Clim. Past Discuss., 10, C965–C977, 2014
www.clim-past-discuss.net/10/C965/2014/

Climate of the Past Discussions

Interactive comment on "Identifying homogenous sub-periods in HadISD" *by* R. J. H. Dunn et al.

R. J. H. Dunn et al.

robert.dunn@metoffice.gov.uk

Received and published: 3 July 2014

We thank Victor Venema for his detailed review and respond to each point individually below.

Major

1) Reviewer

Introduction, page 1570, line 5-15. I would personally call the "off-the-shelf" packages: Modern multiple breakpoint methods. SPLIDHOM is a daily correction methods and does not fit too well to the rest here.

Response

Sentence updated and SPLIDHOM removed from this section and added later.

2) Reviewer

Introduction, page 1570, line 15-25. The main conclusion from HOME is missing, that modern methods (multiple breakpoint methods designed to deal with inhomogeneous references) are clearly better than traditional ones. You might want to add that PHA was recommended by HOME for the homogenization of large datasets.

Response

Added sentences giving main COST-HOME result and recommendation to use PHA for large networks.

3) Reviewer

Introduction, page 1570, line 23. Here SPLIDHOM should be mentioned. You could also mention Quantile Matching (RhTest), wavelet homogenization (Yan and Jones, 2008) and maybe daily detection (Rienzner and GandolïňĄ, 2011, 2013; Wang 2008) here. The sub-daily physical corrections of temperature of Auchmann and Brönnimann may also be something that is worth to mention in this sub-daily article.

Response

SPLIDHOM reference added along with the additional references to other methods.

4) Reviewer

Section 2, page 1572, line 10. If the network wide change is at the same date at all stations also pairwise cannot help, that is a basic limitation of relative homogenization. I would add something about network-wide changes happening during relatively short periods.

Response

Sentences added highlighting issue of network wide changes occurring over a short timescale.

5) Reviewer

Section 2, page 1572, line 15. The number of breaks is not an important quality measure in itself. The break variance would be a better one. (Next to break variance maybe also the trend in the inhomogeneities could be an interesting measure.) A problem for all such measures is that the quality of homogenization depends on the noise level of the reference series. I would thus advice to give your users also the guidance of provide some information on this noise level, maybe the noise level using the nearest reference or the mean of all references used by the PHA.

Response

We have calculated the average RMS noise for each station using the target-station difference series. These figures are now given in the paper (most are in the appendix to save on space) and addressed in the text.

6) Reviewer

Section 2, page 1572, line 25. What does HadISD do in this case? Is such a station withheld?

Response

Stations not processed by PHA are shown in Fig 1 and 11 and will be listed in a separate file available for download on the Hadobs website. As we do not have any information on the homogeneity of these stations we cannot withhold them in case of large magnitude inhomogeneities as they may have none. We leave it up to users to decide how best to use this data. In the analysis in Section 7 we use these stations along with those where PHA could run, but no inhomogeneity was detected.

7) Reviewer

Section 3, page 1575, line 7. Is there really not pattern? It seems that there are more non-processed stations in data sparse areas. The colours in Figure 1(b) make it hard to

estimate the values; it does not have much more dynamic range as a black white colour scale. I had expect to see more breaks where the station density is higher (easier to detect), Figure 1(b) almost suggests the opposite.

Response

This has also been identified by Enric Aguilar and we have updated our discussion appropriately. We have also changed the colourmap used for Fig 1b and the similar panels in Fig 11.

8) Reviewer

Section 3, page 1575, line 25. The bias may be small relative the typical break size, but is climatologically important. Investigating this is scientifically very important, if you could do something here that would be appreciated, but maybe it is more the topic of a follow up study. If you remove the “broader positive wing” in Fig. 2a would there still be a bias? Here is the bias regionally, how does it depend on station density, is there a dependence on the number of readings per day (as proxy for whether the station was always automatic or switched from observations to automatic) such kind of questions come to mind. Idem for Section 4, page 1577, line 21.

Response

We split the distributions of inhomogeneity magnitudes into WMO regions (as for the Records Check of the HadISD QC). The largest bias for temperature is observed in raw data for North America (0.198C) but the Gaussian fit is smaller (0.096C), and Asia has the smallest bias (0.043C) – see Figs. 1 & 2 of this response. The other regions (Africa, South America, Pacific, Europe) all have positive bias as well. For the diurnal temperature range, Africa has the largest negative bias (-0.193C) with the Gaussian fit again being smaller (-0.055) and Asia again the smallest (-0.065C) with an unbiased fit (0.009C). Changes in instrumentation in the USA (from liquid-in-glass to maximum-minimum temperature system) have caused biases in the USHCN, and so could be

part of the cause (Menne, Williams & Vose, 2009).

These differences for the regions are similar for the dewpoint temperatures. Both positive (Europe, 0.084hPa) and negative (Africa, -0.062hPa) biases are observed for the different regions for SLP. For the wind speeds, all regions have negative biases in the maxima (-0.159 m/s, Africa to -0.053 m/s North America), but there is a range for the monthly averages (0.011 m/s South America to -0.118 m/s Pacific).

We have not assessed the effects of reporting frequencies rarer than 3 hourly as these entail very few stations or change points: we only compare hourly and 3 hourly. On the whole we find the largest biases for stations that report hourly for all four meteorological variables. In the monthly average wind-speeds the biases are roughly the same, but using the monthly average maximum wind speed, the hourly stations have a larger bias. This indicates that this bias comes predominantly from automatic stations. Given the clarity of this difference between the station reporting frequencies, we have included some text in the main paper to explain this, along with Figure 18 in the Appendix.

9) Reviewer

Section 3, page 1567, line 18. "The distribution of adjustment values with latitude and longitude show that the largest adjustments are mainly found in regions with large numbers of stations (Fig. 5)" Could this also be an artefact of your scatter plot. Where you have more values you also have more strongly deviating values. I wonder if the 2D histogram would give the same impression.

Response

We agree with Victor on this point, and have changed the text appropriately. We have also updated the text for the 2-D histogram of record length versus change point number.

10) Reviewer

Section 4, page 1577, line 18. Based on the lower correlations between stations one

would expect to find less inhomogeneities in wind.

Response

Although the correlations for wind goes down, PHA uses the best correlating neighbours (min 7, max 40), as long as they do actually correlate. Hence there are still plenty of neighbours with which to do the pairwise analysis. This is to our knowledge the first application of PHA to wind speeds, which are a non-Gaussian variable. We have not performed any scaling of the wind speeds to "Gaussianise" their distribution. The RMS noise maps do show that the noise in the wind is actually very low, which may explain the relatively high number of change points detected.

11) Reviewer

Section 4. Do you have any explanation for the rejections of SLP in Africa and China? The station density should be high enough, especially as SLP is well correlated.

Response

These high rejection rates are the result of very short records (in many cases entirely missing) after conversion to monthly averages for SLP. We have added a sentence explaining this.

12) Reviewer

Section 5 Validation. For all these number about hits and misses, I was always wondering how many you would expect to get by coincidence.

Response

This is an interesting point. What are the chances of coincidental agreement between the metadata and the changepoints? In our validation of the temperature change points, we required that the two dates be within 1 year of each other. For temperature, there are on average around 3 change points per station. Over 41 years of data, this means that the probability of any year containing a change point is $3/41 \approx 0.07$ (as

our validation looks for matches within 12 months). Now assume that all 153 UK stations have full metadata over all of their 41 year record - though this is an overestimate. Although there will be many changes outside of the enclosure (and even some inside) which will not be noted in the metadata, we will assume for this exercise that changes occur on the site which may affect the homogeneity and that are noted in the metadata every 10 years or so. Hence a probability of any given year containing a change is 0.1. For the match to occur, as this can be in +/- 1 year, then the probability that a match will occur in a station is 3 x 0.1 x 0.07 = 0.21. This means that a match between the metadata and the change points will occur in roughly every fifth change point. That is exactly what we found.

13) Reviewer

Section 7, page 1582, line 28. Are you sure there is an improvement in the homogeneity of the stations used?

Response

We agree with Victor that our wording for this sentence was inaccurate. We have reworded this sentence to more clearly indicate what is occurring. By excluding stations with large inhomogeneities, the average magnitude of the remaining inhomogeneities is reduced. However in this part of this assessment the number of change points is not restricted, and so the homogeneity of the stations is not changed.

14) Reviewer

Section 8. Page 1585, line 10. There are some articles on daily detection (Rienzner and GandolïňĄ, 2011, 2013; Wang 2008).

Response

Extra references added.

15) Reviewer

Figure 2. How did you make the (censored) Gaussian fit Because the lines are so thin, the difference between blue and black are small. The statistics for the Gaussian fit may also be worth reporting (or are the top values the ones of the fit and not of the raw detection).

Response

To make the Gaussian fit in Figure 2, we used all the bins further from zero than the peak for each side of the distribution. We have included the mean and standard deviation of the Gaussian fit on the plots, increased the line thickness and used a lighter blue to increase clarity.

16) Reviewer

Figure 7. The thicker tail than a normal distribution would have is not sign yet that the distribution of real inhomogeneities is also not normal. Two breaks in the same direction may be combined by statistical homogenization.

Response

We have added a sentence to this effect in Section 4.

17) Reviewer

I agree with the main comments of Blair Trewin and Enric Aguilar. Except that the small networks in the COST-HOME benchmark were a problem for the PHA. One contribution using the PHA actually performed as the best contribution for the smallest networks (5 stations).

Response

We have added the note about PHA performing well in COST-HOME on a small network into Section 2, but kept the note that it has been primarily designed to work in an automated fashion on large networks.

Minor

1) Reviewer

Title. In science the preferred form seems to be “homogeneous” not “homogenous”.

Response

Title has been changed resulting from Enric Aguilar’s comment.

2) Reviewer

Affiliations: I wonder whether the Latex double ff would make a problem for the email link.

Response

We will check with the editors.

3) Reviewer

Page 1575, line 20. Which *four* methods?

Response

Text updated to "two methods".

4) Reviewer

Page 1576, line 12. Please, add that these are station numbers.

Response

Added, along with station names from ISD database.

5) Reviewer

Page 1577, line 13. “For the SLP, we use the deviations from 1000hPa when calculating the monthly mean values.” Why is this sentence important?

Response

It is useful for those wishing to follow up on this work when re-creating the monthly files for use by PHA.

6) Reviewer

Page 1579. Line 1-2. “There do not appear to be any correlations with geographic features for any of the variables.” Didn’t you just mention these in the previous sentence?

Response

The intended meaning was for topographical and other physical features rather than country borders. This has been clarified.

7) Reviewer

Page 1579. Line 12. Add “(Figure 13)”. All ïňĄgures should be mentioned.

Response

Reference to Figure added.

8) Reviewer

The degree sign in many plots is not printed right.

Response

The degree symbols have been added to Figs. 15, 16 and 17.

9) Reviewer

Figure 6. Is this the distribution of the *absolutely* largest adjustment values?

Response

Yes, it is the largest absolute inhomogeneity magnitude. The caption has been updated appropriately.

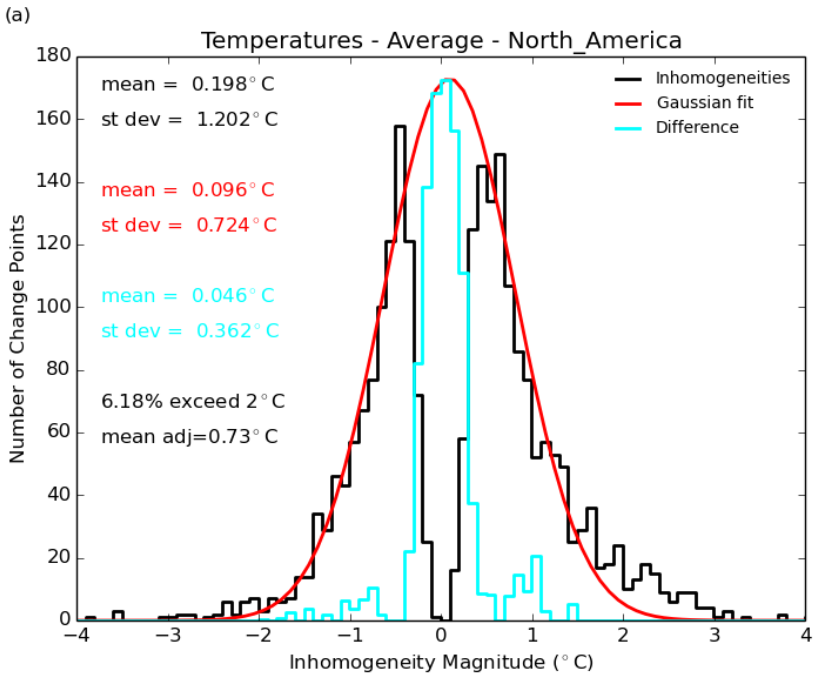

Fig. 1. Distribution of inhomogeneity magnitudes for North America. Colours and lines are the same as for Fig. 2 in the manuscript.

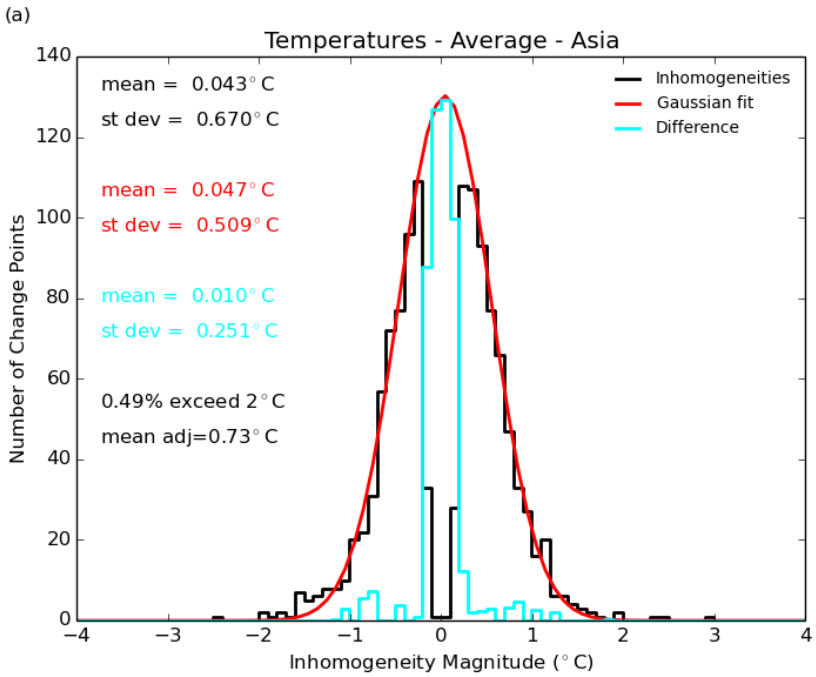

Fig. 2. Distribution of inhomogeneity magnitudes for Asia. Colours and lines are the same as for Fig. 2 in the manuscript.