Chapter 3

The Land Data

SUMMARY

The construction of the land data component of HadCRUH constitutes the largest part of this project. Unlike the marine data, the land component originates from a new and previously unstudied raw data source chosen for its spatial and temporal coverage and regular updates. Humidity data are extracted from a cumbersome database and developed from scratch to create the gridded product. This Chapter discusses: the data source; the creation and implementation of a set of humidity-specific quality control (QC) tests; data homogenisation; and analysis of the gridded product.

3.1 THE ORIGIN OF THE LAND DATA

3.1.1 The ISH Database

The version 2 ISH (Integrated Surface Hourly) dataset, supplied by NCDC (National Climatic Data Center), is the sole basis of the land data for HadCRUH (Lott *et al.*, 2001). The data comprise approximately 20 000 surface weather stations globally, reporting from 1900 to present. These originate from NCDC and the US Navy (ID code TD3280) and Air Force Combat Climatology Center (AFCCC) Datsav 3 (ID code TD9956) surface hourly data with 380 and ~10 000 stations presently active from each set respectively.

The ISH database reports observed surface T and T_{dw}, which for HadCRUH are converted to e, q and RH (Chapter 2). Initially, it was intended to produce HadCRUH in e, q and RH and so much of the quality control process takes place in e. However, it has become apparent that q and RH are of most use and relevance to the scientific community and so it is decided to focus only on these two variables in terms of analysis.

Humidity observations vary widely in instrument type, variable type and conversion algorithm. Unfortunately there are no comprehensive metadata providing instrument or measurement type information with the ISH data. Thus, for the purpose of HadCRUH, the assumption is made that all station measurements are sufficiently similar to be incorporated into a global dataset and that quality control procedures will eliminate the 'bad' data.

Observation frequencies range from sub-hourly (better than hourly) to daily. The sub-hourly data are averaged to hourly for computational convenience. All data 29 or less minutes previous to the hour or 30 minutes or less after the hour are averaged and then assigned to that hour.

All stations are identified by a five digit World Meteorological Organisation (WMO) number where the first two digits refer to country and region respectively and the following three are the station number. This is always followed by a sixth digit, usually a zero, which if non-zero often indicates a new period of reporting from a station in a new location. US run stations also have five digit US Weather Bureau Army Navy (WBAN) identification numbers. These are kept to allow easier identification with 99999 as the default number for those stations without WBANs. WMO country numbers have remained largely consistent over time although some modifications have taken place (Jones & Moberg, 2003). However, there is no comprehensive list available of such changes.

Each hourly observation is recorded with a longitude, latitude and elevation. For most stations (95 %) these are inconsistent within the station record, non-existent or in disagreement with the location and station identification list provided with the ISH database. However, these are nearly all due to small changes in recording precision rather than actual station moves. For example, stations reporting at six hourly periods to the GTS (Global Telecommunications System - http://www.wmo.ch/web/www/TEM/gts.html) may have different precision requirements than would otherwise be recorded. For HadCRUH, small inconsistencies (within 0.1 ^{o} longitude or latitude) are ignored and where possible, missing location information is filled in with WMO record information. In cases of disagreement, the locations supplied simultaneously with the data are used in preference. Notes are made of the dates of any large changes for future reference. These problems have been communicated to NCDC.

The ISH database has a good breadth of coverage both spatially and temporally (Fig. 3.1). A large proportion of the data comes from Europe, the former USSR and North America. Temporally the dataset is far from consistent with a relatively small amount of data pre-1972, a decrease in 1972 and a significant increase thereafter. The drop in 1972 is thought to be linked to the digitisation process (Vose, R *pers. comm.*) such that data exist but have not been digitised yet. Due to the much poorer data coverage pre-1973, HadCRUH begins in 1973, and for the purpose of this project, ends in 2003. However, since ISH is updated yearly there is scope for updating HadCRUH in near real-time in the future.

Notably, the ISH stations are not continuous in their reporting throughout 1973 to 2003. Other datasets have been built combining non-continuous sources regardless of length of record (Dai, 2006) and this approach is by necessity used in the marine component. However, given the large number of stations available for the land component and the desire to ensure high quality within HadCRUH it is decided to use only stations with record lengths sufficient to create a 30-year climatology.

3.1.2 The Climatology Period: 1974 to 2003

A 30-year climatology period is considered sufficient to provide a reference period with which to anomalise (Box 3.1) the data as a recognised method of interpolating over a gridded field (Jones, 1994). This practice is consistent with numerous other datasets (HadCRUT3 – Brohan *et al.*, 2006; global nighttime *MAT* dataset MOHMAT43N – Parker *et al.*, 1995; Rayner *et al.*, 2003; global sea level pressure dataset HadSLP2 – Allan & Ansell, 2006), which use the standard reference period of 1961-1990. However, for HadCRUH land data, using the standard reference period considerably reduces both station density and spatial coverage (Fig. 3.2). While a 1973 to 2002 climatology maximises station density, 1974 to 2003 is preferable in terms of spatial coverage (also shown in Fig. 3.2) giving much better representation over Sweden and Canada. Hence the climatology period for HadCRUH is 1974 to 2003.

Any selection criteria as to the minimum amount of data required to create a station climatology will have a degree of arbitrariness. The criteria used for HadCRUH are loosely based on Jones & Moberg (2003). Minimum criteria for ‘sufficient reporting’ are as follows:

- 4 reporting hours with data per day covering both halves of the diurnal cycle (midnight to midday, midday to midnight) (Box 3.2 - sampling frequency)
- 75 % of days with data per month
- 2 months with data per season
- 3 seasons with data per year
- 5 years with data per decade
- 2 decades with data
- 15 years within the climatology period

Coverage of stations with a 1974 to 2003 climatology (Fig. 3.3) is very good over Europe and South East Asia but poor over central Africa, Amazonia, parts of the Middle East and Antarctica. Surprisingly, coverage is poor over the United States. This information has been communicated to NCDC. On further investigation, there are large amounts of US data in ISH but this is mostly from short station records insufficient to create a 1974 to 2003 climatology. There is future scope for additional data sources in many of these regions to augment coverage.

3.1.3 Duplicate Stations: Finding, Combining and Deleting

There are a number of duplicate stations found within ISH. These are identified by: an identical WMO number and different WBAN number; an identical WBAN number and different WMO number; or a station location match. Stations may change ID numbers for many reasons:

- Reporting requirements - GTS stations may report 6 hourly as one station ID and hourly under another.
- Station/instrument move
- Station closure / re-opening
- Official WMO number changes

All 'duplicates' are double checked by ensuring the longitudes and latitudes are within 0.1 ^{o} and 20 m elevation. This distance represents approximately up to 11.1 km in longitudinal and latitudinal distance which is considered a small region in which to have more than one weather station reporting. Efforts are made to link up stations reporting during different periods of record under different station IDs and to remove exact duplicate stations. In total, 346 combined records of duplicate stations are created, improving coverage over Sweden, eastern North America, Romania and Siberia (Fig. 3.3).

For climate analyses, **anomalies** are preferable to **absolute** values because they largely remove station specific variability (Jones & Briffa, 1992) (due to elevation, observation times, annual cycle etc.) which represents noise clouding any climate change signal. This approach is employed in the vast majority of climate datasets at the surface (e.g. Dai, 2006; HadCRUT3 - Brohan *et al.*, 2006; HadSLP2 - Allan & Ansell, 2006; MOHMAT43N – Rayner *et al*., 2003).

A **pentad** is a five-day mean. There are 73 pentads in a year with six per month except August which has seven. The 12th pentad has the 29th February added in every leap year. They are a useful intermediate unit between hourly/daily and monthly means and are commonly used at the Hadley Centre (Met Office, UK) for marine observational data. Relative to raw data, running QC and homogenisation processes on pentads reduces programming time, essential for large datasets. Also, the signal to noise ratio is reduced but not over damped to the extent that erroneous data are hidden as may be the case with monthly means.

Pentad mean anomalies (**PMA**) are created by first creating **hourly pentad climatologies**. For example, all observations at 00 hours between the 1st and 5th of January from 1974 to 2003 are averaged. **Hourly anomalies** are created for each observation by subtracting the corresponding **hourly pentad climatology**. This removes the diurnal cycle. All **hourly anomalies** within each pentad of each year are averaged to give a **PMA** value. To create a pentad there must be at least three days (section 3.1.2) of data. **Pentad mean climatologies** are made by averaging over each **hourly pentad climatology**. **Absolute pentads** are created by adding back the corresponding **pentad mean climatology** to the **PMA**.

Monthly mean anomalies are made by averaging over the six (seven for August) **PMAs** for each month, where at least three **PMAs** must be present. This gives slightly larger **monthly mean anomalies** than if they were calculated directly (Rayner *et al.,* 2006; Taylor *et al*., 2000), but maintains consistency with the marine data (Chapter 4) and provides the benefits

Box 3.1: Pentad and Monthly Mean Anomalies

The mean diurnal range in humidity averaged over the entire timeseries for each case study station is 25.2 %, 0.87 g kg^{-1} and 1.33 hPa for *RH*, *q*, and *e* respectively. This represents a wide range of latitudes and all seasons, and so diurnal cycles for any given location can be larger. Bias may stem from systematic uneven sampling of the diurnal cycle, especially if sampling times systematically change over time (a source of inhomogeneity). Ideally, to fully capture daily means, data should be hourly. This is not always possible and so the effect of less than hourly data on monthly means is studied at a range of latitudes.

Less than hourly data adds little variance to mid and high latitude stations. However, at low latitudes (likely warmer and wetter), variance is considerable, of the order of 0.5 standard deviations (Fig. 3*i* and Table 3*i*). Hence, calculating monthly means from less than hourly data can increase variance which may bias certain data. To maintain good data coverage while keeping data quality, a compromise is made. HadCRUH will incorporate six hourly (or more frequent) data as long as there are at least four observations per day with at least one observation in each half of the diurnal cycle (midnight to midday, midday to midnight).

Observing Frequency	Minimum Median	Median Median	Maximum Median
3 hourly	-0.2	0.0	0.1
6 hourly	-0.3	0.0	0.2
12 hourly (00, 12)	-0.4	0.0	0.4
12 hourly (06, 18)	-0.5	0.0	0.3
Daily	-0.9	0.05	1.5

Table 3*i*: Median differences (normalised units) between monthly mean *e* from hourly data and from other sampling frequencies for all case study stations.

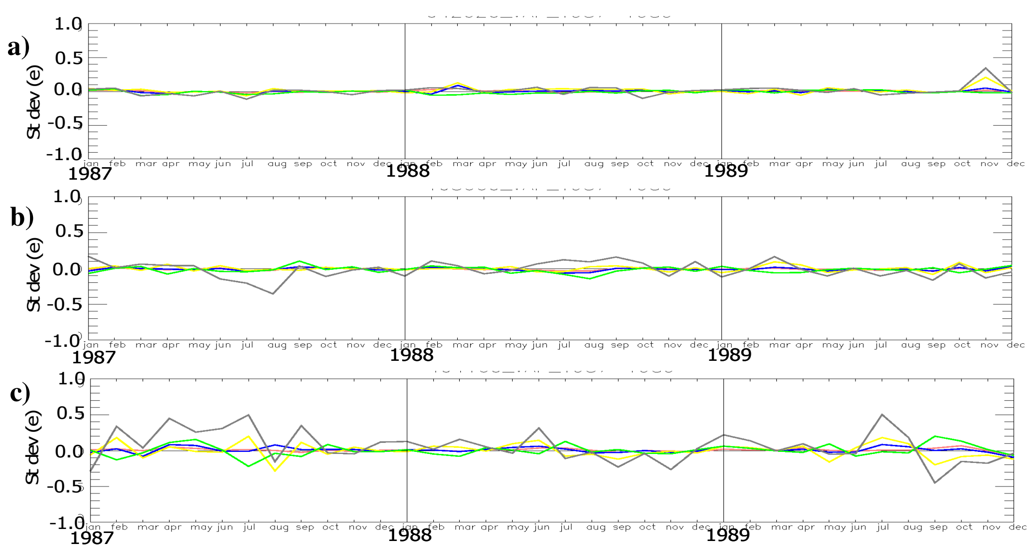


Figure 3*i*: Differences between monthly mean *e* from hourly data and other sampling frequencies. Units are normalised to each station. Monthly means are created from 3 hourly (pink), 6 hourly (blue), 12 hourly (00, 12 hrs) (orange), 12 hourly (06, 18 hrs) (green) and daily (grey) data. a) High latitude station 042020 (Thule, Greenland). b) Mid-latitude station 108660 (Munich, Germany). c) Low latitude station 404160 (Dhahran, Saudi Arabia).

Box 3.2: Observation Sampling Frequency

3.1.4 Comparing ISH with Other Data Sources

As a primary quality check, the raw ISH humidity data are compared with two other sources: the Hahn and Warren (HAHN) dataset reporting from 1971 to 1996 (Hahn & Warren, 1999) and the New (NEW) dataset reporting from 1945 to 2003 (New *et al.*, 2000). Both provide monthly mean *e* data.

Thirty three case study stations are selected for closer investigation (discussed further in section 3.2). There is good general agreement at monthly mean resolution (Box 3.1 for method) between ISH and the other sources in all case study stations. Seven randomly selected examples are shown in Fig. 3.4. Of note is station 718160 (Goose Bat, Canada) where the HAHN dataset gives consistently higher values. However, the ISH data are consistent with the NEW data, implying the problem lies with the HAHN data in this case.

For the purpose of this thesis, only the ISH database will be used in order to fully investigate the potential of ISH as a data source. Ultimately, future versions of HadCRUH would likely benefit from the inclusion of more data both from ISH and other sources.

3.2 REMOVAL OF POOR QUALITY DATA

3.2.1 Creating a Set of Quality Control Tests for Humidity Data

Quality control is a common procedure among dataset builders to eliminate the 'bad' data before undertaking homogenisation (Brohan *et al.*, 2006; Rayner *et al.*, 2006; Thorne *et al.*, 2005b). A substantial but not exhaustive two phase quality control has already been undertaken by NCDC (Lott *et al*., 2001). Phase one checked coincident data originating from different sources to ensure identical location and time before merging, where for each day at least 70 % of the data must be within certain physical thresholds (i.e. 1 ^{o}C for simultaneous *T* observations). A complete inventory of all input and output data was kept in addition to thorough checking of the software and output database for any problems. Phase two applied fifty-seven quality control algorithms (fully automated) to the data consisting of: validity checks; extreme value checks;

internal (within observation) consistency checks and external (versus other observations from the same station) consistency checks. Only two example tests are described in the ISH accompanying literature (Lott *et al*., 2001). The first test removed 'spikes' of data where T values were changed to 'missing' when the difference from the previous and subsequent hour's value was greater than 8 ^{o}C. The original value was saved in a separate section of the data record for future use. The second example test compared present weather with T to ensure consistency. Test data were created and verified to check each algorithm for problems. The NCDC quality control did not include any attempt to homogenise the data with spatial comparisons. Each element of each data record is accompanied by a corresponding flag referring to whether it is 'good', 'suspect', 'erroneous', 'missing' or not checked for quality. Only data flagged 'good' continue through to HadCRUH.

For HadCRUH, further quality control testing more finely tuned to the issues surrounding humidity measurement is necessary prior to homogenisation (section 3.3). The case study stations are used to develop robust quality control criteria. These stations are chosen for geographical coverage (Fig. 3.5) and because they have long records and report with high frequency. Station details are listed in Table 3.1. In total eight hypothesised potential issues of 'quality' are investigated.

ISSUE 1) Physical Constraints on Meteorological Variables: Bad Values QC

H_0: All data are physically plausible within the realms of the climate system and observing practices.

Rationale:

Temperature and humidity are constrained by physical limits such that 'bad' (physically unreasonable) values can be easily identified and removed. These erroneous values may be due to malfunctioning instruments, poor observing practices or transmission errors.

Test and Results:

A simple check list for 'bad' values for each variable is employed. Full results are shown in Table 3.2. In total 0.003 % of case study station humidity data fail these tests.

There is no geographically coherent pattern to these results implying that a one-size fits all approach is valid.

Implications:

Although only a small amount of data is identified by these checks, the null hypothesis is rejected, and this test must, therefore, be part of the quality control process. It is possible that in very extreme conditions some of these criteria may be breached, in particular the 100 % *RH* boundary. However, very few instruments, especially those employed in operational observing stations are capable of measuring such extremes with any accuracy (Makkonen & Laakso, 2005).

ISSUE 2) Strings of Repeated Values: Repeats QC

H_0: Repeated *T* values are not a problem within the ISH data.

Rationale:

Input or instrumental error can lead to sustained periods of identical values, especially at unmanned stations. Given the diurnal cycle and synoptic variability, it is highly unlikely that *T* would remain constant (at 0.5 ^{o}C accuracy) for long periods of time in the vast majority of locations. Therefore it is reasonable to conclude that such occurrences are more likely due to error. One exception is that T_{dw} can remain fairly consistent over time (Ahrens, 2000), certainly within reporting accuracy, and so is not as useful for such a test.

Test and Results:

Each station *T* record is searched for continuous strings (> 12 hours) of identical values. In total 0.37 % of case study data are part of a 12+ hour repeated *T* string (Fig. 3.6). High (>100) frequencies of occurrences are found for all latitudes. Eighteen case study stations have over 50 occurrences identified and three stations (788060, 108660 and 042020) have more than 1 % of data removed by this test.

The Polar North has the highest frequency of occurrences, especially in the seasons of JJA (1.47 %) and SON (September, October and November) (1.15 %). It is possible that this is a real physical signal given the damping of the diurnal cycle during 24 hour daylight. However, further investigations yield no similar pattern in Polar South. On closer inspection of station 042020 (Thule, Greenland, Polar North), string repeats are concentrated in pre-1990 data suggesting instrument or observing practice issues rather than seasonal causes. Therefore it is concluded that the majority of such strings are unlikely to be real events.

Implications:

There are sufficient occurrences of repeated *T* strings to reject the null hypothesis and warrant including this test within the quality control process. The first value of each string repeat is kept in the quality controlled data, but the rest are removed.

ISSUE 3) Events of Continuous Zero Dewpoint Depression: Zero DPD QC

H_0: Wick-drying or screen freezing is not a problem within the data.

Rationale:

The vast majority of humidity data originate from wet-bulb thermometer measurements. There are however, a suite of problems associated with such instruments as described in section 1.3. Screen freezing and wick drying due to reservoir freezing or evaporation leads to continuous periods of precisely 100 % *RH*, which are thought unlikely to occur naturally in the majority of cases and therefore represent, at least in most cases, erroneous data.

Test and Results:

Each station record is searched for events of continuous strings of (>12 hours) 0 ^{o}C dewpoint depression (*DPD*) ($T - T_{dw}$). In total, 0.54 % of the case study humidity data are identified by this test (Fig. 3.7) and 16 case study stations have over 50 events. This is a problem over all latitudes without any particular bias towards regions where sustained rainfall is typical such as the Tropics. This suggests that this test is sufficient to pick out drying or freezing events without biasing the data towards dry conditions by erroneously removing wet condition data. Station 837460 (Galeao, Brazil, Extra-tropical South) has a relatively high frequency of events (544), totalling 3.57 % of the data. These are relatively evenly spread through the timeseries. Seasonally, for Galeao, events are most common during SON and least during MAM (March, April and May), affecting 4.31 % and 2.65 % of the seasonal data respectively. Generally the climate of Galeao lies around 20 ^{o}-30 ^{o}C and never drops below zero. The *RH* does frequently become as low as 20 %, thus wick drying is likely a real problem.

Stations reporting with less than hourly frequency make such an event more difficult to detect. Indeed, this is demonstrated by the fewer occurrences at these stations (stations denoted with an S in Fig. 3.7). This is an argument for only using stations that observe with a high frequency (Box 3.2).

Implications:

Frequent occurrences of > 12 hour strings of 0 ^{o}C *DPD* events are found in most of the case study stations and thus the null hypothesis is rejected. All 0 ^{o}C *DPD* strings of 12 hours or more will be removed from the dataset. In future versions it may prove beneficial to remove entire stations where event frequency exceeds a certain threshold. However, for the first version of HadCRUH, these will be kept to preserve station coverage. As a final check on this, a geographical analysis of data removal due to this test from all stations will be undertaken (section 3.2.2).

ISSUE 4) Recording Problems in Temperature Extremes: Cutoffs QC

H_0: There are no noticeable problems with data recording in extreme temperatures.

Rationale:

It has been common practice for some countries to record extreme values as either a set limit or missing. For instance, in the US radiosonde record (until 1993), T_{dw} was not recorded when *T* fell below -40 ^{o}C. In addition, in the US and surrounding countries, when *RH* fell below 20 % a standard dewpoint depression of 30 ^{o}C was reported (Ross

& Elliott, 1996; Elliott, 1995). Conversion to e, q and RH requires simultaneous recording of T and T_{dw}. Therefore a practice of not reporting T_{dw} or artificially setting T_{dw} values in extreme temperatures will introduce bias into the humidity data.

Test and Results:

Frequency distributions of T and T_{dw} are scrutinised for all case study stations. No evidence is found for T_{dw} values being reported as a set value in extreme temperatures. However, station 702220 (Galena, Alaska, USA) is potentially an example of humidity data not being recorded at low temperatures. Below -37 ^{o}C the distribution of all T values differs to the distribution of T values with simultaneously measured T_{dw} (Figs. 3.8 a and b). This does not occur in all years. Station 042020 (Thule, Greenland) also shows similar ‘cutoffs’ at low temperatures in some years. For comparison, decadally averaged T distributions from nearby US and Greenland stations are scrutinised and comments listed in Table 3.3. Cutoffs appear in these other stations and commonly at a T of ~ -37 ^{o}C although not consistently across the timeseries. Failure to record simultaneous T_{dw} with T may be a problem at other latitudes and temperatures possibly due to episodes of wick drying or instrumental limitations.

To test for cutoffs only, each year of each case study station is checked for the percentage of T data with simultaneously observed T_{dw} by 10 ^{o}C bins. The entire year is considered unusable if for any one 10 ^{o}C bin, less than 90 % of T observations have a simultaneous T_{dw} value. Bins containing less than 10 observations are not tested.

The problem is found to be widespread, affecting 12 of the case study stations (Table 3.4). Although cutoffs are mostly at low temperatures, it is not exclusively so and they affect data at all latitudes. This test results in the exclusion of station 702220 from HadCRUH because there are insufficient problem free years with which to create a climatology. Stations 042020 and 783970 are also greatly affected.

Implications:

Inconsistent recording of T_{dw} observations is evident in the data and so the null hypothesis is rejected and this test is part of the quality control.

ISSUE 5) Reporting Timezones: Timezones QC

H_0: All data are converted to GMT and are correctly sampling the diurnal cycle.

