Clim. Past Discuss., 10, 1567–1607, 2014
www.clim-past-discuss.net/10/1567/2014/
doi:10.5194/cpd-10-1567-2014

Climate of the Past
Discussions
Open Access

Identifying homogenous sub-periods in HadISD

R. J. H. Dunn, K. M. Willett, C. P. Morice, and D. E. Parker

Met Office Hadley Centre, FitzRoy Road, Exeter, EX1 3PB, UK

Received: 20 March 2014 – Accepted: 3 April 2014 – Published: 11 April 2014

Correspondence to: R. J. H. Dunn (robert.dunn@metoffice.gov.uk)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

We report on preliminary steps in the homogenisation of HadISD, a sub-daily, station-based dataset covering 1973–2013. Using temperature, dewpoint temperature, sea-level pressure and wind speeds, change points are detected using the Pairwise Homogenisation Algorithm from Menne and Williams Jr. (2009). Monthly mean values and monthly mean diurnal ranges (temperature and dewpoint temperature) or monthly maximum values (wind speeds) are processed using the full network of 6103 stations in HadISD. Where multiple change points are detected within one year, they are combined and the average date used. Under the assumption that the underlying true population of inhomogeneity magnitudes is Gaussian, adjustments as small as around 0.5 °C, 0.5 hPa or 0.5 $\mathrm{m\,s^{-1}}$ have been successfully detected. No strong biases are present in the distributions of the adjustment values. The change point dates and adjustment values for each of the calculation methods will be provided alongside the dataset to allow users to select stations which have different levels of homogeneity. We give an example application of this change point information in calculating global temperature values from HadISD and comparing these to CRUTEM4. Removing the most inhomogeneous stations results in a better match between HadISD and CRUTEM4 when matched to the same coverage. However, further removals of stations with smaller and fewer inhomogeneities worsens the match.

1 Introduction

To enable the use of a dataset for the study of long-term trends, the raw data have to be first quality controlled to remove random erroneous data arising from instrumental or observer error. After this process is complete, systematic biases will still be present in the data resulting from documented and undocumented station moves, changes to the instruments, incorrect station merges, changes in land-use (urbanization around the station) or even changes in observing practices over time. The process of removing these non-climatic signals from the data is known as homogenisation.

Two main approaches exist for determining the location of change points and the adjustment magnitudes. A reference series can be created from the neighbouring stations, under the assumption that the reference series is free from errors or that they cancel out on averaging; or a pairwise search can be undertaken, assessing each candidate-neighbour pair individually. Then classical statistical tests (e.g. the Standard Normal Homogeneity Test, SNHT, Alexandersson, 1986), regression models (Easterling and Peterson, 1995) or Bayesian approaches (Perreault et al., 2000a, b) have been used to extract the change point locations and magnitudes of the inhomogeneities. There now exist a number of different "off-the-shelf" packages which can be used to homogenise datasets, e.g. HOMER/SPLIDHOM (Mestre et al., 2011, 2013), MISH-MASH (Szentimrey, 1999, 2007), ACMANT (Domonkos, 2011) and PHA (Menne et al., 2009).

Given the variety of options available for homogenising datasets, the COST-HOME project (Venema et al., 2012, www.homogenisation.org) assessed and benchmarked nine different algorithms. Some algorithms had multiple submissions, resulting in 25 contributions which were assessed using real, surrogate and synthetic data. By using benchmark datasets, the relative ability of each implementation of each algorithm in detecting and adjusting for artificial change points could be assessed. The results showed that automatic algorithms can perform as well as manual ones, but that users need training in the use of any of the algorithms to avoid degrading the homogeneity of the data. Following this assessment Mestre et al. (2011, 2013) have released the HOMER (for monthly data) and SPLIDHOM (for daily data) homogenisation packages, which take the recommendations of the COST-HOME assessment into account. Given the general strong performance of PHA (Williams et al., 2012) and its proved success in working on large station networks in an automated fashion, we have used it in our assessment of homogeneity in HadISD.

CPD
10, 1567–1607, 2014

HadISD is a new, sub-daily, station-based dataset from the Met Office Hadley Centre. It is based on the National Climate Data Centre (NCDC) Integrated Surface Dataset (ISD), which contains over 30 000 non-unique station holdings (Smith et al., 2011). After merging candidate stations which were likely to be the same station, Dunn et al. (2012) selected 6103 stations which had primarily hourly or three-hourly data over the period 1973–2011. Temperature, dewpoint temperature, sea-level pressure, wind speed and direction and cloud data were extracted and subject to a detailed suite of quality control tests to identify and remove erroneous observations. A full description of these tests is given in Dunn et al. (2012). HadISD is updated on an annual basis, with each new release being given a unique version number to allow full traceability, with the current version (v1.0.2.2013p) running to the end of 2013. HadISD is an updated, quality controlled dataset, but it has not as yet been homogenised, although it has been used in the creation of a monthly homogeneous specific humidity product (Willett et al., 2013). Initial steps in releasing a homogenised HadISD are detailed in this manuscript. We will use version 1.0.2.2013p of HadISD as this is the most up-to-date version at the time of writing. This homogeneity assessment will be carried out for all future versions of the dataset.

To date, homogenising on an hourly timescale, which entails both detecting the change points and applying adjustments, is not feasible for global station networks. Further research and development is required to be able to automatically detect the change point locations and, more importantly, their characteristics so that adjustments appropriate for each hour of the day are applied. Progress is being made on daily homogenisation for temperature (e.g. Vincent et al., 2002; Brandsma and Können, 2006; Della-Marta and Wanner, 2006; Kuglitsch et al., 2009; Toreti et al., 2010; Trewin, 2013), but so far only for country-scale networks. It is still uncertain how to apply these to daily data on a global scale. As with any data quality process, we do not want to degrade the data and so we will identify change point locations on a monthly scale for this global dataset of 6103 stations only, and for the moment, not apply any adjustments directly

CPD
10, 1567–1607, 2014

to the data. Future developments will detect change points on shorter timescales, and apply adjustments to the hourly HadISD data.

Following the terminology used in the International Surface Temperature Initiative (ISTI, Thorne et al., 2011), we term the locations of inhomogeneities as change points. The values obtained from PHA which would be used to adjust the data to make it homogeneous are termed the adjustment values. In this study, the inhomogeneities are all steps or jumps, with no option for a gradual change or more complicated inhomogeneities.

We outline our method of homogeneity assessment of HadISD on monthly scales in Sect. 2 and present our results on temperature, dewpoint temperature, SLP and wind speed in Sects. 3 and 4. We describe how we will make these data available in Sect. 6 and also outline an example of how these detected change points can be used in a scientific application in Sect. 7. We summarize in Sect. 8.

2 Homogenisation on monthly scales

We have identified the homogeneous sub-periods within the record of each station so that users can select those stations with few change points or small adjustment values when doing sensitive studies. For other analyses, those stations with many or large change points could also be included. As there are 6103 stations in the HadISD, an automated algorithm is required to perform the homogenisation. It had been hoped to run a number of different homogenisation algorithms on HadISD to be able to compare the change point locations and magnitudes. However, the requirement of an automated system which would work reliably on all 6103 HadISD stations limited this study to using the pairwise homogenisation algorithm (PHA) from Menne et al. (2009) and Menne and Williams Jr. (2009).

PHA has been used on NCDC's US Historical Climate Network monthly surface temperature record, and subsequently applied to the Global Historical Climate Network (GHCN) (Lawrimore et al., 2011) and more recently to surface humidity measurements

CPD
10, 1567–1607, 2014

(Willett et al., 2013, 2014). PHA has been designed to run on large networks of stations in an automated fashion, and hence suits the requirements of homogenising HadISD. In the benchmarking analysis from the COST-HOME project where change point locations and magnitudes were known to the coordinators but not the testers, Venema et al. (2012) showed that this algorithm had a low false-alarm rate, in other words few erroneous change points were returned by PHA in locations where none were present. But conversely, it was found to be more conservative in detecting true change points than other algorithms. Using a pairwise approach, testing each candidate-neighbour pair, is also more robust than using a candidate-station vs. (composite)-reference-station approach, as the latter can easily miss network-wide changes, or wrongly attribute them to a single station (Willett et al., 2013).

The PHA has also been subjected to an intensive benchmarking assessment for the US network (Williams et al., 2012) which showed that in all cases it reduced errors in the data without over-adjustments. In our application here, we are less concerned about the estimated values of the adjustments, as for the moment, these are not applied to the data. However the robustness of the change point locations is important, as is their number. As with all change point detection algorithms, PHA is unable to detect the smaller changes, resulting in a "missing middle" in the distribution of adjustment values. Hence the mean absolute adjustment value is over-estimated.

We outline the steps used by the PHA algorithm to find the change point locations and adjustment values:

1. For each candidate station, neighbouring stations which have the highest correlating monthly mean time series are selected (up to 40, more than seven required by PHA[1]). Stations where insufficient neighbours were found are not processed by PHA.

