



# Pairwise homogeneity assessment of 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

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

Received: 20 March 2014 – Published in *Clim. Past Discuss.*: 11 April 2014

Revised: 3 July 2014 – Accepted: 9 July 2014 – Published: 20 August 2014

**Abstract.** We report on preliminary steps in the homogenisation of HadISD, a sub-daily, station-based data set covering 1973–2013. Using temperature, dew point 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 dew point 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 1 year, they are combined and the average date is used. Under the assumption that the underlying true population of inhomogeneity magnitudes is Gaussian, inhomogeneity magnitudes as small as around  $0.5^{\circ}\text{C}$ ,  $0.5\text{ hPa}$  or  $0.5\text{ m s}^{-1}$  have been successfully detected. The change point dates and inhomogeneity magnitudes for each of the calculation methods will be provided alongside the data set 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 worsen the match.

instruments, incorrect station merges, changes in land use (urbanisation 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 inhomogeneity magnitudes. Absolute methods apply statistical tests and, if available, metadata to the target station alone. In relative methods, 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. Relative, multiple-breakpoint detection methods are on the whole more robust and successful at finding inhomogeneities than absolute (without reference series) methods (Venema et al., 2012). Classical statistical tests (e.g. the Standard Normal Homogeneity Test, SNHT; Alexandersson, 1986), regression models (e.g. Easterling and Peterson, 1995) or Bayesian approaches (e.g. 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 modern multiple-breakpoint detection packages which can be used to homogenise data sets, e.g. MASH (Multiple Analysis of Series for Homogenisation; Szentimrey, 1999, 2007), ACMANT (adapted Caussinus–Mestre algorithm for networks of temperature series; Domonkos, 2011), HOMER (HOMogenizatON softwarE in R; Mestre et al., 2013) and PHA (pairwise homogenisation algorithm; Menne et al., 2009). These all have different approaches to the homogenisation and adjustment problem. MASH is a relative method for medium-sized networks (a few hundred stations) using multiple references which are not assumed to be homogeneous. It uses an exhaustive approach, applying corrections until no further change point is found. ACMANT is based on the Caussinus–Mestre method from PRODIGE

## 1 Introduction

To enable the use of a data set 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

(Caussinus and Mestre, 2004) but uses a reference series rather than a pairwise comparison along with some improvements to the detection, correction and interpolation steps. HOMER was constructed to include the best characteristics of other state-of-the-art homogenisation methods after the COST-HOME action, including PRODIGE (Caussinus and Mestre, 2004), ACMANT, CLIMATOL (Guijarro, 2011) and *cghseg* (Picard et al., 2011). It is best suited to medium-sized networks where manual input is possible to finalise the change point locations. In contrast, PHA was specifically constructed to function automatically with large networks, and is described in detail in Sect. 2.

Given the variety of options available for homogenising data sets, the COST-HOME project (Venema et al., 2012; [www.homogenisation.org](http://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 data sets, 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 (SPLine Daily HOMogenization; 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.

HadISD is a new, sub-daily, station-based data set from the Met Office Hadley Centre. It is based on the National Climatic Data Center (NCDC) Integrated Surface Database (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 3-hourly data over the period 1973–2011. Temperature, dew point temperature, sea-level pressure (SLP), 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 (v1.0.2.2013f) running to the end of 2013. HadISD is an updated, quality-controlled data set, but it has not as yet been homogenised, although it the basis of the 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 1.0.2.2013f 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 data set.

The construction of HadISD and also the method employed herein can give rise to inhomogeneities beyond those usually associated with observational data sets (station moves, instrument changes, changes in the external environment). In order to increase the temporal coverage of stations in HadISD, stations were merged which passed certain criteria of similarity (see Dunn et al., 2012, for full details), but it is possible that station records have changed since this assessment was performed so that stations that are currently merged are no longer suitable candidates, e.g. station IDs being reassigned. Issues have also been found with un-documented station moves, which means that, although merging two station IDs is appropriate at one instance in time, it is not at later times (e.g. undocumented station moves). In this work we detect inhomogeneities on a monthly basis (see Sect. 3), and so the sub-daily data are averaged to monthly quantities. If the station reporting frequency changes, then this will have an impact on the daily, and hence monthly, quantities. Also daylight savings changes, which should have been accounted for when taking the observations and also in conversion to Coordinated Universal Time (UTC) by the Integrated Surface Database (ISD) may also have an effect where necessary steps could not be implemented.

Although the data in HadISD have been subject to a detailed suite of quality control tests, there will still be observations which are erroneous. In the case of dew point temperature, for example, if the wet-bulb thermometer wick persistently dries out for fractions of a day, then this will affect the homogeneity of the record. Other data quality issues which could cause inhomogeneities are conversion problems, for example from station level pressure to mean sea-level pressure or from knots to metres per second in wind speed, where the relevant factors have been applied twice. More detailed descriptions of homogeneity issues are discussed by e.g. Thomas et al. (2008), Jakob (2010) and Wan et al. (2010) for wind speed; by Vincent et al. (2007) for temperature and dew point; and by Jones et al. (1987), Young (1993) and Slonosky et al. (1999) for SLP.

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 are 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, but so far only for country-scale networks. In some cases change points are detected and inhomogeneity magnitudes determined on longer timescales (monthly or annual), and then the daily data are adjusted (Vincent et al., 2002; Kuglitsch et al., 2009; Mestre et al., 2011; Trewin, 2013). Others use detailed metadata (Auchmann and Brönnimann, 2012) or statistical methods

(Brandsma and Können, 2006; Della-Marta and Wanner, 2006; Yan and Jones, 2008; Toreti et al., 2010; Rienzner and Gandolfi, 2013) to detect inhomogeneities and also apply the adjustments on a daily basis. It is still uncertain how to apply these methods 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 data set of 6103 stations only and, for the moment, not apply any adjustments directly to the data. Future developments will detect change points on shorter timescales, and apply adjustments to the hourly HadISD data.

Following the widely accepted terminology, adopted 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. As we do not apply adjustments in this study, to avoid confusion we will use the term inhomogeneity magnitudes. 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, dew point 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 summarise 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 inhomogeneity magnitudes when doing sensitive studies. For example, when looking at extreme events (peaks over threshold) or small, restricted regions, the homogeneity information will allow the most homogeneous stations to be selected for use. For other analyses, those stations with many or large change points could also be included, for example when performing continent or climatic zone scale analyses. 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 PHA from Menne et al. (2009) and Menne and Williams Jr (2009).

PHA has been used on NCDC's US Historical Climatology Network (USHCN) monthly surface temperature record, and subsequently applied to the Global Historical Climatology Network (GHCN) (Lawrimore et al., 2011) and more recently to surface humidity measurements (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. Venema et al. (2012) recommended PHA for the homogenisation of large data sets. Using a pairwise approach, testing each candidate–neighbour pair is also more robust than using a candidate station versus (composite) reference station approach, as the latter can easily miss network-wide changes, or wrongly attribute them to a single station. However if network-wide changes occur instantaneously, then relative methods cannot detect them. This may be a problem for large countries where changes in instrumentation may be implemented over a short timescale as, for many stations all neighbours will also be affected by the changes, and hence they cannot be detected.

PHA has 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. The COST-HOME assessment used small networks (5–20 stations) which in some cases may have limited the performance of PHA, which has been designed with large networks in mind. However, one of the submissions to COST-HOME using PHA with a very small network (five stations) performed extremely well (Venema et al., 2012). In our application here, we are less concerned about the estimated values of the inhomogeneity magnitudes, 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 inhomogeneity magnitudes. Hence the mean absolute value of the inhomogeneity magnitudes is overestimated.

We outline the steps used by PHA 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 (PHA requires at least seven neighbours to run and uses the highest correlating 40 if available<sup>1</sup>). Stations where insufficient neighbours were found are not processed by PHA.
2. The 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. The SNHT was chosen when

<sup>1</sup>In some cases fewer neighbours are allowed by the algorithm.

PHA was developed as DeGaetano (2006) and Reeves et al. (2007) showed it had superior accuracy in locating change points under a wide range of scenarios (Menne and Williams Jr, 2009).

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. Otherwise the data are removed for the period by PHA when producing the adjusted monthly series<sup>2</sup>. No data are removed in HadISD as the monthly adjusted data are not used.

The PHA code works on monthly data, and so the hourly data from HadISD have been converted to monthly values (see Sect. 3). 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 data set. 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. Annual and seasonal mean maximum and minimum temperatures were used by Trewin (2013) when homogenising the Australian Climate Observations Reference Network–Surface Air Temperature (ACORN-SAT) data set, which also picked up changes in the diurnal temperature range (DTR). 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 data set with internally consistent change point locations we combine change points from different methods in a post-processing algorithm. To incorporate the extra power of these additional monthly values for temperature and dew point temperature outlined above, we 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 1 year of each other, and use the mean date in the final products, rounded to the first day of the month.

<sup>2</sup>We do not use the adjusted monthly series output by PHA in this work, but rather just the inhomogeneity magnitudes and their dates.

### 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 at least 20 qualifying days within a month, then the monthly mean was calculated. A diurnal temperature range is only calculated for a day when it meets the same completeness requirements. This is therefore not a true diurnal range, but one estimated from the highest and lowest hourly temperature observation of that 24 h period. Again, at least 20 qualifying days are required in a month to obtain the monthly mean. We assess the effect of different completeness criteria later in this section.

The number of change points found using each of these two quantities is 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 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 underestimated 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 1 year of each other. PHA found 12973 change points in 4645 stations, a mean value of 2.79 change points per station over the 41-year period. This is roughly 1 every 15 years, which is at the lower end of the range found by Menne et al. (2009) for the USHCN of 1 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). Stations which could not be processed by PHA are ones with insufficient neighbours or those with very short records<sup>3</sup>. The majority of stations (4645/6103) have at least one change point, and so a 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. A relatively large number of stations are found to be homogeneous compared to other studies (e.g. Menne et al., 2009). Their number is likely to have been augmented by stations with short records and/or few correlating neighbours, as implied above. However, a similar fraction in a densely observed country, the UK, were found to be homogeneous (see Sect. 5).

To assess the effect of the completeness requirements used in this analysis, it was also performed using a more restrictive completeness requirement which matches the “3/5 rule” (no more than five missing days, of which no more than three can

<sup>3</sup>Short records on the monthly scale could arise from missing data so that the completeness criteria are not satisfied.