Rationale:

It is assumed that all data contributing to HadCRUH have accurate reporting times such that the diurnal cycle remains consistent throughout each station record. Also, according to Lott *et al.* (2001) all stations have been converted to Greenwich Mean Time (GMT). Humidity, although to a lesser extent than T and not in all stations, exhibits some degree of diurnal cycle (Robinson, 1998) (Box 3.2). Incorrect or inconsistent reporting times will likely affect the accuracy and homogeneity of the data, especially where the diurnal cycle is strong.

Test and Results:

A mean e diurnal cycle is calculated for each year for each case study station where data are six hourly or more frequent. Each annual mean diurnal cycle is normalised to give a curve with 24 (or less depending on observing frequency of the station) points between -1 and 1. A sine curve is used to give 24 (or less) values describing an ideal diurnal cycle with the maxima at hour zero. The two vectors are then correlated to give an r (correlation coefficient) value. The sine curve is then shifted along by one hour at a time and re-correlated resulting in 24 (or less) r values. This is then plotted as r versus time shift. The peak in the maximum r identifies the annual average maximum of the diurnal cycle at GMT where each time shift form 0 to 23 corresponds to the 24 hour clock. For example, station 037720 (Heathrow, London) is at GMT and the maximum r occurs around the 15th time shift (3pm). Station 723710 (Page, Arizona, USA) is approximately 7.5 hours behind GMT in terms of longitudinal distance (officially at US Mountain Standard time which is 7 hours behind) and so the maximum r occurs at the 22nd time shift (10pm). From these plots it is possible to tell if the station has been converted to GMT and if all years are within good agreement.

Scrutiny of these plots shows that all stations have been converted to GMT. Only two stations (895710 and 284400) show no perceivable common diurnal cycle. However, 29 case study stations have a strong and consistent diurnal cycle with minimal inconsistent years. An attempt is made to create an objective and automated test to detect annual differences in the diurnal cycle using five criteria and then attribute those changes to a period of data within that year. These are as follows:

Test 1: Each annual mean set of *r* values is subtracted from the station mean set of *r* values. If more than 50% of difference values are less than 0.4 or greater than -0.4 the test is failed.

Test 2: If the difference between the maximum and minimum *r* value for each annual mean set (measured in time shifts (hours)) is less than 9 or greater than 15 then the test is failed.

Test 3: If the absolute difference (in time shifts) between each annual mean maximum *r* value and the station mean maximum *r* value is greater than 3 then the test is failed.

Test 4: If the absolute difference (in time shifts) between each annual mean minimum *r* value and the station mean minimum *r* value is greater than 3 then the test is failed.

σ Test: The standard deviations (σ) of the *r* values for each annual mean at each time shift are calculated. If any σ exceeds 0.5 then the test is failed.

Nine case study stations have at least one year failing at least one criterion (Table 3.5). Figure 3.9 shows station 702220 as an example of a station where all but two years show accurate GMT conversion and strong agreement. On further analysis, the data in station 702220 become significantly moister (~4 σ from the mean) and more erratic from December 2000 to February 2002. Ideally, each failed year would then be further tested to identify the period or cause of the problem in an automated way. Unfortunately, it has not been possible to create a test that performs satisfactorily in the timeframe available. Notably, the problem with years 2000 to 2002 at station 702220, and others like it, should be picked up during the homogenisation process.

Implications:

The first part of the null hypothesis is accepted as all case study stations are converted to GMT. However, the diurnal cycle is not always consistent and so the second part of the null hypothesis is rejected. While this test proves unsuccessful in applying improvements, the σ Test will be used to list suspect stations for further scrutiny in later versions of HadCRUH. The later homogenisation process should account for any major discontinuities caused by changes to diurnal cycle sampling or representation, and use of gridded monthly mean anomalies should minimise any effects on the end product.

ISSUE 6) Wind Speed: Wind QC

H_0: Poor ventilation at low wind speeds does not have a significant effect on humidity.

Rationale:

All wet-bulb thermometers and some other hygrometers such as the Dewcel depend on adequate aeration to enable evaporation. Non-artificially aerated wet-bulbs depend on good location and sufficient natural air flow through the Screen, known as the ventilation rate (v) which is in ms^{-1}. Figure 3.10 shows the linear relationship between mean wind speed (U) in ms^{-1} and the ventilation rate in ms^{-1} in the Screen which is described by the following equation (Folland, 1977):

$$v = -0.08 + 0.15U \qquad \text{Eq.3.1}$$

The accepted mean ventilation rate is 1.25 ms^{-1}, which would require a mean wind speed of ~9 ms^{-1} (Folland, 1977). For stations that observe wet- and dry-bulb temperatures a psychrometer coefficient is required to convert $T - T_w$ to other humidity variables. This value is commonly 0.8×10^{-3} K^{-1} which is suitable for a ventilation rate of 1 to 1.5 ms^{-1} (mean wind speed of 7.2 -10.5 ms^{-1}). At some stations the psychrometer coefficient is increased for low ventilation rates and decreased for higher ones, thus accounting for the inhibited evaporation due to near stationary air within the Screen at low wind speeds and over-ventilation at high wind speeds. To complicate things further, the actual psychrometer coefficient differs with instrument type. For example, at v = 1.25 ms^{-1} the psychrometer coefficients for the Mercury-in-glass and Resistance thermometers were found to be 0.87×10^{-3} K^{-1} and 0.70×10^{-3} K^{-1} respectively (Folland, 1977). Humidity values derived from the common psychrometer coefficient and the Mercury-in-glass thermometer T_w were always overestimated. Those derived from the Resistance thermometer T_w were overestimated when $v < 0.3$ ms^{-1} (U = ~2.5 ms^{-1}) and underestimated when $v > 0.3$ ms^{-1}. Table 3.6 lists the approximate error in each variable for both instruments assuming that RH = 75 % and $T + T_w$ = 18 °C based on findings from Folland (1977). Error increases with decreasing RH but the relationships are non-linear preventing simple linear extrapolation over the values shown in Table 3.6.

As the proportion of stations using naturally aerated wet-bulb thermometers is not known, there can be no assumption about the potential scale of this problem. Compiling information on instrument type and psychrometer coefficient where used for an entire global dataset is a task beyond the scope of this thesis and likely to prove near-impossible for all land stations. Thus detecting problems that result in errors of either sign (therefore effecting no systematic bias when considering grid-box monthly mean anomalies), such as: using an inaccurate psychrometer coefficient for actual wind speed at the time; using a psychrometer coefficient that differs from that of the instrument employed; or over-ventilation, is considered unfeasible at least for version 1 of HadCRUH. Detecting under-ventilation (U < 2.5 ms^{-1}) which always results in a positive bias, may be easier.

Test and Results:

An ideal test would be to compare stations that use naturally aerated wet-bulb thermometers with neighbouring stations employing a different instrument type. Due to paucity in readily available metadata, this is not possible. Given the results from the two common wet-bulb thermometers shown in Table 3.6, a bias in e due to under-ventilation should be less than 1 hPa. However, the range of e values calculated at low wind speeds (<2.5 ms^{-1}) is too large in all case study stations, even when analysed seasonally, for such a small error to be detected. Station 042020 (Thule, Greenland), which has the smallest seasonal range, but this still exceeds 5 hPa – an order of magnitude larger than any likely effect. As a further example (Fig. 3.11) , station 404160 (Dhahran, Saudi Arabia) exhibits a relatively large range when compared to the other case study stations which is smallest in DJF (~20 hPa) and largest in JJA (~45 hPa). All stations are scrutinised in this manner but no effects of spurious moistening in low wind speeds are visible to the eye. This is, however, expected considering the likely magnitude of the effect.

Implications:

It has not been possible to either identify any stations where measurement errors due to low wind speed are apparent or to quantify the extent of this problem within the land data. The effect of these relatively small errors in hourly station data should be greatly reduced by using grid-box monthly mean anomalies. Furthermore, if it is assumed that mean wind speed remains largely constant over time at any one station then errors will be consistent over time, and so although having some small effect on the climatology, they should not affect any analysis on changes in humidity such as trend fitting. Finally, simultaneous wind speed observations do not always accompany the ISH humidity data and ISH wind data quality is as yet unknown and would likely require its own set of quality control tests. Therefore it is deemed that: further investigation; ensuring the quality of the ISH wind data; and the creation of automated detection tests suitable to be applied to 3000+ stations are beyond the focus of this project but may be desirable for future versions.

ISSUE 7) Problems with Station Elevation

H_0: The homogeneity of a grid-box will not be compromised by containing both high and low elevation stations.

Rationale:

The effect of elevation on spatial continuity in humidity is complex and relatively unstudied. A single 5 ^{o} by 5 ^{o} grid-box may incorporate a number of stations at a wide spectrum of elevations. If the large scale variance in station timeseries is very different, at different altitudes, resulting in poor correlation of station timeseries, the homogeneity of the grid-box could be compromised. This has previously been considered a problem for *T* resulting in stations above 2500 m being excluded from previous versions of HadCRUT3. Such stations are now however included (Brohan *et al.*, 2006).

Test and Results:

Four pairs of neighbouring high/low elevation stations are analysed (Fig. 3.12, Table 3.7). Both the statistical mean and variance are smaller at the higher elevation stations.

Using simple linear Pearson regression, the timeseries for each pair are correlated. An assumption is made that an r of less than $1/e$ (0.37) shows poor correlation that would compromise the homogeneity of a grid-box if both stations are included. This criterion is based on that used by Briffa & Jones (1993) to define ‘spatial correlation decay lengths’ for meteorological variables.

The European pair correlate well, despite their large horizontal distance (greater than that of a 5 ^{o} by 5 ^{o} grid-box), with an r value of 0.81. However, they are within 500 m vertically. The Japanese pair have a large vertical distance (1286 m) but are horizontally close and correlate well with an r value of 0.87. The USA pair also have a large vertical distance and are further apart horizontally than the European pair yet still correlate with a low but sufficient r value of 0.38. The Colombian pair have the greatest vertical distance but are relatively close (within a 5 ^{o} by 5 ^{o} grid-box). They have a poor correlation with an r value of 0.24. However, these stations are on different sides of the Cordillera Mountain Range with Cali on the moister seaward side and Bogota on the more arid landward side.

These correlations compare well with the globally averaged correlation decay distances (CDDs) calculated by New *et al.* (2000) using the same $1/e$ threshold as Briffa & Jones (1993). This was the distance over which any given station can be expected to bear some climatic relation to its neighbours. For precipitation, diurnal temperature range and mean T these distances were 450 km, 750 km and 1200 km respectively. Two points can be drawn from this analysis. Humidity can have a strong spatial consistency horizontally, in agreement with predefined CDDs for precipitation, diurnal temperature range and mean T, despite possibly large vertical ranges. However, this can be strongly influenced by features of topography other than just elevation (e.g. slope aspect, prevailing wind direction etc.).

Fig. 3.13 maps the stations meeting the climatology criteria (section 3.1.2) colour coded by elevation. It is necessary to retain stations over 1000 m to maintain good coverage. Although very high elevation stations (>2500 m) make up a very small amount of the data, excluding these will leave large gaps over the Himalaya and Tibetan Plateau which is undesirable.

Implications:

High and low elevation stations can correlate sufficiently. However, there is a degree of ambiguity depending on aspect and other topographical features within each grid-box and so the null hypothesis is rejected. Very high elevation stations have added problems with pressure conversion algorithms performing less well. This is particularly important for humidity as conversions for e and q use pressure converted from sea level to actual level using the given station elevation.

As there is no clear or simple way of dealing with this problem all stations will be kept regardless of their elevation for now. Homogenisation will take account of elevation difference as only those stations within 1000 m vertical distance and correlating sufficiently will be considered as 'neighbours'. Furthermore, ultimately only those with at least five neighbours will be considered in the homogenisation and hence be used in HadCRUH (section 3.3). Ultimately, the first version of HadCRUH will allow a more detailed study of humidity at high elevations enabling later versions to employ a more robust and sensitive approach.

ISSUE 8) Outlier Removal: Outliers QC

H_o: All data are within a reasonable measure of the climatology and have no unrealistic jumps in the timeseries.

Rationale:

An 'outlier' check is a common and valuable quality control test for timeseries data to avoid bias caused by a few erroneous observations (Parker *et al.*, 1995; Rayner *et al.*, 2006; Brohan *et al.*, 2006; Dai, 2006). The removal of outlying values, that are most likely inaccurate, prevents skewing of mean values towards these 'extremes'. These outliers may be individual or clusters of observations. They can occur for a number of reasons:

- Measurement unit: *T* may have been recorded as Fahrenheit or Kelvin rather than Celsius.
- Input error: Values may have been input with the incorrect decimal place or just incorrectly.

- Digitisation and transmission errors
- Instrumental errors: Malfunctioning instruments may give random values and go unnoticed for short periods of time.

Test and results:

This test is undertaken in *e* at the pentad (five-day mean – Box 3.1) resolution. To avoid removing gradual and more likely real deviations from the climatology a pentad mean anomaly first difference ($t - t_{-1}$) and standard deviation (σ) timeseries are created for each station. Four tests are applied to identify pentads to remove from the dataset:

1. Individual/group outliers: Pentad mean anomalies greater than 3 σ from the mean and with a first difference series value greater than 2 σ are removed. The next pentad (and so on) is also removed if it too is greater than 3 σ from the mean.
2. Sudden spikes: Pentads with two adjacent first difference series values greater than 3 σ are removed.
3. Post-gap outliers: Pentad mean anomalies greater than 3 σ and preceded by missing data are removed. The next pentad (and so on) is also removed if it too is greater than 3 σ from the mean.
4. Sandwich outliers: If both pentad $_{t-2}$ and pentad $_{t0}$ are removed then pentad $_{t-1}$ will be removed too because the whole period is suspect.

As *e* is approximately proportional to *q* (Peixoto & Oort, 1996; Barry & Chorley, 1998; McCarthy & Willett, 2006), an outlier test in both variables is unnecessary. In the marine data (section 4.2) the removal of data by the Outlier QC test using *RH* is proportionally much less (< 50 %) than for the Outlier QC tests using *e* and *q*. Given this information and that *RH* is derived from the same humidity value, a separate Outlier QC in *RH* land data is not deemed necessary either. Furthermore, *RH* is characterised by large ranges over relatively short temporal periods, especially in dry regions. For example, in the Canadian prairies, summer *RH* can range from 40 % in the afternoon to over 80 % at night (van Wijngaarden & Vincent, 2005). As a result, a three σ outlier threshold could potentially be very large thus reducing both the effectiveness and value of an *RH* specific outlier test.

Station 843770 (Iquitos, Peru) is shown as a representative sample of the case study stations with a relatively large amount of data removal (Fig. 3.14a). All identified pentads are considered appropriate for removal. This has the effect of dampening of high resolution variance in the timeseries (Fig. 3.14b) which is as expected. Overall, the case study station data removal is 0.2 %, 0.07 %, 0.005 % and 0.05 % for tests 1 to 4 respectively. There are no previous assessments of how much data should be removed by these types of tests with which to compare.

On investigation of each of the case study station timeseries alongside pentads identified by this test for removal, the Outlier QC is found to be satisfactory in identifying and removing suspect data. No seasonal patterns of removal are apparent. As a second test, each station distribution (of hourly *e* values) is scrutinised before and after data removal. The majority of stations have data removals from the middle or whole spectrum of the range of *e* values. As this test is conducted at a pentad resolution it is quite possible that a pentad is skewed high or low due to only a proportion of hourly values within it such that hourly values from well within the normal distribution are removed too. In addition to removing data some distance from the mean this test also looks for sudden spikes or clusters of data that differ considerably from rest of the data for that period. For example, if a station timeseries has a period where humidity is slightly higher than the mean and then one (or a group) of pentads drops sharply back down to the mean or below and then returns to this higher level, this small period may be removed even though it is perhaps close to the mean because such a sudden and large deviation from the surrounding data is physically unlikely.

Data removal appears to be randomly spread across all latitudes (Fig. 3.15). Removal is high in station 702220 (Galena, Alaska, USA) largely due to the highly suspect period between 2000 and 2002 discussed earlier in *issue 5* (this section). Harsh weather conditions, Screen and wick icing and reservoir freezing associated with Polar latitude stations may be responsible for the relatively high data removal apparent in these stations. Comparatively high data removals also occur in some stations in the Tropics which adds support to this test because no bias is shown towards larger removals from certain regions.

Implications:

Most stations incur data removals due to this test so the null hypothesis is rejected and this test included in the quality control. There are no systematic patterns emerging from data removal by this test and for the case study stations data removals are considered valid. Consequently, this test will be used for the whole dataset. A further outlier check will be run during homogenisation incorporating neighbour comparison (section 3.3). Later versions of HadCRUH may require a more sensitive test, possibly more specific to latitude.

3.2.2 Running the Quality Control Process: From Raw Hourly Data to Pentad Mean Anomalies

The final quality control process compiles the six issues discussed above that were found to be important and successfully tested for (see flow diagram in Fig. 3.16). All stations with a sufficient record length (section 3.1.2) and sampling frequency (Box 3.2) that pass through the first quality control (Bad Values, Repeats, Zero *DPD* and Cutoffs QC tests) are then averaged up to pentad climatologies and pentad mean anomalies (Box 3.1). There is then a second quality control (Outlier and Timezones QC tests) followed by recalculation of pentad climatologies and pentad mean anomalies.

In total 7.68 % of case study station data are removed by the quality control (Fig. 3.17). This is dominated by Cutoff QC removals (6.62 %). Station 702220 is completely removed by this test. Four stations (042020, 404160, 726650 and 783970) are analysed in greater detail (Fig. 3.18) with 19.0 %, 0.3 %, 10.5 % and 32.7 % of data removed respectively. These data removals have the effect of slightly lowering the mean in stations 042020 and 726650 making the timeseries appear dryer. Conversely, station 783970 displays a more positive mean giving a moistening effect of quality control data removals. Although undoubtedly there are occasions where the quality control removes ‘good’ data, this is inevitable and outweighed by the benefit of removing ‘bad’ data.

The quality control process as described is applied to run through all 4760 stations. Removals are similar in scale and magnitude to that of the case study stations with the Cutoff QC responsible for the largest data removal (Fig. 3.17). Overall, the percentage removals from all data confirm the case study stations as being a broadly representative sample set of stations. In total the quality control procedure removes 5.74 % of the raw data. In addition to this, those stations passing through to a pentad dataset must contain sufficient remaining data to meet the sampling thresholds listed in section 3.1.2. This leaves 3514 'good' stations.

As mentioned earlier, there is a possibility of bias resulting from the erroneous removal of truly 100 % *RH* data due to the Zero *DPD* QC. If this is the case then it might be expected that proportionally more data are removed from typically wet regions such as the Tropics and monsoonal regions and fog prone regions. No such pattern is apparent for the Tropics and monsoonal regions (Fig. 3.19). However, on comparison of Figure 3.19 with the annual mean (over the period 1971 to 1996) percentage of sky obscured by fog for the global land area (Hahn & Warren, 2006) there may well be some correlation with fog prone regions. Over Europe especially, a region where sky obscuring due to fog is relatively higher, stations with > 1 % data removal are common. The large removal of Romanian data, however, is likely due to poor data quality because Romanian stations incur high data removals from the other quality control tests.

The decision to include this test is justified based on findings from the case study stations and overall data removal is not large (1.25 %) when compared to the 13 % of data removed due to wick drying events in the 2003-2004 GCOS plastic screen trial at three British stations (Elms & Hatton, 2005)). Some real 100 % *RH* events are undoubtedly removed and there is a risk, although likely small in effect (due to the small amount of data removed), of underestimating the climatology and any moistening trends. Further refinement of this test is necessary for any future versions of HadCRUH.

Coverage of stations passing the quality control is good in most land areas excluding high latitudes (>70 ^{o} N and >60 ^{o} S), South America, North and Central Africa, Saudi Arabia and the Middle East (Fig. 3.20 a and b). The US is fairly sparsely covered for reasons discussed in section 3.1.2. The majority of the data removals clustered in Austria, Romania and Japan are due to either the station length being too short or reporting frequency less than six hourly.

3.3 HOMOGENISATION OF THE LAND DATA

3.3.1 Why Homogenise?

If variability within an observational timeseries is attributable solely to weather and climate, then that timeseries can be considered ‘homogenous’ (Conrad & Pollack, 1962, in: Jones *et al.*, 1999). Inhomogeneities may be abrupt or gradual discontinuities within the timeseries (Easterling & Peterson, 1995). Abrupt discontinuities to humidity are regarded as highly unlikely as real signals of global climate change. They can be caused by a number of factors, identical to those affecting *T* (Jones *et al.*, 1985):

- INSTRUMENTS: Changes in type or exposure.
- RECORDING PRACTICES: Changes in precision, measurement technique and switches to automation.
- STATION: Change of instrument location.
- CONVERSION ALGORITHM: Changes in the method for converting the measured humidity variable to the desired reported variable or temporal resolution (Elliott & Gaffen, 1993).

As examples of the above, Canadian conversion algorithms, in the form of psychrometric tables, were refined in the 1960s and the metric system was adopted in 1977. In addition, from the early 1970s psychrometers were replaced by the Dewcel leading to a decreasing shift in *RH* at a number of stations (van Wijngaarden & Vincent, 2005). Other factors can directly affect the local climate. For example, at the microscale, the growth of a nearby tree may affect the local climate through shade and evapo-transpiration. Furthermore, changes to the hydrology of an area such as the introduction or cessation of irrigation practices or the construction or draining of a local reservoir (Elliott, 1995) can effect real changes in humidity. The latter of these is an example of a gradual discontinuity which is much harder to detect.

Where possible, it is desirable to remove these inhomogeneities of non-climatic origin to enhance signal detection in climatic analyses (Easterling & Peterson, 1995; Jones *et al.*, 1999; Vincent *et al.*, 2002). Although little work has been done on homogenising humidity data (Vincent *et al.*, 2002), the principal ideas are the same and so methods of homogenisation should be largely transferable.

Of the many methods of homogenisation (see Peterson *et al.*, 1998 for full review), a modified version of the reference station method (Easterling & Peterson, 1995; Jones *et al.*, 1999; Potter, 1981; Alexandersson, 1986; Young, 1993; Peterson *et al.*, 1998) has been chosen for use with the land data. This is a reasonably common method of detection where the candidate station is compared with a quasi-consistent independent background field, assumed largely free of inhomogeneities. A discontinuity appearing in the candidate station series but not the background field is treated as a breakpoint. The series before the breakpoint can then be adjusted appropriately.

There are four major caveats with this technique:

- an abrupt discontinuity may appear in both the candidate station and background field and thus go undetected;
- gradual discontinuities will not be well detected;
- homogenisation may introduce errors itself (structural uncertainty – Thorne *et al.*, 2005a); and
- it is a subjective and non-replicable test.

The first may occur if, for example, a change to instrument, algorithm or recording practice is implemented simultaneously at large number of stations. Despite these caveats, the benefits of homogenisation in terms of improving the stationarity of the dataset for climate studies, are recognised across the scientific community (Peterson *et al.*, 1998; Thorne *et al.*, 2005b; Brohan *et al.*, 2006; Vincent *et al.*, 2002; Jones *et al.*, 1985; Jones *et al.*, 1986).