[1] In some cases fewer neighbours are allowed by the algorithm.

CPD
10, 1567–1607, 2014

2. The standard normal homogenisation test (SNHT, Alexandersson, 1986) is used iteratively on the candidate-neighbour difference series to locate the change points. These are noted in the candidate and the neighbour stations.

3. The resulting large array of potential change points is resolved iteratively to determine which station is common to most change points at a given date, and so is the cause of the change point. The final date of the change point is calculated from all of the neighbour pairs.

4. The change point is assessed to see if it is part of a local trend or a step change. If it is a step change, and a reliable magnitude can be determined, then this is applied to the monthly time series. Else the data is removed for the period by PHA when producing the adjusted monthly series[2].

The PHA code works on monthly data, and so the hourly data from HadISD have been converted to monthly values. The seasonal cycle is automatically removed by PHA and the monthly actuals are converted to anomalies. There are a number of different monthly quantities which could be used when assessing the homogeneity of the dataset. Apart from standard monthly mean values, the mean diurnal range has been used for temperature by Wijngaard et al. (2003) and in some cases change points are clearer than in the mean temperatures. Monthly mean maximum and minimum temperatures were used by Trewin (2013) when homogenising the Australian Climate Observations Reference Network – Surface Air Temperature (ACORN-SAT) dataset. PHA as yet cannot homogenise using two variables at the same time, e.g. monthly mean temperature and monthly mean diurnal temperature range. Hence, to produce a final dataset with internally consistent change point locations we will have to combine change points from different methods in a post-processing algorithm. To keep things relatively simple but yet be able to incorporate the extra power of these additional monthly values for temperature and dewpoint temperature outlined above, we

[2]We do not use the adjusted monthly series output by PHA in this work, but rather just the adjustments and their dates.

CPD
10, 1567–1607, 2014

use monthly mean diurnal ranges along with the monthly means for the homogeneity assessment.

The variation in sea-level pressure is sufficiently small that we will only use the monthly mean values for this variable. However, for wind speeds we shall use the monthly-mean daily-maximum wind speed as well as the monthly mean wind speed. When two methods (monthly mean and either monthly mean diurnal range or monthly mean maximum value) are used, we merge change points together if they occur within one year of each other, and use the mean date in the final products, rounded to the first day of the month.

3 Temperatures

To allow PHA to process the HadISD stations, the hourly data were converted to monthly means. First, daily means were created for all days which had more than four observations spread over at least a 12 h time-span. If there were more than 20 qualifying days within a month, then the monthly mean was calculated. These completeness criteria are the same for calculating the monthly mean diurnal temperature range.

The number of change points found using each of these two quantities are shown in Table 1 along with the number per station and other details about the homogenisation process.

In the COST-HOME analysis it was found that the PHA is conservative and identified fewer change points than other algorithms. In cases where none were present, this results in a low false-alarm rate, but it also means that the number of change points could be under-estimated in other circumstances. We merge the dates for the change points found using the monthly mean and monthly mean diurnal range if they are within one year of each other. PHA found 12 973 change points found in 4645 stations, a mean value of 2.79 change points per station over the 41 year period. This is roughly one every 15 years, which is at the lower end of the range found found by Menne et al. (2009) for the USHCN of one every 15–20 years.

There are 252 temperature stations which could not be processed by PHA, and a further 1206 have no change points detected (Fig. 1a). The majority of stations (4645/6103) have at least one change point and so an homogeneity assessment is an important part of any analysis to ensure that no spurious results arise because of non-climatic changes. In Sect. 7 we show for one application that coverage is a greater issue than station quality, but this will not always be the case.

There is no pattern to the stations which could not be processed by PHA. Small clusters can be found in South America and Africa, which may be due the completeness requirements when calculating monthly values from the sub-daily HadISD, resulting in short or sporadic monthly series. The stations with no change points are mainly found in Eurasia, with concentrations in Germany, Northern European Russia and Ukraine. These concentrations can also be seen in Fig. 1b, which shows the number of change points found in the 5851 temperature stations which were processed by PHA. Clusters of stations with particularly large numbers of breaks are seen along the US coasts, Italy, and the Maritime Continent. In the US, the areas correspond to the most populated parts of the country, and so greater than mean number of change points may arise because of repeated station moves or more zealous improvements to station instruments. Europe, especially the aforementioned regions from Fig. 1a, has large regions which have relatively few change points.

Figure 2 shows the distribution of the adjustment values for each of the four methods. Under the assumption that the underlying true adjustment values follow a Gaussian distribution, we do not detect those change points which have very small adjustment values. This "missing middle" is also seen when using other homogenisation processes. However, these plots show that change points with adjustment values down to around 0.5 °C have been found. The means of the distributions are very close to zero, but adjustment values based on the monthly mean have a slight positive bias and for the diurnal range have a slight negative bias. The spreads of the distributions are also very similar. For the missing change points (the blue histograms in Fig. 2), the mean values are very close to zero, with a relatively similar standard deviation for both methods.

CPD
10, 1567–1607, 2014

The largest adjustment values constitute only a very small faction of the total population (below three and four per cent for the mean and diurnal range respectively). The occurrence of change points in time appears to be relatively constant over the period of record, but with a possibly larger number occurring in the mid-to-late 1990 s (Fig. 3). Note that change points cannot be detected within two years of the start or end of a data series.

Unsurprisingly few change points are found in stations with very short records (Fig. 4). Many of the stations within HadISD have records which are 41 years long, and the most common number of change points for these stations is two. The distribution of the number of change points has a relatively smooth decay from the 1333 stations which have only one change point detected (excluding those on which PHA was not able to run) to the four which have 11 change points (485650-99999, 722265-13821, 723235-13896 and 725825-24121). There is a concentration of stations which have 41 years of record, and between zero and four change points, which matches the average number quoted above. It is also clear that there are fewer change points for stations with medium length records (25–40 years) than those with shorter records (< 25 years).

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). These regions also tend to have greater spatial and temporal variability than the tropics. These distributions also show that PHA is more able to find change points with small amplitudes in regions which have a high station density. This may be due to the close proximity of a large number of neighbouring stations, allowing for a wide choice of highly correlating neighbours resulting in higher signal-to-noise difference series. With longitude, there appears to be a bias towards larger positive adjustment values in the Americas. In both panels of Fig. 5 it should be noted that in areas with more stations there is a greater chance of finding change points with very high or very low adjustment values because of the increased number. In Fig. 2a there is a broader positive wing for the monthly mean temperature calculation. In contrast, in the mean diurnal

CPD
10, 1567–1607, 2014

range, there is a negative bias, and this contrast is apparent when comparing the two panels in Fig. 6. As the largest adjustment values have the most intense colour, the effect of these biases in the wings of the distribution stand out in Fig. 6. Particularly large positive adjustment values are observed in the US and the area around the Adriatic Sea when using monthly mean temperatures. Smaller adjustment values are seen in Western Europe and Asia. When using the diurnal temperature range, clusters of negative adjustments are seen in Western Europe, and the western half of the US has larger adjustment values than the eastern half. Similar small biases were found in the USHCN by Menne et al. (2009, Fig. 6).

4 Dewpoint temperatures, sea level pressure and wind speeds

We also carry out an identical set of calculations for the dewpoint temperatures, sea-level pressure (SLP) and wind speed measurements within HadISD. For the SLP, we use the deviations from 1000 hPa when calculating the monthly mean values. We find 16 785 change points in 5051 stations for dewpoint temperatures (3.32 per station), 5658 change points in 2781 stations for SLP (2.03 per station) and 23 781 change points in 5496 stations for wind speeds (4.32 per station). This immediately shows that the wind speed records appear to be more inhomogeneous than the other variables, and the sea-level pressure ones more homogeneous.

The distributions of the adjustment values for the dewpoint temperatures and SLP are shown in Figs. 7 and 8 respectively. As for the temperature results (Fig. 2), the means are close to zero but have a small positive bias, but the diurnal range has a small negative bias. Again, under the assumption that the distribution of adjustment values is Gaussian, most of the change points have been detected down to a limit of around 0.5 °C or 0.5 hPa. The distribution of the adjustments proposed for the monthly mean dewpoint temperatures is much broader than that for the monthly diurnal range, whereas for the temperatures they were very close (Fig. 2). The means of the missing change points are all close to zero.

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

CPD
10, 1567–1607, 2014

Homogenous sub-periods of HadISD

R. J. H. Dunn et al.

The distribution of the number of change points per station against the record length for the dewpoint temperatures (Fig. 9) is very similar to that from the temperatures (Fig. 4). There is a concentration of stations with 41 year records and one to two change points, with a smooth decline to greater numbers of change points, with station 442840-99999 having 13 change points. The distribution for SLP is much flatter, as would be expected from Table 1. Two stations, 483270-99999 and 577760-99999, have eight change points, and the decline is very steep from the 2364 stations which have no detected change points.