**Table 1.** Statistics of inhomogeneity detection for HadISD stations.

	Diagnostic	Temperature	Dew point	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	265
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	6493	9904	5647	15912
	DR	7735	8771	–	–
	Maximum	–	–	–	15092
Total combined change points		12973	16785	5658	23781
Change points/station		2.79	3.32	2.03	4.33
Inhomogeneity magnitude*	Mean	0.725	0.979	0.719	0.562
	DR	0.805	0.690	–	–
	Maximum	–	–	–	0.849

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

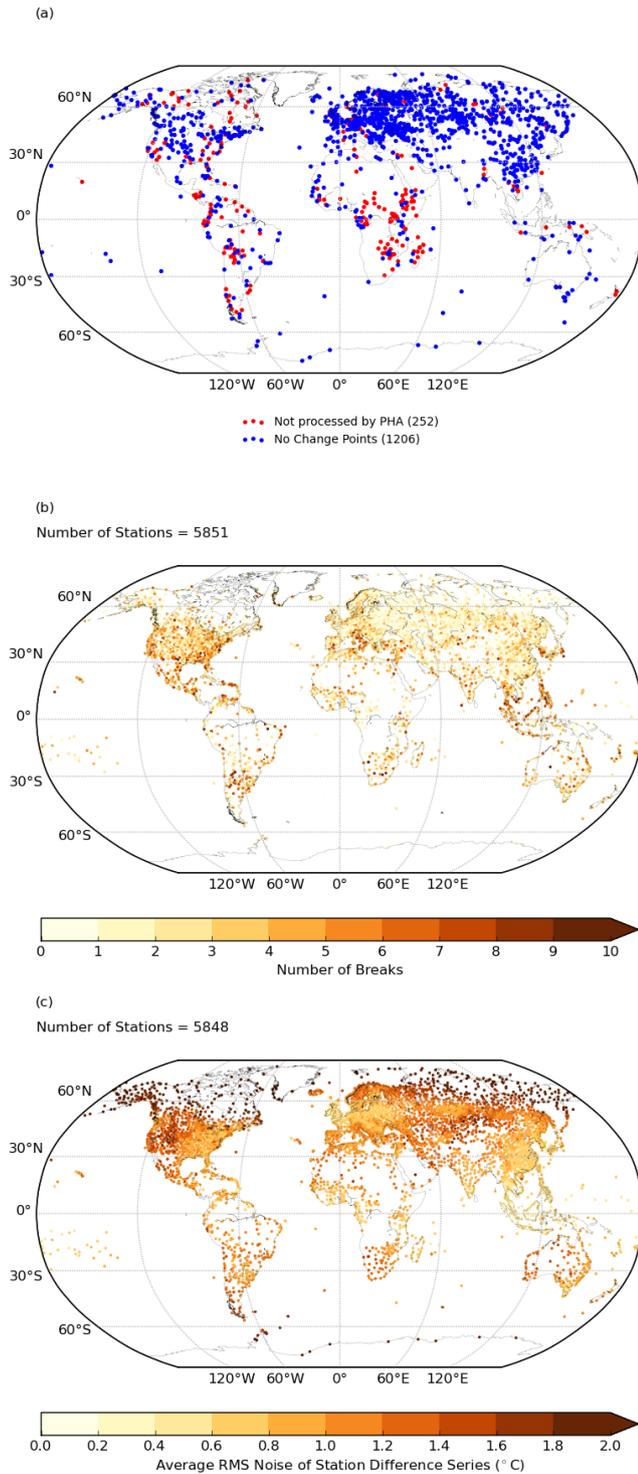
be consecutive) from the World Meteorological Organization (WMO) for calculating monthly averages from daily values (WMO, 1998). This results in 2457 stations which were not processed by PHA, compared to the 252. In a further 2340 stations PHA does not find any change points, compared to 1206, and there are therefore only 1306 stations in which PHA finds 2221 change points. Although the final monthly averages are likely to be more homogeneous and robust by using a more restrictive completeness requirement, omitting 40 % of the stations from the homogeneity assessment of HadISD makes the assessment less useful for users.

There is no simple pattern to the stations which could not be processed by PHA. This is because a lack of neighbours can arise because of a low density of stations (e.g. Africa) or because of complex topography which reduces the correlations between neighbouring stations (e.g. South America). Furthermore, for stations which report sporadically in HadISD (or have had large portions of data removed by the quality control) the completeness requirements may result in very few monthly values being calculated. 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 a 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. Overall, the number of change points is not especially high in regions with a high station density, where it might be expected that the change points are easier to detect.

We also show the average root mean square (rms) differences for all neighbours of each target station in Fig. 1c. This shows clearly that in areas with high station density and relatively simple topography the noise is low, whereas in complex topography (e.g. the Rocky Mountains) or regions with sparse station coverage (Arctic high latitudes) the noise increases. The rms noise has been calculated from the difference series of each target–neighbour station pair. The noise level gives an indication of the smallest inhomogeneity magnitude that can be detected using PHA. The figure for the diurnal temperature range is given in the Appendix, and shows a decrease in the noise for the Arctic high latitudes, showing that this measure is more useful in these areas for detecting change points.

Figure 2 shows the distribution of the inhomogeneity magnitudes for both of the two methods. Under the assumption that the underlying distribution of the inhomogeneity magnitudes is Gaussian, we do not detect those change points where the inhomogeneity magnitudes are very small. This “missing middle” is also seen when using other homogenisation processes. We fit a Gaussian to the distribution, ignoring the smallest inhomogeneity magnitudes (those smaller than the value corresponding to the peak on each side of the distribution), to estimate the population of change points that have not been detected by PHA (Brohan et al., 2006). Figure 2 shows that change points with inhomogeneity magnitudes down to around 0.5 °C have been found. The means of the distributions are very close to zero, but inhomogeneity



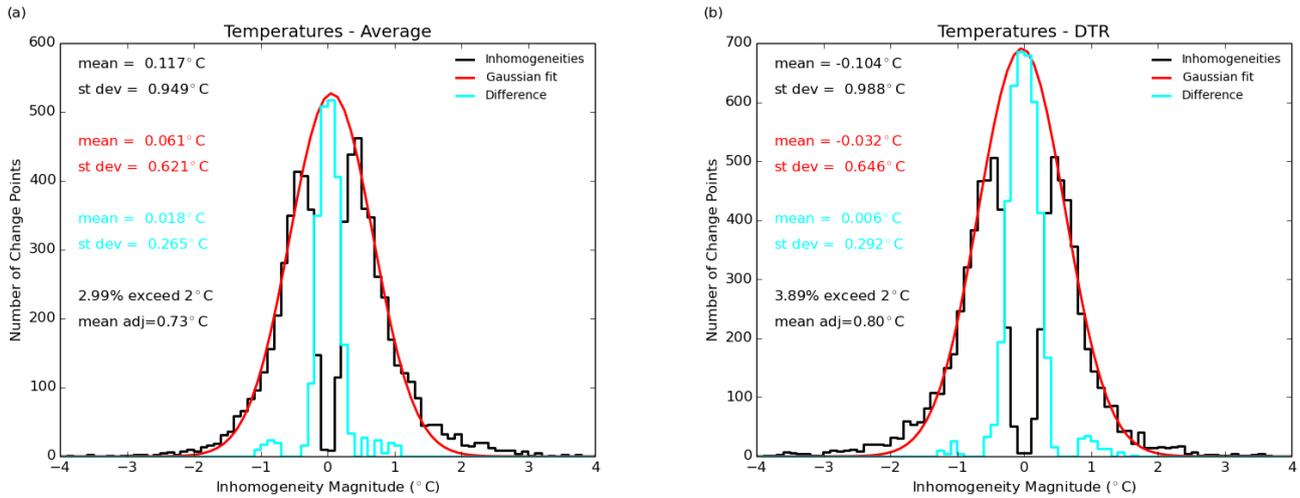
**Figure 1.** (a) Stations where temperatures could not be processed by PHA (red) and those stations where no temperature change point was found in the entire record (blue). (b) The number of change points detected for each station. (c) The average rms difference for the neighbours from each station for the monthly-average temperatures.

magnitudes based on the monthly mean have a slight positive bias ( $0.12\text{ }^{\circ}\text{C}$ ) and for the diurnal range have a slight negative bias ( $-0.10\text{ }^{\circ}\text{C}$ ). The biases of the Gaussian fits are smaller than the raw inhomogeneity magnitudes, but are still non-negligible ( $0.061$  and  $-0.032\text{ }^{\circ}\text{C}$ , respectively). Although these biases are small relative to the typical break size, they are important climatologically. As most stations on average have a change point every 15–20 years, and if this systematic bias is present for all stations, then this could sum to  $\sim 0.2\text{ }^{\circ}\text{C}$  over the period of the data set. However, by comparing the Gaussian curve to the binned inhomogeneities, these biases are likely to arise from the large positive (monthly average) and negative (monthly DTR) tails around inhomogeneity magnitudes of 2 and  $-2\text{ }^{\circ}\text{C}$ , respectively, and so only affect a subset of stations. As positive inhomogeneity values indicate that the earlier period was warmer (had a larger diurnal range), these tails would be consistent with stations which reduced the effect of urbanisation. Moving a station from an urban(ised) to a rural site would reduce the monthly-average temperature and increase the monthly-average DTR (Karl et al., 1988).

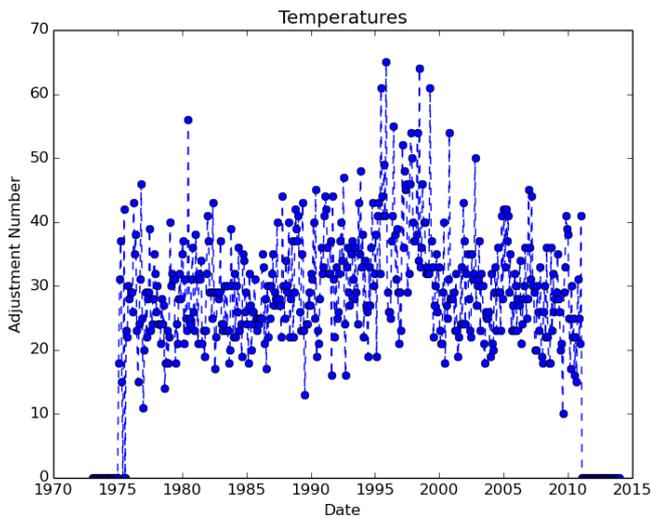
To further check on the possible origin for this bias, we created the distribution separately for the stations which report predominantly hourly versus those that report every 3 hours (Fig. A1 Appendix). The bias is much stronger in the hourly stations than in the 3-hourly ones, indicating that the underlying cause may be more common in automatic meteorological stations. This difference in the bias magnitude is also seen when using the diurnal temperature range. We also created distributions for different WMO regions for the monthly-average temperature. The distribution for North America has the largest bias ( $0.198\text{ }^{\circ}\text{C}$ ), with Asia having the smallest ( $0.043\text{ }^{\circ}\text{C}$ ). Biases in the inhomogeneity magnitudes relating to changes in instrumentation have been reported for the USHCN by Menne et al. (2009), which could be another source.

