavia and the Netherlands are not part of the 'dry' class, but they still find increasing trends (and this is also the case after correcting for the biases discussed in Section 2). Please reword to avoid the impression of cherry-picking.

As pointed out above, we have extended the discussion of the definition of dry regions: This discussion mentions now that also dryness definitions based on precipitation totals are in use, and discusses Scandinavia and the Netherlands only in terms of the dryness definition (thus, there should not be the impression of cherry-picking).

Page 3, lines 5 and 12: The statements about changes in spread of the spatial distribution do not seem to be relevant since only means are included in the analysis (not e.g. variance). These statements should be removed, or it should be explicitly stated that they are not relevant to the current analysis.

The spatial variance would be important for trend slopes, for example if confidence intervals would be obtained from the trend. However, we agree, and have separated the discussion. Also, please see our short analytical argument that allows to derive a first-order estimate of the magnitude of the normalisation-induced biases. We hope the analytical argument/correction might be useful if observations up to the present would be compared for example with model simulations for the future.

Page 4, lines 6-9: Over which time period where these alternative masks (2,3,4) defined? If not 1951-2010, you need to clarify that they may introduce the "regression to the mean" bias.

The Köppen-Geiger classification is based on temperature and precipitation taken from the CRU TS 2.1 and the Global Precipitation Climatology Centre (GPCC), respectively, for the time period 1951-2000. Although this is not the full period, the period is (1) fairly long, and (2) two independent datasets are combined (temperature and precipitation), such that any potential "regression to the mean" effect should be negligible.

Page 4, Line 9: What is the rationale behind including transitional regions when studying precipitation in dry regions?

The rationale is simple: Our intention for this paper is to explore a range of different choices in order to test the sensitivity for different trend slopes and period increments of extreme precipitation - to this end, we believe that a combination of "arid" and "semi-arid" region can indeed provide additional insights.

Page 4, lines 15/16: large parts of these 'subsidence regions' with no or little precipitation changes are located over the ocean. Water availability can clearly not be a limiting factor here, so this is unrelated to the discussion of different definitions of ‘dry’.

We do not claim causality here - i.e. the statement does not imply that the reduced trend slopes in precipitation extremes in arid and semi-arid regions are due to water availability. This statement is just a short plausibility discussion of our results - given that the section is now entitled "Sensitivity of changes in precipitation totals and extremes to the definition of a dry region" we think this is appropriate.

Page 4: Lines 17-21 give a hint of a balanced discussion, but unfortunately lead to a highly biased conclusion (lines 22-24), again appealing to what is supposedly ‘commonly understood’ and suggesting arid would be a conventional definition for dry.

We have clarified and extended the conclusion: We report about the reduction in trend slopes, and indicate that there is a significant increase if the datasets are merged.

References

Allan, R. P., Soden, B. J., John, V. O., Ingram, W. J., Good, P.: Current changes intropical precipitation. Environ. Res. Lett. 5, 025205, 2010.

Donat, M. G. , Alexander, L.V., Yang, H., Durre, I., Vose, R., Dunn, R., Willett, K., Aguilar, E., Brunet, M., Caesar, J., et al.: Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: the HadEX2 dataset, Journal of Geophysical Research: Atmospheres, 118, 2098–2118, 2013a.

Donat, M. G., Alexander, L. V., Yang, H., Durre, I., Vose, R., Caesar, J.: Global land-based datasets for monitoring climatic extremes, Bulletin of the American Meteo- rological Society, 94, 997-1006, 2013b.

Donat, M. G., Lowry, A. L., Alexander, L. V., O’Gorman, P. A., and Maher, N.: More extreme precipitation in the world’s dry and wet regions, Nature Climate Change, 2016.

Greve, P., Orlowsky, B., Mueller, B., Sheffield, J., Reichstein, M., and Seneviratne, S. I.: Global assessment of trends in wetting and drying over land, Nature Geoscience, 7, 716–721, 2014.

Liu, C. and Allan, R. P.: Observed and simulated precipitation responses in wet and dry regions 18502100. Environ. Res. Lett. 8, 034002, 2013.

Sun, F., Roderick, M. L., Farquhar, G. D: Changes in the variability of global land precipitation. Geophys. Res. Lett.

39, L19402, 2012.

Figure Caption (complete caption as the online system seems to cut the cap- tion after the second sentence):

Figure 1: Extreme precipitation changes in arid and humid regions. Time se- ries of Rx1day (the annual-maximum daily precipitation) for dry/arid (a) and wet/humid (b) regions as identified by Greve et al., 2014. Area-weighted average time series are shown for HadEX2 and the ensemble mean and spread of CMIP5 simulations. Precipitation indices were first normalized by calculating annual values as a fraction of the 1951–2010 local mean before calculating the dry- and wet-region averages. Black lines, annual values from observations and ensemble mean; red lines, linear trend; blue dashed lines, 30-yr averages for 1951–1980 and 1981–2010; grey shading, ± one ensemble standard deviation. dRx1day indicates the difference between the averages during 1981–2010 and 1951–1980; slope is the linear trend Sen-slope estimate (unit, decade−1); and the p-value is the trend significance using a Mann–Kendall test. (c) The mask indicates the locations of the grid cells contributing to the average of the dry (red) and wet (blue) regions, and the number n of grid cells contributing to the area averages of dry and wet regions is given. Land grid cells that are less complete than 90