The distributions for the wind speeds are shown in Fig. 10. All of the mean adjustment values have a small negative bias, but are close to zero. Here, the distribution of the adjustments proposed for the monthly mean daily maximum wind speeds is broader than that of the monthly mean. Again, it is reasonable to deduce that most of the change points whose adjustment values are $> 0.5\,\mathrm{m\,s^{-1}}$ have been detected. The distributions of the missing adjustment values also have no strong biases. It is clear from Fig. 9 that relatively few stations with long records are break free. Most stations have around four change points over their record, with 161340-99999 having 13 change points in total. Across all variables, there are two regions in Figs. 4 and 9 which contain most of the stations, firstly, those with 41 years of record and having roughly the average number of change points for that variable. The other cluster is between 10 and 25 years, in a band which extends from zero change points up and right to longer station records and more change points, which is clearest in the wind speeds panel of Fig. 9. This follows the average increase in the number of change points within a station record as the record length increases.

The stations which were not processed by PHA cluster in southern and eastern Africa, western China and western South America (Fig. 11). The number of change points for dewpoints is relatively uniform across the globe, but large values are found for Italy, Korea and the southern tip of Argentina (Fig. 11). The largest number of change points for SLP are found in the Maritime Continent and South East Asia, whereas for wind speeds, large numbers are found in China through to Indochina, in the eastern

CPD
10, 1567–1607, 2014

USA and in the south of Argentina. There do not appear to be any correlations with geographic features for any of the variables.

Compared to the temperatures, the dewpoint temperatures have a more even distribution of adjustment values across the globe (Fig. 12). The adjustment magnitudes are on average larger for the monthly mean dewpoint temperatures than for the monthly mean dewpoint ranges. Again, the western half of the US appears to have larger adjustment values than the eastern half, but the region around the Mediterranean and also Scandinavia also stand out.

There are fewer stations which have detected change points in SLP, with a reduction in the station number being clearest in Africa and western China. Positive adjustments are seen in the eastern USA, in Siberia and down through China into the Maritime Continent, with other smaller clusters in Europe. Negative adjustments are seen in Central and South America.

The pattern of adjustment values for the wind speeds is not strong when using the monthly mean values (Fig. 14). There is a fairly uniform mix of positive and negative adjustment values, but with clusters of more positive values in the USA and more negative ones through Eurasia. When using the monthly mean daily maximum wind speed, the adjustment values are larger.

5 Validation

To validate whether the change point locations detected correspond with documented breaks we look at subsets of the 153 UK stations in the HadISD database, of which 102 contain the 196 change points in the temperature observations. Of these stations 18 have been merged during the creation of HadISD, and contain 31 change points in total. Although great care was taken when selecting stations to merge, it is likely that in some cases this process has created or left a discontinuity in the station record. The locations of change points identified from the temperature records were compared to the dates where input station identifiers changed. In six stations these dates were

CPD
10, 1567–1607, 2014

in close agreement accounting for eight of the change points (Table 2). This indicates that the change in input station identifiers may have represented a change of observing protocol or instruments, albeit at the same site.

There have been changes in the station reporting accuracy (between 0.1 °C and single degree precision) over time as stations were up- or downgraded from climatological stations. Of the 153 stations, 34 had no change in reporting accuracy, 13 had changes in reporting accuracy corresponding to a change point, and the remaining 106 had changes in reporting accuracy but no corresponding change point was detected by PHA.

As station metadata are more complete in later years, we initially focus on the change points detected in UK stations only after 2000, 35 stations containing 45 change points. Seven change points were close (within 12 months) in date to notes in the metadata, detailed in Table 2.

We also looked at the 25 UK stations which had three or more change points (some of which overlapped with the 35 stations with change points post 2000) and account for 105 change points. Each of the five change points detailed in Table 2 are within 12 months of the change noted in the station metadata, though the dates are never very close. Finally, the 29 UK stations which have two change points were also assessed (see Table 2), resulting in five stations where change points could be linked to metadata information, accounting for seven change points. However, metadata was not available for all stations, and in others did not appear to cover the years in which change points were identified. In most cases change point dates have no corresponding change noted in the station metadata, or the metadata itself are not sufficiently complete.

Although this investigation – which has accounted for 40 (20 %) of the 196 change points – demonstrates that using the PHA to check for inhomogeneities does find change points which could correspond to documented changes in some stations, there are still many undocumented change points which have no explanation from the station metadata. In fact most of the change points proposed by PHA do not have any corresponding change documented in the station metadata.

6 Provision of the change points

The change point dates and values determined by this study will be made available on the HadISD webpage (http://www.metoffice.gov.uk/hadobs/hadisd/) as text files in the first instance. These are easily readable by computer and human and so can be quickly implemented into analysis schemes. This homogeneity assessment will become part of the annual update of HadISD, so that change points are available for the most recent version of the dataset. In due course the information will also be included in the netCDF files created when HadISD is updated on an annual basis. Diagnostic information and the relevant plots will also be included on the website as well as the analysis scripts where possible.

7 Example application: global temperatures

As the adjustments calculated for each change point have not been applied to the data during the course of this work, there are a number of choices left up to the user as how best to use this information. One option, outlined here, is to progressively exclude poor stations, i.e. those with more change points or larger adjustment values, beginning at the “worst”. Alternatively, the “best” stations can be selected first, those where PHA did not find a change point, then including progressively “worse stations”. This would allow the effect of inclusion of heterogeneous stations to be assessed to see whether this has a significant effect on the final results. However, for global analyses, the coverage will change as more stations are added in, and this will need to be taken into account in the assessment.

What we will do here is inspired by the work of Callendar (1938, 1961) which recently celebrated its 75th anniversary (Hawkins and Jones, 2013). Callendar (1938) used a relatively small number of stations (147) to estimate the global land-surface air-temperature record in the early twentieth century, and his results agree very well with the latest global land-surface air-temperature datasets (Hawkins and Jones, 2013). So,

CPD
10, 1567–1607, 2014

if our stations with the largest adjustment values are excluded, then the global mean temperature should be at least as accurate, because Callendar's work suggest that the spatial sampling error is insensitive to the loss of a few stations when there are $\gg 150$ (widely distributed) stations. We calculate the gridded global temperature series from a number of subsets of the full HadISD station listing and compare this to CRUTEM4 (Jones et al., 2012).

To calculate the global temperature series from HadISD, firstly, daily mean temperatures are obtained, requiring that there are at least four observations in a day, spread over at least 12 h. Monthly mean temperatures are calculated if there are at least 20 qualifying days within a month. A climatology is calculated over 1975–1994 requiring at least 16 years to be present, and this is used to calculate monthly anomalies. All stations' anomalies within each $5 \times 5^{\circ}$ grid box are averaged on a monthly basis, producing a set of gridded monthly anomaly fields. If there are more than eight valid months present, then annual mean anomalies are calculated, and finally, a cosine weighted global mean temperature series is calculated. The gridded monthly CRUTEM4 anomaly fields are also converted to annual global anomalies relative to 1975–1994, matching the coverage of the gridded HadISD data in each year.

We firstly show the results of the full 6103 stations, and compare this to versions where we have only taken stations which have maximum adjustment values (in any of the four calculation methods) of less than 2 °C, 1 °C and 0.5 °C. We initially place no restrictions on the number of change points within any station series. In the upper panels of Fig. 15 we show the global trend from HadISD in black, the coverage-matched CRUTEM4 in blue, and the full CRUTEM4 in red, along with the range determined from the uncertainty information given in CRUTEM4. The lower panels show the differences between HadISD and the matched CRUTEM4 in blue, and HadISD and the full CRUTEM4 in red.

There are two competing changes occurring in the four panels in Fig. 15. There is an improvement in the homogeneity of the stations used for the global assessment, but with this, there is a reduction in the number of stations available, and hence the global

CPD
10, 1567–1607, 2014

coverage. By using CRUTEM4 as a comparison, it is possible to show how close the HadISD version is to a "truth". There are two comparisons to be made, one to the full coverage of CRUTEM4, shown in red, and one to CRUTEM4 where the coverage has been matched to that of the HadISD sample, shown in blue.

Focusing firstly on the difference between HadISD and the matched CRUTEM4, by restricting the stations to those with smaller and smaller inhomogeneities, this difference reduces, especially from 1996 onwards. But when only stations with adjustments $< 0.5\,^{\circ}$C are retained the differences start to increase again. We also fit linear trends using the median of pairwise trends method of Sen (1968). These are shown in the top left. The linear trends also become closer as the stations are restricted, but the uncertainty in the linear trends increases as the number of stations reduce. Using the RMS error as a measure of the difference of the global mean between HadISD and the two versions of CRUTEM4 shows this more clearly. When comparing to a matched CRUTEM4, the $e_{\rm RMS}$ reduces to a minimum when adjustments are restricted just to $< 1\,^{\circ}$C but then increases again. Despite matching the coverage of CRUTEM4 to that of the gridded HadISD, the differences increase, indicating that there are changes to individual grid box values which become more important as the station number reduces. The difference to the full CRUTEM4 remains steady when restricting to $< 2\,^{\circ}$C, but then increases thereafter. This shows that when trying to obtain a global picture, the coverage is an important factor, and that station quality does not have as much of an impact. In fact, when taking only the 1458 stations in which no change points were, or could be, detected the linear trends still agree within the uncertainties, but the $e_{\rm RMS}$ are large, especially in the last decade of data (Fig. 16).