The spreads of the two 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. The largest inhomogeneity magnitudes constitute only a very small fraction of the total population (below 3 and 4% 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 1990s (Fig. 3). Note that change points cannot be detected within 2 years of the start or end of a 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 overall 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



**Figure 2.** The distribution of the inhomogeneity magnitudes 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).



**Figure 3.** The number of change points found in each year from both the calculation methods (monthly-mean temperature and monthly-mean diurnal temperature range).

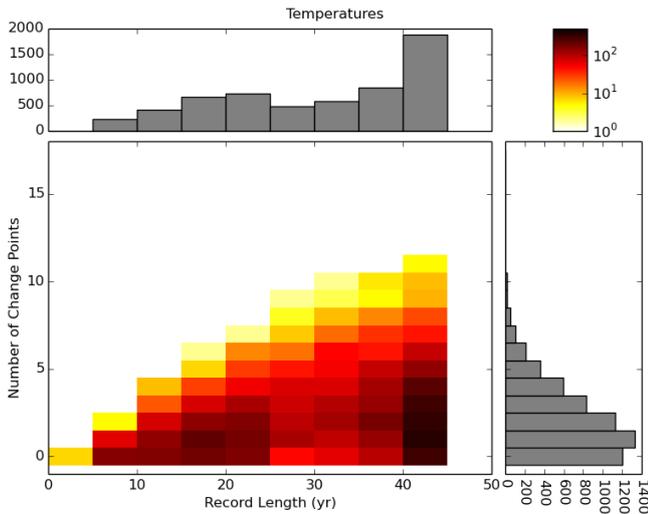
points (station numbers: 485650-99999 – Phuket Airport; 722265-13821 – Maxwell AFB; 723235-13896 – Northwest Alabama Regional Apt; and 725825-24121 – Elko Regional Apt<sup>4</sup>). There is a concentration of stations which have 41 years of record, and between zero and four change points, which matches the average number quoted above. There are no untoward patterns in the number of change points with

<sup>4</sup>The stations in HadISD are numbered using a quasi WMO number for the first six digits and a WBAN (Weather Bureau, Air Force and Navy) number for the final five. For full details see Smith et al. (2011) and references therein.

the station record length: the peaks in the main panel follow the uneven distribution of the station record lengths and the maximum number of change points rises quasi-linearly with record length until around 30 years, where it flattens slightly. The maximum number of change points also occurs at around five times the median number for that record length.

The distribution of homogeneous stations (no change points detected) with record length follows that of the complete population of stations. Therefore it is not the record length which is restricting the ability of PHA to find change points in these stations.

The distribution of inhomogeneity magnitudes with latitude and longitude show that the largest inhomogeneities are mainly found in mid-latitude and Euro-American regions with large numbers of stations (Fig. 5). This is because having more stations in a zonal or meridional band increases the chances of finding a large inhomogeneity within that population even though 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. This also implies that under-detection of inhomogeneities occurs in regions with low station densities. With longitude, there appears to be a bias towards larger positive inhomogeneities in the Americas. Of course, in areas with more stations there is a greater chance of finding change points with very large or very small inhomogeneity magnitudes merely because of the increased number of stations. In Fig. 2a there is a broader positive tail



**Figure 4.** Main panel: the distribution of the stations with number of change points and record length. Histograms show the distribution of stations with record length (top) and with change point number (right), which are the projections onto the  $x$  and  $y$  axes of the main panel, respectively. The colour bar (top right) is on a logarithmic scale.

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

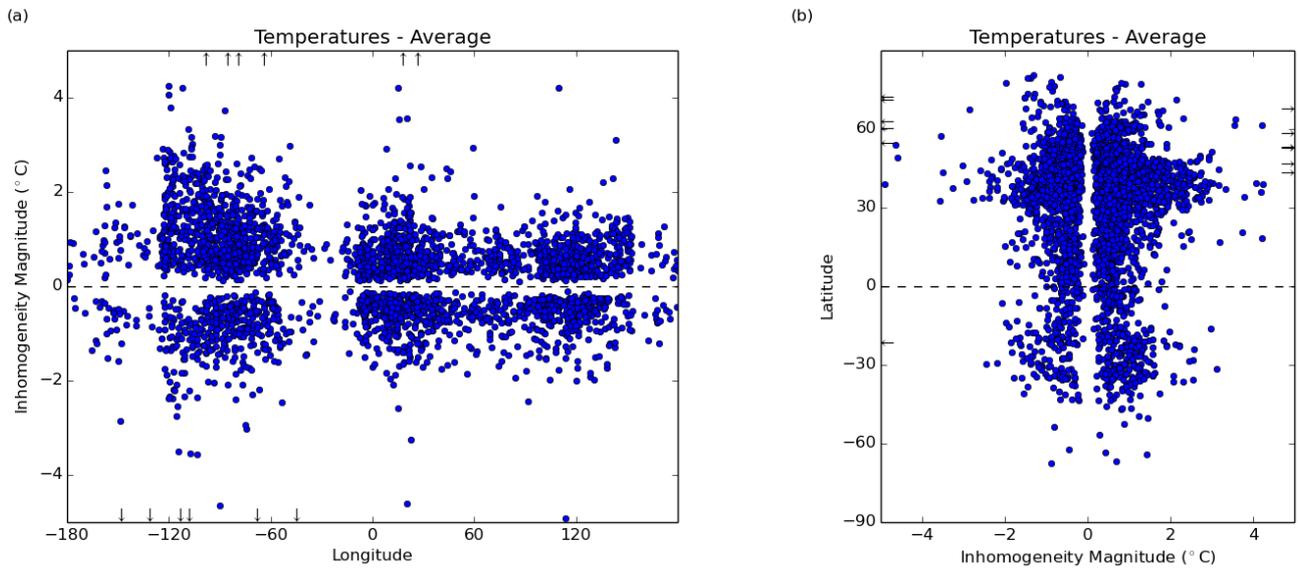
#### 4 Dew point temperatures, sea-level pressure and wind speeds

We also carry out an identical set of calculations for the dew point temperatures, sea-level pressure and wind speed measurements within HadISD, using the same completeness criteria when calculating monthly averages. For SLP, we use the deviations from 1000 hPa when calculating the monthly-mean values. We find 16785 change points in 5051 stations for dew point temperatures (3.32 per station), 5658 change points in 2781 stations for SLP (2.03 per station) and 23781 change points in 5496 stations for wind speeds (4.32 per station). Thus the wind speed records appear to be more inhomogeneous than the other variables, and the sea-level pressure records more homogeneous.

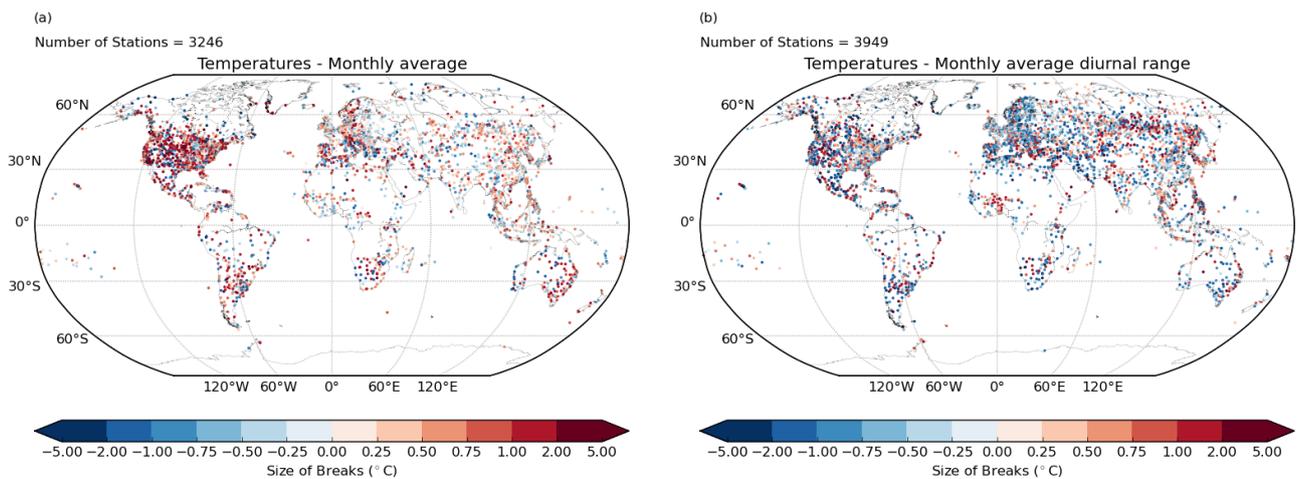
The distributions of the inhomogeneity magnitudes for the dew point temperatures and SLP are shown in Figs. 7 and 8, respectively. As for the temperature results (Fig. 2), the means are close to zero but have a small positive bias, and the diurnal range has a small negative bias. These biases in the dew point inhomogeneity magnitudes are smaller for the Gaussian curve, and in both cases smaller than for temperature. For the SLP, the raw data and the fitted curve have similar small biases. As for the temperature measurements, the hourly stations have larger biases than the 3-hourly stations. The overall excess number of large inhomogeneity magnitudes seen in the dew point does not necessarily indicate that the underlying distribution is non-Gaussian. Large final inhomogeneities may result from the combination of two or more smaller changes close in time (in the same direction) during the statistical homogenisation procedure. Therefore, under the assumption that the distribution of inhomogeneities is Gaussian, most of the change points have been detected down to a limit of around  $0.5^{\circ}\text{C}$  or  $0.5\text{ hPa}$ . The distribution of the inhomogeneities proposed for the monthly-mean dew point 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.

The distribution of the number of change points per station against the record length for the dew point temperatures (Fig. 9) is very similar to that from the temperatures (Fig. 4). There is a concentration of stations with 41 years of records and one to two change points, with a smooth decline to greater numbers of change points, with station 442840-99999 (Galut, Mongolia) having 13 change points. The distribution for SLP is much flatter, as would be expected from Table 1. Two stations, 483270-99999 (Chiang Mai, Thailand) and 577760-99999 (Nanyue, China), 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. The inhomogeneities have only a very small negative bias. Although the hourly stations have a larger bias than the 3-hourly ones for the monthly-average daily-maximum wind speed, the biases are equal for the monthly-average wind speed. The overall distribution of the inhomogeneities proposed for the monthly-mean daily-maximum wind speeds is broader than that of the monthly mean. It is reasonable to deduce that most of the change points whose inhomogeneities are  $> 0.5\text{ m s}^{-1}$  have been detected. The distributions of the missing inhomogeneities 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 (Monte Cimone, Italy) 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,



**Figure 5.** The distribution of inhomogeneity magnitudes with (a) longitude and (b) latitude using the monthly-mean temperatures. Arrows indicate values which fall outside of the plotted area.



