Journal: Climate of the Past (CP)
Title: Investigating uncertainties in global gridded datasets of climate extremes
Author(s): R.J.H. Dunn, M.G. Donat, and L.V. Alexander
MS No.: cp-2014-53; MS Type: Research Article

This paper systematically investigates the robustness/sensitivity of global gridded datasets of temperature and precipitation extremes to various parametric and structural changes in the dataset construction, with emphasis on HadEX2, but results are illuminating also for other similar datasets. It certainly addresses relevant scientific questions within the scope of CP. It is mostly well written and is well worth publishing, though some revision is needed. Basically, I agree with comments made by Anonymous Referee #1, so I don't need to repeat issues raised in that report.

General comments:

1. Throughout the manuscript, much of the numeric analysis and discussion focus on data after 1950 while global time-series plots start in 1900. However, many of the figures (also in the supplement) reveal tremendous differences in global trends, between the different experimental choices, before 1950. These differences clearly relate to situations when the global time series are based on few data in the beginning but more data later on. The authors do acknowledge this in their discussion, but it could be emphasized at relevant places more how tremendously sensitive several indices can be to changes in the station network.

2. The Taylor diagrams may be very useful, but they appear to have been thrown in in the very last minute with almost no explanation of how to interpret them and no discussion about what they show. If they are not needed for the conclusions, then they could be omitted. It seems better, though, to explain and discuss them more in detail. I suggest to add some discussion of all panels in Fig. 20, to point out differences and similarities and highlight what the main message is. I see no particular reason to put them in an appendix. They could rather be shown and discussed in section 5.

Technical comments:

p 2106, l 6: Change "or which" to something else

p 2110, l 11: Change "normalised" to "anomalised"

p 2110, l 18. Explain better what is meant by "grid boxes with 90 % completeness"

p 2111, l 6: Insert "are" after "1960"

p 2111, l 16: What kind of percentage is considered here? "60 per cent" of what?

p 2113, l 19: Is it really likely that changes in individual grid box values are always expected to be "small"? Perhaps they are large?

p 2114, l 8: Insert "larger" before "regions"

p 2114, l 9: Indonsesia could be added here.

p 2115, l 3: Unexpectedly "regional workshops" are mentioned here. Why? What is that?

p 2116, l 2: In which indices do you have "high confidence" in the trends...? I don't see the same

thing in Fig 6b.

p 2116, l 4-9: Do you talk about precipitation indices here?

p 2116, l 21-22: I don't agree that there is "little" effect on the overall trend. Rather "quite large".

p 2119, l 9: Insert "parts of" before "South America"

p 2119, l 9-11: The sentence is incomplete.

p 2120, l 7 (and other related places): Why are not all curves in Fig. 11 forced to pass through the point (0,1)? Physically, the correlation is definitely expected to be 1 and distance 0. The only curve that is close to the observations when the correlations have dropped to 1/e is the blue (exp+off). All others have a very poor fit. I can accept that the different curve-fitting approaches has a usefulness just to study how sensitive the results are to different ways to estimate DLS, but it is clear that some choices are physically unrealistic and give much too large DLS values (for the case shown in Fig. 11). This should be commented upon. Also, as Referee #1 pointed out, the polynomial fit is very unrealistic at long distances. I assume it has to be used here because it was used in HadEX2, but a motivation and discussion about its usefulness could be added for clarity.

p 2121, l 1: What is meant by "only the level differs"? I see no difference in mean level in Fig. 12b.

p 2123, l 23-27: The assumption that "issues with the quality and inhomogeneity are minor" is a bit dangerous. I don't think every single station record has been sufficiently "carefully checked" yet to allow this assumption. Consider a more "careful" statement.

p 2124, l 26-27. The sentence here would be more true if "relatively good" is changed to "relatively poor". This is one example where the authors tend to paint a too rosy picture. There are BIG differences in trends between different methodological approaches in some cases when the entire period is considered.

p 2125, l 1: It is good when short timescale variations occur at the same time and in the same direction, but from a climage-change viewpoint it is even more important that long-term trends are about the same. This is another case where a too rosy picture is painted.

p 2125, l 6-8: I would like to see a more expanded discussion on when and where the match with HadEX2 is "reasonably well" and when it is not so well – and also why it is not so well in those cases.

p 2125, l 14: You probably mean "grid box values" rather than "grid boxes".

p 2126, l 17-19: But if you include pre-1920 data, very different trends may result randomly depending on how the jack-knifing is realised.

p 2126, l 22: Insert "in particular" before "in the 25 % runs"

p 2128, l 3-5: Mathematically, I don't agree with the statement in this sentence. A strong correlation can be obtained even in the absence of a trend. Take for example an arbitrary time series with no trend, and a copy of the same time series but with 10 times larger amplitude. The correlation between these two series is 1, but the RMS is large. Modify your explanation on what the Taylor diagram shows.

p 2128, l 12: Delete "that" after "stipulates"

p 2128, l 22: Delete "is" after "grid box"

p 2130, l 12-13: Change "almost all the choices fall within the statistical trend of HadEX2" to something like "the trends of HadEX2 fall within the envelopes of trends for almost all the choices". I guess this is what you really mean.

p 2130, l 15: Can you somehow quantify what you mean with "only slightly less robust"?

p 2131, l 3: Delete one "s" in "ass" at the end of this line.

p 2131, l 17: The fact that the quality of the station data has not been investigated could be highlighted more clearly in relevant places. Why hiding this information?

p 2132, l 8-9: This sentence is ridiculous. Delete it.

p 2132, l 12: I can see several instances in the supplement when changes in methods introduce a strong trend or a drastic change in variability! Be more honest and clear about this. Don't hide where there are problems, but do point out where there is robustness.

p 2133, l 6-7: Similar comment as above for p 2130, l 12-13.

p 2133, l 7-8: The very last sentence is an overstatement that does not strictly follow from the previous one. Be more careful when trying to summarize your main findings. What is robust? What is not roubust?

p 2133, l 11. As I pointed out above in my second point: Expand the discussion on Taylor diagrams and consider move it to Sect. 5.

p 2137. Caption to Table 1. Change "Sen (1968), Lanzante (1996)" to "(Sen, 1968; Lanzante, 1996)"

p 2137. Delete the second "temperature" in the definition of TNn

p 2138. The name of SDII should probably not be "Simple daily index index"

p 2140. Caption to Fig. 1. I cannot remember that you have define how you exactly define e_RMS and sigma^2 anywhere. Have you?

p 2148. Caption to Fig. 8. Change "stations per grid box choices" to "choices of stations per grid box".