In the set of four versions shown in Fig. 17, we keep the maximum value of adjustment allowed fixed at $< 1\,^{\circ}$C, as this has the lowest $e_{\rm RMS}$ between HadISD and the matched CRUTEM4 from the analysis of the adjustment values above. However, we place restrictions on the number of change points that occur within the record of the station, from five to one. In this case, despite the increase in average station quality as the number of change points are restricted, the $e_{\rm RMS}$ between HadISD and both

CPD
10, 1567–1607, 2014

versions of CRUTEM4 increase, with the clearest deviations relative to the matched CRUTEM4 being visible in the post-2005 period.

What is clear from all eight versions shown in Figs. 15 and 17 is that the linear trends all agree, within the uncertainties. Although by excluding more and more inhomogeneous stations will improve the fidelity of the remaining stations, the reduction in coverage will eventually cause larger deviations from the true underlying value. The exact point where these two competing effects balance will depend on the observed variable, and also the study being performed. Also, as the number of valid grid boxes reduces, differences in individual grid boxes become more prominent, leading to the increase in the scatter between the matched HadISD and CRUTEM4 versions. Therefore, as including the stations which have many change points or large adjustment values does not have a large effect on large scale analyses, it may be best to use all the data available rather than worry too much about station quality, depending on the application.

8 Summary

In this work we have started the process of homogenising HadISD, a sub-daily, multivariate, station-based dataset covering 1973–2013. Using the PHA homogenisation code of Menne and Williams Jr. (2009) we have determined the locations of change points on a monthly scale using the monthly mean values, diurnal ranges and maximum values for the temperature, dewpoint temperature, sea-level pressure and wind speed variables. Change point locations have been combined when they occur within a year of each other. The final number of stations which could be processed by PHA along with the average change point properties are given in Table 1. We show that there is no strong bias in the adjustments proposed by any of the methods for all of the variables, and this is also true for the distribution of change points we have not been able to detect (assuming that the true underlying population is Gaussian). There are some geographical patterns in the stations which could not be homogenised, or in which no

CPD
10, 1567–1607, 2014

breaks were found. The main concern regarding the former is the lack of sufficient target-station and neighbour-station data.

We use the change point locations and adjustment magnitudes to guide alternative estimates of global temperature from HadISD stations, and compare these to CRUTEM4. Removing the most inhomogeneous stations results in an improvement in the scatter between global mean land-surface air-temperature from HadISD and CRUTEM4 with matched coverage. However, further removals of stations with smaller and fewer inhomogeneities increases the scatter.

Future work will focus on detecting change points on a daily level, with then the application of adjustments onto the hourly data. Daily homogenisation of maximum and minimum temperatures has already been successfully accomplished (e.g. Vincent et al., 2002; Trewin, 2013). However, the issue of automated scaling of adjustment magnitudes across all hours of the day has not yet been solved.

Acknowledgements. The authors were supported by the Joint DECC/Defra Met Office Hadley Centre Climate Programme (GA01101).

RJHD thanks Claude Williams and Victor Venema for helpful discussions during the course of this work and Colin Chatters of the Land Networks Team for providing the UK station metadata.

References

Alexandersson, H.: A homogeneity test applied to precipitation data, J. Climatol., 6, 661–675, 1986. 1569, 1573

Brandsma, T. and Können, G.: Application of nearest-neighbor resampling for homogenizing temperature records on a daily to sub-daily level, Int. J. Climatol., 26, 75–89, 2006. 1570

Callendar, G. S.: The artificial production of carbon dioxide and its influence on temperature, Q. J. Roy. Meteorol. Soc., 64, 223–240, doi:10.1002/qj.49706427503, 1938. 1581

Callendar, G. S.: Temperature fluctuations and trends over the earth, Q. J. Roy. Meteorol. Soc., 87, 1–12, doi:10.1002/qj.49708737102, 1961. 1581

Della-Marta, P. and Wanner, H.: A method of homogenizing the extremes and mean of daily temperature measurements, J. Climate, 19, 4179–4197, 2006. 1570

Domonkos, P.: Adapted Caussinus-Mestre Algorithm for Networks of Temperature series (ACMANT), Int. J. Geosci., 2, 293–309, 2011. 1569

Dunn, R. J. H., Willett, K. M., Thorne, P. W., Woolley, E. V., Durre, I., Dai, A., Parker, D. E., and Vose, R. S.: HadISD: a quality-controlled global synoptic report database for selected variables at long-term stations from 1973–2011, Clim. Past, 8, 1649–1679, doi:10.5194/cp-8-1649-2012, 2012. 1570

Easterling, D. R. and Peterson, T. C.: A new method for detecting undocumented discontinuities in climatological time series, Int. J. Climatol., 15, 369–377, 1995. 1569

Hawkins, E. and Jones, P. D.: On increasing global temperatures: 75 years after Callendar, Q. J. Roy. Meteorol. Soc., 139, 1961–1963, doi:10.1002/qj.2178, 2013. 1581

Jones, P., Lister, D., Osborn, T., Harpham, C., Salmon, M., and Morice, C.: Hemispheric and large-scale land-surface air temperature variations: an extensive revision and an update to 2010, J. Geophys. Res., 117, D05127, doi:10.1029/2011JD017139, 2012. 1582

Kuglitsch, F. G., Toreti, A., Xoplaki, E., Della-Marta, P. M., Luterbacher, J., and Wanner, H.: Homogenization of daily maximum temperature series in the Mediterranean, J. Geophys. Res., 114, D15108, doi:10.1029/2008JD011606, 2009. 1570

Lawrimore, J. H., Menne, M. J., Gleason, B. E., Williams, C. N., Wuertz, D. B., Vose, R. S., and Rennie, J.: An overview of the global historical climatology network monthly mean temperature data set, version 3, J. Geophys. Res., 116, D19121, doi:10.1029/2011JD016187, 2011. 1571

Menne, M. J. and Williams Jr., C. N.: Homogenization of temperature series via pairwise comparisons, J. Climate, 22, 1700–1717, 2009. 1568, 1571, 1584

Menne, M. J., Williams Jr., C. N., and Vose, R. S.: The US historical climatology network monthly temperature data, version 2, B. Am. Meteorol. Soc., 90, 993–1007, 2009. 1569, 1571, 1574, 1577

Mestre, O., Gruber, C., Prieur, C., Caussinus, H., and Jourdain, S.: SPLIDHOM: A method for homogenization of daily temperature observations, J. Appl. Meteorol. Clim., 50, 2343–2358, 2011. 1569

Mestre, O., Domonkos, P., Picard, F., Auer, I., Robin, S., Lebarbier, E., Böhm, R., Aguilar, E., Guijarro, J., Vertachnik, G., Klancar, M., Dubuisson, B., and Stepanek, P.: HOMER: a homogenization software–methods and applications, Q. J. Hung. Meteorol. Serv., 117, 47–67, 2013. 1569

Perreault, L., Bernier, J., Bobée, B., and Parent, E.: Bayesian change-point analysis in hydrometeorological time series, Part 1. The normal model revisited, J. Hydrol., 235, 221–241, 2000a. 1569

Perreault, L., Bernier, J., Bobée, B., and Parent, E.: Bayesian change-point analysis in hydrometeorological time series, Part 2. Comparison of change-point models and forecasting, J. Hydrol., 235, 242–263, 2000b. 1569

Sen, P. K.: Estimates of the regression coefficient based on Kendall's Tau, J. Am. Stat. Assoc., 63, 1379–1389, 1968. 1583

Smith, A., Lott, N., and Vose, R.: The integrated surface database: recent developments and partnerships, B. Am. Meteorol. Soc., 92, 704–708, 2011. 1570

Szentimrey, T.: Multiple analysis of series for homogenization (MASH), in: Proc. Second Seminar for Homogenization of Surface Climatological Data, WCDMP-No. 4, WMO, Budapest, 27–46, 1999. 1569

Szentimrey, T.: Manual of homogenization software MASHv3.02, Hungarian Meteorological Service, Budapest, Hungary, 65 pp., 2007. 1569

Thorne, P. W., Willett, K. M., Allan, R. J., Bojinski, S., Christy, J. R., Fox, N., Gilbert, S., Jolliffe, I., Kennedy, J. J., Kent, E., Klein Tank, A., Lawrimore, J., Parker, D. E., Rayner, N., Simmons, A., Song, L., Stott, P. A., and Trewin, B.: Guiding the creation of a comprehensive surface temperature resource for twenty-first-century climate science, B. Am. Meteorol. Soc., 92, ES40–ES47, 2011. 1571