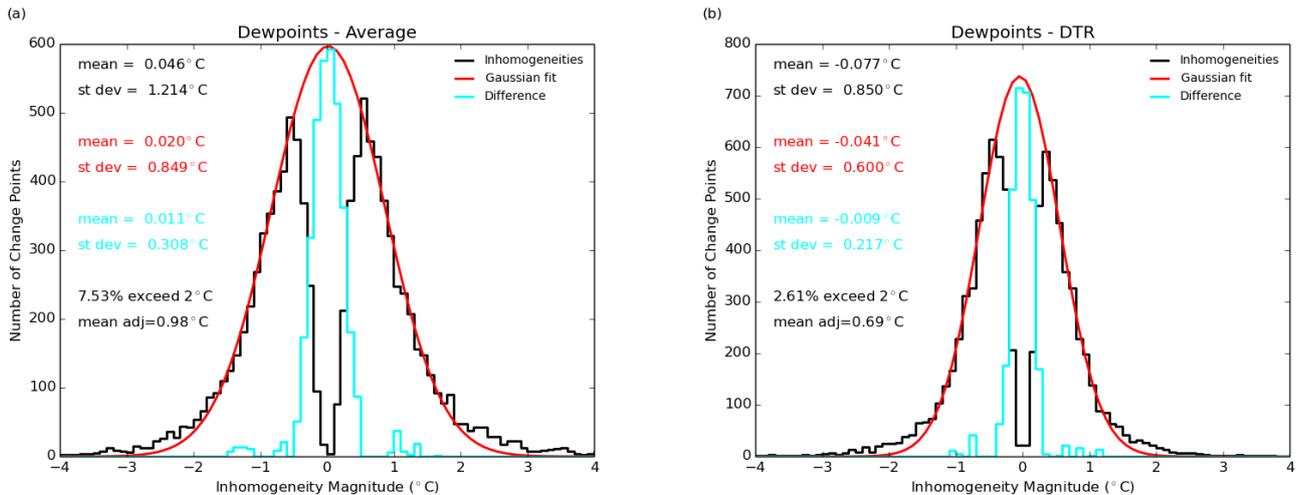
**Figure 6.** The distribution of the largest absolute inhomogeneity found for each station when using the (a) monthly-mean temperature and (b) the monthly-mean diurnal temperature range. Only those stations which have at least one change point are shown, resulting in the different numbers of stations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

in a band which extends from zero change points right up 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.

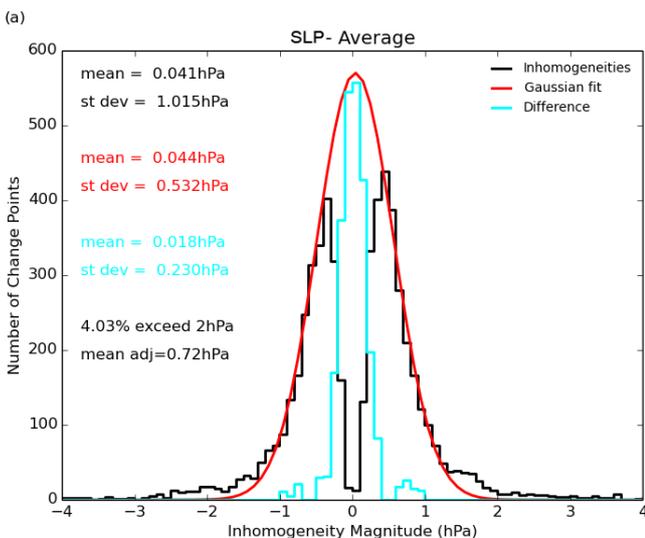
The stations which were not processed by PHA cluster in southern and eastern Africa and western South America for all three variables, and also western China for SLP (Fig. 11). For SLP, this is the result of very short records after conversion to monthly averages in these areas. For many stations, no monthly means could be calculated at all. The number of change points for dew points is relatively uniform across the globe, but large values are found for Italy, Korea and the tip

of Argentina (Fig. 11). The largest number of change points for SLP are found in the Maritime Continent and southeast Asia; whereas for wind speeds, large numbers are found in China through to Indochina, in the eastern US and in Argentina. There do not appear to be any links with physical features for any of the variables.

The rms difference maps are given in the Appendix. The dew points have high noise in high or dry areas (the Rocky Mountains, the Sahara, central Australia) as well as high latitudes, whereas the sea-level pressure picks out just the high latitudes with high rms noise. The wind speeds have a comparatively uniform noise distribution across the globe, with some tendency to be higher on coasts and islands, and with



**Figure 7.** The distribution of the inhomogeneity magnitudes for monthly-mean (a) dew point temperatures and (b) diurnal dew point temperature range. The distributions have been fitted with a Gaussian (red), and the difference between the data and the Gaussian is shown in blue.



**Figure 8.** The distribution of the inhomogeneity magnitudes 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.

a lower level for the monthly averages than for the average maxima. The low noise has enabled the detection of the most change points for one variable (see Table 1 and Fig. 9)

Compared to the temperatures, the dew point temperatures have a more even distribution of inhomogeneity magnitudes across the globe (Fig. 12). The magnitudes are on average larger for the monthly-mean dew point temperatures than for the monthly-mean dew point ranges. As with temperature, the western half of the US appears to have larger magnitudes than the eastern half, but the region around the Mediterranean and also Scandinavia also stand out.

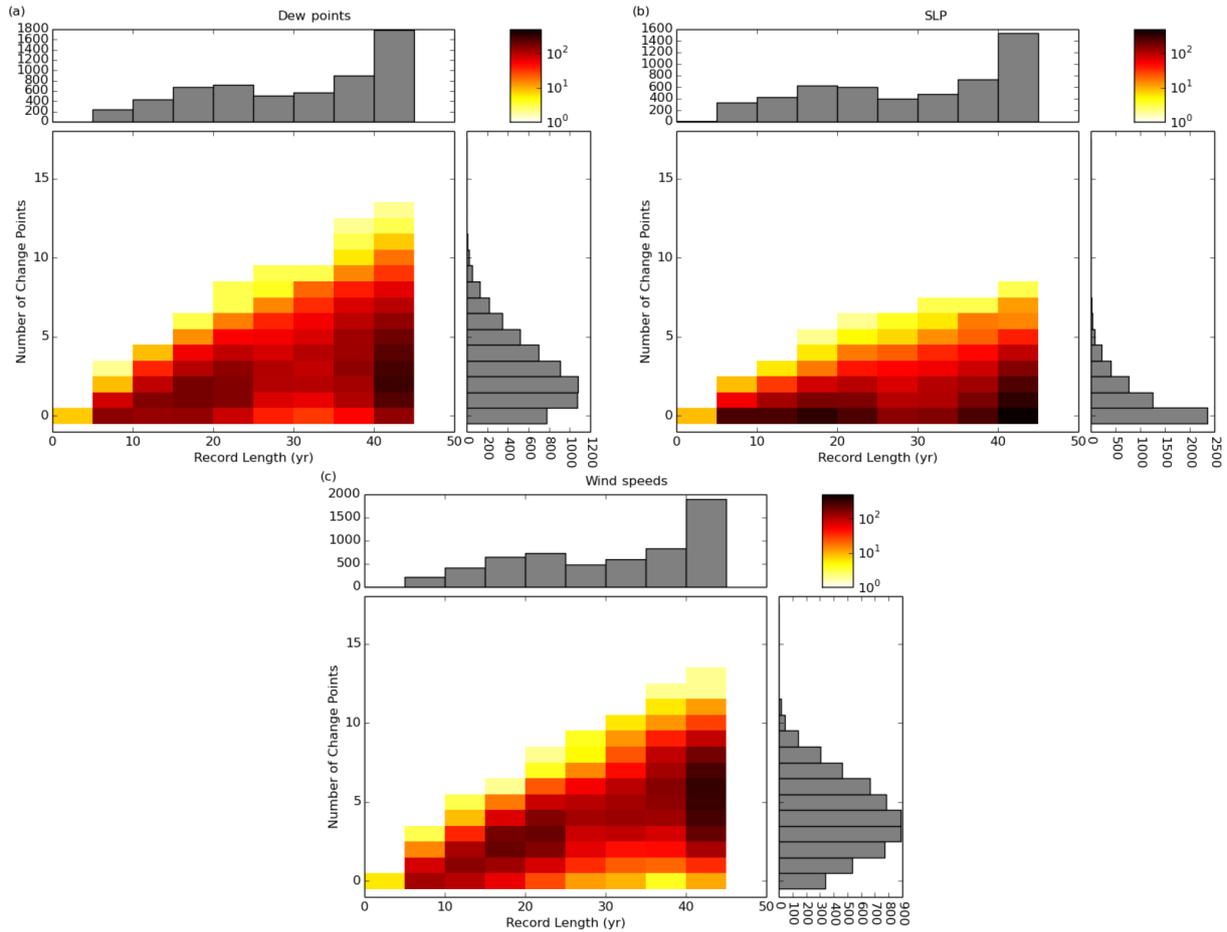
There are fewer stations which have detected change points in SLP, with a reduction in the station number being clearest in Africa and western China (Fig. 13). Positive inhomogeneities are seen in the eastern US, parts of Europe, in Siberia and down through China into the Maritime Continent, with other smaller clusters in Europe. Negative inhomogeneities are seen in Central and South America.

The pattern of inhomogeneity magnitudes 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 inhomogeneities, but with clusters of more positive values in the US and more negative ones through Eurasia. When using the monthly-mean daily-maximum wind speed, the inhomogeneities are larger. Adjustment of positive inhomogeneities would weaken apparent declines in wind strength (McVicar and Roderick, 2010).

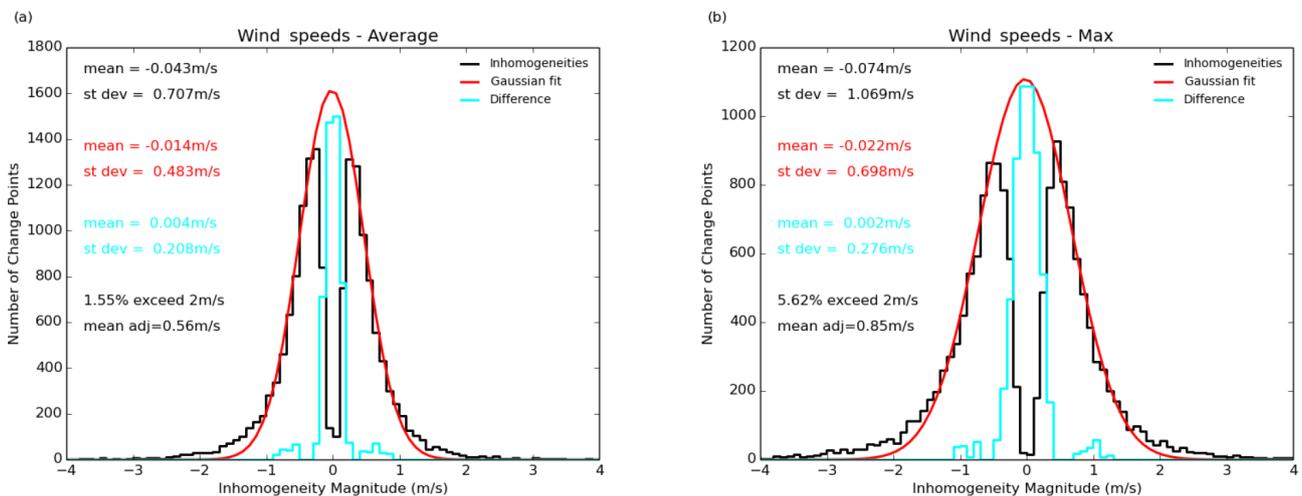
## 5 Validation

To validate whether the change point locations detected using PHA correspond with documented breaks in the station metadata, we look at the 153 UK stations in the HadISD database. Of these 153 stations, 102 contain the 196 change points in the temperature observations and no change points are detected in the remaining 51 (10 of which were not processed by PHA).