3.3.2 Homogenisation of Specific Humidity

The size of the land dataset after the quality control (3514 stations) requires a largely automated approach to homogenisation but with some manual input necessary for breakpoint identification. The technique is modified from that used in the creation of the radiosonde temperature dataset HadAT (Thorne *et al.*, 2005b). Homogenisation takes place at the pentad mean anomaly (Box 3.1) resolution. This avoids very high resolution noise and enables a second outlier test. A flow diagram of the process is presented in Fig. 3.21.

Neighbour composites are used as a consistent but independent background field. For the *q* data 3391 stations are able to have a neighbour composite and difference series (candidate station – neighbour composite) created. The methodology is described in Box 3.3. The data series are then passed through the homogenisation process as described in Fig. 3.21. As humidity is generally spatially consistent over 100s of km (section 3.2.1), any real spikes in humidity seen at one station can be expected to appear in a sufficiently correlating but independent background field. Thus the creation of independent background fields for homogenisation provides an opportunity for further and more sensitive outlier checking. This is done simultaneously with the homogenisation. Any pentad in the difference series greater than four standard deviations from the mean is removed.

Neighbour composites are created using a similar convention to that of Briffa & Jones (1993) and Thorne *et al.*, (2005b). Each station is assigned a 1 ^{o} by 1 ^{o} grid-box. All stations within a **potential correlation region** of 50 ^{o} longitude and 10 ^{o} latitude on either side of the **candidate station** are **potential neighbours**.

NCEP reanalysis data of monthly mean (*q* and *RH*) anomalies for 1973-2003 (Kalnay *et al.*, 1996) are used to create an **actual correlation region** for each station. This consists of all grid-boxes (1 ^{o} by 1 ^{o}) within the **potential correlation region** that correlate with the **candidate station** grid-box timeseries with an *r* value greater than 1/*e*. All stations within those grid-boxes and within 1000 m elevation of the **candidate station** become **actual neighbour stations** and are given the grid-box *r* value as a **weighting coefficient**. The elevation requirement accounts for the poorer spatial continuity of humidity vertically than horizontally and the possibility that NCEP reanalyses may not represent this accurately.

A **neighbour composite** is made by creating a weighted average over all **actual neighbours** with a caveat that there must be at least 5 neighbours to avoid spurious breakpoint assignment as much as possible. A **difference series** is also created of the **candidate station** minus **neighbour composite**.

Box 3.3: Creating Neighbour Composites

Breakpoints are identified using a Kolmogorov-Smirnov (K-S) test on two years of difference series data before and after each pentad where at least one third of data are present. The two year subsections are matched for temporal coverage to avoid the effect of differential sampling across the seasons. This test is chosen because it is nonparametric, more liberal than a t-test (tried and found to be unsuitable) and does not assume a normalised distribution. Pentads where the K-S test gives at least 0.01 % significance and that remain significant at this level for at least six months (37 pentads) thereafter, are labelled as potential breakpoints. This period of six months sustained significance is used to avoid over-identification of breakpoints which was a problem with the other methods tried.

All potential breakpoints are then scrutinised manually alongside its neighbour composite, difference series, K-S test results and potential breakpoints. The panel for candidate station 747320 (Holloman, New Mexico, USA) is shown with decisions and reasoning annotated (Fig. 3.22). Decisions are based on the K-S test value, approximate means (as guessed by eye) of the sections either side of the breakpoint and whether the discontinuity genuinely appears to originate from the candidate and not the neighbour series. Only obvious breakpoints are accepted. All accepted breakpoints are adjusted for by adding to all pre-breakpoint data the difference in medians of the difference series before and after the breakpoint (described in Fig. 3.21). When all adjustments have been made each station anomaly timeseries climatology is recalculated and the anomaly series adjusted accordingly (later referred to as re-normalising). Inevitably, there are both erroneously accepted and rejected breakpoints. However, in the case study stations at least (Fig. 3.23), adjustments effectively moderate or in many cases entirely remove discontinuities in the q record without completely removing trends implying that the method works sufficiently well.

Before commencing actual homogenisation, a test run through all stations is conducted in an effort to standardise manual decision making. The homogenisation process is reiterated and each anomaly, neighbour and difference timeseries recalculated each time until the number of adjustments made is less than 25 % of that made in the first iteration.

As all station timeseries are scrutinised manually, the opportunity is taken to delete stations that look unsatisfactory. Although unavoidably subjective, this is a useful extra quality control. These removed stations are then scrutinised in more detail in regions of poor data coverage and added back to the dataset if deemed possible. Final deletions are clustered in the Chita Oblast (a federal subject of Russia, eastern Siberia), the east coast of China, Croatia and Romania (Figs. 3.24 a and b).

After the second iteration, the number of adjustments is less than 25 % compared to that in the first iteration, and hence no further iterations are deemed necessary. The results for *q* are summarised in Table 3.8. In theory, adjustments should be fairly evenly distributed both over time and around zero, but perhaps with peaks in specific regions at times of broad scale instrument change. A geographical analysis of adjustment timing and sign demonstrates this to be true, especially in regions with larger numbers of adjustments (Figs. 3.25, 3.26, - second iteration results are similar and therefore not shown). There is a peak in US first iteration adjustments in 1985. Here, the US National Weather Service introduced a new T_{dw} sensor in the mid- to late 1980s in addition to moving over to the Automated Surface Observing System from 1987 onwards (Gaffen & Ross, 1999). This link is not conclusive. Overall, there is no pattern or peak in the adjustments sufficient to give doubt to the value of the homogenisation process.

3.3.3 Homogenisation of Relative Humidity

Ideally, a full homogenisation would be undertaken on *RH* independently. However, as *RH* is considered of secondary importance relative to *q* in terms of this project output and analysis, a faster fully automated, semi-independent approach is used. Correlation regions, neighbour composites and adjustment quantities are derived from *RH* directly and the outlier checking is conducted on *RH* directly. Breakpoint locations are taken from those accepted in *q* for each iteration respectively.

This is preferable to direct conversions from *q* adjustments because deriving humidity away from point source values introduces small errors due to the non-linear relationship between *T* and humidity and between humidity variables themselves (McCarthy & Willett, 2006). The chosen method has two issues: non-discontinuities may be adjusted for; and discontinuities may go un-adjusted. In the first case, introduced errors should be minimal because if no discontinuity exists then the difference series medians either side of the breakpoint should be very similar thus effecting very little adjustment. The second case is more serious. Although, breakpoints in humidity alone (in T_{dw} – the common source variable for *q* and *RH*) should occur simultaneously in both *q* and *RH*, *RH* is also dependent on *T* and would be affected by *T* inhomogeneities. It is assumed for the purpose of this homogenisation that *T* and T_{dw} have identical breakpoints. This assumption is highly unlikely to hold in all cases.

There are slightly fewer stations with *RH* neighbour composites than for q due to typically smaller correlation decay distances (section 3.2.1 *issue 7*). However, results (summarised in Table 3.8) are similar to q in that adjustments are generally normally distributed around zero (Fig. 3.27). There are no outliers found in the *RH* stations. This implies that either *RH* has a lower variability relative to q or that first order variability is higher such that four standard deviations is a very large quantity and therefore never exceeded. As for q, in the case study stations at least (Fig. 3.28), adjustments effectively moderate or in many cases entirely remove apparent discontinuities in the record. This supports the assumption of generally identical T and T_{dw} breakpoint locations to a point. Interestingly, adjustments in *RH* tend to be more conspicuous (i.e. 1996 in station 837460 (Galeao, Brazil)). All breakpoint locations originate from identification in the q timeseries and so it could be concluded that discontinuities in humidity affect *RH* more than q.

3.4 AN ANALYSIS OF SURFACE HUMIDITY OVER LAND

3.4.1 The Climatology of Surface Humidity over Land

The land data are converted to monthly mean anomalies (Box 3.1) and gridded to 5 ^{o} by 5 ^{o} grids using simple (non-weighted) averaging over each grid-box. Many grid-boxes consist of two or more stations. Outside the Northern Hemisphere mid-latitudes however, coverage is sparse (Fig. 3.29).

Monthly climatology grids are created by averaging the station pentad climatologies over each month and then averaging all station monthly mean climatologies for each grid-box (Figs. 3.30 and 3.31 for q and *RH* respectively). Globally, q ranges from ~0 to 20 g kg^{-1} and has strong zonal continuity. The moistest q values are in the Tropics followed by the Summer Hemisphere. These decrease meridionally towards each Pole. The structure of *RH* is much more region specific (zonally discontinuous) with the lowest values in desert regions and highest values in the Northern Hemisphere in winter months and a global range of ~10 - 90 %.

3.4.2 Recent Changes in Surface Humidity over Land

For trend and timeseries analysis the data are regionally averaged at the monthly mean and seasonal mean anomaly resolution for four regions: the Globe (70 ^{o}S - 70 ^{o}N), the Northern Hemisphere (20 ^{o}N - 70 ^{o}N), the Southern Hemisphere (70 ^{o}S - 20 ^{o}S) and the Tropics (20 ^{o}S - 20 ^{o}N). This is done by weighting each monthly/seasonal grid-box value by the cosine of its latitude and then calculating the weighted mean for each month / season over the whole region. Trend fitting methods are described in Box 3.4.

The *q* timeseries shows peak positive anomalies in the Tropics for 1982/83, 1987/88 and 1997/98 in common with the dates of recent large El Niño events (WMO, 1999). The 1998 peak is also apparent in all regions (Fig. 3.32). These features are not so clear in the *RH* timeseries and negative in the Southern Hemisphere and Tropics (Fig. 3.33). For *q*, all trends are positive and highly significant (at 1 %), except for the Southern Hemisphere, with a Global trend of 0.11 g kg^{-1} $10yr^{-1}$. The strongest trend is found over the Tropics at 0.16 g kg^{-1} $10yr^{-1}$. This follows the Clausius-Clapeyron relation under the assumption of near-constant *RH* (section 1.2), where changes in absolute humidity are expected to be largest in regions of higher ambient *T* and where surface water sources are effectively unlimited. It is commonly assumed that *RH* stays largely constant over time and this is apparent in climate models (Allen & Ingram, 2002). The observed *RH* trends demonstrate this to some extent because the Global trend is very small (-0.03 % $10yr^{-1}$) and not significant. However, the Southern Hemisphere trend is considerable (-0.34 % $10yr^{-1}$), and significant at the 5 % level.

Seasonal trends are largely similar (Table 3.9). Of interest is the occurrence of the largest *q* trends in the Summer season for both the Northern (highly significant at 1 %) and Southern Hemisphere, when, for the Northern Hemisphere, trends are larger than in the Tropics. As discussed earlier, trends are expected to be larger in regions of higher ambient *T* if water is not limiting. Seasonal trends in *RH* differ widely in both magnitude and sign across the seasons. The largest *RH* trends for each region are all significant (at the 5 % level) implying that *RH* cannot always be assumed to remain constant over smaller temporal scales.

Trends are used to compare apparent changes in HadCRUH over time, between regions and with other datasets. This construct of a linear change is hypothetical and not an ideal model of the climate system. Therefore trend 'significance' is relative to timeseries variability rather than a physical truth and has no implications for cause and effect. Significance at 1 % and 5 % levels are referred to as highly significant and significant in the text respectively.

Trends for large-scale regions are fitted using a Restricted Maximum Likelihood estimation (henceforth referred to as REML) (Diggle *et al.*, 1996), method in common use at the Hadley Centre (Met Office, UK) (Parker, D *pers. comm.*) and employed in IPCC Chapter 2 (Folland *et al*., 2001b). The data are first pre-whitened assuming good representation by an AR1 noise process. There is no dependence on the number of elements within the timeseries, nor heavy influence from end points. The user supplies estimated ranges for the autocorrelation and variance of the residuals. These ranges are chosen by guess work and experience from previous tests. If the output autocorrelation and variance of the residuals are close to either end of the estimated ranges input by the user then the program is re-run with new ranges until output values lies near the mid-point of the estimated ranges. A trend is estimated and likelihood calculated for 121 different sets of values within the given ranges to find the most likely trend, and significance levels calculated.

The REML method employed here is computationally demanding and seasonal and regional (Chapter 5) trends are calculated and significance tested using a faster but slightly less robust least squares regression (henceforth referred to as LSR) method. The data are also first pre-whitened assuming good representation by an AR1 noise process thereby allowing for autocorrelation (Cochrane & Orcutt, 1949; Wei, 1990). Estimates for the autocorrelation and variance of the residuals are created within the program. When cross-compared with trends fitted by the REML method, the LSR performs identically for most timeseries but tends to give more weight to the end points of the timeseries causing problems with some trends, especially in the Tropics.

The third method uses the median of pairwise slopes (henceforth referred to as MPS) (Lanzante, 1996) to fit a trend but without significance. The slopes of all possible pairs of points within the timeseries are calculated with the median value as the trend estimate. This is employed for looking at grid-box trends as it is quick to apply and can cope with missing data (set to a maximum of 50 %) better than the first two methods. It is also used later in Chapter 5 for analyses of model output with HadCRUH where trend comparison between the two is felt to be more important than the significance of the trend itself.

Box 3.4: Three Methods of Trend Fitting

At the grid-box scale, trends in *q* are mostly positive (87 % of grid-boxes), with negative clusters in the western US, western South Africa and southern Australia. Trends range from -0.33 to 0.65 g kg^{-1} $10yr^{-1}$, and for seasonal averages, from -0.70 (in MAM) to 1.14 (in DJF) g kg^{-1} $10yr^{-1}$ (Fig. 3.34). Trends in *RH* are more evenly mixed in sign with the majority negative (52 %) and range from -2.46 to 2.34 % $10yr^{-1}$, and for seasonal averages, from -5.82 (in MAM) to 5.32 (in SON) % $10yr^{-1}$ (Fig. 3.35). There is no general consistent spatial pattern between trends in *q* and in *RH*. Positive trends in *q* are widespread and spatially coherent whereas *RH* trends have less spatial coherence.

Seasonal grid-box trends in *q*, at least for the Northern Hemisphere, are clearly strongest and most consistent in the summer months. In all seasons, trends are positive in the Tropics. Although mostly consistent, a few regions exhibit trends of opposite sign across the seasons such as northern North America, western Russia, parts of Australia, Madagascar and Saudi Arabia. For *RH*, seasonal changes of trend sign are of large spatial scale and magnitude, notably over Europe, Saudi Arabia, Asia, parts of the US, and west Africa.

3.5 CONCLUSIONS

The HadCRUH land component is transformed from raw sub-hourly data to a quality controlled, homogenised monthly mean anomaly 5 ^{o} by 5 ^{o} gridded product. This process is described in Fig. 3.36. The quality control procedure has been designed using a sub-set of case study stations from the ISH data set, specifically for humidity data. Key findings are summarised below:

- Climatologically, *q* varies from ~0 to 20 g kg^{-1} decreasing meridionally from the Tropics with strong zonal continuity. The peak in the Tropics shifts polewards towards the Summer Hemisphere. *RH* ranges from ~10 to 90 % with little zonal continuity or meridional pattern relative to *q* but a strong regional structure.
- In absolute terms, atmospheric surface moisture (*q*) has increased (highly significant at the 1 % level relative to dataset variability) since 1973 when averaged over: the Globe (0.11 g.kg^{-1} $10yr^{-1}$); the Northern Hemisphere (0.12 g.kg^{-1} $10yr^{-1}$); and the Tropics (0.16 g.kg^{-1} $10yr^{-1}$).

- Trends in q are largest in the Tropics and Northern Hemisphere Summer, where ambient T is high and water is more readily available relative to other regions/seasons. Thus, over large scales at least, this is in accordance with the Clausius-Clapeyron relation under the assumption of near-constant *RH*.
- In relative terms (*RH*), trends are not significant except when analysed for the Southern Hemisphere or seasonally (Global MAM, Northern Hemisphere DJF, Tropics and Southern Hemisphere JJA). Hence, the assumption of constant *RH* at least on large spatial and temporal scales holds true.

For version one, spatial and temporal coverage provided by ISH and the data building process, are sufficient to provide a near-global surface humidity land dataset from 1973 to 2003 suitable for climate analyses. There is scope for improvements at all levels in future versions of HadCRUH.

3.6 TABLES AND FIGURES FOR CHAPTER 3

Table 3.1 page 1
(3Table3_1.pdf)

Table 3.1 page 2
(3Table3_1.pdf)

Test	Polar North	Polar South	Extra-Tropical North	Extra-Tropical South	Tropical North	Tropical South
$T_{dw} > T$	0.002	0	0.005	0.0004	0.003	0.001
T_{dw} > 60° C	0	0	0.0001	0	0	0
T_{dw} < -80° C	0	0.003	0.003	0	0	0
T > 60° C	0	0	0.00003	0	0	0
T < -80° C	0	0	0	0	0	0
RH < 0%	0	0	0	0	0	0
e < 0 hPa	0	0	0	0	0	0
Total failing	0.002	0.003	0.005	0.0004	0.003	0.001

Table 3.2: Results of the Bad Values QC for the case study stations. Results show the percentage of humidity data failing each test for all thirty three case study stations by latitude band where Polar covers from 60 ° to 90 °, Extra-tropical covers from 20 ° to 60 ° and Tropical covers from 0 ° to 20 ° in each hemisphere.

WMO ID	Station Name	Latitude	Longitude	Elevation (m)	Analysis
042310	Kangerlussuaq	67.017	-50.700	53	cut offs at -37 °C in 1965-74 for spring and winter T
042700	Narsaruaq	61.133	-45.417	5	cut offs at -37 °C in 1965-74 for spring and winter T
702720	Anchorage	61.250	-149.800	59	rarely goes below -30 °C
703210	Dillingham	59.050	-158.517	29	T_{dw} cut off at -37 °C in 1995-2003 winter

Table 3.3: Greenland and Alaskan stations analysed for evidence of 'cutoffs' at low temperatures. Greenland stations begin with 04 and Alaskan stations begin with 70.

Table 3.2 page 1
(3Table3_2.pdf)

Table 3.2 page 2
(3Table3_2.pdf)

WMO ID	Year	Test 1	Test 2	Test 3	Test 4	σ Test
284400	1988	X	X	---	X	---
284400	1990	X	X	---	X	---
284400	1994	X	X	---	---	---
637080	1983	X	X	---	---	---
683680	1974	X	X	---	---	---
683680	1975	X	X	---	---	---
683680	1977	X	X	---	X	---
683680	1982	X	X	---	X	---
683680	1993	X	X	---	---	---
702220	2000	---	X	X	X	---
702220	2001	---	X	X	X	---
726650	1979	---	X	X	X	---
726650	1997	X	X	---	---	---
879250	1986	X	X	---	X	---
879250	1988	X	X	---	---	---
895710	1979	X	---	X	X	X
916800	1983	X	X	---	---	---
916800	1990	X	X	---	---	---
916800	1991	X	X	---	X	---
916800	1992	X	X	---	X	---
943260	1977	X	X	---	---	---
943260	1980	---	---	X	X	---
943260	1981	X	---	---	X	---

Table 3.5: Timezone QC test criteria and results. All tests check for similarity of the annual mean diurnal cycle to the station mean diurnal cycle using annual mean *r* value vectors described in the text. All five tests are described in section 3.2.1 *issue 6*. X denotes a fail. Grey shading separates station year groups.

Ventilation Rate (ms^{-1})	Wind Speed (ms^{-1})	Dewpoint Temperature (°C)		Vapour Pressure (hPa)		Specific Humidity ($g\ kg^{-1}$)		Relative Humidity (%)	
		MIG	R	MIG	R	MIG	R	MIG	R
0.0	0.5	1.0	0.5	0.66	0.33	0.59	0.21	4.8	2.0
1.0	7.2	0.2	-0.4	0.13	-0.25	0.08	-0.15	1.0	-1.0
2.0	13.9	0.1	-0.5	0.07	-0.31	0.05	-0.19	0.5	-1.8

Table 3.6: Approximate errors in measuring humidity due to differences in wind speed. Measurements are taken from two common wet-bulb thermometers, the Mercury-in-glass (MIG) and Resistance (R) thermometers, in comparison to assumed high accuracy measurement from a Dewpoint hygrometer at three ventilation rates. Results are obtained from Folland (1977). Measurements are observed at 75 % *RH* where $T + T_w = 18$ °C such that $T_{dw} = 5.8$ °C, $e = 9.26$ hPa and $q = 5.70$ g kg^{-1}. Errors increase with decreasing *RH*.

Station 1	Elevation (m)	Station 2	Elevation (m)	Vertical Distance	Horizontal Distance
802590 Cali, Colombia	969	802220 Bogota, Colombia	2548	1579	~400km
726556 Redwood, US	312	726650 Gillette, US	1230	918	~800km
064510 Brussels, Belgium	58	108660 Munich, Germany	529	471	~600km
476710 Tokyo, Japan	8	476900 Nikko, Japan	1294	1286	~125km

Table 3.7: Station pairs for the Elevation QC.

	q	*RH*
No. of Stations before Homogenisation	3514	3514
No. of Stations with ≥ 5 Neighbours	3391	3268
No. of Adjustments in 1st Iteration	4864	4680
No. of Adjustments in 2nd Iteration	684	651
No. of Outliers in 1st Iteration	4737	0
No. of Outliers in 2nd Iteration	4081	0
Stations Deleted Due to Poor Data Record Quality	180	---
Deleted Stations Revisited to Retain Data Coverage	58	---
Deletions Kept after Revisit	32	---
No. of Stations after Homogenisation	3243	3128

Table 3.8: Results of homogenisation and second outlier check on the land data.

REGION	***q* (g kg^{-1} 10yr^{-1})**			
	DJF	**MAM**	**JJA**	**SON**
GLOBAL (70°N-70°S)	0.11**	0.09**	**0.15****	0.11**
NORTHERN HEMISPHERE (20°N-70°N)	0.09**	0.08**	**0.20****	0.11**
TROPICS (20°S-20°N)	**0.19****	0.15**	0.13**	0.18**
SOUTHERN HEMISPHERE (70°S-20°S)	**0.02**	-0.00	-0.02	0.01
	***RH* (% 10yr^{-1})**			
GLOBAL	0.09	**-0.19***	-0.08	-0.03
NORTHERN HEMISPHERE	**0.25***	-0.18	0.09	0.12
TROPICS	-0.12	-0.16	**-0.22***	-0.10
SOUTHERN HEMISPHERE	-0.24	-0.31	**-0.59***	-0.53

Table 3.9: Regionally averaged trends by season for *q* and *RH*. Values of largest magnitude for each region and each variable are in bold. Trends are created using the LSR method (Box 3.4). Significance at 5 % is shown with a * and at the 1 % with **.