Toreti, A., Kuglitsch, F. G., Xoplaki, E., Luterbacher, J., and Wanner, H.: A novel method for the homogenization of daily temperature series and its relevance for climate change analysis, J. Climate, 23, 5325–5331, 2010. 1570

Trewin, B.: A daily homogenized temperature data set for Australia, Int. J. Climatol., 33, 1510–1529, 2013. 1570, 1573, 1585

Venema, V. K. C., Mestre, O., Aguilar, E., Auer, I., Guijarro, J. A., Domonkos, P., Vertacnik, G., Szentimrey, T., Stepanek, P., Zahradnicek, P., Viarre, J., Müller-Westermeier, G., Lakatos, M., Williams, C. N., Menne, M. J., Lindau, R., Rasol, D., Rustemeier, E., Kolokythas, K., Marinova, T., Andresen, L., Acquaotta, F., Fratianni, S., Cheval, S., Klancar, M., Brunetti, M., Gruber, C., Prohom Duran, M., Likso, T., Esteban, P., and Brandsma, T.: Benchmarking homogenization algorithms for monthly data, Clim. Past, 8, 89–115, doi:10.5194/cp-8-89-2012, 2012. 1569, 1572

Vincent, L. A., Zhang, X., Bonsal, B., and Hogg, W.: Homogenization of daily temperatures over Canada, J. Climate, 15, 1322–1334, 2002. 1570, 1585

Wijngaard, J., Klein Tank, A., and Können, G.: Homogeneity of 20th century European daily temperature and precipitation series, Int. J. Climatol., 23, 679–692, 2003. 1573

Willett, K. M., Williams Jr., C. N., Dunn, R. J. H., Thorne, P. W., Bell, S., de Podesta, M., Jones, P. D., and Parker, D. E.: HadISDH: an updateable land surface specific humidity product for climate monitoring, Clim. Past, 9, 657–677, doi:10.5194/cp-9-657-2013, 2013. 1570, 1572

Williams, C. N., Menne, M. J., and Thorne, P. W.: Benchmarking the performance of pairwise homogenization of surface temperatures in the United States, J. Geophys. Res., 117, D05116, doi:10.1029/2011JD016761, 2012. 1569, 1572

CPD
10, 1567–1607, 2014

Homogenous sub-periods of HadISD

R. J. H. Dunn et al.

Table 1. The number of stations used and the number which had too few neighbouring stations for each of the four variables. The number of change points detected for the calculation methods used for each variable, along with the number per station (excluding those with too few neighbours).

	Diagnostic	Temperature	Dewpoint	SLP	Wind Speeds
			Number of Stations		
input station number		6103	6103	6103	6103
not processed by PHA:	mean	255	276	958	265
	DR	262	273	–	–
	maximum	–	–	–	265
Not Tested		252	273	958	266
Tested		5851	5830	5145	5837
	No Change Points	1206	779	2364	341
	With Change Points	4645	5051	2781	5496
			Number of Change Points Detected		
	mean	6488	9903	5658	15 914
	DR	7744	8781	–	–
	maximum	–	–	–	15 092
Total Combined Change points		12 973	16 785	5658	23 781
Change Points/Station		2.79	3.32	2.03	4.33
Adjustment Magnitude*	mean	0.733	0.989	0.719	0.562
	DR	0.815	0.697	–	–
	maximum	–	–	–	0.849

* Although no adjustments were made, the values were still extracted.

CPD
10, 1567–1607, 2014

Homogenous sub-periods of HadISD

R. J. H. Dunn et al.

Table 2. Proposed change points from PHA which are close in date to changes in the station location or instrumentation as indicated in the metadata.

Station ID	Name	Change Point Date	Metadata Date	Type
		Merged Stations		
035580-99999	Bedford Airport	1 Nov 1994	1 Apr 1994	Merger
037010-99999	Gawlish	1 Mar 1983	1 Jun 1994	Merger[a]
		1 Aug 1984	1 Jun 1994	Merger[a]
		1 Dec 1988	1 Mar 1989	Merger
037260-99999	Bristol Weather Centre	1 Sep 2001	16 Sep 2001	Merger
038140-99999	Lizard	1 Mar 1988	19 Jul 1988	Merger
038560-99999	Portland Bill	1 Oct 1991	1 Mar 1992	Merger
038840-99999	Herstmonceux	1 Aug 1992	30 Nov 1992	Merger
		Change points post-2000		
030080-99999	Fair Isle	1 Oct 2004	15 Sep 2004	Instrument Change
032810-99999	Fylingdales	1 Mar 2009	11 Mar 2009	Instrument Change
033180-99999	Blackpool	1 Mar 2010	12 Feb 2010	New screen and instruments
033340-99999	Manchester Ringway	1 Apr 2005	1 Nov 2004	Instrument &Site Change
033730-99999	Scampton	1 Jul 2001	31 Jan 2001	Instrument Change
038390-99999	Exeter Airport	1 Dec 2009	3 Nov 2009	Station Move
039170-99999	Belfast Aldergrove	1 Aug 2003	24 Jan 2003	Station Move
		$\geq$ 3 Change points		
033180-99999	Blackpool	1 May 1991	1 Oct 1991	Instrument Change
		1 Jan 1994	1 Oct 1994	Instrument Change
031110-99999	Machrihanish	1 Jan 1993	8 Aug 1992	Instrument Change
036720-99999	Northolt	1 Oct 1994	12 Jan 1995	Instrument Change
038270-99999	Plymouth Mountbatten	1 Feb 1991	17 Jul 1991	Instrument Change
		2 Change points		
032570-99999	Leeming	1 May 1995	30 Nov 1995	Instrument Change[b]
		1 Jun 1996	30 Nov 1995	Instrument Change[b]
036040-99999	Milford Haven	1 May 1995	28 Apr 1995	Instrument Change
037170-99999	Cardiff Weather Centre	1 Feb 1991	8 Nov 1991	Instrument Change[b]
		1 Mar 1992	8 Nov 1991	Instrument Change[b]
037610-99999	Odiham	1 Apr 1993	2 Apr 1993	Instrument Change
038530-99999	Yeovilton	1 Mar 1995	10 Nov 1994	Instrument Change

[a] These two change points occur either side of a gap before which one station stops and after which the other starts. The change points have not been merged as they are too far apart. [b] For these stations the change points surround the date in the metadata, but have not been merged as they are more than 12 months apart.

CPD
10, 1567–1607, 2014

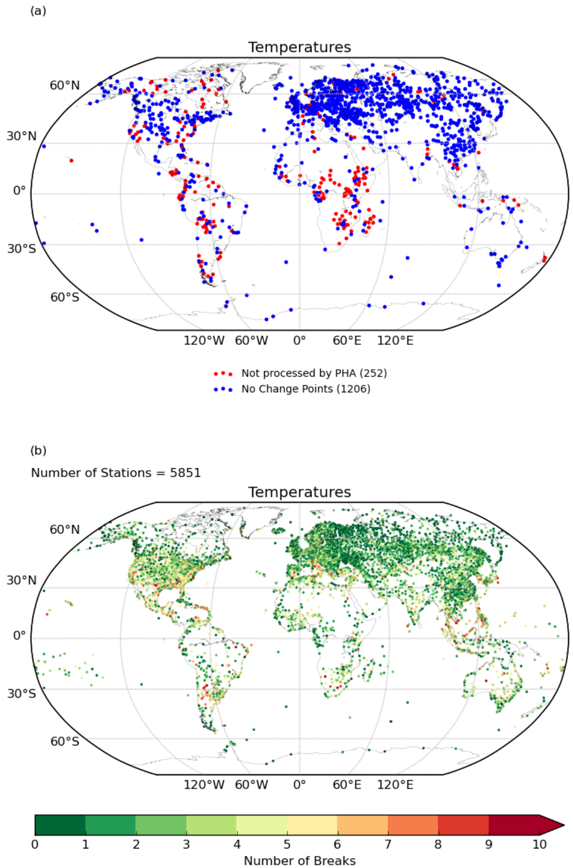

Fig. 1. (a) The location of stations which could not be processed by PHA (red) and those stations where no temperature change point was found in the entire record (blue). **(b)** The number of change points detected for each station.

CPD
10, 1567–1607, 2014

Homogenous sub-periods of HadISD

R. J. H. Dunn et al.

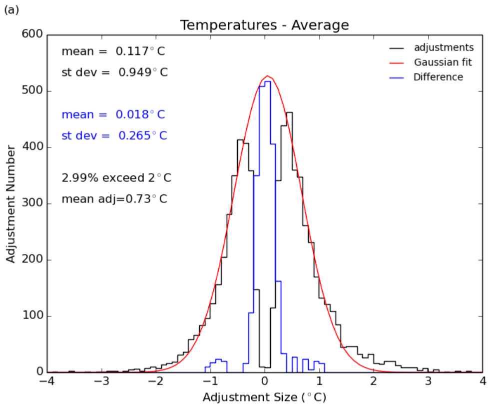

Fig. 2. The distribution of the adjustment values for monthly mean **(a)** temperatures and **(b)** diurnal temperature range. The distributions have been fitted with a Gaussian (red), and the difference between the data and the Gaussian is shown in blue. A positive step means that the earlier homogeneous period (before the break) is spuriously warmer or has a spuriously larger diurnal temperature range than the later homogeneous period (after the break).