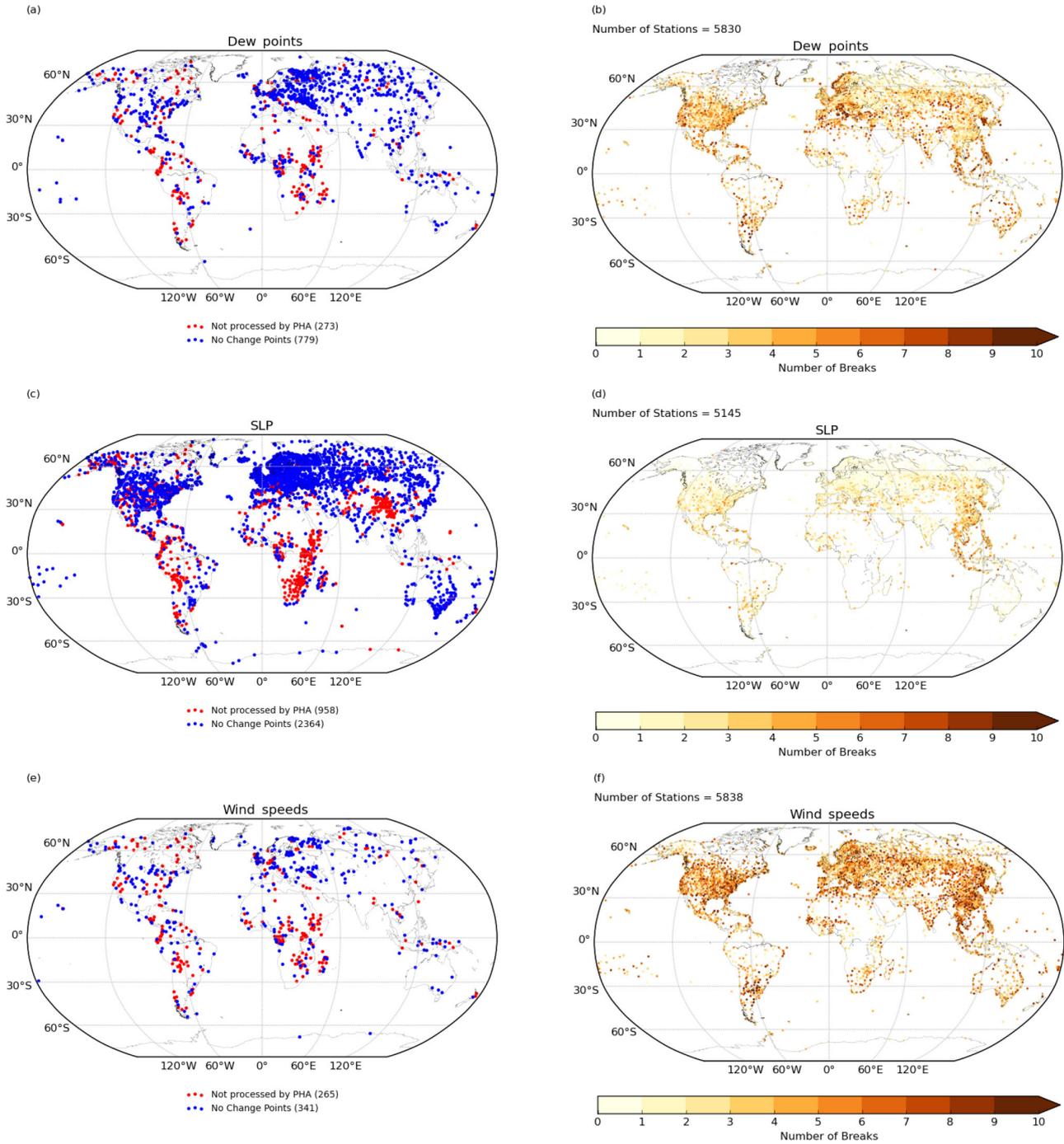
Of these 153 stations, 18 have been merged during the creation of HadISD and contain 31 of the 196 change points. Although great care was taken when selecting stations to merge, it is likely that in some cases this process has created or left a discontinuity in the station record. The locations of change points identified by PHA from the temperature records were compared to the dates where input station



**Figure 9.** The distribution of the number of change points within a station record against the number of years of data for (a) dew point temperature, (b) sea-level pressure and (c) 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.



**Figure 10.** The distribution of the inhomogeneity magnitudes for (a) monthly-mean wind speeds and (b) monthly-mean daily-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.



**Figure 11.** (a, c, e) 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 (a) dew points, (c) SLP and (e) wind speeds. (b, d, f) The number of breaks detected for each station.

identifiers changed in the HadISD Network Common Data Form (NetCDF) headers<sup>5</sup>. In six stations these dates were in

<sup>5</sup>When creating the HadISD NetCDF files, the input station IDs are retained as a field precisely for this purpose. For more details see Dunn et al. (2012).

close agreement, and accounted for eight of the change points (Table 2).

There have been changes in the station reporting accuracy (usually between 0.1°C and single-degree precision) as stations were up- or downgraded over time. Of the 153 UK HadISD stations, 34 had no change in reporting accuracy

**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.

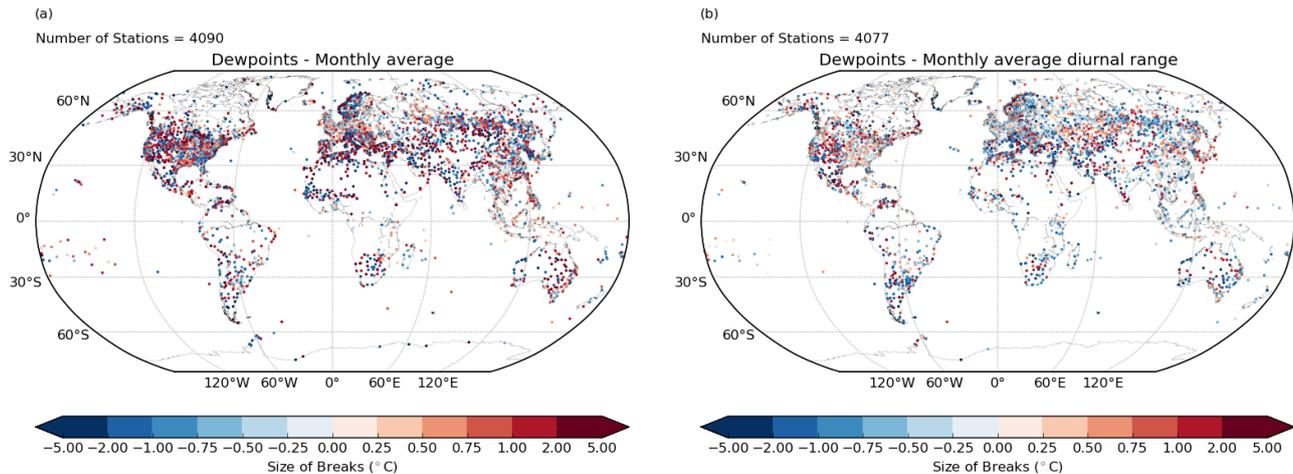
Station ID	Name	Change point date	Metadata date	Type
Merged stations				
035580-99999	Bedford Airport	1 Nov 1994	1 Apr 1994	Merger
037010-99999	Gawlish	1 Mar 1983	1 Jun 1994	Merger <sup>a</sup>
		1 Aug 1984	1 Jun 1994	Merger <sup>a</sup>
		1 Dec 1988	1 Mar 1989	Merger
037260-99999	Bristol Weather Centre	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
≥ 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 Mount Batten	1 Feb 1991	17 Jul 1991	Instrument change
2 change points				
032570-99999	Leeming	1 May 1995	30 Nov 1995	Instrument change <sup>b</sup>
		1 Jun 1996	30 Nov 1995	Instrument change <sup>b</sup>
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 <sup>b</sup>
		1 Mar 1992	8 Nov 1991	Instrument change <sup>b</sup>
037610-99999	Odiham	1 Apr 1993	2 Apr 1993	Instrument change
038530-99999	Yeovilton	1 Mar 1995	10 Nov 1994	Instrument change

<sup>a</sup> These two change points occur on 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. <sup>b</sup> 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.

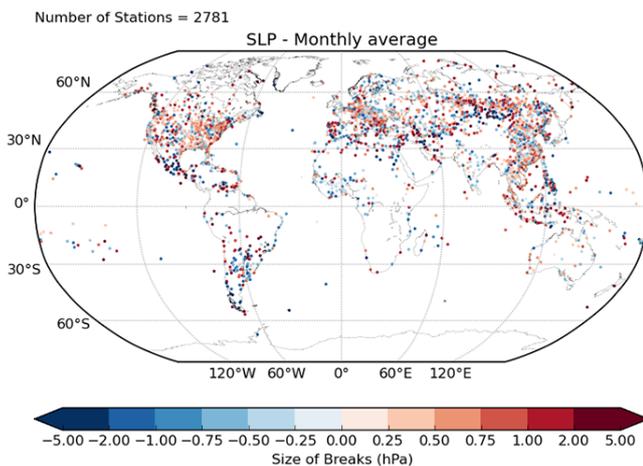
during their record. A further 13 stations had changes in reporting accuracy where the date corresponded to a change point detected by PHA, and the remaining 106 stations had changes in reporting accuracy during their record, but the dates of these changes did not correspond to change points detected by PHA.

Using the station metadata is another way to validate the detected change points. As station metadata are more complete in later years, we initially focus on the UK stations where detected change points occurred after 2000. This was a subset of 35 stations containing 45 change points. In all, 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 were 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



**Figure 12.** As for Fig. 6 but for (a) monthly averages of dew point temperatures and (b) monthly-average diurnal ranges of dew point temperatures.



**Figure 13.** As for Fig. 6 (a) but for SLP.

the station metadata, or the metadata themselves are not sufficiently complete.