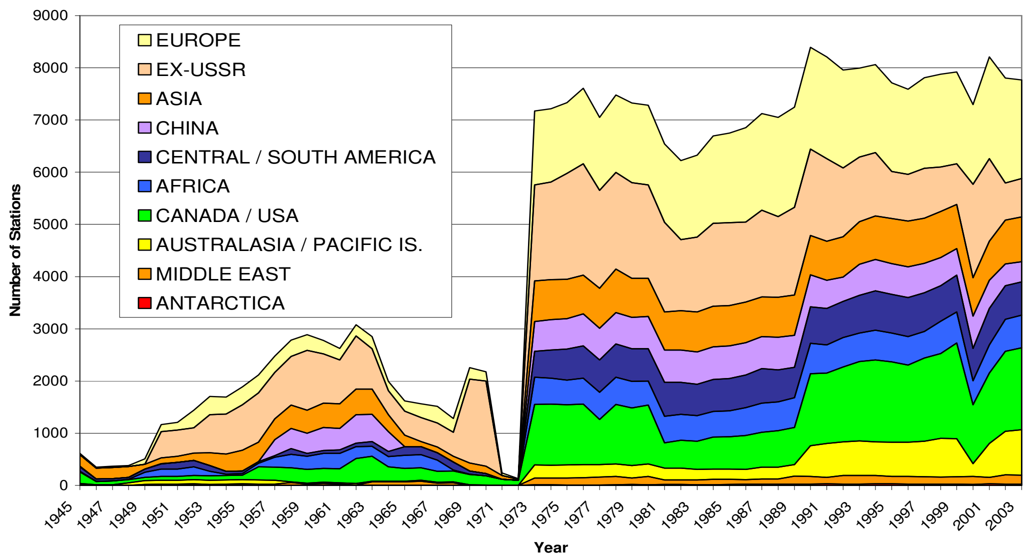


Figure 3.1: Spatial and temporal coverage of all stations reporting humidity in the ISH dataset. Regions are listed in order of appearance in the plot from top to bottom.

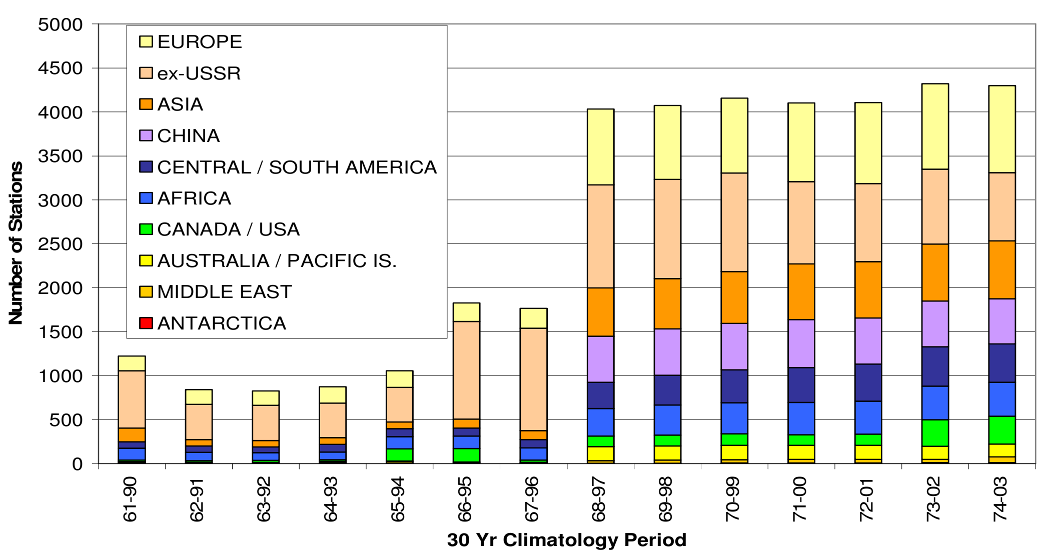


Figure 3.2: ISH stations reporting humidity with sufficient longevity to create 30-year climatologies by climatology period. Regions are listed in order of appearance in the plot from top to bottom.

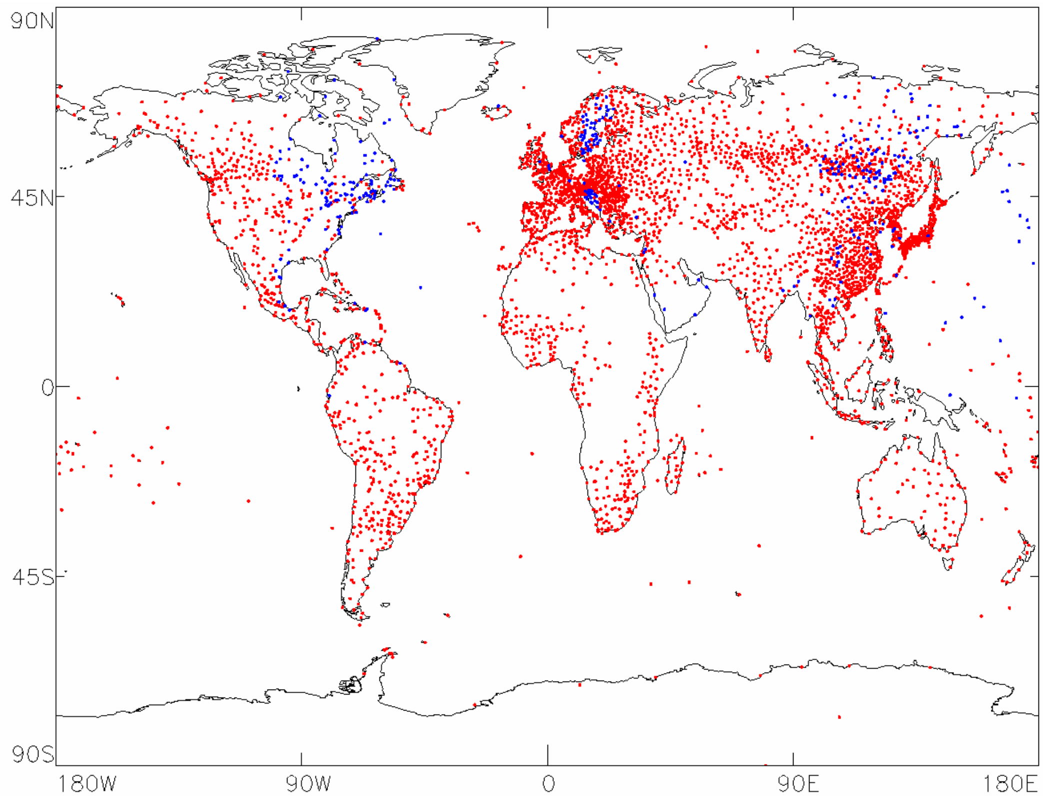


Figure 3.3: Spatial coverage of ISH stations with sufficient humidity data to create a 1974 to 2003 climatology. Stations shown in blue are those combined from identical stations reporting for shorter periods under different IDs (346 in total).

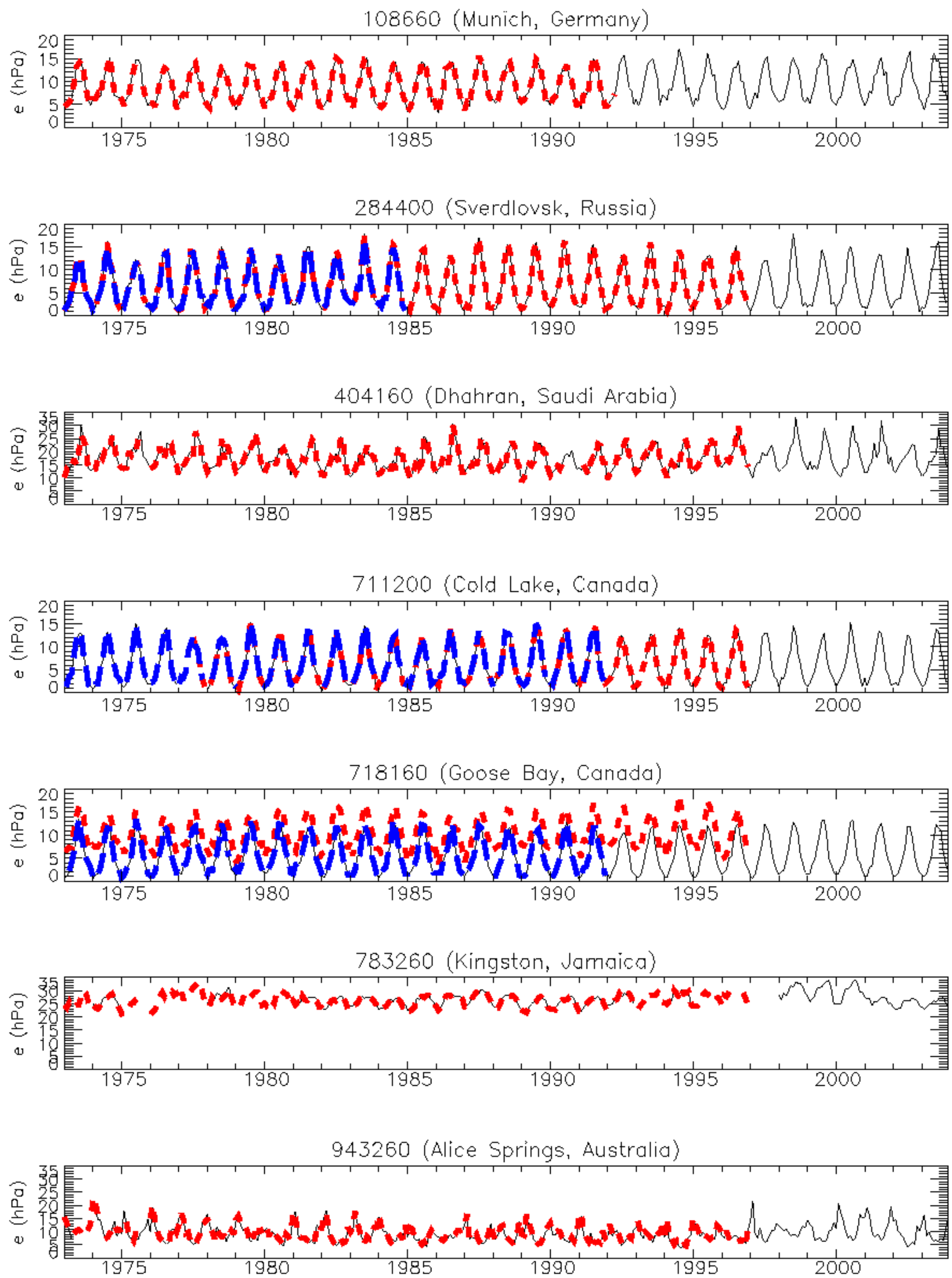


Figure 3.4: ISH monthly mean *e* compared to other sources for seven case study stations. Solid black lines represent the raw ISH data, short dashed red lines represent the HAHN data and long dashed blue lines represent the NEW data. Each station is identified by its six digit WMO number, the corresponding details of which can be found in Table 3.1.

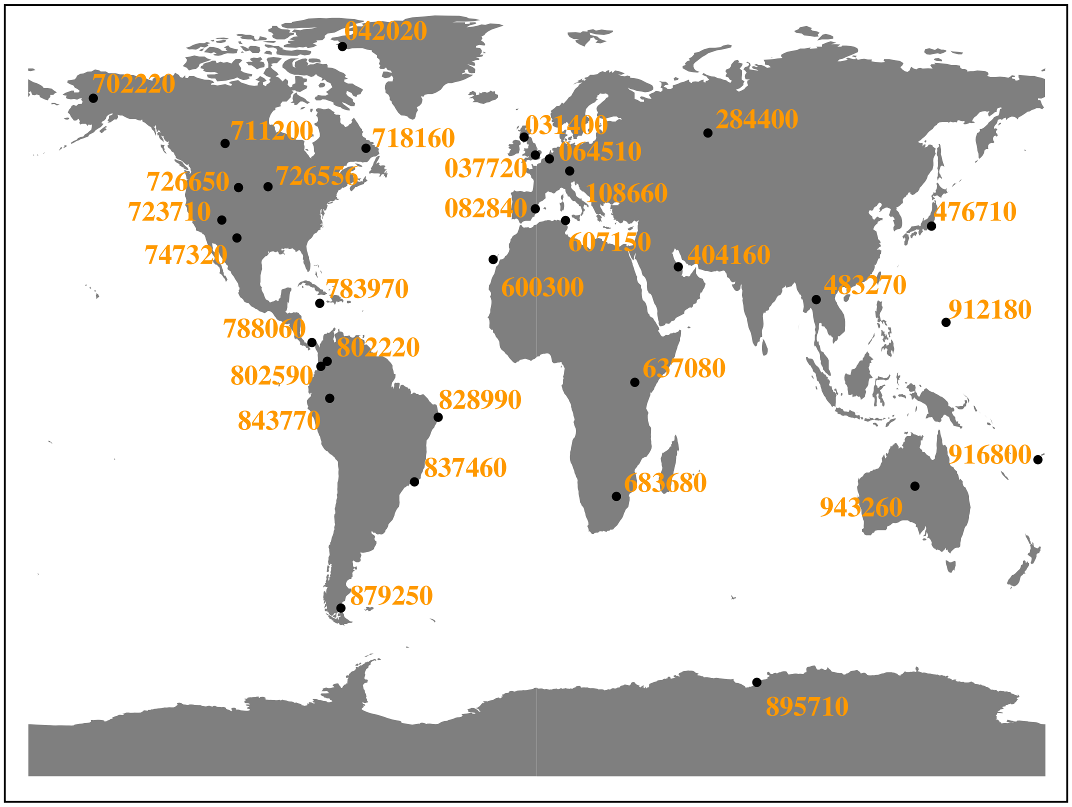


Figure 3.5: Station locations and WMO IDs for the thirty three case study stations.

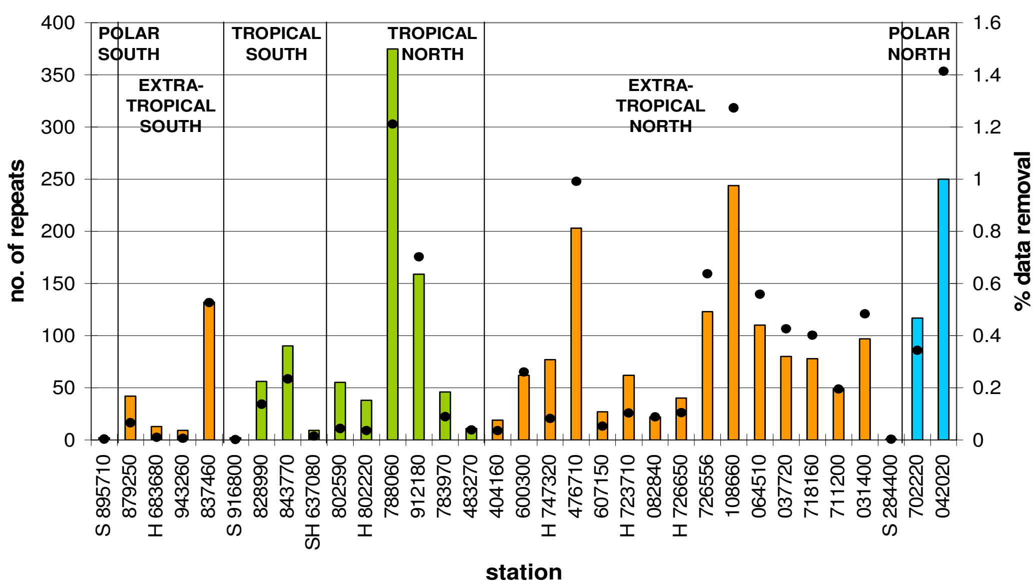


Figure 3.6: Occurrences of >12 hours continuous repeated values of *T* for all case study stations from 1973 to 2003. Stations are grouped latitudinally where green represents the Tropics (0 ° – 20 °), orange represents the Extra-Tropics (20 ° – 60 °) and blue represents the Poles (60 ° – 90 °). Black spots mark the percentage of data removed from each station. Stations over 1000m elevation are prefixed with H and stations reporting only 6 hourly are prefixed with S.

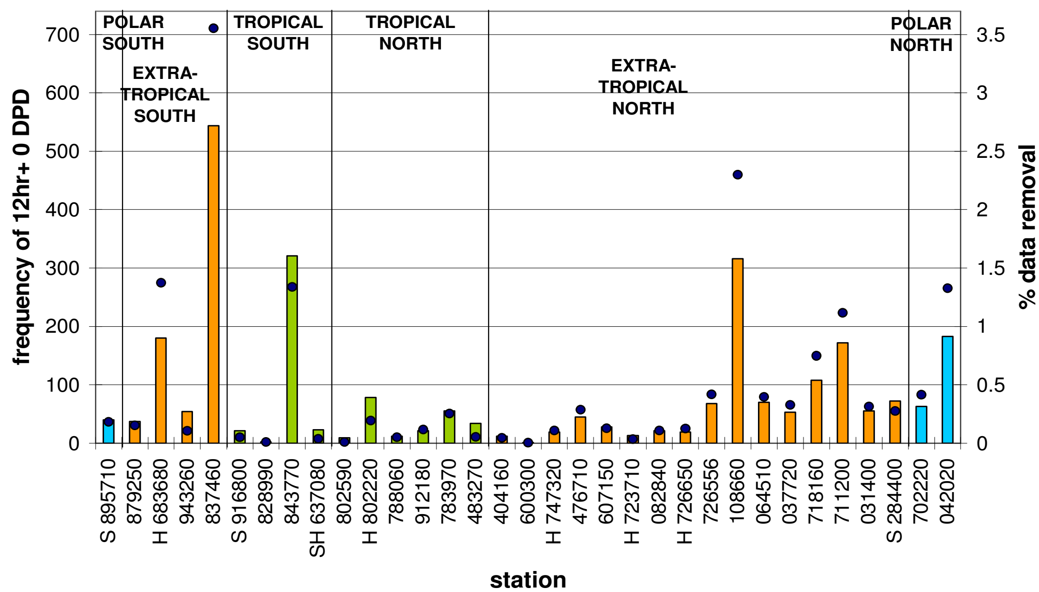


Figure 3.7: Occurrences of >12 hours continuous zero *DPD* for all case study stations from 1973 to 2003. Stations are grouped latitudinally as described in Fig. 3.6. Black spots mark percentage of data removed. Stations over 1000m elevation are suffixed with H and stations reporting only 6 hourly are suffixed with S.

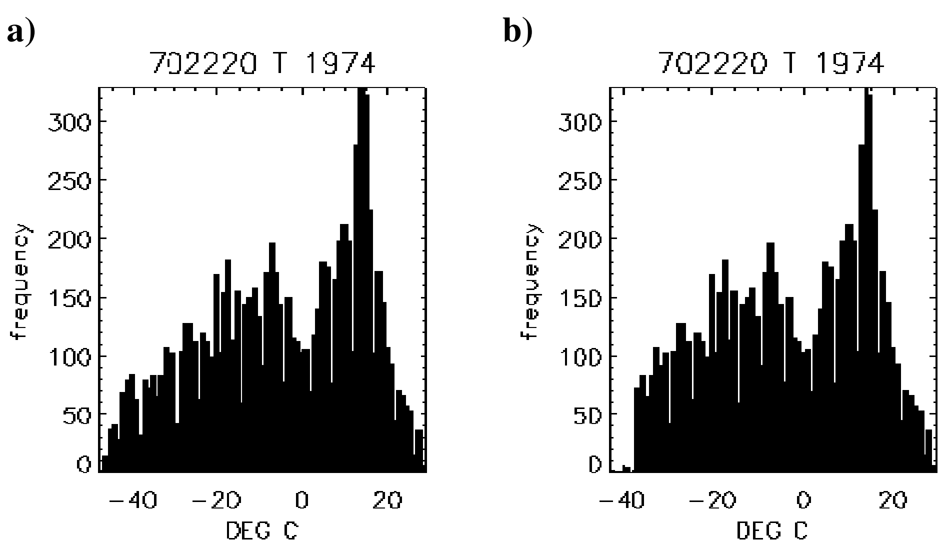


Figure 3.8: Frequency distribution of T values for case study station 702220 (Galena, Alaska, US) in 1974. Plot a) shows all T values and plot b) shows only T values that have simultaneously recorded T_{dw} values.

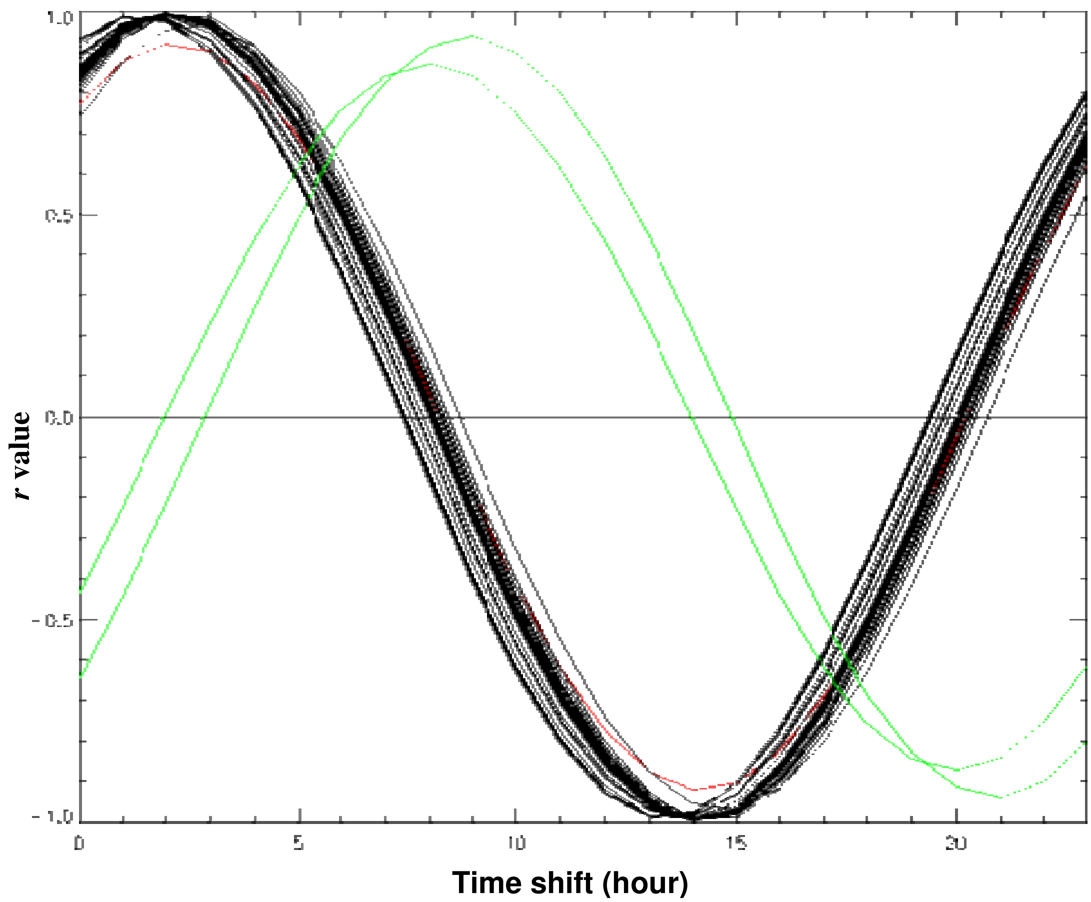


Figure 3.9: Example of the Timezones QC test for station (702220, Galena, Alaska, USA). Each black line represents an annual mean diurnal cycle as described in the text (section 3.2.1 *issue 5*) where the time shift of the maximum *r* value corresponds to the time of daily maximum when converted to GMT. The red line is the station mean diurnal cycle. The green lines show years that have failed one or more of the five tests (see section 3.2.1) that detect differences from each annual mean diurnal cycle to the station mean diurnal cycle (in this case 2000 and 2001, see Table 3.5).