CPD
10, 1567–1607, 2014

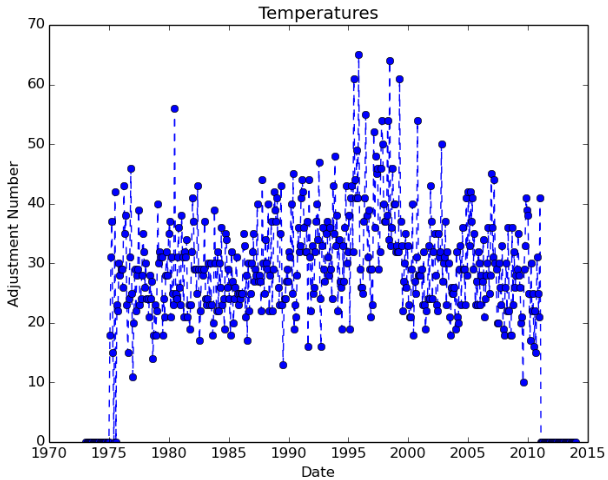

Fig. 3. The number of change points found in each year from both calculation methods combined.

CPD
10, 1567–1607, 2014

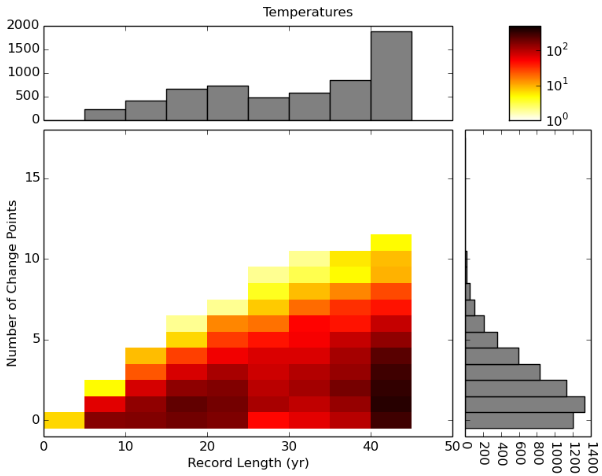

Fig. 4. The distribution of the number of change points within a station record against the number of years of data. The histograms on the top and to the right of the grid plot show the projections onto the *x* and *y* axes respectively. The colour bar is on a logarithmic scale.

CPD
10, 1567–1607, 2014

Homogenous sub-periods of HadISD

R. J. H. Dunn et al.

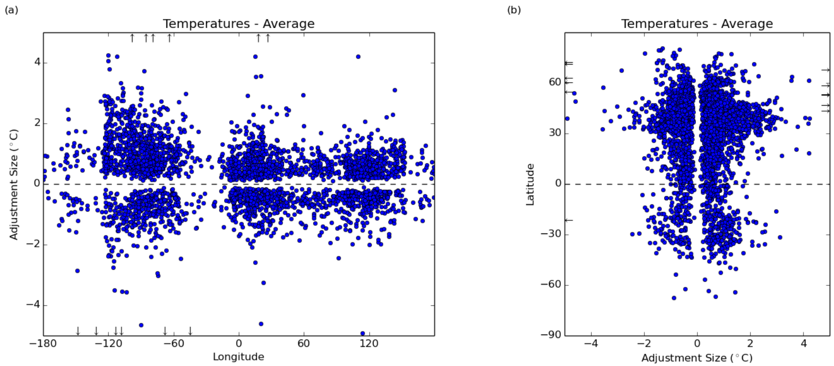

Fig. 5. The distribution of adjustment values with **(a)** Longitude and **(b)** Latitude using the monthly mean temperatures. Arrows indicate values which fall outside of the plotted area.

CPD
10, 1567–1607, 2014

Homogenous sub-periods of HadISD

R. J. H. Dunn et al.

Title Page
Abstract | Introduction
Conclusions | References
Tables | Figures
◄ | ►
◄ | ►
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

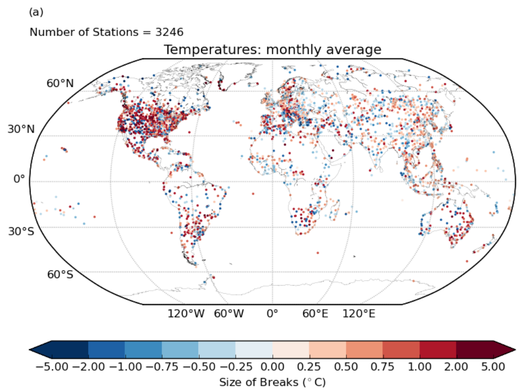

Fig. 6. The distribution of the largest adjustment value found for each station when using the **(a)** monthly mean temperature and **(b)** the monthly mean diurnal temperature range. Only those stations which have at least one change point are shown, resulting in the different numbers of stations for each panel. Note the non-linear colour scale.

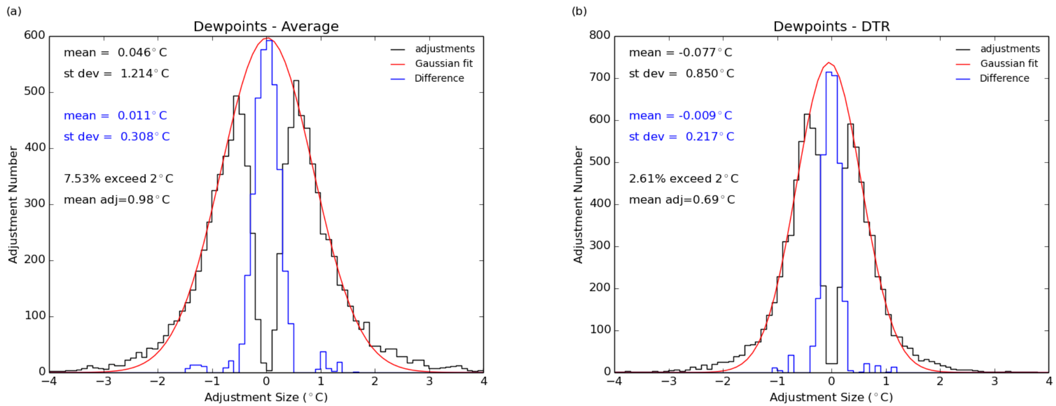

Fig. 7. The distribution of the adjustment values for monthly mean **(a)** dewpoint temperatures and **(b)** diurnal dewpoint temperature range. The distributions have been fitted with a Gaussian (red), and the difference between the data and the Gaussian is shown in blue.

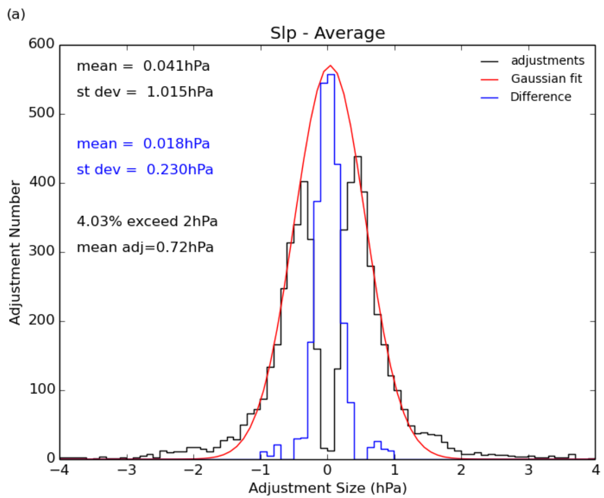

Fig. 8. The distribution of the adjustment values for monthly mean SLP. The distributions have been fitted with a Gaussian (red), and the difference between the data and the Gaussian is shown in blue.