Although this investigation – which has accounted for 40 (20%) of the 196 change points – demonstrates that using 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. This is likely to be for a combination of reasons. The metadata are known to be incomplete and in some cases do not cover the entire station record. Also changes can occur inside and outside of an observation enclosure that would not be noted in the metadata (new paths in the enclosure, change in crop type, tree growth/felling, new buildings) but could result in an inhomogeneity in the record.

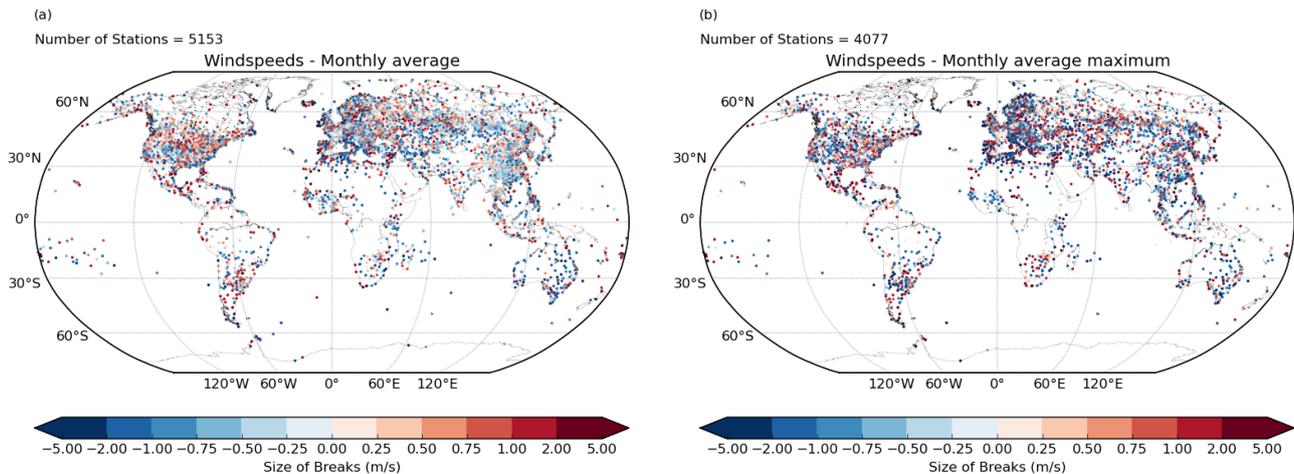
## 6 Provision of the change points

The change point dates and values determined by this study will be made available on the HadISD webpage (<http://www.metoffice.gov.uk/hadobs/hadis/>) 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 data set. 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 inhomogeneities 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 inhomogeneity magnitudes, 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. For large-scale 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



**Figure 14.** As for Fig. 6 but for (a) monthly-average wind speeds and (b) monthly-average daily-maximum wind speeds.

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 data sets (Hawkins and Jones, 2013). So, if our stations with the largest inhomogeneity magnitudes are excluded, then the global mean temperature should be at least as accurate, because Callendar's work suggests 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).

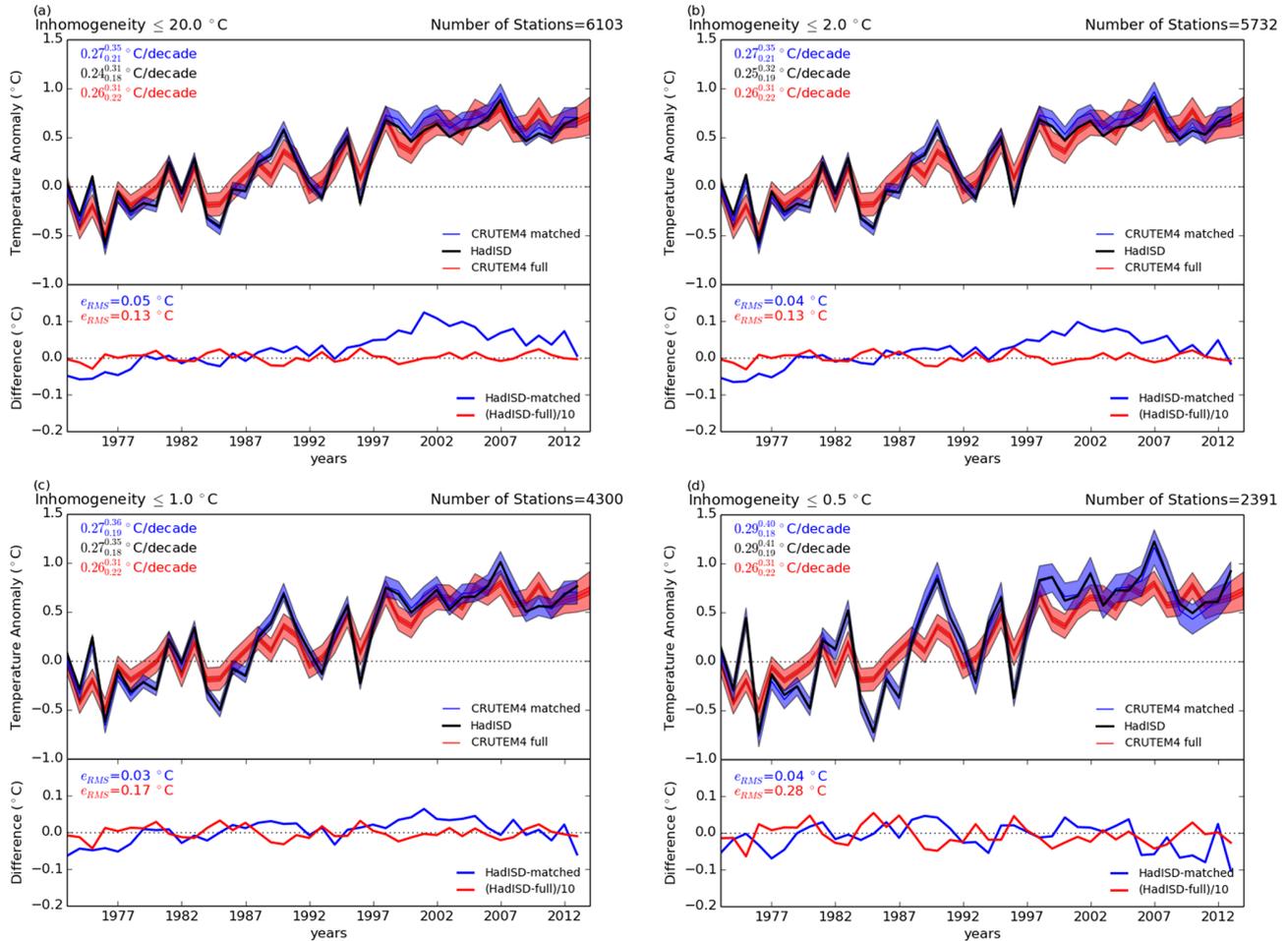
To calculate the global temperature series from HadISD, firstly, daily mean temperatures are obtained, requiring that there are at least four observations in a day, spread over at least 12 h. Monthly-mean temperatures are calculated if there are at least 20 qualifying days within a month. A climatology is calculated over the period 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. As no adjustments have been applied to HadISD during this homogeneity assessment, the global averages are of the unadjusted HadISD data. The gridded monthly CRUTEM4 anomaly fields are also converted to annual global anomalies relative to the period 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 taken on those HadISD stations which have maximum inhomogeneity magnitudes (in either of the two calculation methods – mean or diurnal range) of less than 2, 1 and  $0.5^\circ\text{C}$ . 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 CRUTEM in blue, and HadISD and the full CRUTEM in red.

There are two competing changes occurring in the four panels in Fig. 15. Stations with inhomogeneities above a certain magnitude are progressively excluded, reducing the average magnitude of the inhomogeneities left in the remaining stations; but with this, there is a reduction in the number of stations available, and hence the global coverage. By using CRUTEM as a comparison, it is possible to show how close the HadISD version is to a widely used global air-temperature data set. There are two comparisons to be made, one to the full coverage of CRUTEM, shown in red, and one to CRUTEM 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 CRUTEM, by restricting the stations to those with smaller and smaller inhomogeneities, this difference reduces, especially from 1996 onwards. But when only stations with inhomogeneity magnitudes  $< 0.5^\circ\text{C}$  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 of each panel of Fig. 15. The linear trends also become closer as the stations are restricted, but the uncertainty in the linear trends increases as the number of stations reduces. Using the rms difference as a measure of the difference of the global mean between HadISD and the two versions of CRUTEM shows this more clearly. When comparing to a matched CRUTEM, the  $e_{\text{rms}}$  reduces to a minimum when inhomogeneity magnitudes are restricted just to  $< 1^\circ\text{C}$  but then increases again. Despite matching the



**Figure 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) shows 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. **(a)** No restrictions on the magnitudes of the inhomogeneities (all stations); **(b)** magnitudes  $< 2^\circ\text{C}$ , **(c)**  $< 1^\circ\text{C}$  and **(d)**  $< 0.5^\circ\text{C}$ . The trends for each of the three curves are given in the top left of each panel, with the 5th- and 95th-percentile 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 ( $\div 10$ , red) along with the rms differences. In this plot there is no restriction on the number of change points allowed in a station record.

coverage of CRUTEM 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 between the full CRUTEM and HadISD remains steady when restricting to  $< 2^\circ\text{C}$ , but then increases thereafter. This shows that, when trying to obtain a global quantity, 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  $e_{rms}$  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 inhomogeneity allowed fixed at  $< 1^\circ\text{C}$ ,

as this has the lowest  $e_{rms}$  between HadISD and the matched CRUTEM from the analysis 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  $e_{rms}$  between HadISD and both versions of CRUTEM increase, with the clearest deviations relative to the matched CRUTEM 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 obtained from the median pairwise algorithm. Although by excluding more and more inhomogeneous stations will improve the fidelity of the remaining stations, the reduction

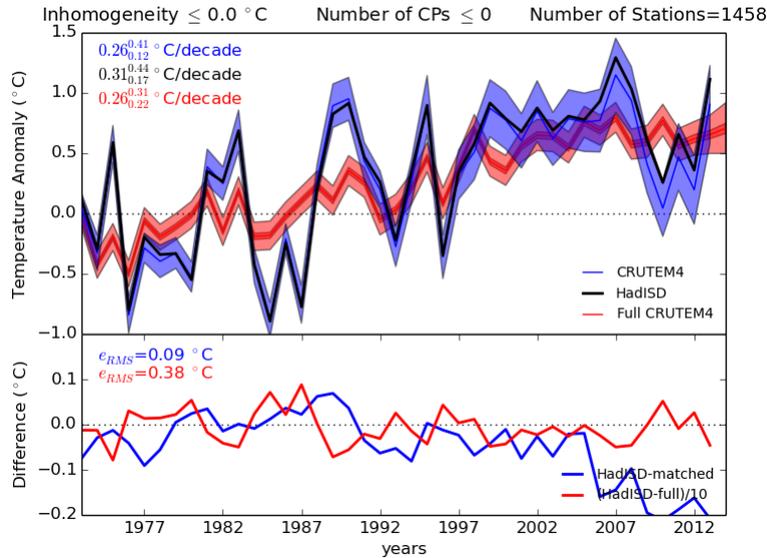


Figure 16. As for Fig. 15 but for the 1458 stations with no detected change points.