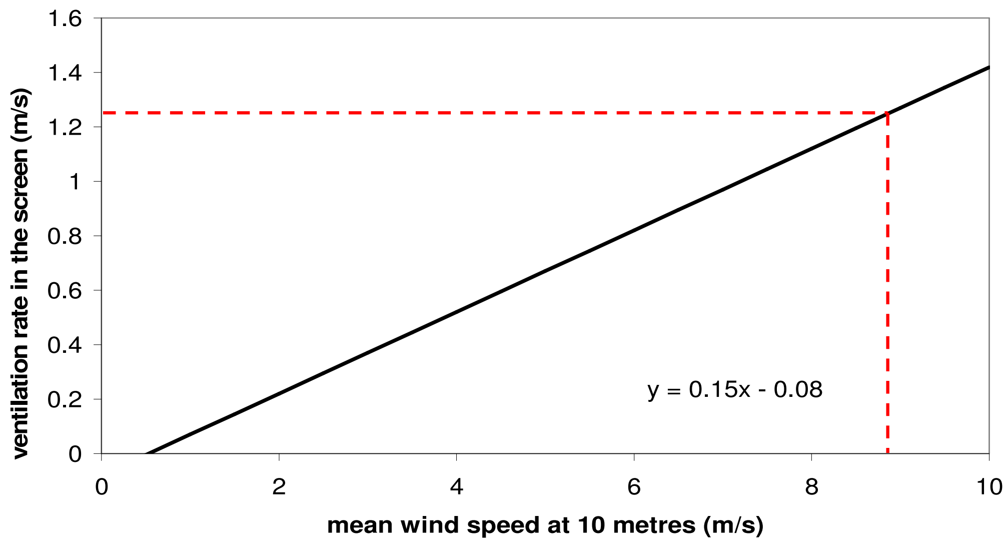


Figure 3.10: Relationship between ventilation rate of a Stevenson Screen and mean wind speed (Folland, 1977). The red dashed line shows the commonly used ventilation rate of 1.25 ms^{-1} and corresponding wind speed.

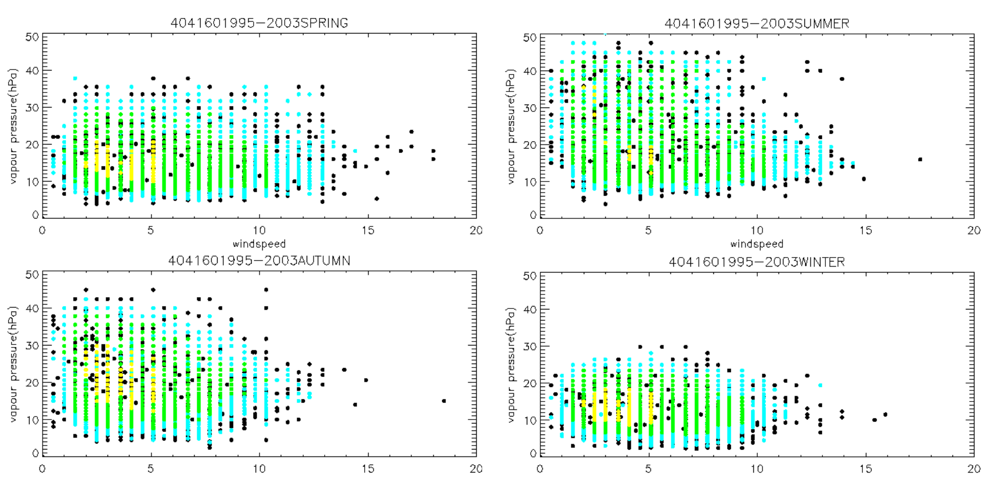


Figure 3.11: Seasonal relationship between wind speed and observed *e* for the period 1995 to 2003 at case study station 404160 (Dhahran, Saudi Arabia). The colour of dots signifies the number of observations where black, blue, green and yellow represent 1, 2-10, 11-100 and >100 respectively. Wind speed is in ms^{-1}.

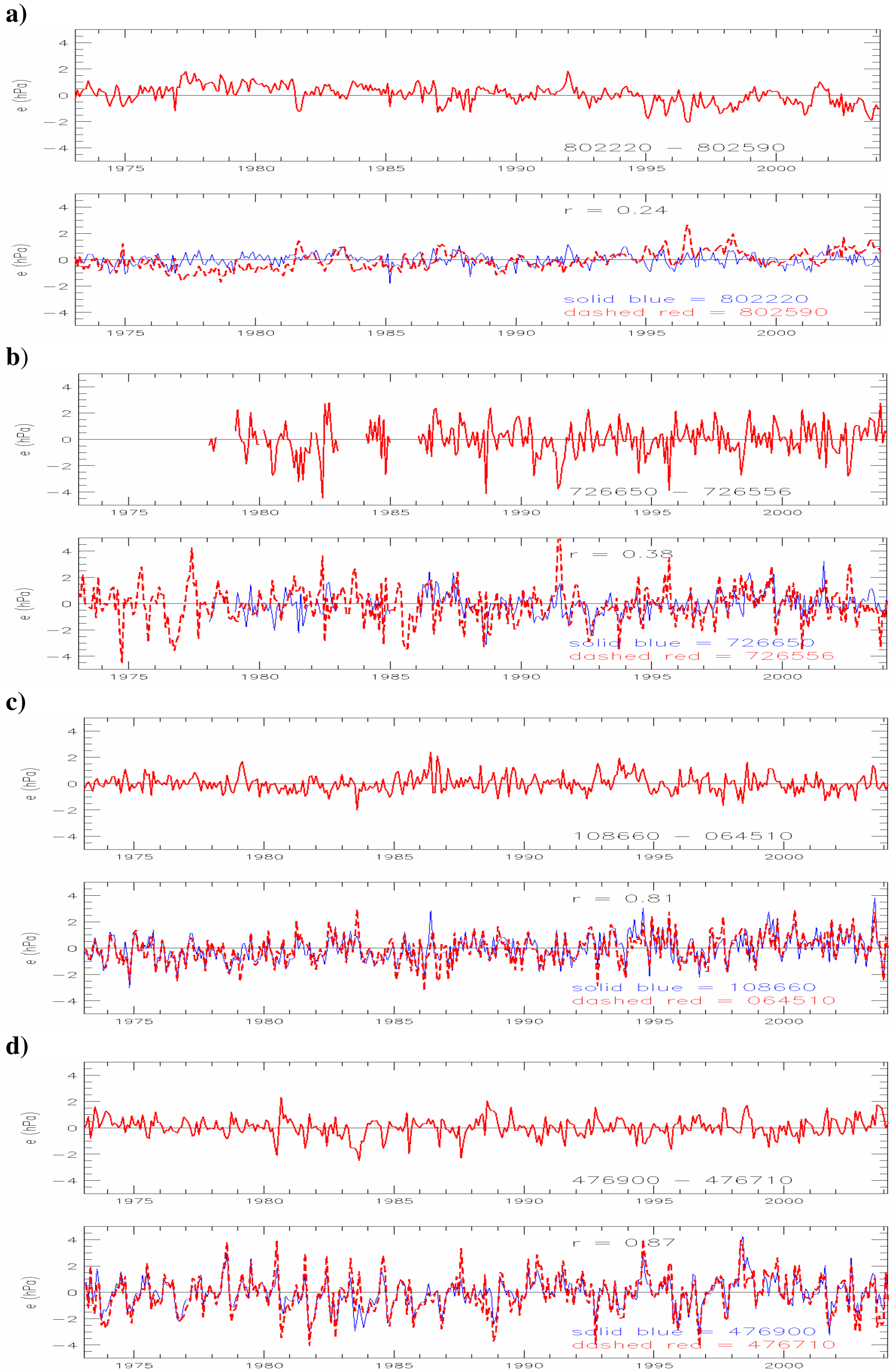


Figure 3.12: Comparisons of monthly mean anomaly timeseries between neighbouring high and low elevation case study stations. Plots a) to d) show Columbia, the USA, Europe and Japan respectively. Upper plots in each pair show high minus low elevation difference timeseries. Lower plots show both high and low elevation timeseries with the linear Pearson correlation coefficient *r* value.

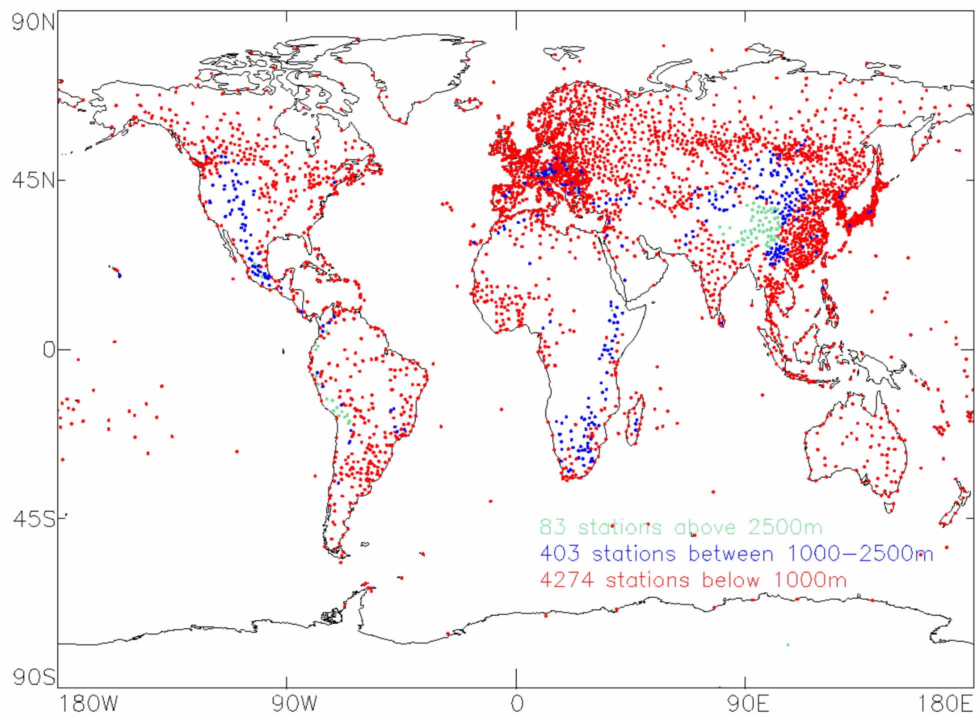


Figure 3.13: Spatial coverage of all stations with sufficient data to create a 1974 to 2003 climatology coloured by elevation. Red dots represent low elevation stations (>1000 m), blue dots represent medium elevation stations (1000-2500 m) and aqua dots represent high elevation stations (>2500 m).

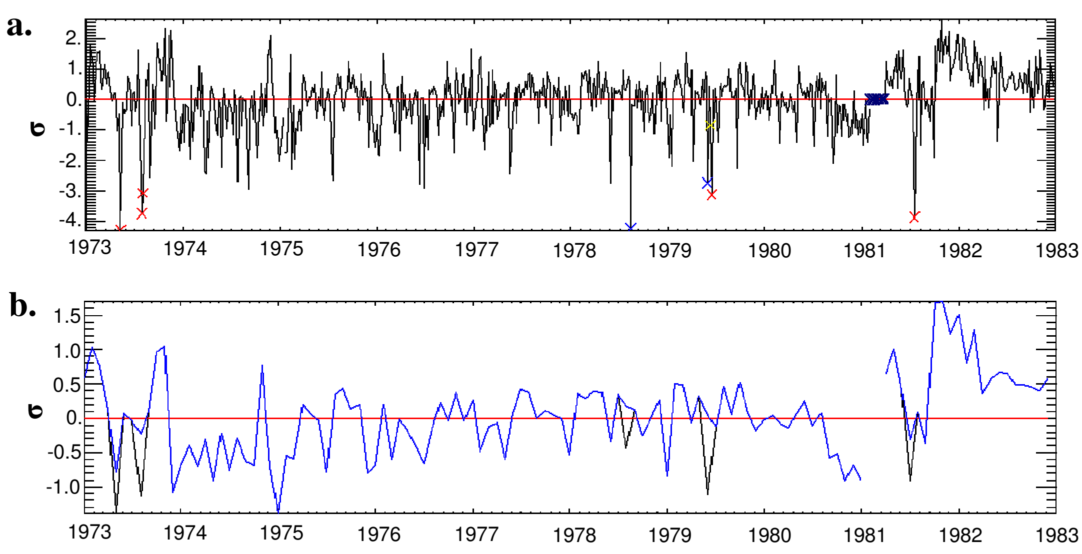


Figure 3.14: Identification and effect of outlier removal on the monthly mean anomaly timeseries using case study station 843770 (Iquitos, Peru) as an example. Panel a. shows the pentad mean anomaly *e* timeseries and identified outliers. Crosses mark pentads failing tests as described in section 3.2.1 where red crosses are test 1 (outlying individuals and groups), blue crosses are test 2 (sudden spikes), and yellow crosses are test 4 (sandwich outliers). There are no examples of test 3 in this timeseries section. Black crosses show gaps in the data. Panel b. shows the monthly mean anomaly timeseries before (black) and after (blue) the Outlier QC.

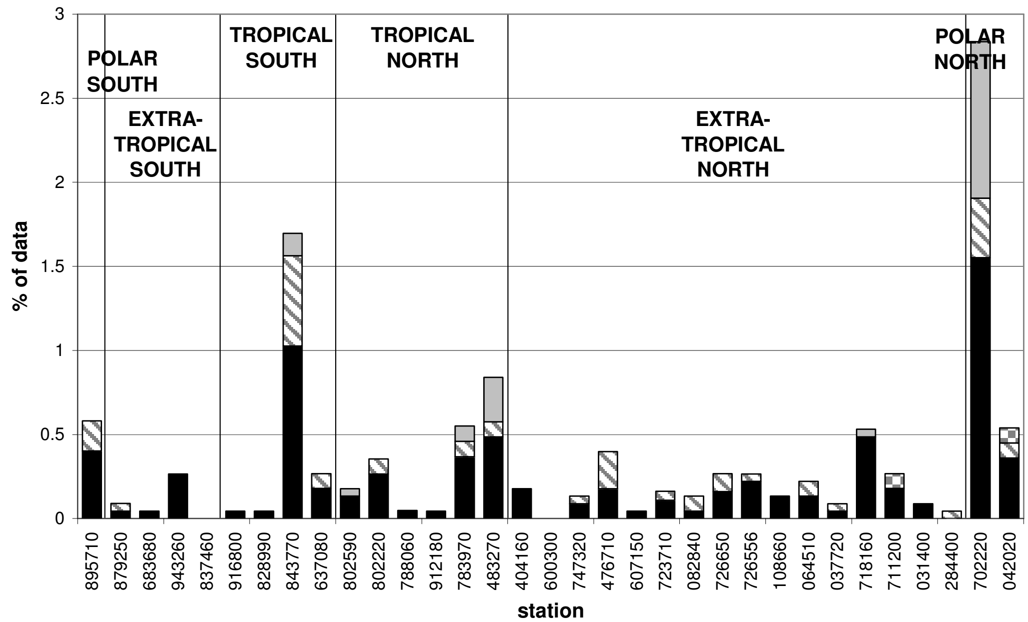


Figure 3.15: Percentage of data removed by the Outlier QC for all case study stations. Stations are grouped latitudinally as in Fig. 3.6. The proportion of data removal due to each of the four tests (section 3.2.1 and Fig. 3.14) is also shown where black is test 1, grey (slash) is test 2, grey (chequered) is test 3 and grey (solid) is test 4.

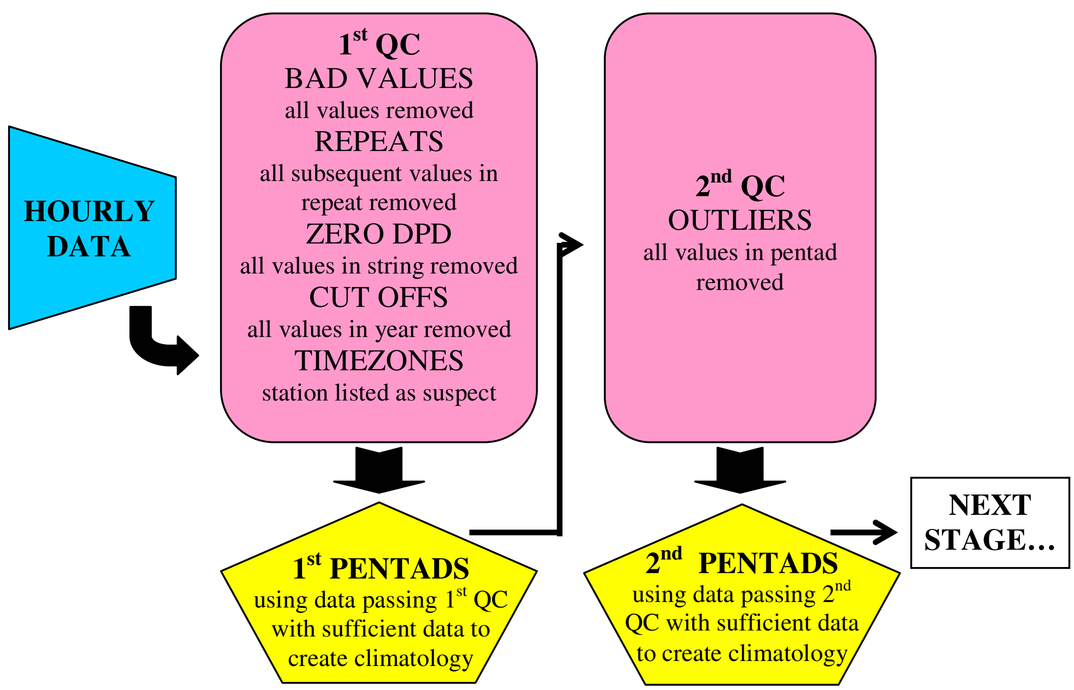


Figure 3.16: Flow diagram of the land quality control process.

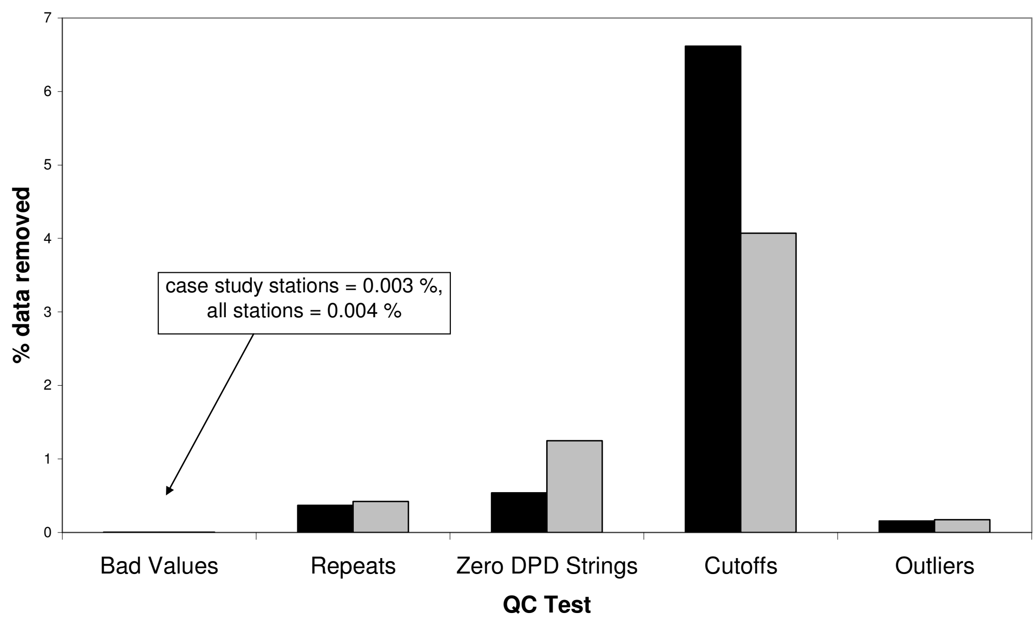
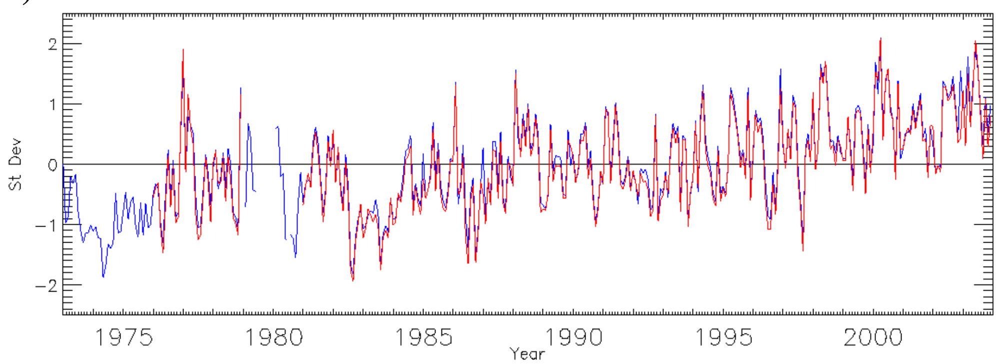
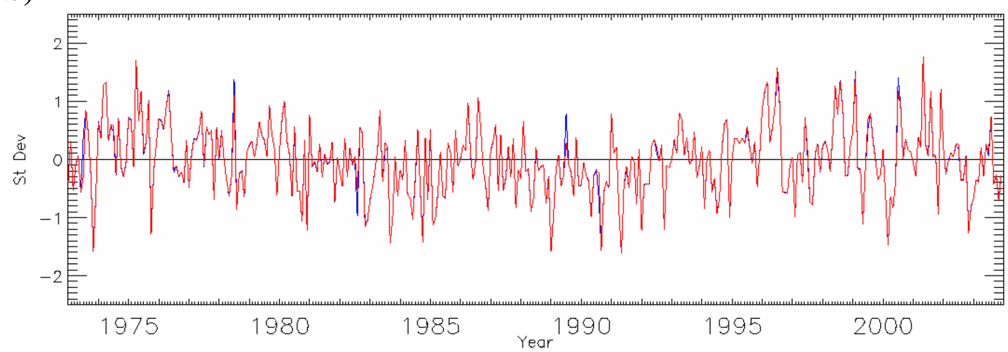


Figure 3.17: Percentage of data removed by the quality control process for all case study stations and all ISH stations. Black bars show percentages for the case study stations and grey bars show percentages for all ISH stations.

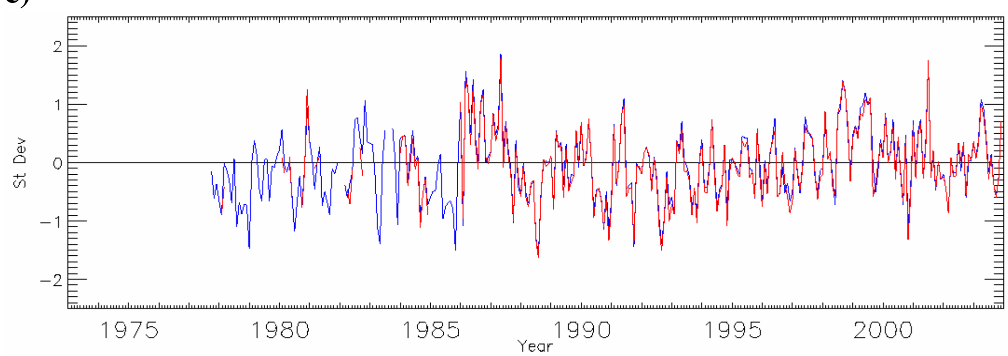
a)



b)



c)



d)

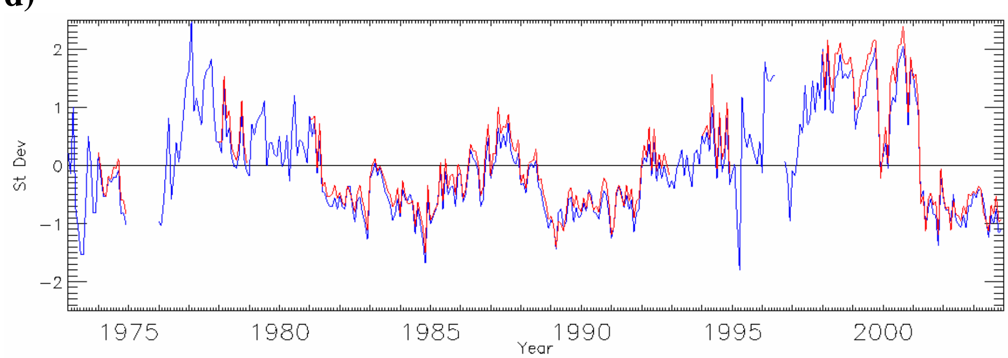


Figure 3.18: Monthly mean anomaly *e* timeseries before and after the quality control for four case study stations. Panels a) to d) show stations 042020 (Thule, Greenland – Polar), 404160 (Dhahran, Saudi Arabia – Extra-tropical), 726650 (Gillette, USA – Extra-tropical) and 783970 (Kingston, Jamaica – Tropical) respectively. Blue and red lines show the raw and quality controlled data respectively.

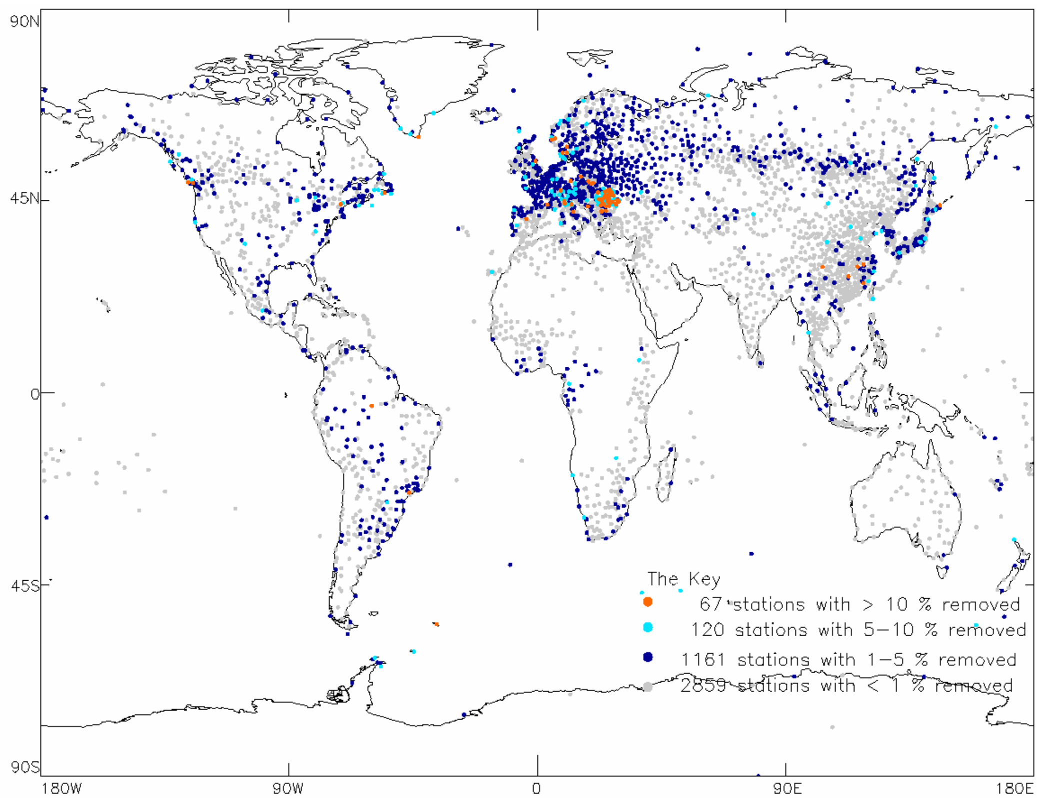


Figure 3.19: Geographical distribution of data removal due to the Zero *DPD* QC test by percentage per station.

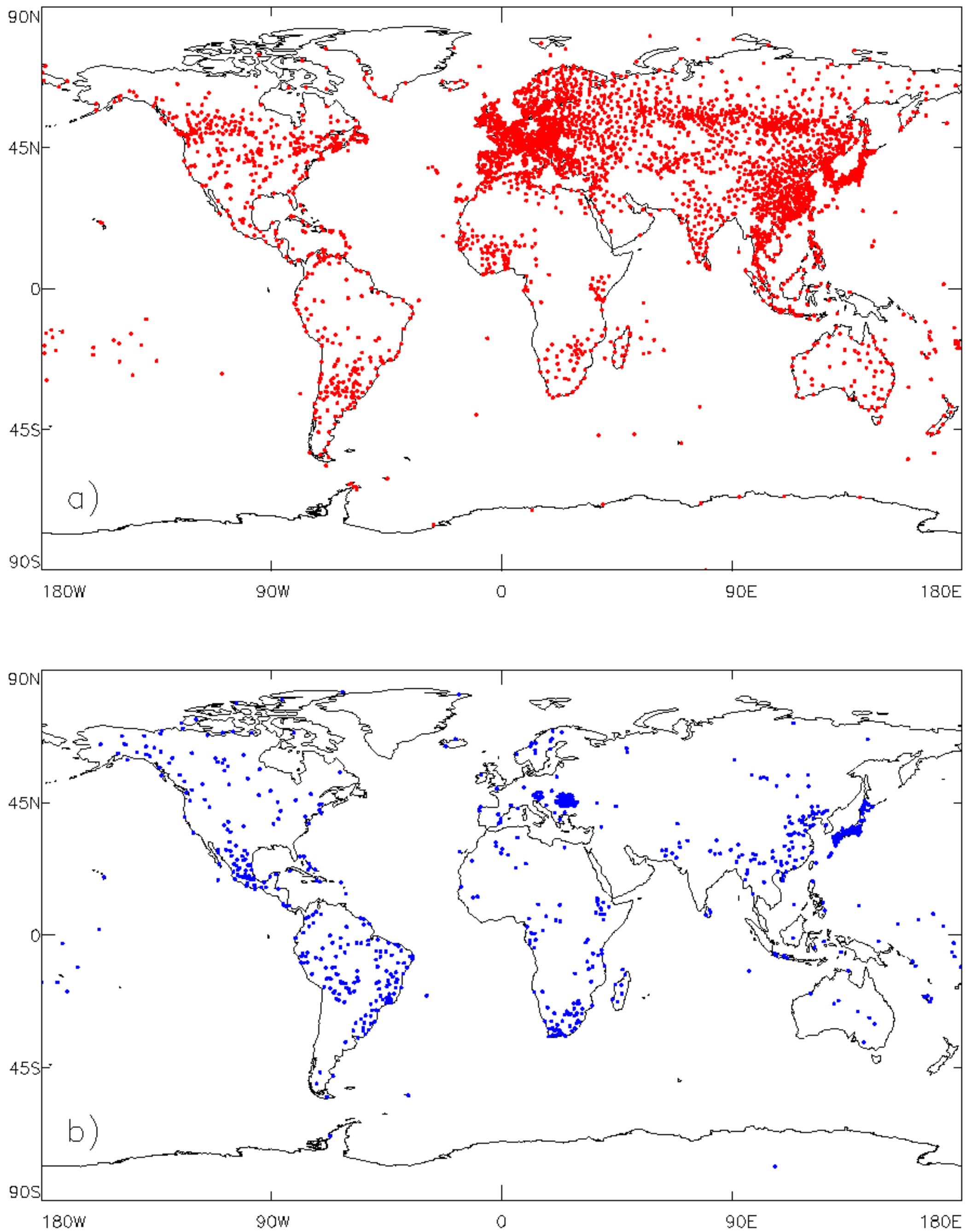


Figure 3.20: Spatial coverage of the land data after quality control. Panel a) shows all stations passing quality control and remaining in pentad dataset (3514). Panel b) shows stations that after the quality control have insufficient data to create a 1974 to 2003 climatology (1246).

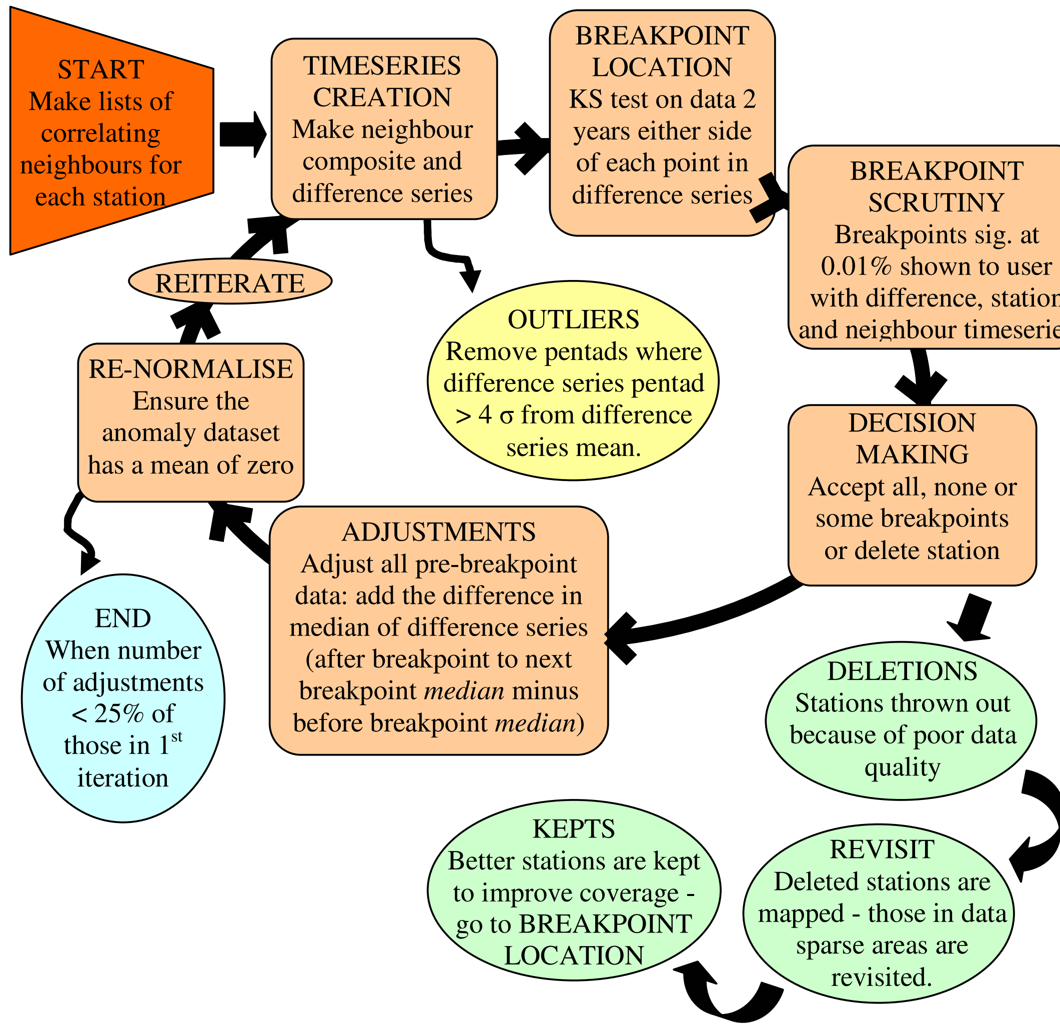


Figure 3.21: Flow diagram of the homogenisation process for the land data.

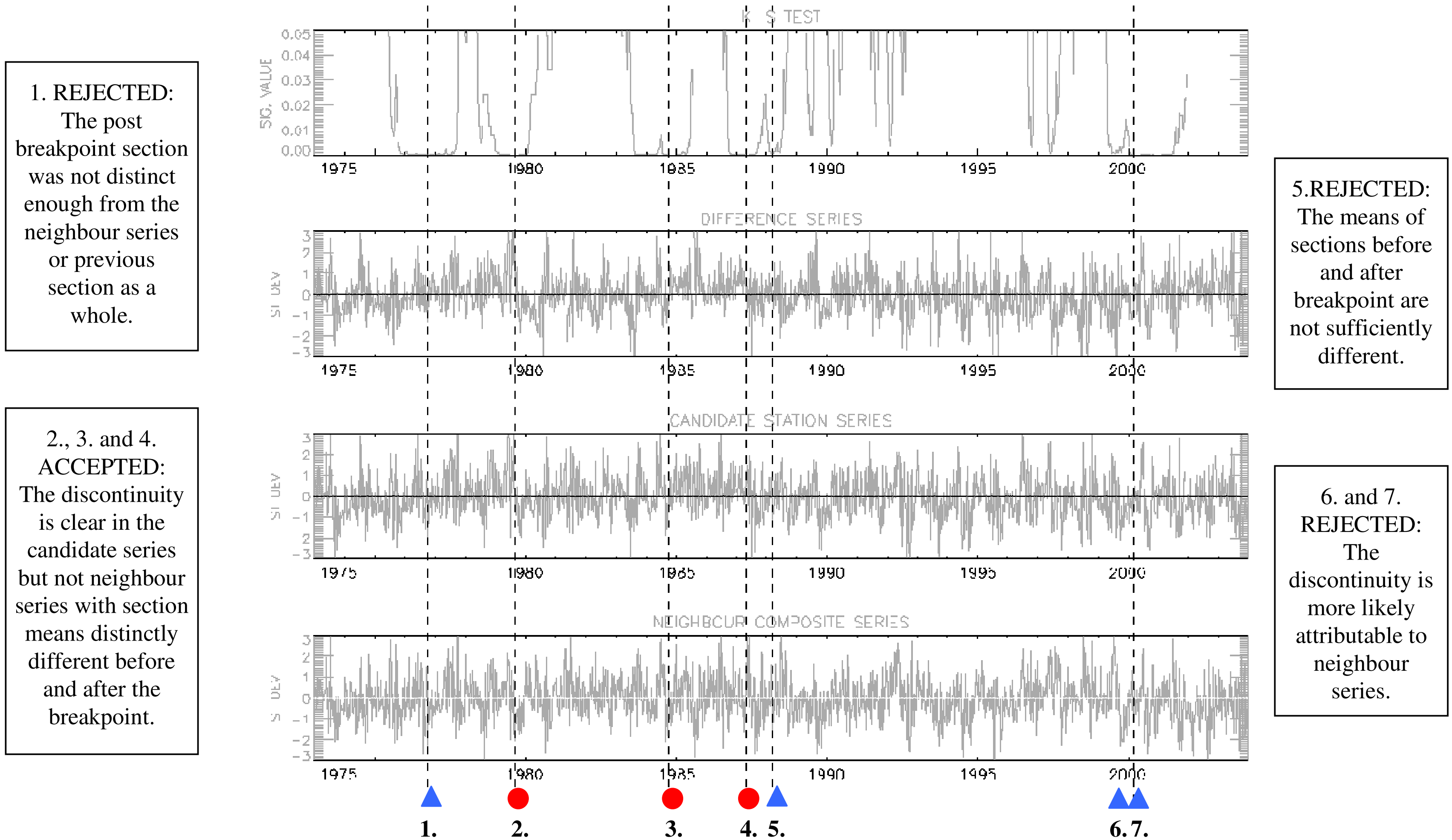


Figure 3.22: Example of the homogenisation process panel for manual scrutiny for case study station 747320 (Holloman, New Mexico, USA). Dashed lines show potential breakpoints. Those with red circles are accepted and those with blue triangles are rejected.

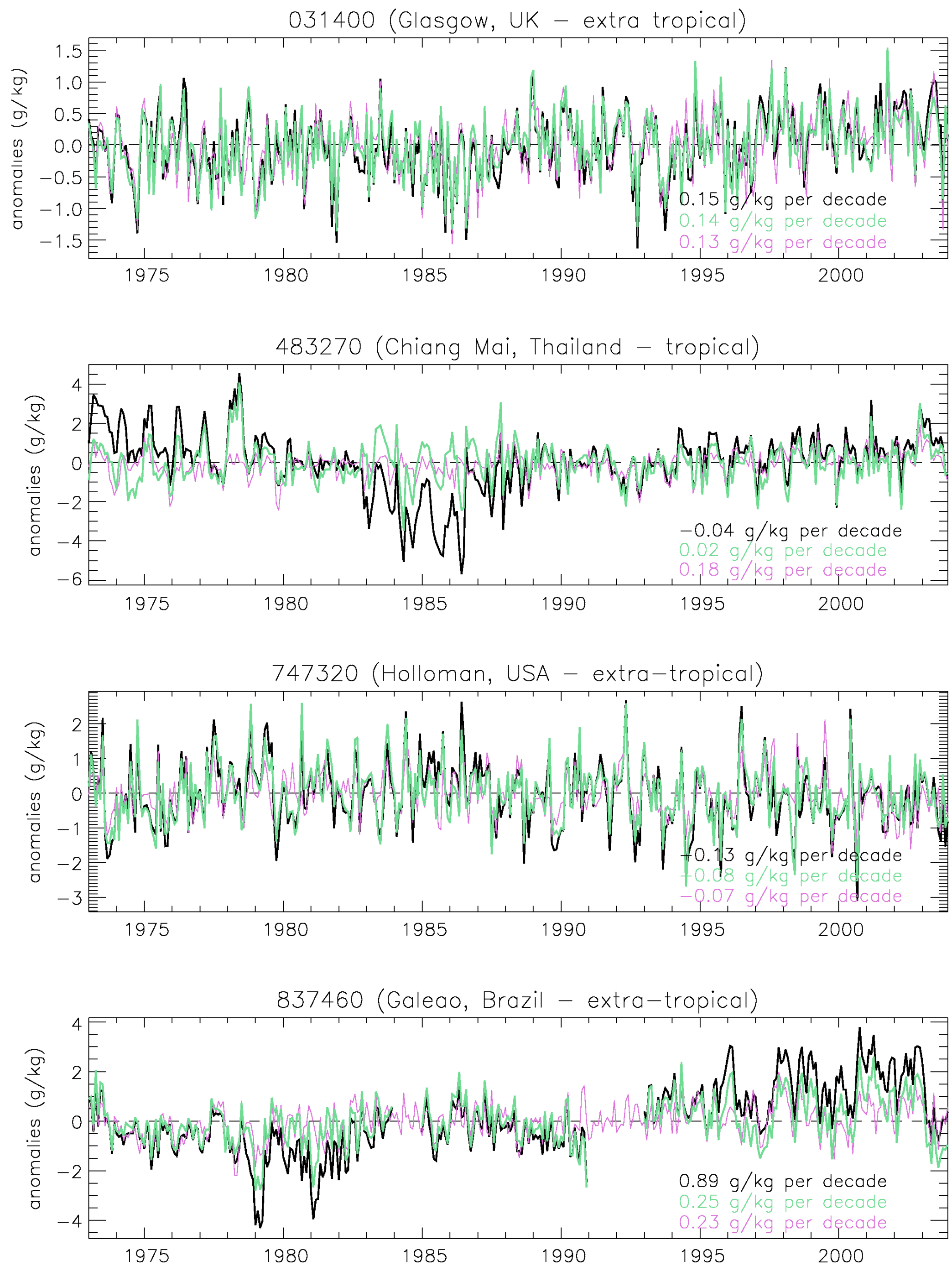


Figure 3.23: Comparison of monthly mean anomaly *q* before and after homogenisation and with the neighbour composite timeseries for four case study stations. The pre-homogenised, post-homogenised and neighbour composite timeseries are shown in black, green and pink respectively. Trends are calculated using a least-squares polynomial fit routine.