CPD
10, 1567–1607, 2014

Homogenous sub-periods of HadISD

R. J. H. Dunn et al.

Title Page
Abstract | Introduction
Conclusions | References
Tables | Figures
◀| |▶
◀ ▶
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

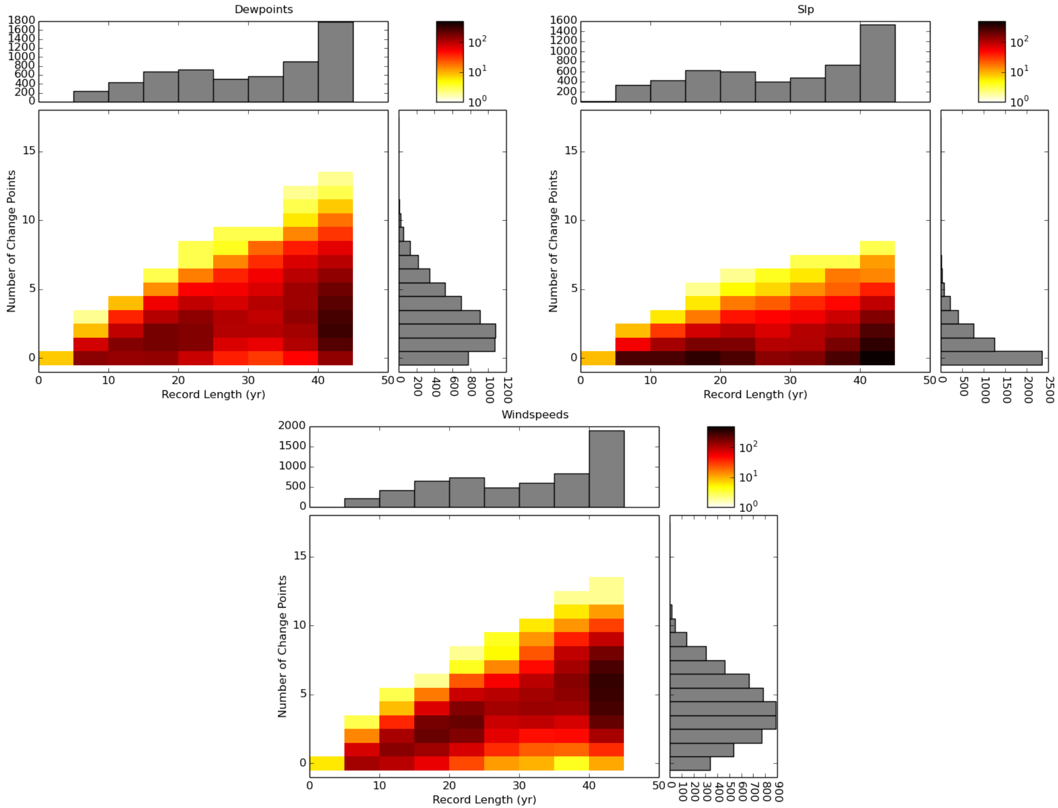

Fig. 9. The distribution of the number of change points within a station record against the number of years of data for left dewpoint temperature and right sea-level pressure and bottom wind speeds. The histograms on the top and to the right of the grid plot show the projections onto the x and y axes respectively. The colour bar is on a logarithmic scale.

CPD
10, 1567–1607, 2014

Homogenous sub-periods of HadISD

R. J. H. Dunn et al.

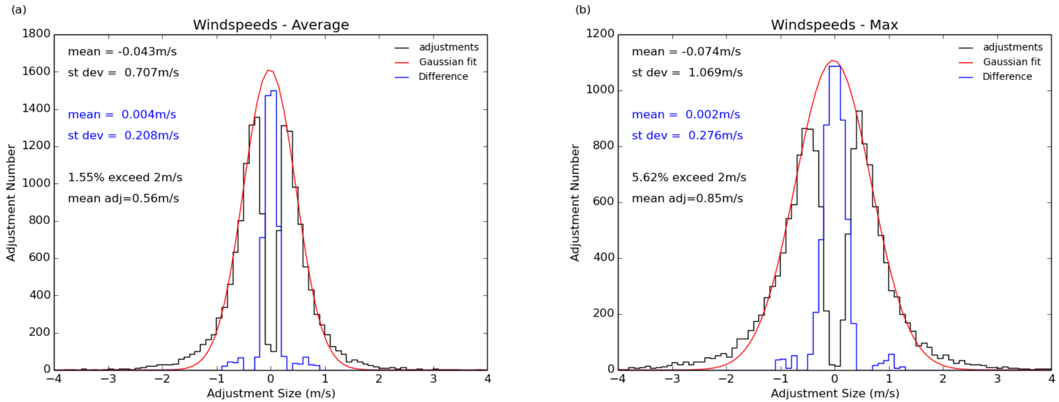

Fig. 10. The distribution of the adjustment values for monthly **(a)** mean wind speeds and **(b)** mean dialy maximum wind speed. The distributions have been fitted with a Gaussian (red), and the difference between the data and the Gaussian is shown in blue.

CPD
10, 1567–1607, 2014

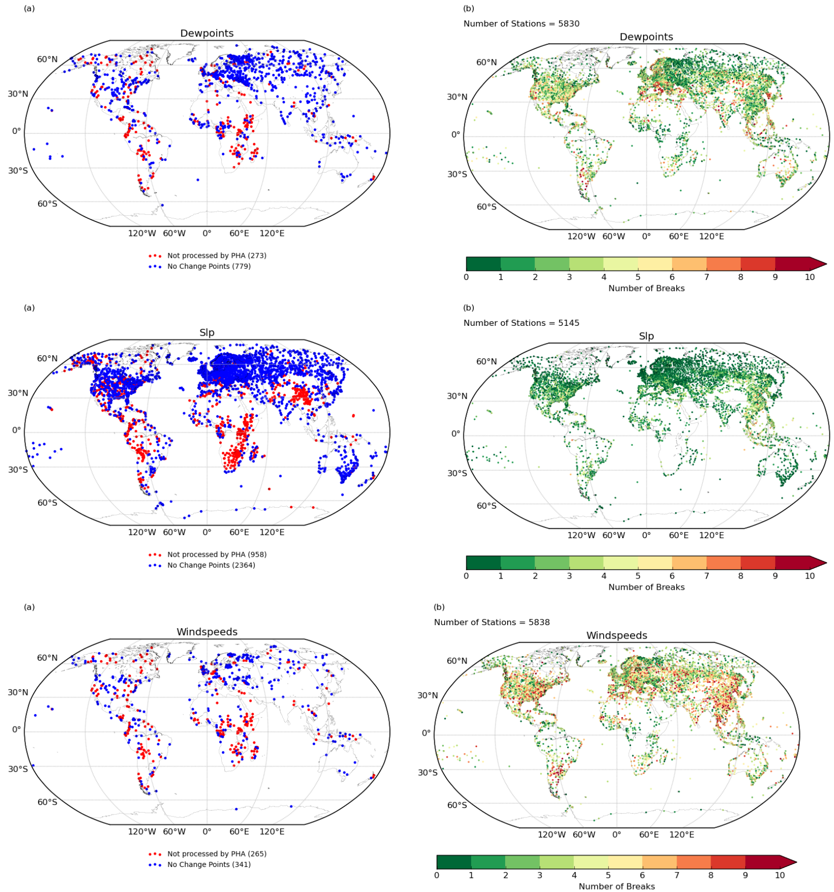

Fig. 11. Left panels: the location of stations which could not be processed by PHA (red) and those stations where no change points were found in the entire record (blue) for dewpoints (top panels), SLP (middle panels) and wind speeds (bottom panels). Right panels: the number of breaks detected for each station.

CPD
10, 1567–1607, 2014

Homogenous sub-periods of HadISD

R. J. H. Dunn et al.

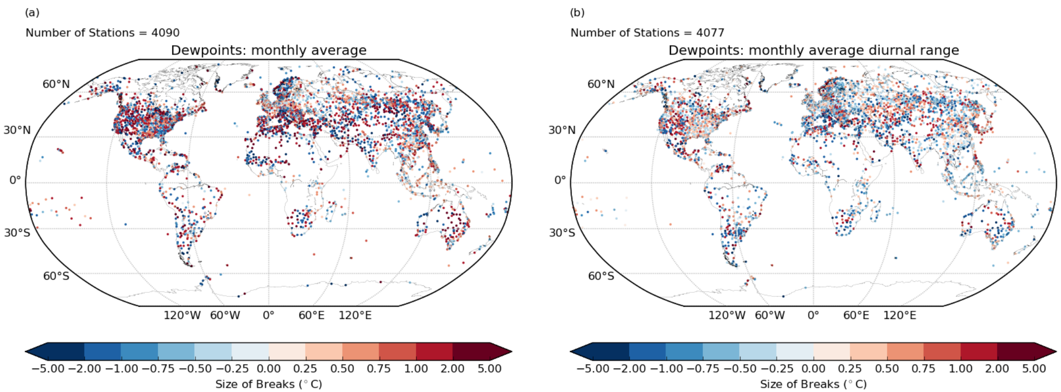

Fig. 12. As for Fig. 6 but for dewpoints.

CPD
10, 1567–1607, 2014

Homogenous sub-periods of HadISD

R. J. H. Dunn et al.

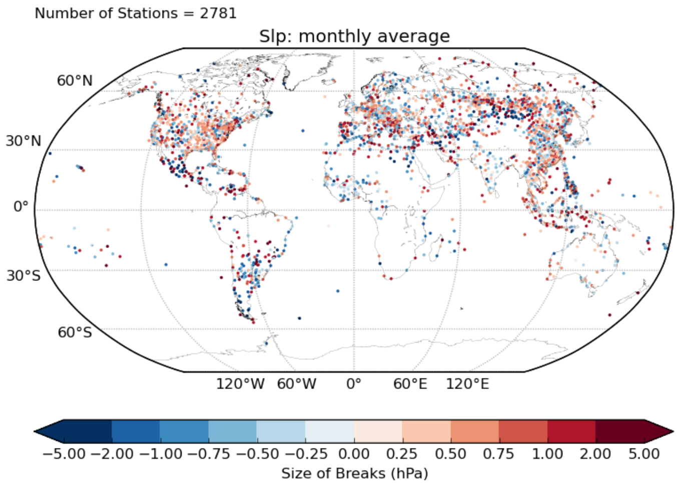

Fig. 13. As for Fig. 6 but for SLP.