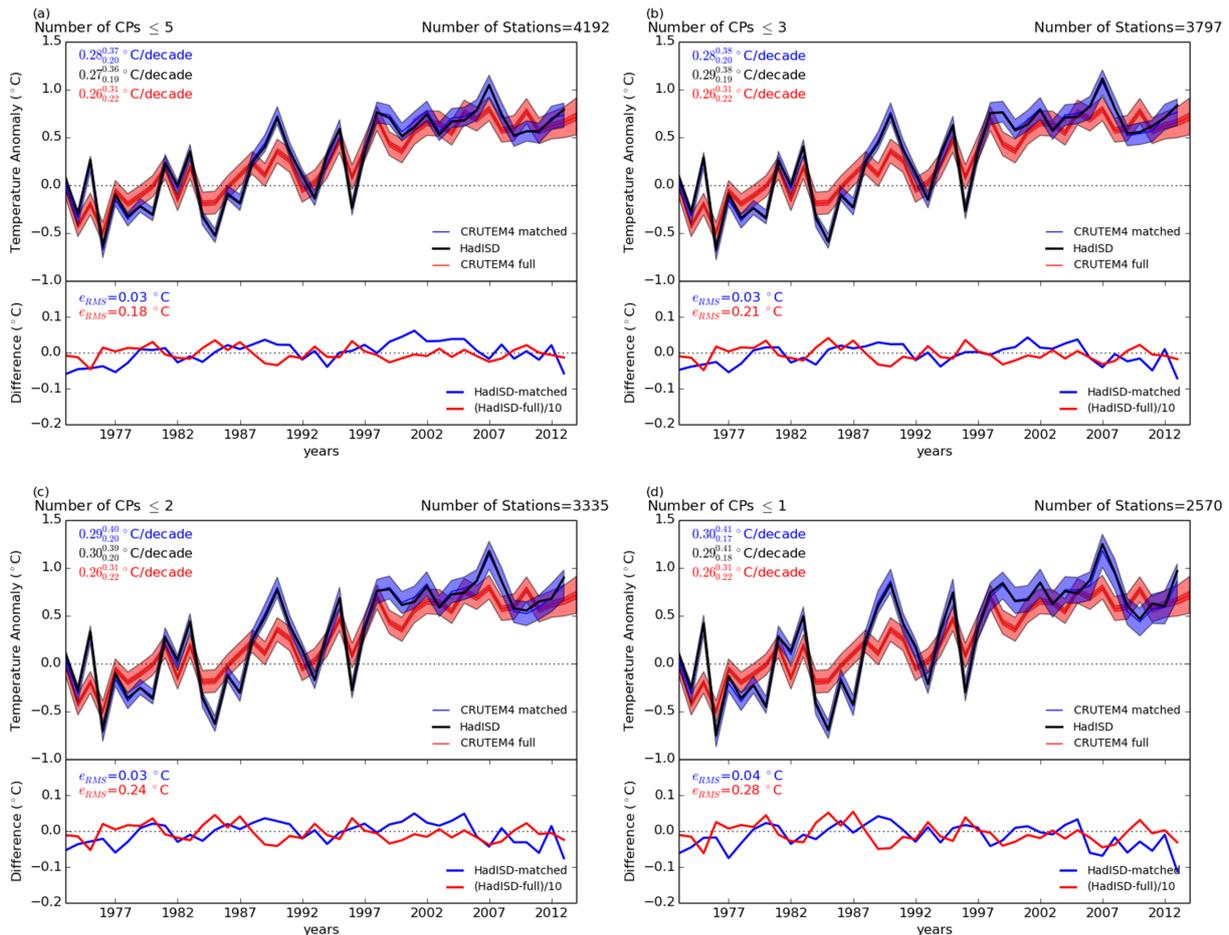


Figure 17. As for Fig. 15 but for stations with inhomogeneity magnitudes  $< 1\text{ }^{\circ}\text{C}$ . (a) No more than five, (b) three, (c) two and (d) one inhomogeneity allowed in the full record of the station.

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 inhomogeneity magnitudes 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 data set 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 and diurnal ranges or maximum values for the temperature, dew point temperature, sea-level pressure and wind speed. 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. There are some geographical patterns in the stations which could not be homogenised, or in which no 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 inhomogeneity 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 increase the scatter because the coverage is degraded.

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; Della-Marta and Wanner, 2006; Trewin, 2013; Rienzner and Gandolfi, 2013) on country-scale networks. However, the issue of automated scaling of inhomogeneity magnitudes across all hours of the day has not yet been solved.

Appendix A

We show the distributions of the inhomogeneity magnitudes for stations which report hourly and those which report 3-hourly (Fig. A1).

We show the maps of the average rms noise for each station calculated from the target–neighbour difference series for each variable and monthly averaging method (Figs. A2–A5).

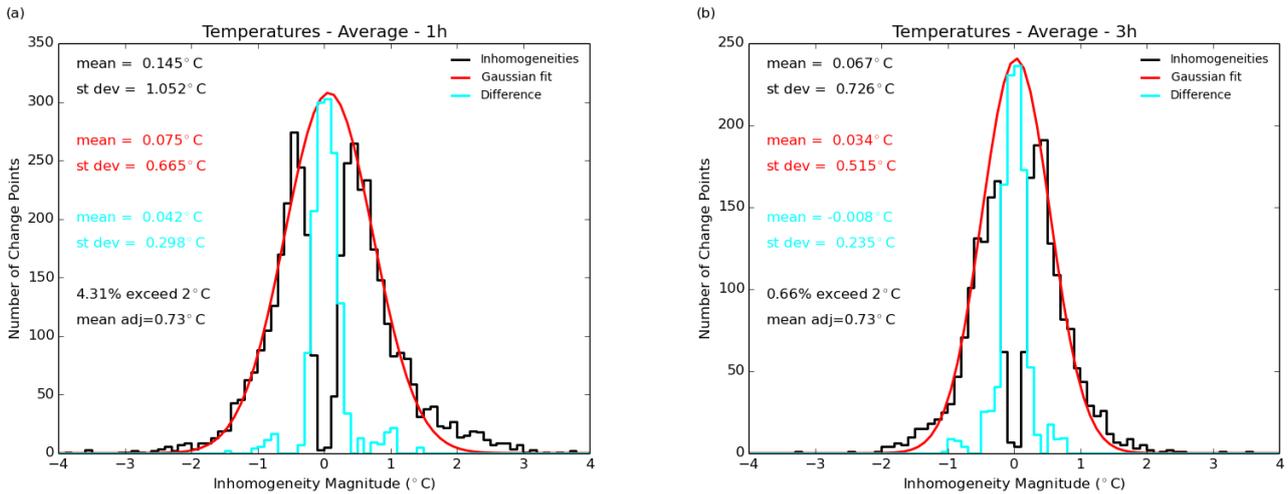


Figure A1. The distribution of the inhomogeneity magnitudes for monthly-mean temperatures for stations which report (a) hourly and (b) 3-hourly. The distributions have been fitted with a Gaussian (red), and the difference between the data and the Gaussian is shown in blue.

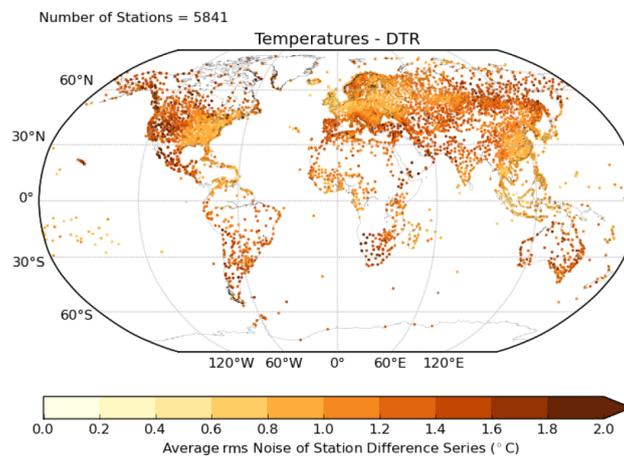
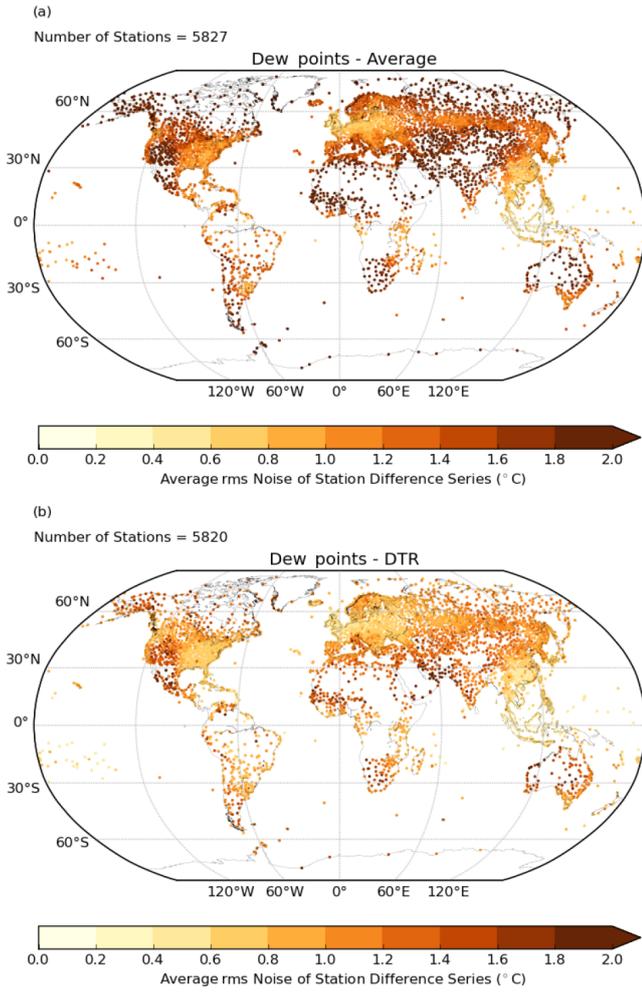
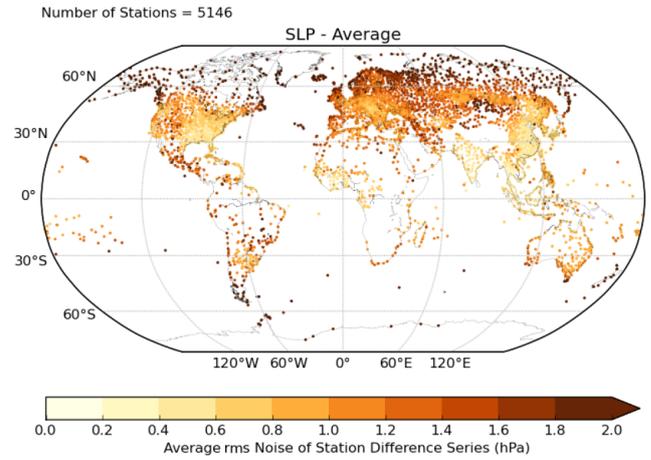


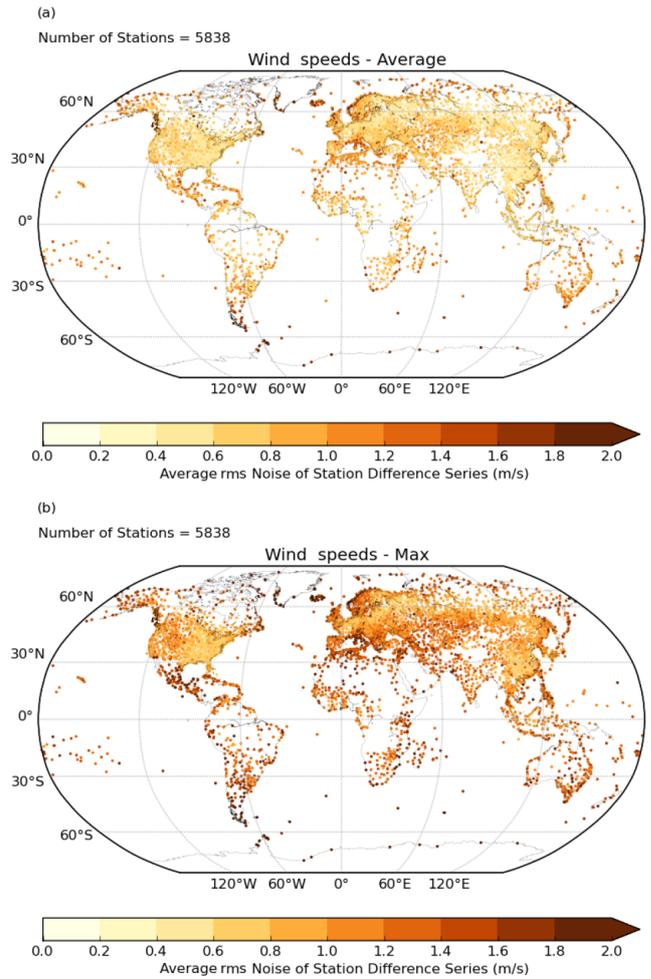
Figure A2. The average rms difference for the neighbours from each station for the monthly-average diurnal temperature range.



**Figure A3.** The average rms difference for the neighbours from each station for (a) the monthly-average dew point and (b) the monthly-average diurnal dew point range.



**Figure A4.** The average rms difference for the neighbours from each station for the monthly-average sea-level pressure.



**Figure A5.** The average rms difference for the neighbours from each station for (a) the monthly-average wind speed and (b) the monthly-average maximum wind speed.

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

R. J. H. Dunn thanks Claude Williams for helpful discussions during the course of this work and Colin Chatters of the Met Office Land Networks Team for providing the UK station metadata. The authors thank the three referees – Enric Aguilar, Blair Trewin and Victor Venema – for helpful and thorough reviews which improved this manuscript.

This work is distributed under the Creative Commons Attribution 3.0 License together with an author copyright. This license does not conflict with the regulations of the Crown Copyright.

Edited by: S. Bronnimann

## References

- Alexandersson, H.: A homogeneity test applied to precipitation data, *J. Climatol.*, 6, 661–675, 1986.
- Auchmann, R. and Brönnimann, S.: A physics-based correction model for homogenizing sub-daily temperature series, *J. Geophys. Res. Atmos.*, 117, D17119, doi:10.1029/2012JD018067, 2012.
- 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.
- Brohan, P., Kennedy, J. J., Harris, I., Tett, S. F., and Jones, P. D.: Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850, *J. Geophys. Res. Atmos.* (1984–2012), 111, 2006.
- Callendar, G. S.: The artificial production of carbon dioxide and its influence on temperature, *Quart. J. Roy. Meteorol. Soc.*, 64, 223–240, 1938.
- Callendar, G. S.: Temperature fluctuations and trends over the earth, *Quart. J. Roy. Meteorol. Soc.*, 87, 1–12, 1961.
- Caussinus, H. and Mestre, O.: Detection and correction of artificial shifts in climate series, *J. Roy. Statist. Soc. Series C (Applied Statistics)*, 53, 405–425, 2004.
- DeGaetano, A. T.: Attributes of several methods for detecting discontinuities in mean temperature series., *J. Climate*, 19, 838–853, 2006.
- Della-Marta, P. and Wanner, H.: A method of homogenizing the extremes and mean of daily temperature measurements, *J. Climate*, 19, 4179–4197, 2006.
- Domonkos, P.: Adapted Caussinus-Mestre Algorithm for Networks of Temperature series (ACMANT), *Int. J. Geosci.*, 2, 293–309, 2011.
- 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.
- 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.
- Guijarro, J.: User’s guide to climatol. An R contributed package for homogenization of climatological series, Report, State Meteorological Agency, Balearic Islands Office, Spain, 2011.
- Hawkins, E. and Jones, P. D.: On increasing global temperatures: 75 years after Callendar, *Quart. J. Roy. Meteorol. Soc.*, 139, 1961–1963, 2013.
- Jakob, D.: Challenges in developing a high-quality surface wind-speed data-set for Australia, *Australian Meteorological and Oceanographic Journal*, 60, 227–236, 2010.
- Jones, P. D., Wigley, T., and Briffa, K.: Monthly mean pressure reconstructions for Europe (back to 1780) and North America (to 1958), Carbon Dioxide Research Division, US DoE, Tech Rep TR037, 1987.
- 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. Atmos.* (1984–2012), 117, 2012.
- Karl, T. R., Diaz, H. F., and Kukla, G.: Urbanization: Its detection and effect in the United States climate record, *J. Climate*, 1, 1099–1123, 1988.
- 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. Atmos.* (1984–2012), 114, 2009.
- 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.
- McVicar, T. R. and Roderick, M. L.: Atmospheric science: Winds of change, *Nature Geosci.*, 3, 747–748, 2010.
- Menne, M. J. and Williams Jr, C. N.: Homogenization of temperature series via pairwise comparisons, *J. Climate*, 22, 1700–1717, 2009.
- Menne, M. J., Williams Jr, C. N., and Vose, R. S.: The US Historical Climatology Network monthly temperature data, version 2, *Bull. Am. Meteorol. Soc.*, 90, 993–1007, 2009.
- Mestre, O., Gruber, C., Prieur, C., Caussinus, H., and Jourdain, S.: SPLIDHOM: A method for homogenization of daily temperature observations, *J. Appl. Meteorol. Climatol.*, 50, 2343–2358, 2011.
- Mestre, O., Domonkos, P., Picard, F., Auer, I., Robin, S., Lebarbier, E., Boehm, R., Aguilar, E., Guijarro, J., Vertachnik, G., Klancar, M., Dubuisson, B., and Stepanek, P.: HOMER: a homogenization software—methods and applications, *Quart. J. Hung. Meteorol. Serv.*, 117, 47–67, 2013.
- 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.
- 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.
- Picard, F., Lebarbier, E., Hoebcke, M., Rigai, G., Thiam, B., and Robin, S.: Joint segmentation, calling, and normalization of multiple CGH profiles, *Biostatistics*, 12, 413–428, 2011.
- Reeves, J., Chen, J., Wang, X. L., Lund, R., and Lu, Q.: A review and comparison of changepoint detection techniques for climate data., *J. Appl. Meteorol. Climatol.*, 46, 900–915, 2007.
- Rienznner, M. and Gandolfi, C.: A procedure for the detection of undocumented multiple abrupt changes in the mean value of daily temperature time series of a regional network, *Int. J. Climatol.*, 33, 1107–1120, doi:10.1002/joc.3496, 2013.

- Sen, P. K.: Estimates of the Regression Coefficient Based on Kendall's Tau, *Journal of the American Statistical Association*, 63, 1379–1389, 1968.
- Slonosky, V., Jones, P., and Davies, T.: Homogenization techniques for European monthly mean surface pressure series., *J. Climate*, 12, 2658–2672, 1999.
- Smith, A., Lott, N., and Vose, R.: The integrated surface database: Recent developments and partnerships, *Bull. Am. Meteorol. Soc.*, 92, 704–708, 2011.
- Szentimrey, T.: Multiple analysis of series for homogenization (MASH), in: *Proc. Second Seminar for Homogenization of Surface Climatological Data*, pp. 27–46, 1999.
- Szentimrey, T.: Manual of homogenization software MASHv3.02, p. 65, 2007.
- Thomas, B. R., Kent, E. C., Swail, V. R., and Berry, D. I.: Trends in ship wind speeds adjusted for observation method and height, *Int. J. Climatol.*, 28, 747–763, 2008.
- 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., et al.: Guiding the creation of a comprehensive surface temperature resource for twenty-first-century climate science, *Bull. Am. Meteorol. Soc.*, 92, ES40–ES47, 2011.
- 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.
- Trewin, B.: A daily homogenized temperature data set for Australia, *Int. J. Climatol.*, 33, 1510–1529, 2013.
- Venema, Mestres, Aguilar, Auer, Guijarro, Domonkos, Venema, V. K. C., Mestres, 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.
- Vincent, L. A., Zhang, X., Bonsal, B., and Hogg, W.: Homogenization of daily temperatures over Canada, *J. Climate*, 15, 1322–1334, 2002.
- Vincent, L. A., van Wijngaarden, W. A., and Hopkinson, R.: Surface Temperature and Humidity Trends in Canada for 1953–2005., *J. Climate*, 20, 5100–5113, 2007.
- Wan, H., Wang, X. L., and Swail, V. R.: Homogenization and Trend Analysis of Canadian Near-Surface Wind Speeds., *J. Climate*, 23, 1209–1225, 2010.
- 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.
- 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.
- 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. Atmos.* (1984–2012), 117, 2012.
- WMO: Report of the meeting of the WMO CCI Task Team on climate aspects of resolution 40 (Geneva, 10–11 June 1998): final report, 36, *World Climate Programme Data and Monitoring*, 1998.
- Yan, Z. and Jones, P.: Detecting inhomogeneity in daily climate series using wavelet analysis, *Adv. Atmos. Sci.*, 25, 157–163, 2008.
- Young, K. C.: Detecting and removing inhomogeneities from long-term monthly sea level pressure time series, *J. Climate*, 6, 1205–1220, 1993.