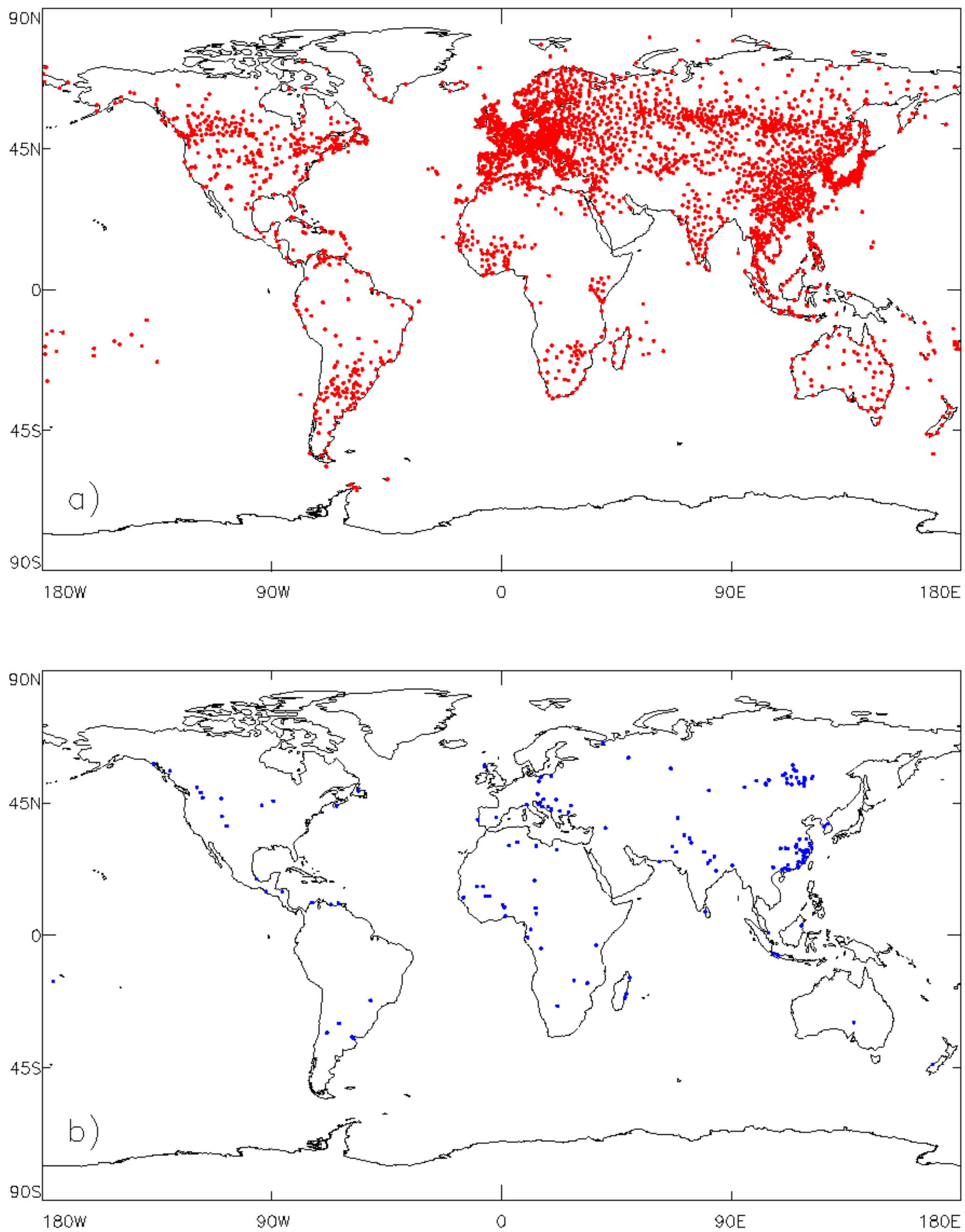


Figure 3.24: Spatial coverage of stations after homogenisation of *q*. Panel a) shows stations remaining after homogenisation (3243 stations) and panel b) shows stations deleted during homogenisation (148 stations).

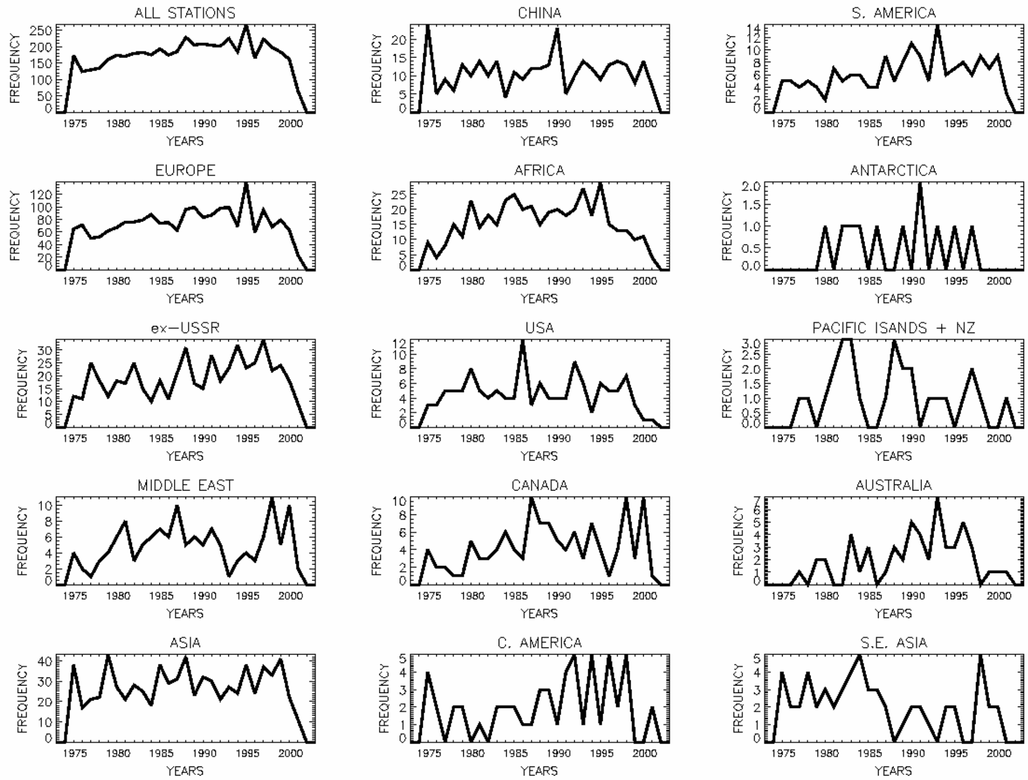


Figure 3.25: Frequency distribution of accepted breakpoints over time for the 1st iteration *q* homogenisation by region.

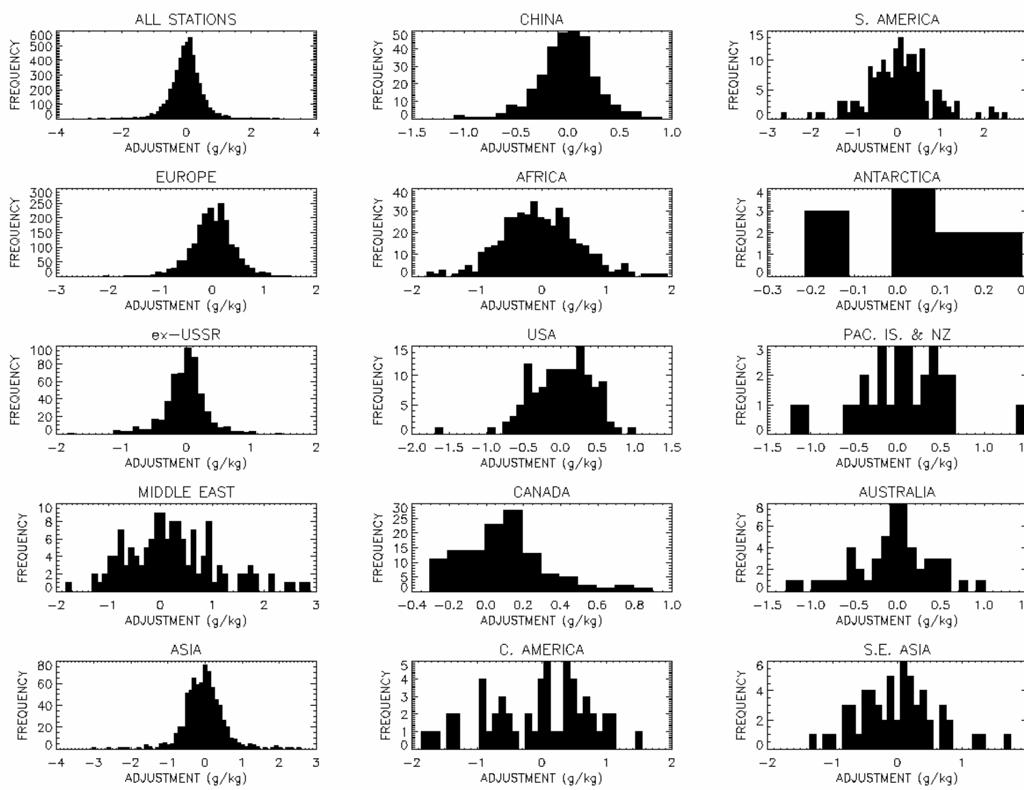
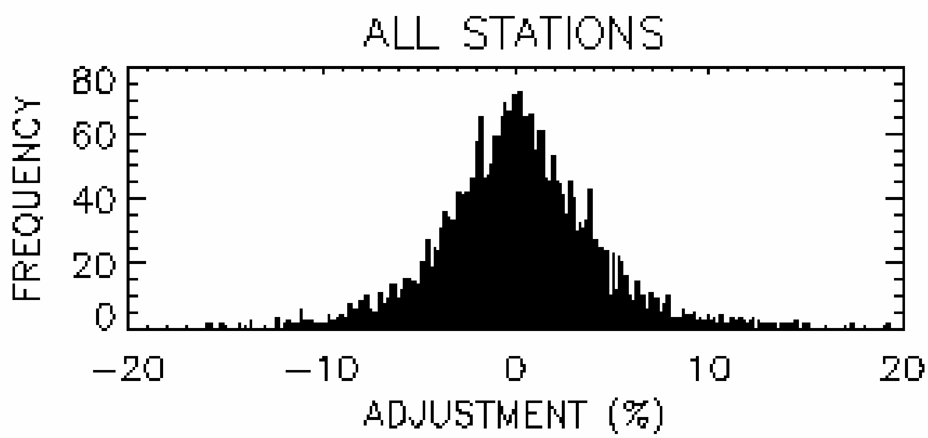


Figure 3.26: Frequency distribution of adjustment quantities for the 1st iteration *q* homogenisation by region.

a)



b)

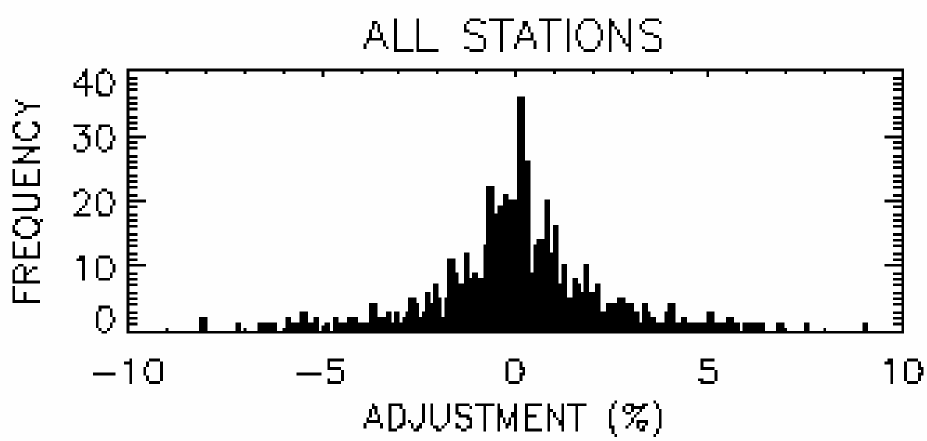


Figure 3.27: Frequency distribution of adjustment quantities for the 1st and 2nd iteration *RH* homogenisation for all stations. Plot a) shows 1st iteration adjustments and panel b) shows 2nd iteration adjustments.

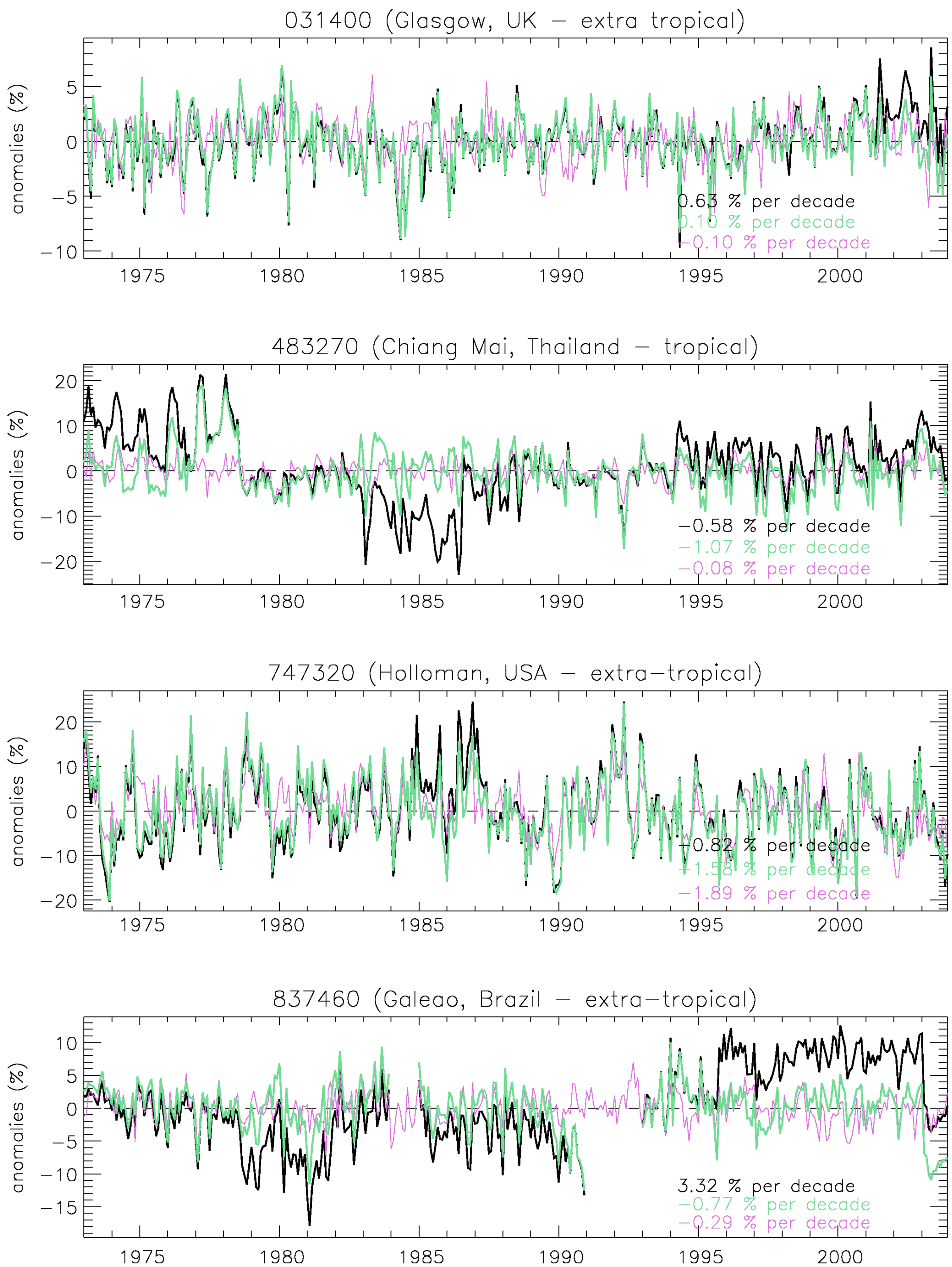


Figure 3.28: Comparison of monthly mean anomaly *RH* before and after homogenisation and with the neighbour composite timeseries for four case study stations. The pre-homogenised, post-homogenised and neighbour composite timeseries are shown in black, green and pink respectively. Trends are calculated using a least-squares polynomial fit routine.

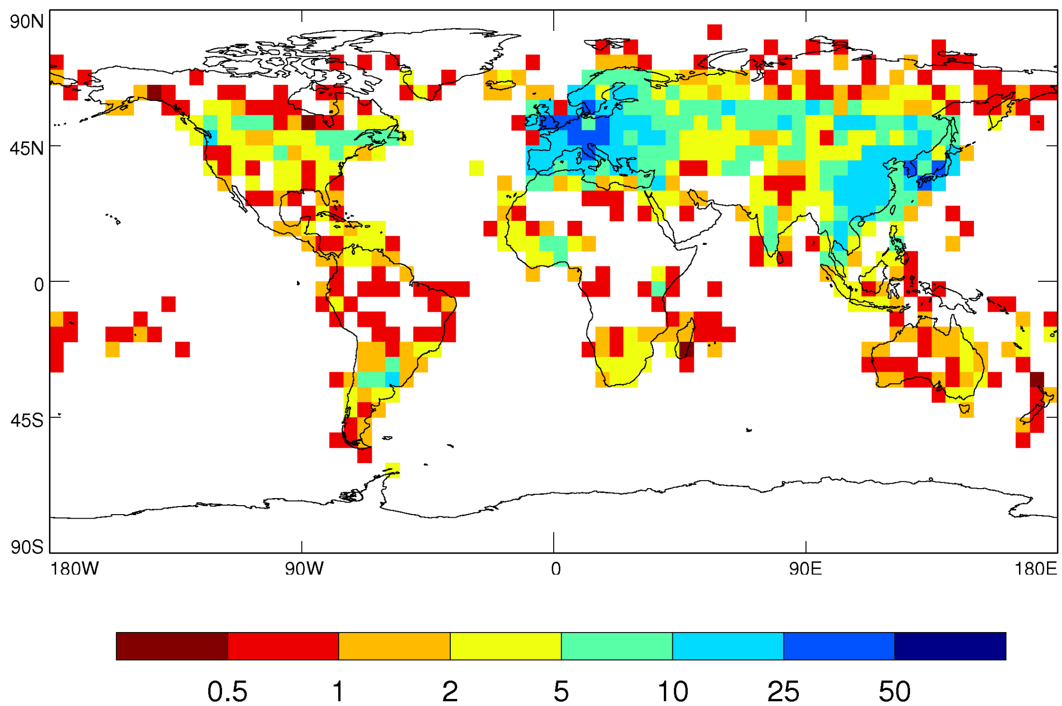


Figure 3.29: Average monthly station density per grid-box for the quality controlled and homogenised land data 1973 to 2003.

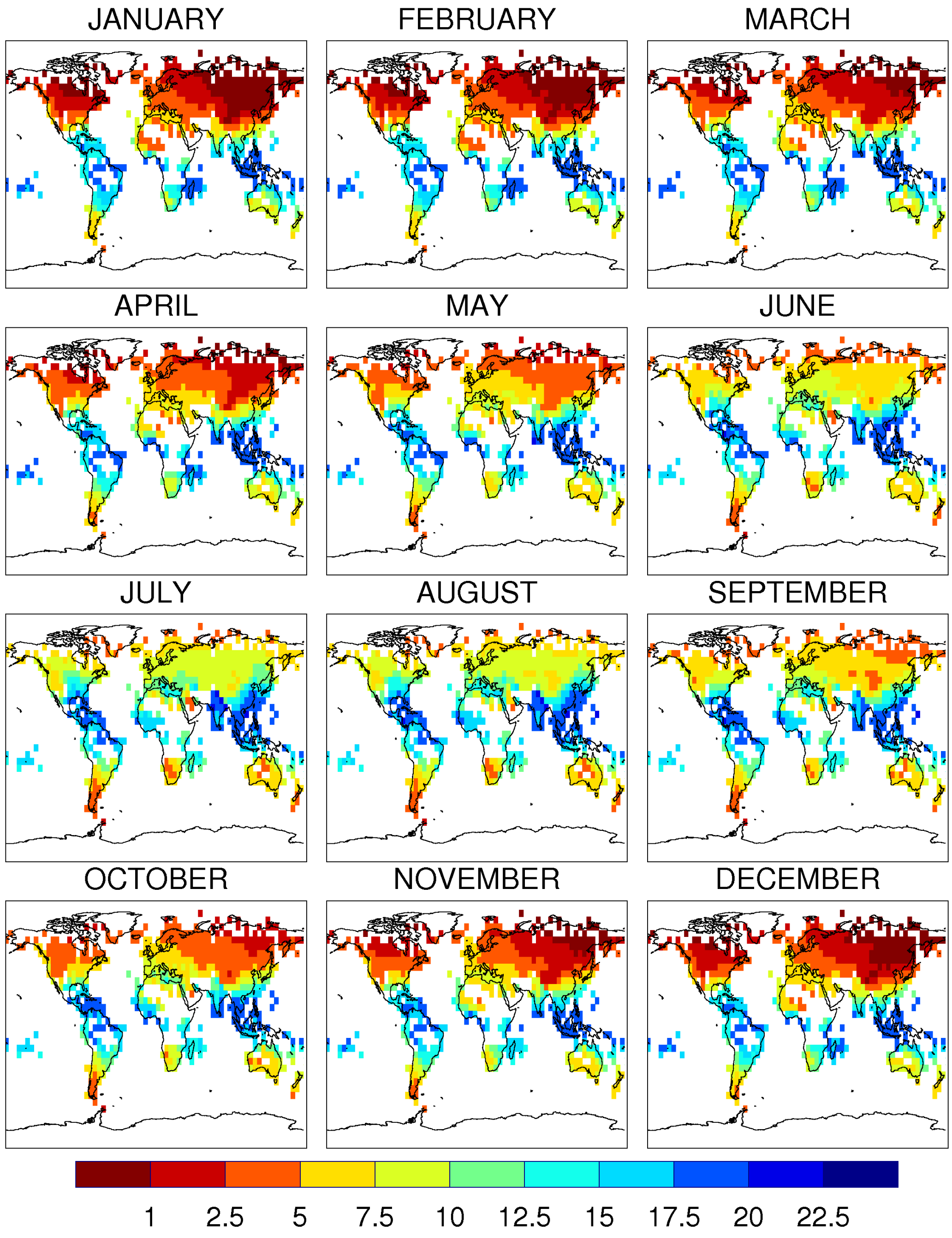


Figure 3.30: Monthly climatology for *q* over the period 1974 to 2003. Units are in g kg^{-1}.

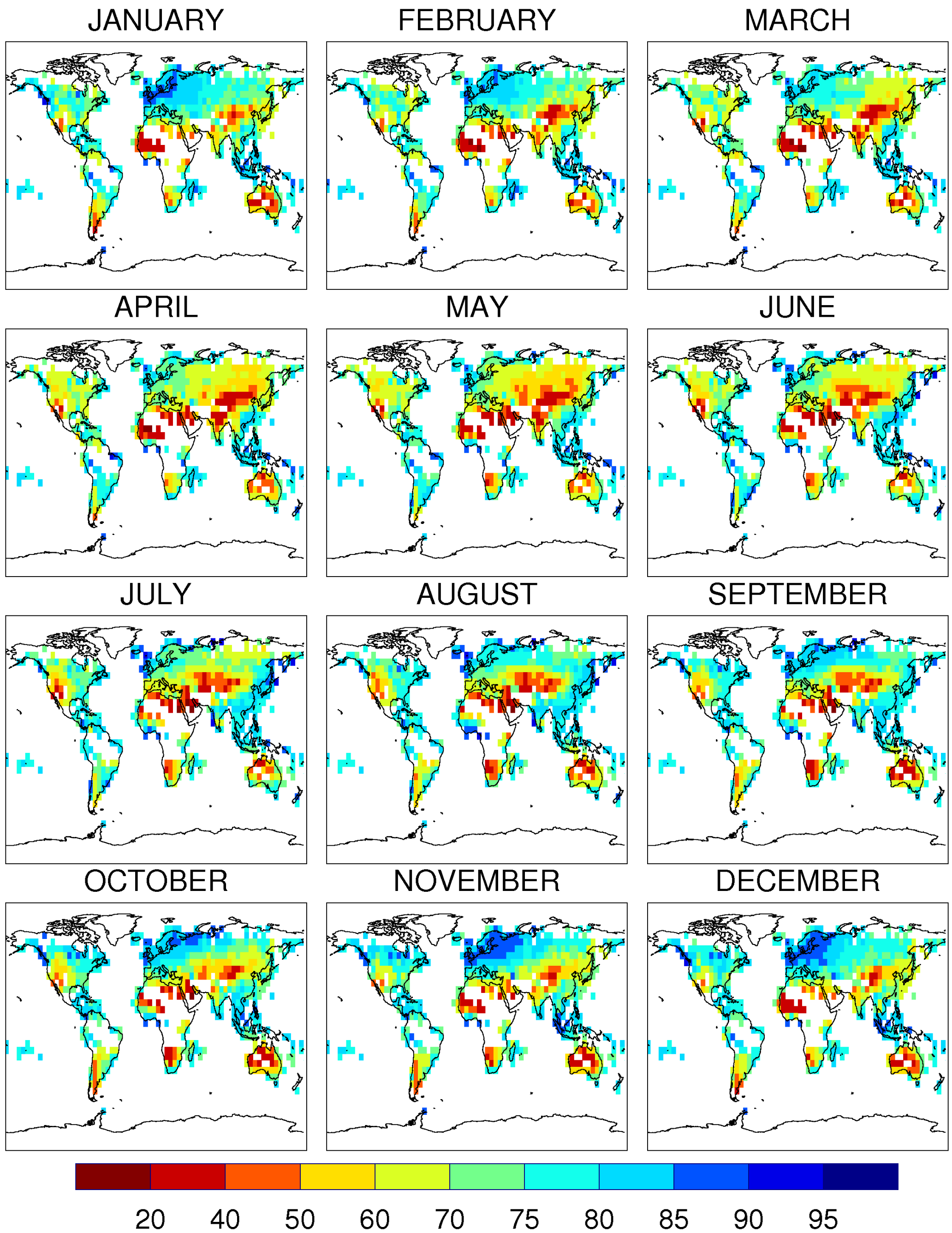


Figure 3.31: Monthly climatology for *RH* over the period 1974 to 2003. Units are in %.

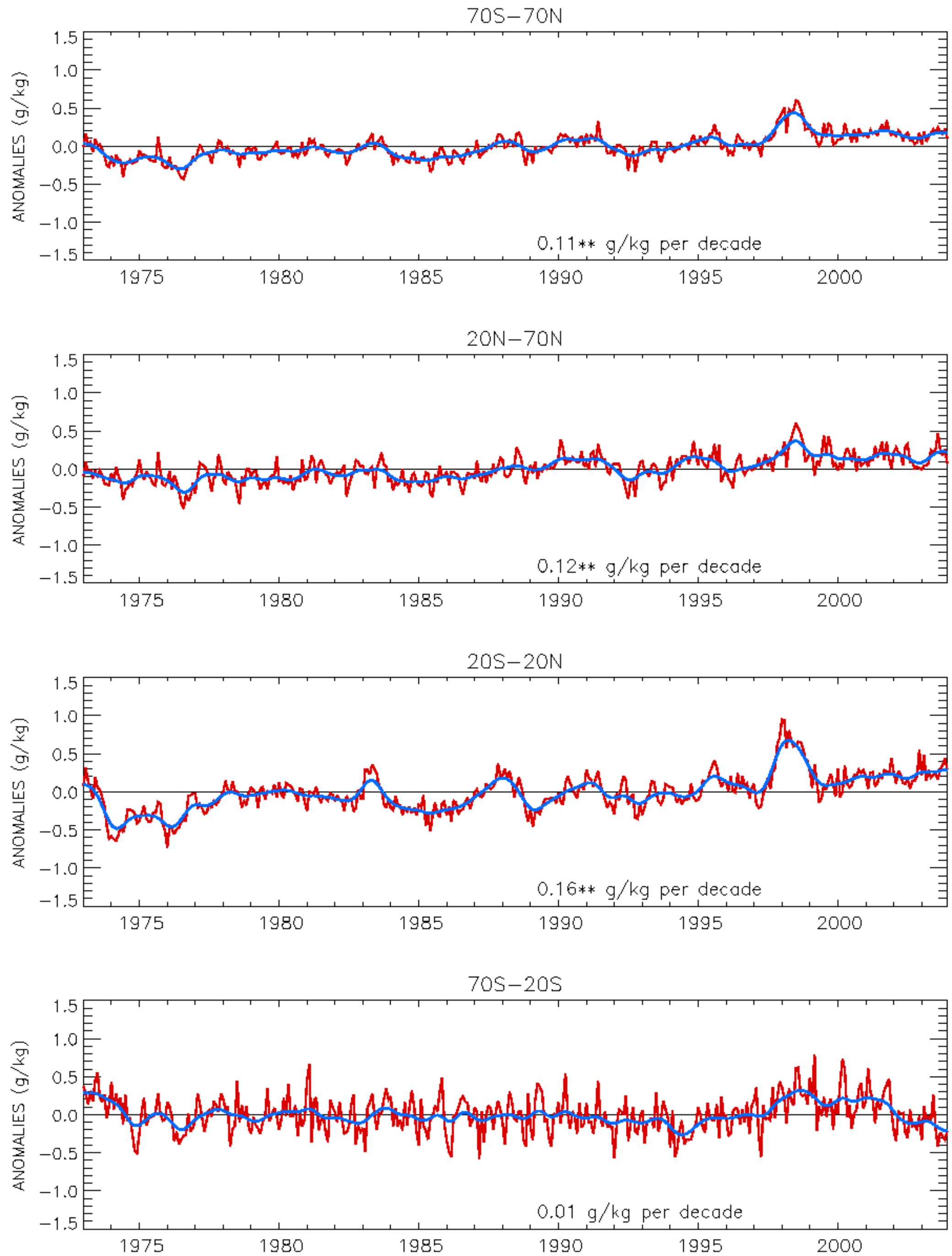


Figure 3.32: Regionally averaged monthly mean anomaly *q* timeseries and trends from 1973 to 2003. Trends are fitted and significance tested using the REML method (Box 3.4) where ** denotes significance at the 1 % level and * denotes significance at the 5 % level. Monthly mean anomaly timeseries are in red and the blue lines are a 21 point Gaussian weighted filtering for the low frequency component of the timeseries.

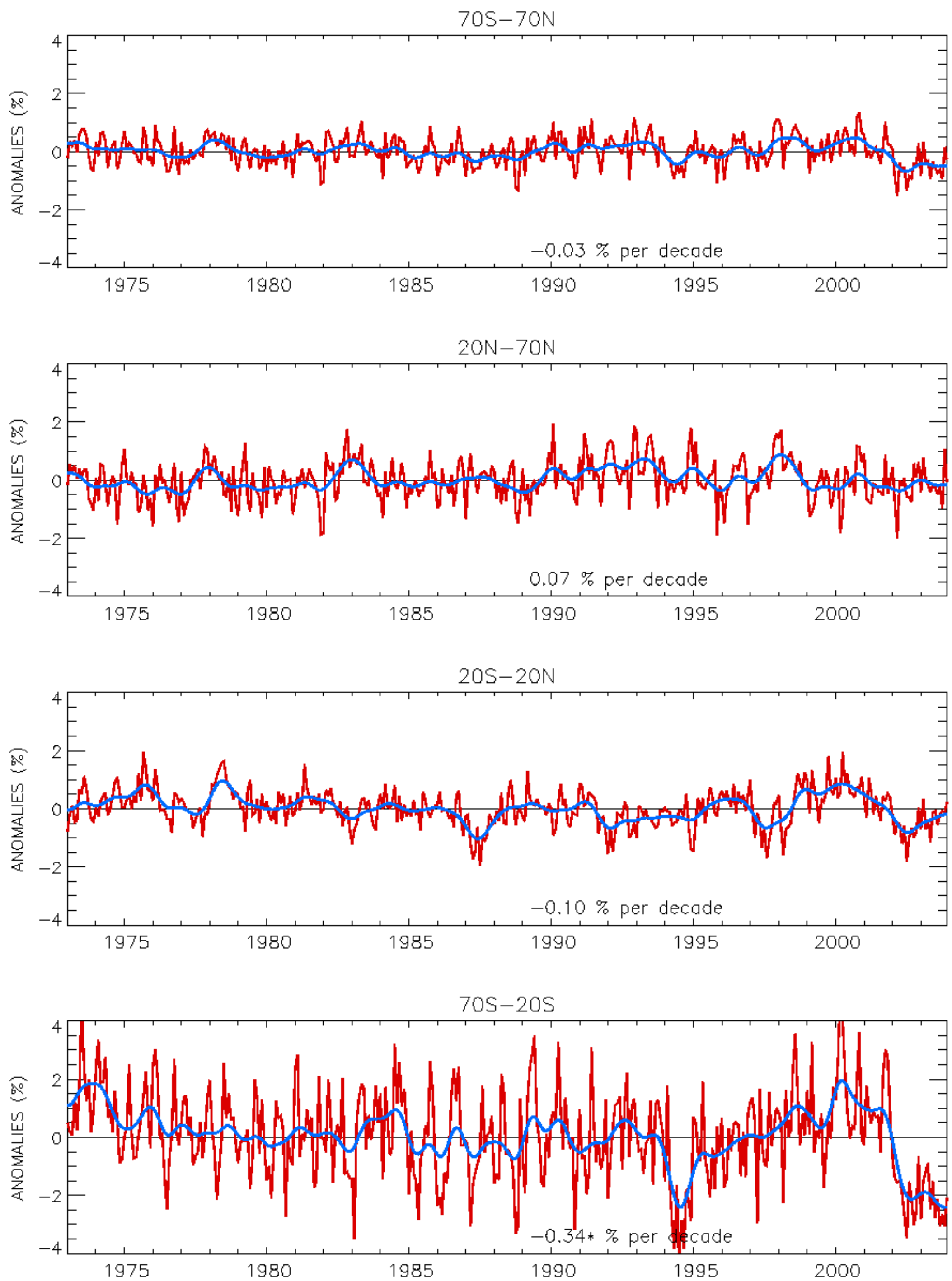


Figure 3.33: Regionally averaged monthly mean anomaly *RH* timeseries and trends from 1973 to 2003. Trends are fitted and significance tested using the REML method (Box 3.4) where ** denotes significance at the 1 % level and * denotes significance at the 5 % level. Monthly mean anomaly timeseries are in red and the blue lines are a 21 point Gaussian weighted filtering for the low frequency component of the timeseries.

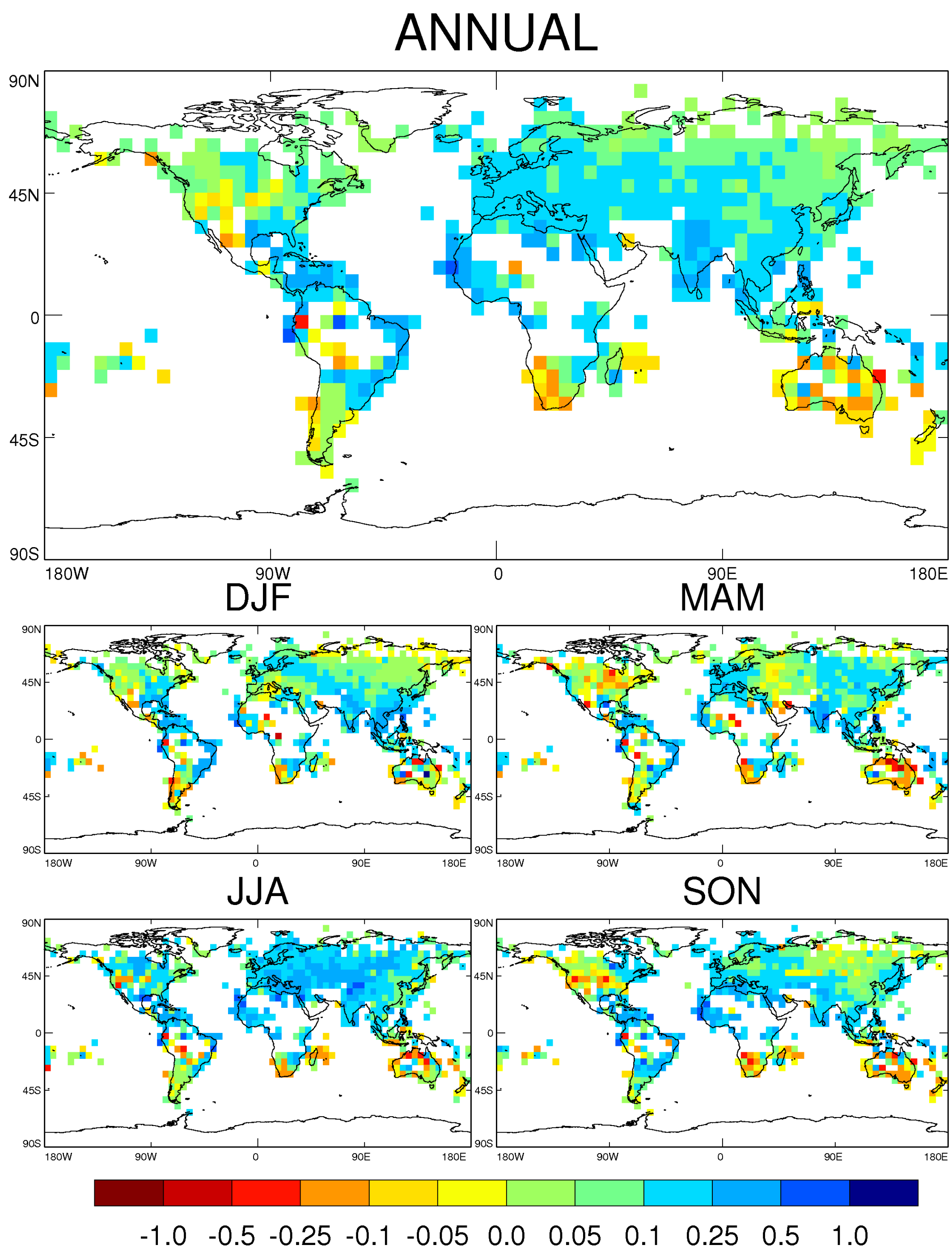


Figure 3.34: Grid-box trends for *q* for the period 1973 to 2003 calculated for the whole annual cycle and seasonal averages. Units are in g kg^{-1} $10yr^{-1}$. Trends are calculated using the MPS method (Box 3.4). At least 50 % of months (seasons) must be present to calculate trends for the whole (seasonal) timeseries. As only two months per season are required to create seasonal averages a grid-box may have a trend present in all seasons but not when trends are calculated over the whole dataset.

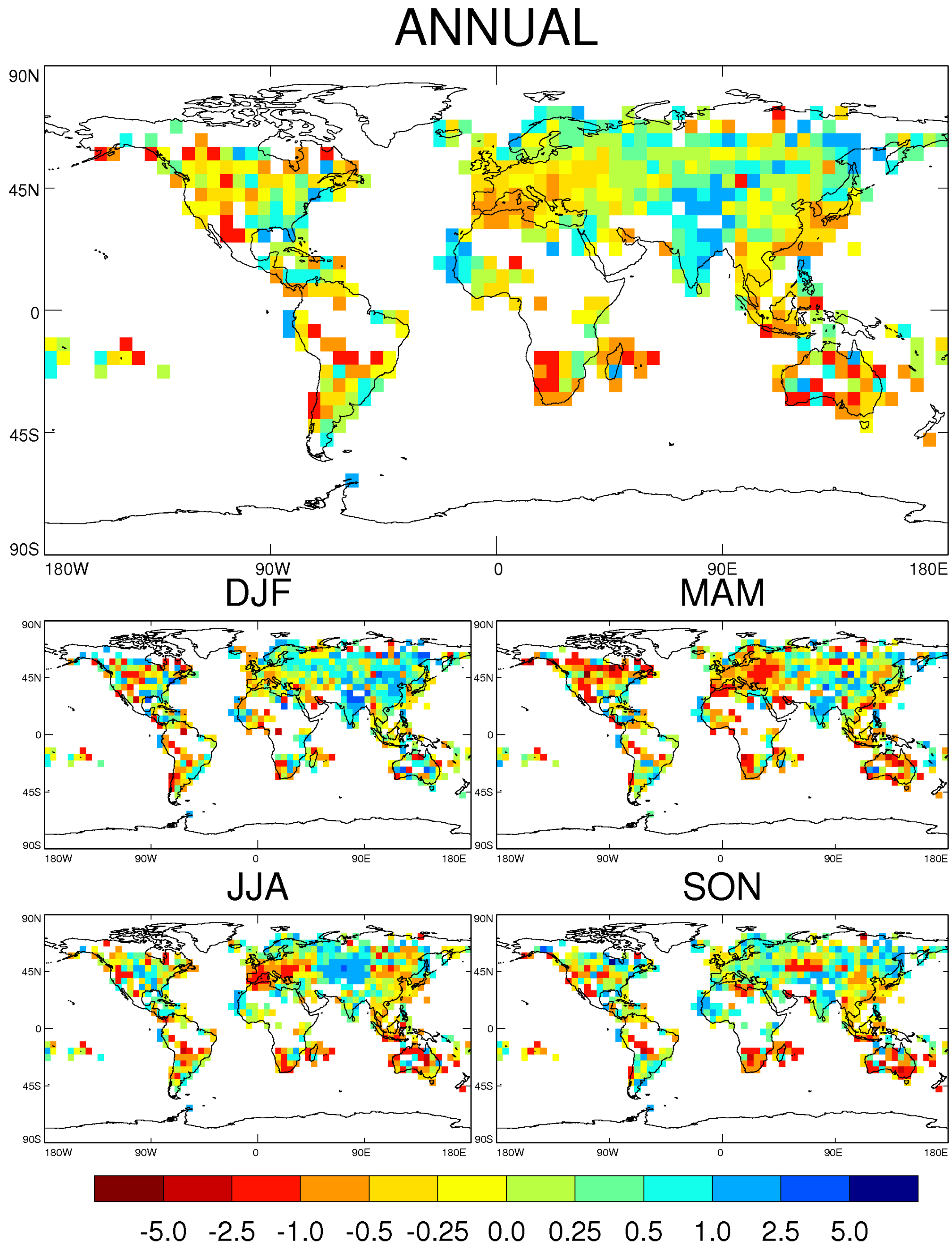


Figure 3.35: Grid-box trends for *RH* for the period 1973 to 2003 calculated for the whole annual cycle and seasonal averages. Units are in % $10yr^{-1}$. Trends are calculated using the MPS method (Box 3.4). At least 50 % of months (seasons) must be present to calculate trends for the whole (seasonal) timeseries. As only two months per season are required to create seasonal averages a grid-box may have a trend present in all seasons but not when trends are calculated over the whole dataset.

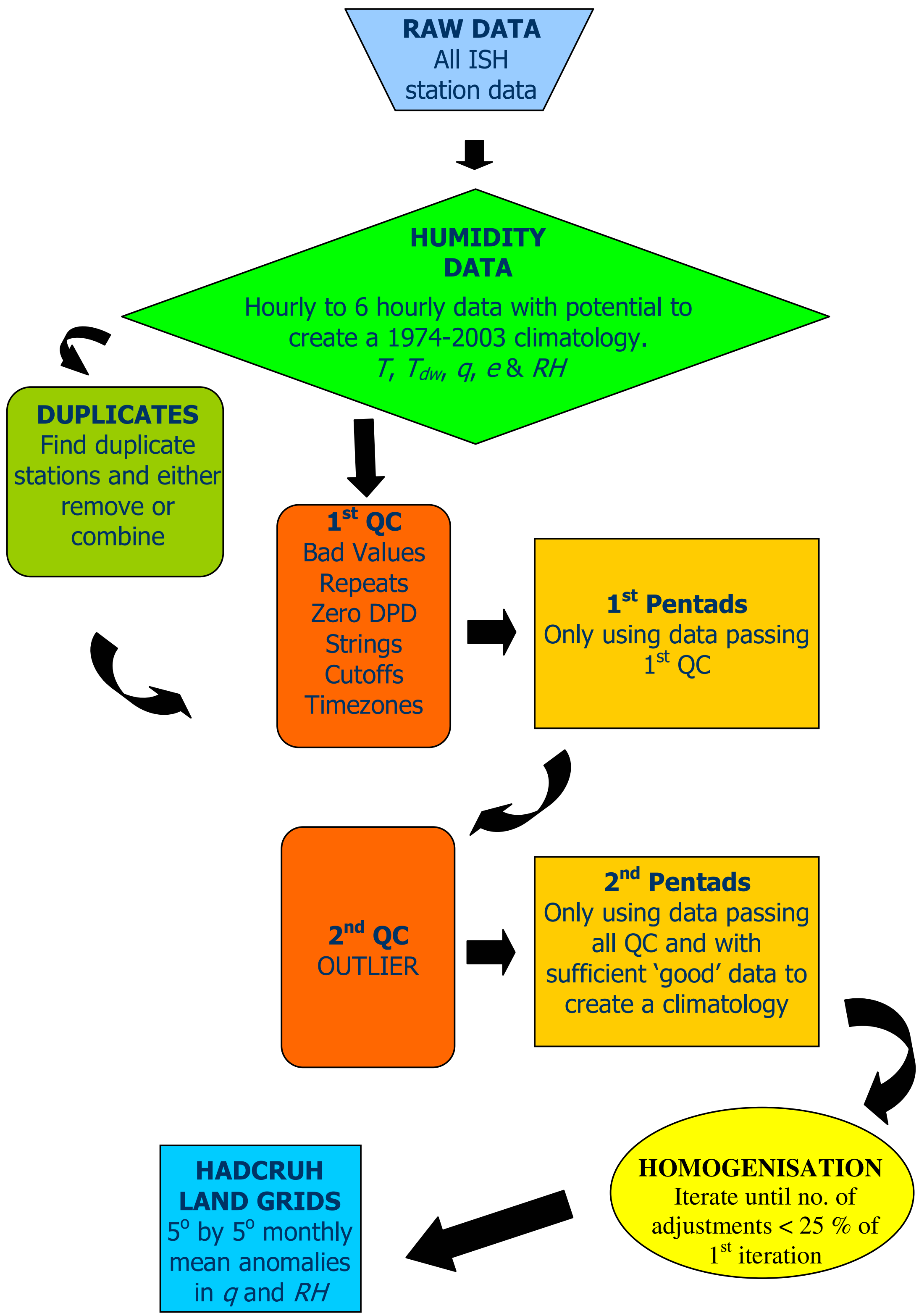


Figure 3.36: Flow diagram for building the land component of HadCRUH.