CPD
10, 1567–1607, 2014

Homogenous sub-periods of HadISD

R. J. H. Dunn et al.

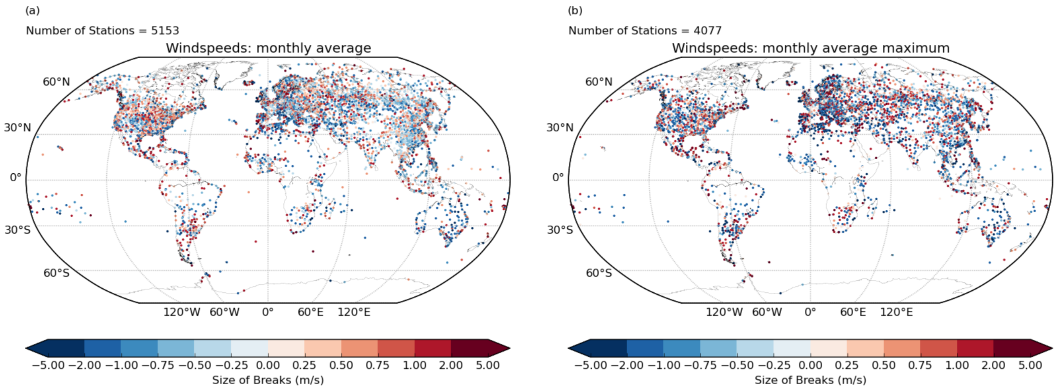

Fig. 14. As for Fig. 6 but for wind speeds.

CPD
10, 1567–1607, 2014

Homogenous sub-periods of HadISD

R. J. H. Dunn et al.

Title Page
Abstract Introduction
Conclusions References
Tables Figures
◀◀ ▶▶
◀ ▶
Back Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

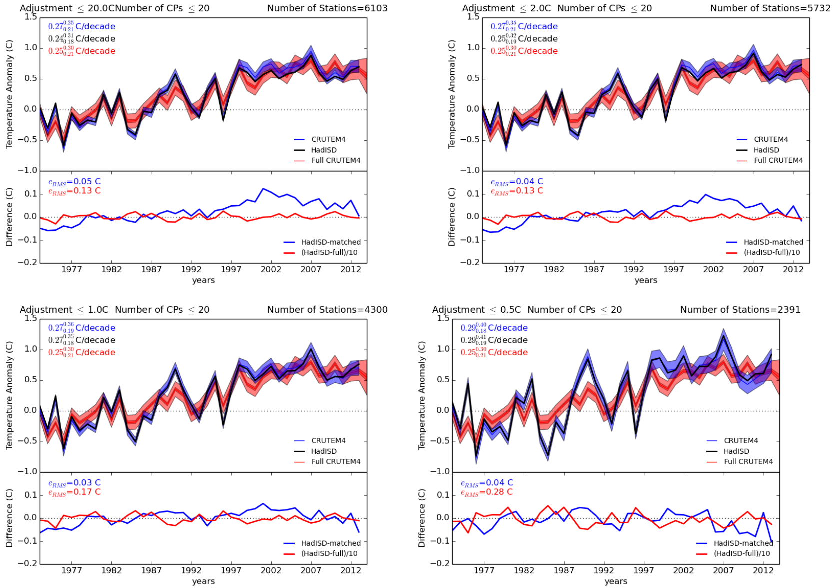

Fig. 15. The CRUTEM4 global temperature series (red) compared with a CRUTEM4 series matched to the HadISD coverage (blue) and the global series as calculated from HadISD (black). The shading surrounding the matched CRUTEM4 (blue) show the combined station, grid-box sampling and bias uncertainties. The dark shading for the full CRUTEM4 (red) shows the combined station and grid-box sampling uncertainties, with the light shading including the bias and coverage uncertainties as well. Top left panel: no restrictions on the values of the adjustments (all stations), top right panel: adjustments < 2 °C, bottom left panel: < 1 °C and bottom right panel: < 0.5 °C. The trends for each of the three curves are given in the top left of each panel, with the 5th and 95th values shown in the sub- and superscripts respectively, as calculated using the median of pairwise slopes algorithm. The bottom panels show the differences between HadISD and CRUTEM4 matched to the HadISD sample (blue) and HadISD and the full CRUTEM4 (÷10, red) along with the RMS errors.

CPD
10, 1567–1607, 2014

Homogenous sub-periods of HadISD

R. J. H. Dunn et al.

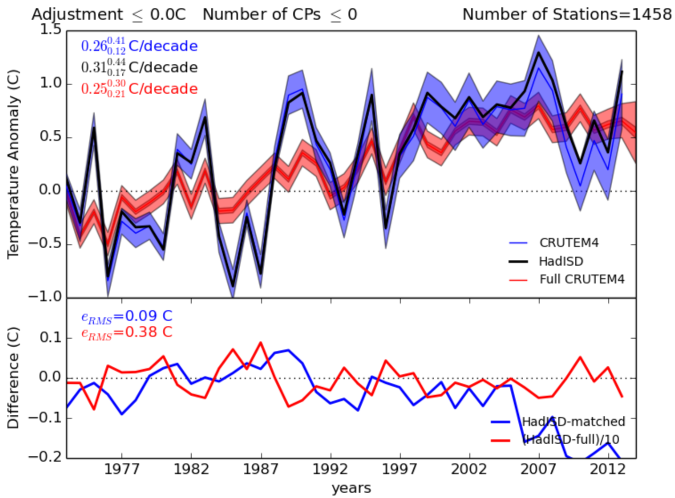

Fig. 16. As for Fig. 15, but for the 1458 stations with no detected change points.

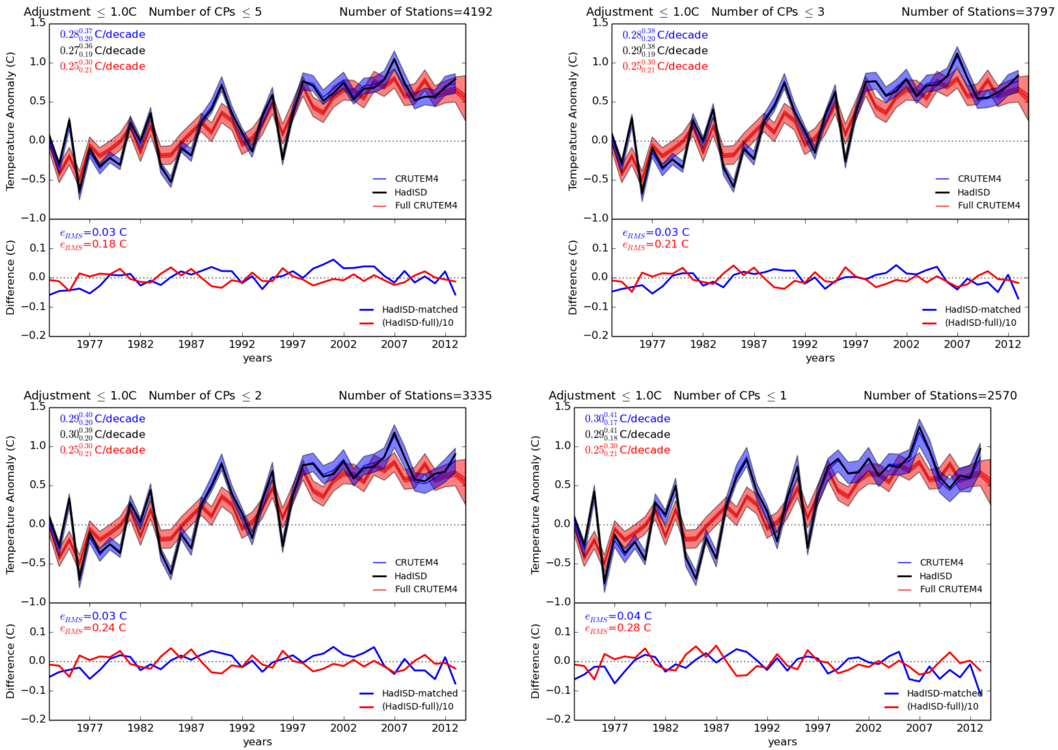

Fig. 17. As for Fig. 15, but for stations with adjustment values < 1 °C. Top left panel: fewer than five, top right panel: three, bottom left panel: two and bottom right panel: one adjustment in the full record of the station.

CPD
10, 1567–1607, 2014

Homogenous sub-periods of HadISD

R. J. H. Dunn et al.

Title Page
Abstract | Introduction
Conclusions | References
Tables | Figures
◀| |▶
◀ ▶
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion