Clim. Past Discuss., 11, C2856–C2861, 2016
www.clim-past-discuss.net/11/C2856/2016/

Climate of the Past Discussions

Interactive comment on "Expanding HadISD: quality-controlled, sub-daily station data from 1931" *by* R. J. H. Dunn et al.

R. J. H. Dunn et al.

robert.dunn@metoffice.gov.uk

Received and published: 6 January 2016

We thank Blair for his review of the manuscript, and have responded to each comment in line below.

Reviewer

- The paper does not explain anywhere why the first data release started in 1973 but the new version extends to 1931 - I think this would be useful for readers to know. (Is it because ISD itself has expanded ? and if so, is it known why? (has additional digitisation contributed?) - or because more work has been done to incorporate older ISD data in the HadISD dataset?).

Response

We have added a paragraph in the introduction to highlight this point, and also refer to it in Section 2. The issue is highlighted by Figure 2, showing the large drop in station numbers prior to 1973.

Reviewer

- I think it would be useful to indicate the coverage of the earlier data - this could be done, for example, by having maps showing snapshots of the data coverage in some representative years (say, 1940, 1960, 1980, and 2000). It is useful information as to whether the older data drawn mostly/entirely from the developed world, or whether there is also some coverage in Africa, South America and Asia.

Response

This is a good suggestion, and we have added figures for the four years suggested as well as the first and last year. The images show that in 1931 most stations are in Europe, with a few scattered elsewhere. More stations are available in the Americas and Australia in 1940, with a high concentration in central Europe and a few other countries. By 1960 the coverage is almost world wide, with gaps in Africa (which occur throughout), South America and Arabia. There is a doubling in the station number to 1980, with a all gaps bar those in Africa filled in. Thereafter there are no large-scale changes in the distribution of stations through until the end of the dataset. We will include these plots in an appendix to the paper.

Reviewer

- There is a sharp drop in the number of available stations between 1965 and 1972, and especially around 1972. Is there a reason for this?

Response

The steep drop in 1972 was from changes in the GTS which has meant that records have been lost from the ISD. They may still exist in other archive holdings and so we hope that with future developments in the ISD that they will be filled in over time.

Reviewer

- A potential issue with annual updates is that sometimes stations close and are replaced (for example, a city station might be replaced with an airport station, sometimes with a period of overlapping observations, sometimes not). Is there any process to manage this? (e.g. to search for potential composites if a station has closed, even if the 'new' station does not have 15 years in its own right?).

Response

At the moment we do not allow for the creation of merged stations from individuals where all have less than 15 years of data. The primary station has to have at least 15 years worth of data and a median observing interval of 6 hours - but secondary stations that are merged into it need not have this much. Therefore, in the case you mention, if either the city or the airport station have 15 years of data, and they are identified to be merged, then the resulting merged station will include data from both. If neither have 15 years of data, then neither would be included - even if the resultant merged station would have sufficient data.

This is different to the method used in HadISD.1.0.x - where merged stations made up of lots of short periods of record were included. We have clarified this section in the manuscript to highlight these points.

The station selection and merging process has been adapted slightly in light of the comments of the first reviewer.

Reviewer

- P4573 1st paragraph - somewhere here it would be worth noting potential limitations in the accuracy or precision of coordinates (if coordinates are only available to one decimal place, as is sometimes the case, this implies precision in the order of +/- 5 km). Could also be worth mentioning (line 10) that differences in spelling could arise from language differences (especially in countries with multiple languages, or those

which do not use the Roman alphabet).

Response

This is a good point, and we have added sentences to highlight these limitations and possibilities.

Reviewer

- P4574 line 4 - 'in HadISD.2.0.0 there are fewer stations in central and southern Africa...' - is this in comparison to Europe and North America, or in comparison to earlier HadISD versions? This needs to be clarified.

Response

The comparison was against HadISD.1.0.x, which has now been clarified in the text.

Reviewer

- Section 3.1 distributional gap - while specific processes for tropical cyclones are mentioned later, it still appears to me as if the distributional gap check might produce false positives for extreme low MSLP values in tropical cyclones (especially if there is only one instance of the station concerned receiving a direct hit from a tropical cyclone). I'm not sure if there is a good automated way to address this, but it might be useful to have some kind of process in place where users can report data which they believe have been inappropriately flagged (either positively or negatively).

Response

There are extra parts of the distributional gap check which account for tropical storms, we have now included a summary of these in this section. Tropical storms are mentioned in section 3.4 on the unusual variance check. For the distributional gap check, observation times with low anomalies are checked to see if the wind speeds have high anomalies - if so no flag is set.

However, you are right that if no wind speeds are present, the neighbour check may throw out the observation. We have attempted to mitigate this effect by doing the following. If any candidate-neighbour pair has a separation of over 100km, then if more then 2/3 of the flags are from negative anomalies (indicating that the candidate station is regularly affected by storms) then only the positive differences are counted towards the neighbour removals. Although this does retain many storm signals, direct hits, especially at coastal stations are still prone to removal. We agree that a reporting back procedure would be useful, and in the first instance encourage contact by email for users who have found this the case. A paragraph has been added to Section 5 detailing where users can find out more information on updates and also encouraging them to get in touch should they find issues in the dataset, for example observations that have been erroneously flagged.

Reviewer

- P4581 lines 6-8 - would the higher density of observations (and hence presumably more effective neighbour checks) explain the high frequency of flagged observations in eastern North America and western and central Europe?

Response

We will provide maps showing the percentage of data removed by each test on the website - as we have for each version of HadISD.1.0.x. These maps show that it is not specifically the neighbour outlier check which removes large fractions of observations in these areas. The final check removing months which already have large removals or large gaps left removes most, with the test removing strings of repeated values also removing relatively many observations.

Other tests also have higher removal rates in the areas mentioned - which we believe arises from the better completeness and higher time resolution of the data from these stations. This increase in data quantity makes it easier to accurately characterise distributions, and hence highlight data that fall outside of them.

Reviewer

- Figure 3 - this figure doesn't seem to match the caption - the legend suggests that the bottom figure shows both composites and non-composites but the caption only mentions composites? Also, does the overwhelming 'red' signal in North America and Europe indicate that almost all stations in those regions are composites, or is it because the mapping tools used overwrite blue dots with red ones?

Response

We have clarified the caption as the lower panel does show both composites and non-composites. The red dots are plotted over the blue dots and so overwrite the areas where the station density is high. We have also clarified this in the caption.

Interactive comment on Clim. Past Discuss., 11, 4569, 2015.