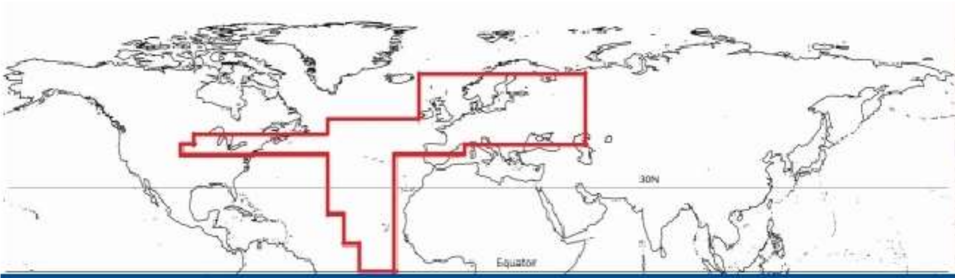
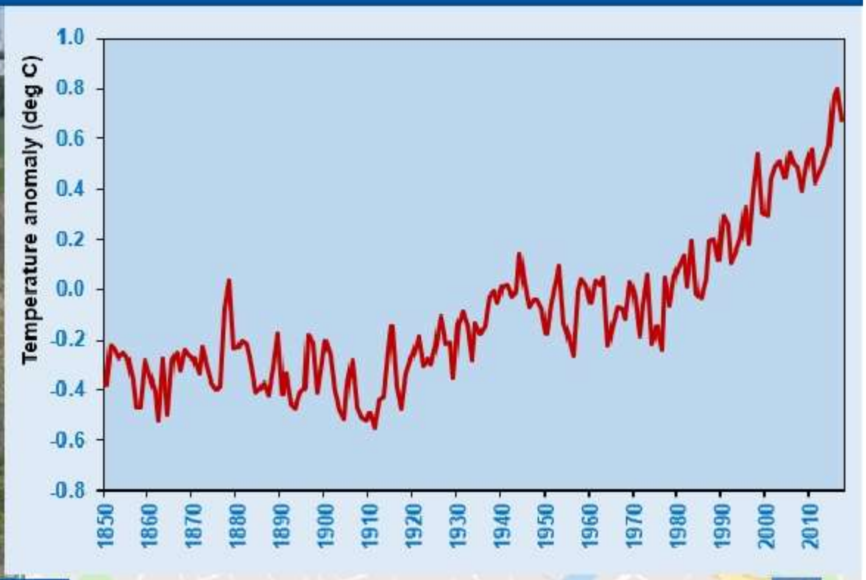


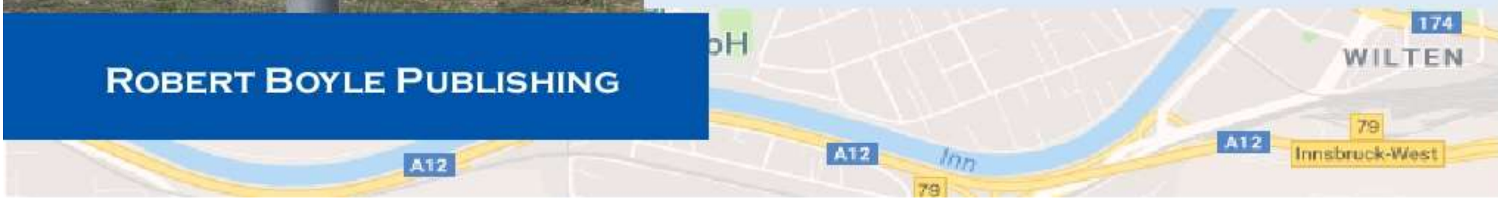
Location	Country	Start	End
ı Mayen	NORWAY	1921	2017
·JORD RADIO	NORWAY	1912	2004
·ALESUND	NORWAY	1969	2017
albard Lufthavn	NORWAY	1898	2017
IDENES	NORWAY	1867	2014
RKENES	NORWAY	1963	2004
KKINGEN (R)	NORWAY	1940	2017
BOSTAD	NORWAY	1936	1991
RDUFOSS	NORWAY	1933	2017

An Audit of the Creation and Content of the HadCRUT4 Temperature Dataset

John McLean, PhD

October 2018





ROBERT BOYLE PUBLISHING

An Audit of the Creation and Content of the HadCRUT4 Temperature Dataset

John McLean, PhD

October 2018

ROBERT BOYLE PUBLISHING

robert-boyle-publishing.com

ISBN 978-0-9876361-0-2 (PDF format)

Copies of this document have been lodged with the National Library of Australia and the State Library of Victoria.

This document was initially published without an ISBN, the only differences between that and this version being the details on this page and the addition of the copyright notice on the previous page.

EXECUTIVE SUMMARY

As far as can be ascertained, this is the first audit of the HadCRUT4 dataset, the main temperature dataset used in climate assessment reports from the Intergovernmental Panel on Climate Change (IPCC). Governments and the United Nations Framework Convention on Climate Change (UNFCCC) rely heavily on the IPCC reports so ultimately the temperature data needs to be accurate and reliable.

This audit shows that it is neither of those things.

More than 70 issues are identified, covering the entire process from the measurement of temperatures to the dataset's creation, to data derived from it (such as averages) and to its eventual publication. The findings (shown in consolidated form Appendix 6) even include simple issues of obviously erroneous data, glossed-over sparsity of data, significant but questionable assumptions and temperature data that has been incorrectly adjusted in a way that exaggerates warming.

It finds, for example, an observation station reporting average monthly temperatures above 80°C, two instances of a station in the Caribbean reporting December average temperatures of 0°C and a Romanian station reporting a September average temperature of -45°C when the typical average in that month is 10°C. On top of that, some ships that measured sea temperatures reported their locations as more than 80km inland.

It appears that the suppliers of the land and sea temperature data failed to check for basic errors and the people who create the HadCRUT dataset didn't find them and raise questions either.

The processing that creates the dataset does remove some errors but it uses a threshold set from two values calculated from part of the data but errors weren't removed from that part before the two values were calculated.

Data sparsity is a real problem. The dataset starts in 1850 but for just over two years at the start of the record the only land-based data for the entire Southern Hemisphere came from a single observation station in Indonesia. At the end of five years just three stations reported data in that hemisphere. Global averages are calculated from the averages for each of the two hemispheres, so these few stations have a large influence on what's supposedly "global".

Related to the amount of data is the percentage of the world (or hemisphere) that the data covers. According to the method of calculating coverage for the dataset, 50% global coverage wasn't reached until 1906 and 50% of the Southern Hemisphere wasn't reached until about 1950.

In May 1861 global coverage was a mere 12% - that's less than one-eighth. In much of the 1860s and 1870s most of the supposedly global coverage was from Europe and its trade sea routes and ports, covering only about 13% of the Earth's surface. To calculate averages from this data and refer to them as "global averages" is stretching credulity.

Another important finding of this audit is that many temperatures have been incorrectly adjusted. The adjustment of data aims to create a temperature record that would have resulted if the current observation stations and equipment had always measured the local temperature. Adjustments are typically made when station is relocated or its instruments or their housing replaced.

The typical method of adjusting data is to alter all previous values by the same amount. Applying this to situations that changed gradually (such as a growing city increasingly distorting the true temperature) is very wrong and it leaves the earlier data adjusted by more than it should have been. Observation stations might be relocated multiple times and with all previous data adjusted each time the very earliest data might be far below its correct value and the complete data record show an exaggerated warming trend.

The overall conclusion (see chapter 10) is that the data is not fit for global studies. Data prior to 1950 suffers from poor coverage and very likely multiple incorrect adjustments of station data. Data since that year has better coverage but still has the problem of data adjustments and a host of other issues mentioned in the audit.

Calculating the correct temperatures would require a huge amount of detailed data, time and effort, which is beyond the scope of this audit and perhaps even impossible. The primary conclusion of the audit is however that the dataset shows exaggerated warming and that global averages are far less certain than have been claimed.

One implication of the audit is that climate models have been tuned to match incorrect data, which would render incorrect their predictions of future temperatures and estimates of the human influence of temperatures.

Another implication is that the proposal that the Paris Climate Agreement adopt 1850-1899 averages as “indicative” of pre-industrial temperatures is fatally flawed. During that period global coverage is low – it averages 30% across that time – and many land-based temperatures are very likely to be excessively adjusted and therefore incorrect.

A third implication is that even if the IPCC’s claim that mankind has caused the majority of warming since 1950 is correct then the amount of such warming over what is almost 70 years could well be negligible. The question then arises as to whether the effort and cost of addressing it make any sense.

Ultimately it is the opinion of this author that the HadCRUT4 data, and any reports or claims based on it, do not form a credible basis for government policy on climate or for international agreements about supposed causes of climate change.

Table of Contents

Preface

This report is based on a thesis for my PhD, which was awarded in December 2017 by James Cook University, Townsville, Australia. The thesis[1] was based on the HadCRUT4 dataset and associated files as they were in late January 2016. The thesis identified 27 issues of concern about the dataset.

The January 2018 versions of the files contained not just updates for the intervening 24 months, but also additional observation stations and consequent changes in the monthly global average temperature anomaly right back to the start of data in 1850.

The report uses January 2018 data and revises and extends the analysis performed in the original thesis, sometimes omitting minor issues, sometimes splitting major issues and sometimes analysing new areas and reporting on those findings.

The thesis was examined by experts external to the university, revised in accordance with their comments and then accepted by the university. This process was at least equivalent to "peer review" as conducted by scientific journals.

John McLean
June 2018

[1] Thesis title: "An audit of uncertainties in the HadCRUT4 temperature anomaly dataset plus the investigation of three other contemporary climate issues"

General Notes

Latitude and Longitude

Under the convention used by the HadCRUT4, HadSST3 and CRUTEM4 datasets discussed in this report, positive values for latitude apply for North, negative to south, and positive longitudes to West, negative to East. In some tables the data appears in the format.

Abbreviations used in this report

CRUTEM*n*	The near-surface atmospheric temperature anomaly dataset produced by the Climatic Research Unit at the University of East Anglia, Norwich, UK. The *n* indicates the version. This dataset is based on data from observation stations.
HadCRUT*n*	Composite temperature anomaly dataset created using the same data as for CRUTEM (see above) and HadSST (see below). The *n* indicates the version.
HadSST*n*	Sea surface temperature anomaly dataset created by the UK Met Office Hadley Centre. The *n* indicates the version.
ICOADS	International Comprehensive Ocean-Atmosphere Data Set
IPCC	Intergovernmental Panel on Climate Change
SST	Sea surface temperature
WMO	World Meteorological Organization

1. Introduction

Western governments are spending a small fortune addressing what was originally called "global warming" and then became "climate change". Not only are attempts being made to reduce carbon dioxide emissions, which are claimed to be the cause of rising temperatures, but developed countries are handing over significant sums of money to less developed countries with the aim of fighting "climate change".

This expenditure by governments and consumers rests in part on the historic global temperature record. The question that seems to have never been addressed is whether the global temperature record is accurate. and whether Is the claimed extent of warming in fact correct? If the temperature record is incorrect then so too are climate models because they have been adjusted so that their output closely matches that record (although of course the models might be incorrect for other reasons).

The key temperature data used by the Intergovernmental Panel on Climate Change (IPCC) is the HadCRUT dataset, now in its fourth version and known as HadCRUT4. When I was an Expert Reviewer of the IPCC's 2013 Climate Assessment report I raised questions as to whether the HadCRUT4 dataset and the associated HadSST3 dataset had been audited. The response both times was that it hadn't.

Further indication that no-one has independently audited the HadCRUT4 dataset came early in my analysis, when I found that certain associated files published simultaneously with the main dataset contained obvious errors. Given the nature of the errors and the years in which some of the errors occurred, it seemed that they probably existed for at least five years. (At the time I notified the relevant people and the files have since been corrected.)

It seems very strange that man-made warming has been a major international issue for more than 30 years and yet the fundamental data has never been closely examined. One can understand that individuals, organizations and even countries that benefit in some way from the claims about manmade warming might be reluctant to audit the HadCRUT4 in case it reveals the possible exaggeration of the change in temperatures and they lose some of those benefits. It is less understandable that individuals, organizations and countries that might suffer under actions based on the interpretations of the temperature records would not try to determine whether the record was in fact accurate.

Almost all of the published papers about the HadCRUT4 dataset and its two associated datasets were written by people involved in the construction and maintenance of them, which hardly makes for unbiased analysis. McKitrick (2010) is probably the most extensive independent audit to date, albeit not of the HadCRUT4 dataset but the previous version, HadCRUT3. Some comments in that paper have been made redundant by changes of basic procedures for HadCRUT4 but others are still pertinent. Cowtan and Way (2014) focussed on data coverage but this is only the tip of a very large iceberg of issues.

This report is an initial audit of the HadCRUT4 dataset, and of various issues that relate to its creation. It also looks further upstream to the measurement of temperatures and the adjustment of temperature data, because if these are inaccurate then the HadCRUT4 dataset will likewise be inaccurate. HadCRUT4 data is updated each month, not only to add data for the latest month, but to also add any newly discovered or revised relevant data for early years, and occasionally to correct errors in the data. This report uses the HadCRUT4 data, and associated material, as it was on 25 January 2018.

Data analysis for this report was largely by purpose-built computer software that read the available files and at times calculated simple statistics (e.g. mean and standard deviation) or determined coverage or average temperatures over large regions using the system of weighting described in section 2.6.

To provide context to some of the discussions in this report Figure 1.1 shows the HadCRUT4 annual average global temperature anomaly as it was when the data was downloaded. As with other issues in this report annual average values are used in this figure to illustrate the situation. This is because the data varies every month both in its supply (location and reporting) and the small variations known as data "noise" that come from the chaotic minor variations in weather.

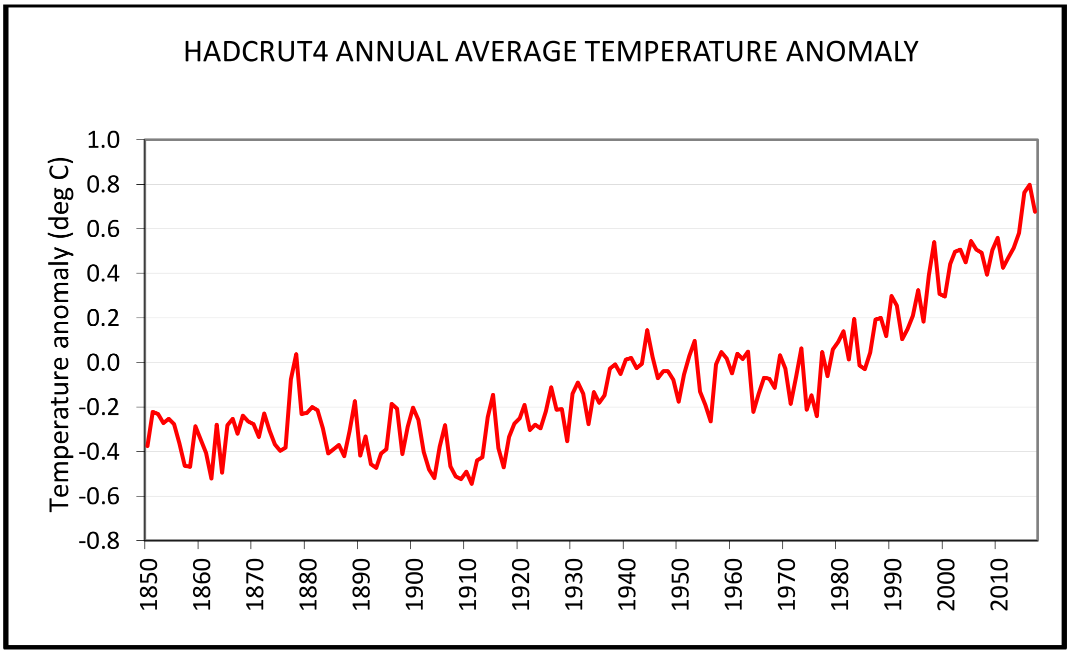


Figure 1.1 HadCRUT4 global annual average temperature anomalies relative to a baseline of the average 1961-1990 temperatures

Some issues in this study focus on individual situations, such as a single observation station, that would have negligible impact on global average values. Similar issues could exist elsewhere in the data and processing, perhaps less obviously, and the fact that issues can be identified at all suggests a variety of problems including lack of attention to detail and possible problems with fundamental procedures or processing. Above all they show that considerable uncertainty exists about the accuracy of the HadCRUT4.

The PhD candidature on which this work is based was funded on the normal "per candidate" basis by the Australian government and had no additional funding. The creation of this report itself had no funding whatsoever.

2 Background information

2.1 Introduction

This section presents a brief summary of the basic concepts of the HadCRUT4 dataset so that readers might better understand the sections that follow.

2.2 Data and information sources

The HadCRUT4 dataset is derived from the data used in the HadSST3 dataset and the CRUTEM4 dataset. The HadSST3 dataset is derived from sea surface temperatures. It is created and maintained by the UK's Met Office Hadley Centre for Climate Science and Services (more commonly just the "Hadley Centre"). The CRUTEM4 dataset is derived from temperature data from observation stations that are almost exclusively on land. It is created by the Climatic Research Unit of the University of East Anglia (and commonly known as "CRU"). Internet home pages for the two are given in footnote[2] below.

For details of the HadCRUT4 dataset we have Morice et al (2012) and Jones (2016). The definitive papers for details about the CRUTEM4 dataset are Jones et al (2012) and Osborn and Jones (2014). There appears to be no single comprehensive description of the HadSST3 dataset and its creation. According to the web (see footnote) the main reference is Kennedy et al (2011b) but it refers to other papers for certain details (e.g. Rayner et al 2006), and those papers refer to yet more papers.

2.3 Grid-based system

Like the CRUTEM4 and HadSST3 datasets, the HadCRUT4 dataset is a grid-based system with data values for each grid cell. The three datasets all use grid cells of 5° latitude x 5° longitude, covering all of the Earth's surface, requiring 36 grid cells from north to south and 72 from west to east, a total of 2592 grid cells. A 5° x 5° grid cell near the equator is about 550km by 550 km and the width of cells reduces as we move away from the equator into higher latitudes.

Because the values for each grid cell are calculated independently, provided that data is supplied for the area covered by a grid cell, in the final product there is no bias towards grid cells with a greater amount of data over those cells with less.

[2] Climatic Research Unit CRUTEM4 - https://crudata.uea.ac.uk/cru/data/temperature/
Hadley Centre HadSST3 - https://www.metoffice.gov.uk/hadobs/hadsst3/
(both as at 10 April 2018). The latter says that the data is based on version 2.5 of the ICOADS database but given that version 3 of ICOADS was published in 2016 (Freeman et al, 2017) and that one author of the associated paper was from the Hadley Centre, it seems likely that HadSST3 now uses version 3 of ICOADS.

2.4 Temperature anomalies

The HadCRUT4, CRUTEM4 and HadSST3 datasets are based on temperature anomalies. A temperature anomaly is the difference between a given temperature and a base or standard temperature that's appropriate for the situation. The advantage of using anomalies over actual or absolute temperatures is that it takes into account the fact that temperatures generally vary with latitude, warmer near the equator and cooler towards the poles. Temperature can also be affected by altitude, exposure to certain weather, ocean current and proximity to the coast, and so on.

In general terms, all temperature anomalies are calculated according to

$$T_{anom} = T_s - T_{base}$$

where T_{anom} is the calculated temperature anomaly, T_s is a specific temperature (i.e. recorded temperature or some value derived from it) and T_{base} is the base or reference temperature.

For the HadCRUT4, CRUTEM4 and HadSST3 datasets the three values are in reference to specific months with T_s being the mean temperature in a given month and T_{base} the long-term average for the same calendar month, and the anomaly, T_{anom}, being for the given month.

For CRUTEM4 data (i.e. data obtained from observation stations) both the base temperature and the anomaly are calculated from the mean monthly temperatures, which are the average of the mean daily minimum temperature and the mean daily maximum temperature across the month. The base temperature, T_{base}, is the average of the monthly mean temperatures for the same calendar month across the period from 1961 to 1990 inclusive. For example, when calculating a temperature anomaly for January the base temperature is the average in each January from 1961 to 1990 inclusive. More than one station might be located in any given CRUTEM4 grid cell and the cell's value in a given month is the average of the temperature anomalies for all reporting observation stations within that grid cell in that month.

A different approach is used to calculate the sea surface temperature anomalies for the HadSST3 dataset. While the dataset is expressed on a grid cell size of 5° x 5° and by month, the values are derived from temperature anomalies calculated using 1° x 1° grid cells (here called sub-cells for convenience) and 5-day intervals (also known as pentads). Each calendar month has a notional six pentads except for August, which is assigned seven, making 73 pentads per 365-day year. Baseline average temperatures are derived for each sub-cell and pentad, first by estimation, based on geometry and numerous assumptions and then by modifying those estimates using actual SST measurements made during the period from 1961 to 1990 within the sub-cell and pentad. The monthly mean temperature anomalies that appear at 5° x 5° grid cell resolution in the HadSST3 dataset were derived by the interpolation and extrapolation of the sub-cell and pentad data.

2.5 Grid cell types

While the HadCRUT4 data are from two basic sources, fixed observation stations (almost exclusively on land) or measurements of sea surface temperature, the grid cells are of three types – Land, Sea, and Coastal/Island. The first covers regions that are only of land and the data is the same as used in the CRUTEM4 dataset. The second covers regions that are only of sea grid and the data is the same as the HadSST3 dataset. The third are coastal and island grid cells that cover both land and sea and whose temperature data is derived from land or sea temperature measurements or both, depending on data availability.

Using Figure 8 of Brohan et al (2006) as a guide and supplementing it with any further grid cells whose data was not always from one source (i.e. land and sea), cells can be identified as land, sea or costal. The details for all three types of cells are shown in Table 2-1 and an indicative map in Figure 2.1.

The coverage shown in Table 2-1 is calculated according to the methodology associated with using grid cells. After taking into account the overlap of the two types of data in coastal/island grid cells the maximum potential global coverage of sea surface temperature data and data from observation stations over land are 81.87% and 46.15% respectively.

	Cell Type	Cell Count	Percent of cells	Percent of Earth's Surface
Northern Hemisphere	Land	293	22.61%	24.32%
	Sea	560	43.21%	41.79%
	Coastal	443	34.18%	33.89%
Southern Hemisphere	Land	270	20.83%	11.94%
	Sea	752	58.02%	65.91%
	Coastal	274	21.14%	22.15%
Global	Land	563	21.72%	18.13%
	Sea	1312	50.62%	53.85%
	Coastal	717	27.66%	28.02%

Table 2-1 Grid cell types and the number grid cells, percentage of grid cells and the percentage of the Earth's surface they cover.

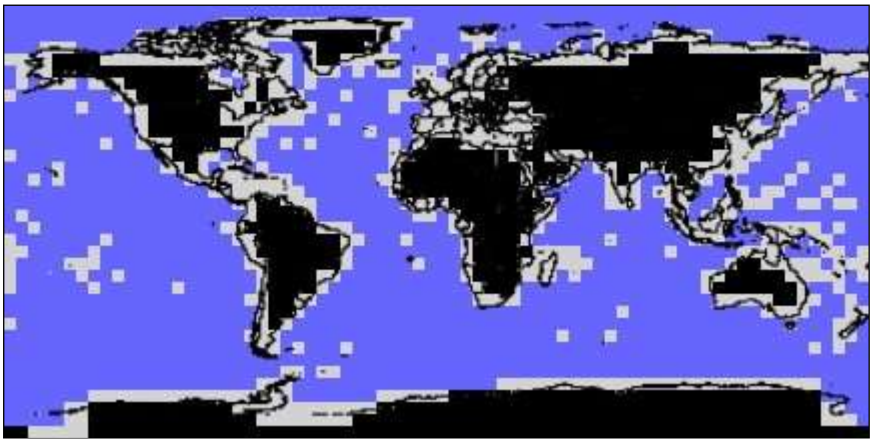


Figure 2.1 Indicative map of the three types of grid cells, with 'land' grid cells in black, 'sea' grid cells in blue and 'coastal/island' grid cells in grey. (Based on Brohan et al, 2006)

2.6 Data weighting

The Earth's near-spherical shape means that lines of longitude become closer as they move towards the North and South poles, which means that the area of each 5° latitude x 5° longitude grid cell decreases. To compensate for this in calculations of coverage and average global, hemispheric or regional temperature the data is weighted by the cosine of the latitude of the centre of the grid cell. This produces two basic equations:

(a) the percentage coverage for a given area in a given month:

$$\%cover = \frac{\sum (\text{cosine } (x)) \text{ for all reporting grid cells in the region}}{\sum (\text{cosine } (x)) \text{ for all grid cells in the region, reporting or not}} \times 100.0$$

where x is the degrees of latitude of the centre of the grid cell and both the numerator and denominator refer to grid cells over the same area (e.g. hemisphere or region)

(b) The average temperature anomaly for a given region:

$$TA_{avg} = \frac{\sum (\text{cosine}(x) * T_{cell})}{\sum \text{cosine}(x)}$$

where T_{cell} is the grid cell value, x is the latitude of the centre of the grid cell and both numerator and denominator are for only the grid cells that reported mean monthly temperature data.

The calculation of the HadCRUT4 global average attempts to take into account different extents of coverage in the two hemispheres by calculating hemispheric averages separately and then averaging the two to produce the global average.

On the basis of the above, the key point being the diminishing size of grid cells as we move from the equator to the poles, 50% of the surface area of a hemisphere is from the equator to 30° latitude, 36.6% from 30° to 60° and 13.4% from 60° to 90°. (The global percentages are the same if we talk about the total of the corresponding bands in the northern and southern hemispheres.) Further, a grid cell with one side along the equator is slightly less than double (actually 1.86 times) the size of one with its polar-most side at 60° and just under 23 times the size of a grid cell that touches the north or south pole.

2.7 Data adjustment

Data from observation stations and measurements of sea surface temperature are often adjusted after their recording. For land-based observation stations this is to make historical data (notionally) consistent with being recorded at the current location, with the current instruments in the current screening at the current time of day, all of which might have previously changed over time. For sea surface temperatures the adjustments are to (notionally) make the data consistent with being measured by the same technique, which is no easy task when the technique was not always recorded and might even vary on a single ship.

Jones (2016) tries to imply that the land and sea surface temperatures are independently adjusted, saying "*Related to this, adjustments for land data are estimated completely independently from the marine series, so these two components mutually support each other*" (pg271)

This is refuted just two pages later when Jones (2016) says "*If the [SST] adjustments were not applied then … there would be a major discrepancy between the land and marine components prior to about 1940.*" (p.273)

Shortly after, Jones (2016) reiterates the point saying "*If the latter had not been adjusted for the large bias due to the change from bucket measurements, then the agreement with the land record would not have been produced.*" (p.275). On this basis, if observation station data is incorrectly adjusted then HadSST3 will likewise be incorrect.

Jones (2016) also states in relation to adjustments to data from observation stations "*At the hemispheric and global scale, however, because adjustments of both signs occur with similar frequencies, the adjustment factors tend to cancel.*" (p.275) The frequency of the upward or downward adjustments are irrelevant on these scales; it is the size of the adjustment that matters. For example, five adjustments downwards by 1.0°C are not cancelled out by five adjustments upwards by 0.2°C.

Various specific issues related to data adjustment will be discussed later in this report but one conceptual issue warrants mention at this point. The HadSST3 dataset is based on 100 variants (i.e. different datasets), each derived from different assumptions about the methods used to measure sea surface temperature and the associated adjustments necessary to bring the data from those different methods to a common base (i.e. to conceptually make it all consistent with being measured in the same way). The data is independently processed for each of the 100 variants until finally an "ensemble" version of HadSST3 is created by taking the median[3] value.

The same technique is applied to the HadCRUT4 dataset, using firstly the 100 variants of the HadSST3 dataset and then a set of artificially created variants of the observation station data used in the CRUTEM4 dataset which was not derived from variants.

Jones (2016) says "*Morice et al. (2012) introduced the concept of multiple, but equally plausible, realizations of the past. The HadCRUT4 dataset has developed 100 such realizations with a best guess, the median value for each grid box, and the median of the 100 realizations of global and hemispheric averages*." (p.271) (my underlining).

Nowhere does Jones (2016) offer any explanation of why the median is in fact a better guess than any other value when at most only one variant of a dataset would be correct. The presence of incorrect variants that predominantly resulted in temperature anomalies either above or below the correct value will push the median in one direction or the other.

[3] The 'median' is the value at the centre of a sorted list or in the case where the number of entries is an even number the average of the two values either side of the centre.

2.8 Notes on an important statistic

The 'Standard deviation' of a set of data is a statistical measure of how spread out the data is, assuming that the data distribution follows the "bell-curve" of so-called "normal" or Gaussian distribution. The name "bell-curve" arises because more data is found near the mean or average value and less as we move out in either direction from there. Temperature data for a given time of year in a given location generally follows this distribution pattern. Under normal distribution 68% of the data will be found within one standard deviation from the mean, 95% within two standard deviations and 99.7% within three standard deviations. (Later this report will discuss how much data is excluded if the cut-off is at a certain number of standard deviations. Note that according to the above, only 3 in 1000 values could be expected to be beyond three standard deviations.)

Figure 2.2 is a scatterplot of the long-term mean monthly temperatures and standard deviations for each month as supplied for all observation stations used in the HadCRUT4 dataset (except for three omitted data pairs that are obviously incorrect). The standard deviation is inversely linked to the temperature, indicating that temperatures over land vary more widely under cold conditions than they do under warm conditions. As Figure 2.3 indicates, the CRUTEM4 dataset, with its bias towards the Northern Hemisphere due to the greater land area there, shows greater standard deviations during the months of colder weather in that hemisphere.

The standard deviations calculated from the temperature data are used to determine if a mean monthly temperature from an observation station is so unusual that that it should be rejected as a possible error. The use of a threshold of five standard deviations from the mean will be discussed in more detail later. For now, it is sufficient to appreciate that a far greater range of values is acceptable under cold conditions than under warm conditions, where a difference of just over 1.0°C from the mean temperature will be regarded as an outlier.

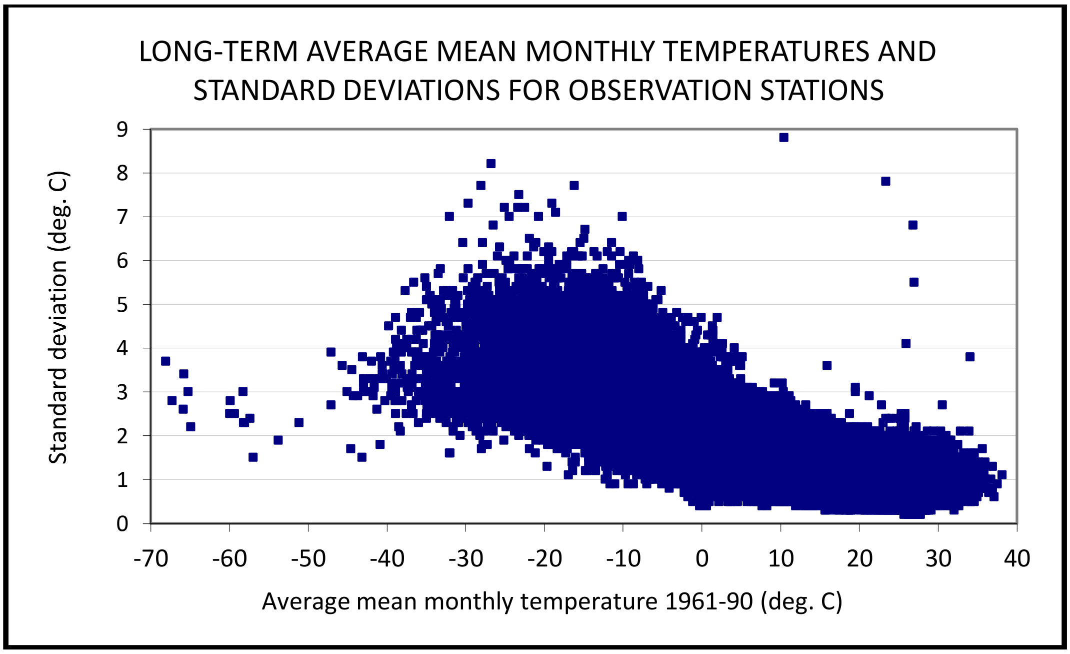


Figure 2.2 Long-term average mean monthly temperatures and standard deviations for each calendar month for each observation station used to construct the HadCRUT4 dataset.

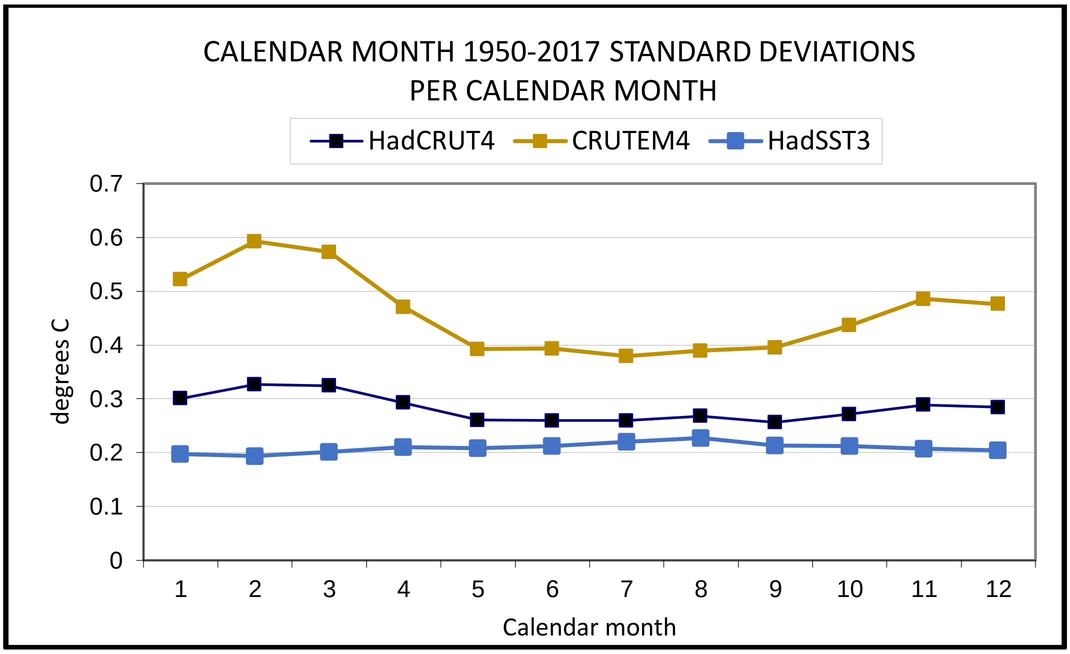


Figure 2.3 The standard deviations in each calendar month according to data from the three datasets based on all monthly data from 1950 to 2017.

3 Issues with data management, processing and concepts

3.1 Introduction

The HadCRUT4, HadSST3 and CRUTEM4 datasets suffer from a variety of problems with data management, data acquisition and the general structure of how the data is processed.

A key issue in this is data quality. A brief review of the concepts might be useful. The term "data quality" has a variety of definitions depending on the business and purpose of the data. International standards such as ISO 8000 are useful in that they define names and meanings but the standards are very generalised because data is used in a very wide range of applications.

The key elements of data quality are (1) whether the data is fit for purpose, and (2) able to be verified. For digital data the characteristics of good data quality are[4]:

(a) Sufficiently comprehensive
(b) Meaningful and relevant
(c) Unambiguous
(d) Accurate and correct
(e) Completely described (so that software can be created)
(f) Consistent (in format, level of precision, use of special values etc.)
(g) Meets all requirements specific to the processing of the data
(h) Traceable (verifiable and able to be audited)
(i) If supplied by others then preferably verified by them, because they are best placed to correct any issues, and the data transmission verified.

And for the proper management of data

(j) The set of data is to be clearly and unambiguously identified
(k) Data ownership must be assigned to an individual or entity (team, department etc.).

Having the processing software reject data that fails to meet certain criteria is a common practice with the advantage of simplicity. The approach does however have two disadvantages. Firstly, it can be difficult to verify that the rejection process operates correctly under all circumstances because the software would need to include diagnostic output and the volume of such output is likely to be considerable. Secondly, any other software, whether created at the same establishment or not, that uses the same data will need to be identical in its filtering of the data and the rejection of unacceptable values.

3.2 Identification of specific releases of datasets

The versions of the HadCRUT, HadSST and CRUTEM datasets used in this study are 4.6.0.0, 3.1.1.0 and 4.6.0.0 respectively, these being the versions available on January 25, 2018.

[4] This list is based on a number of sources including my 40 years' experience in Information Technology.

According to the website of the UK Met Office these versions of the datasets were first published in September 2016, August 2014 and August 2017 respectively, the names having remained the same since those months.

All three datasets are updated each month with the latest data and any corrections to earlier data, but the dataset name is unchanged, which means there is no simple method of associating a dataset with the month in which it was published. Users must either rely on the creation date of the computer file they downloaded or read the entire file and report year and month of the last data found in it. This is in breach of item (j) of the characteristics of good data quality noted above.

Finding 1 - The HadCRUT, HadSST and CRUTEM datasets lack clear identification as to when they were created (or published) and the period of data that they contain. The file names should clearly identify when they were created and information describing the contents be available and particularly note the changes since the previous release.

3.3 Inconsistent data file formatting

In the ASCII versions of the files we find that HadSST3 and HadCRUT4 use the same format with each month's data having a heading record that includes Month and Year in that order, with a generous eight characters allowed for each, and missing values indicated by -99.99. In contrast the CRUTEM4 dataset has Year then Month, each of only six characters each, and missing values shown as -1.000e+30. Likewise, the files that contain the number of observations (for HadSTT3) or number of reporting stations (for CRUTEM4), which are conceptually similar descriptions of the amount of source data, have different formats.

While the HadSST3 and CRUTEM4 datasets are created by different organisations it could reasonably be expected that the generally common purpose of the datasets would mean consistency between the two. This is an irritant rather than an issue but it illustrates poor co-ordination between the organisations.

Finding 2 - The three datasets are related so it would be logical to format them in identical fashion for ease of processing. A single piece of software could read and process each file if the formats were the same. In terms of data quality this would address the issue of consistency.

3.4 Management issues with sea surface temperature data

3.4.1 Uncertainty about the date of acquisition from ICOADS

The HadSST dataset is created from data in the International Comprehensive Ocean-Atmosphere Data Set (ICOADS) of meteorological and sea surface observations made from ships and buoys. The ICOADS database appears to be updated almost constantly so a thoroughly audit of the HadSST dataset would require knowing the date and time at which the ICOADS data was downloaded and a means of identifying the data in the ICOADS database that was present at that time. Neither the HadSST database or accompanying files provide the date and time of extract.

Finding 3 - The date and time of the download from ICOADS that was used to create the HadSST dataset is essential information for a thorough audit of the dataset but this is not provided. In terms of data quality, the description of the data is incomplete.

3.4.2 The source SST data is taken on trust and not checked

It will be shown in Chapter 8 that there are good reasons to question the accuracy of data in the ICOADS database, including some obvious errors such as the latitude and longitude of the sea temperature observations indicating that the measurements were taken more than 100 km inland from the coast.

The creators and maintainers of the ICOADS dataset are of course responsible for its accuracy but the Hadley Centre should have verified that it was error free before using it to create the HadSST dataset.

Finding 4 - The SST data has not been properly audited prior to the creation of the HadSST3 and HadCRUT4 datasets. The carrying forward of errors from a data supplier is contrary to good data quality practices.

3.4.3 The published SST data is poorly described and inconvenient

The principal HadSST3 dataset is different to earlier versions (e.g. HadSST2) in that each grid cell value is the median of 100 different versions of the dataset, each of which is created using certain assumptions. The different versions of the dataset are not described in a way that distinguishes one version from another.

Finding 5 - Clear descriptions sufficient for others to reproduce them, and therefore verify their accurate construction, are not available for the 100 different variants of the HadSST3 dataset.

3.4.4 HadSST3 variants indicate implausible SSTs

The Hadley Centre website provides the SST anomalies and climatologies for each of the 100 variants of the HadSST3 dataset, the latter being the supposed average sea surface temperature across the period from 1961 to 1990 for the given grid cell and calendar month. The commonly used version of HadSST3, the “ensemble” version” contains the median SST anomalies for each grid cell from each of the 100 variants. For this report a median version of the climatology was created and the mean monthly sea surface temperature for each grid cell derived by summing the median climatology and the ensemble version of the HadSST3 data for the given cell and calendar month.

Among the derived mean monthly sea surface temperatures were 2351 instances of less than -2.0°C and 424 of more than 32.5°C. Note that these are monthly mean values and that the individual temperature measurements were either exactly those amounts or a mixture of values above and below the mean.

At the average salt content of 35 parts per thousand (ppt) by weight, sea water freezes at -1.8°C. This figure falls by 0.28°C for each 5 ppt increase in salt content but the tropical ocean has a higher salt content that polar waters where freezing is far more likely.
A widely accepted reference for the coldest measurement of seawater in a liquid state is Sylte, Gudrun Urd (2010) in which a temperature of -2.6°C is reported, on or near the sea bed under the Antarctic’s Ronne ice shelf. Cold water will sink to the bottom and therefore the water above the point at which that coldest measurement was made would have been warmer, suggesting that the temperature at the sea surface cannot be that low.

The mean monthly SSTs derived from the HadSST3 data showed more than 2351 instances of sea surface temperatures being measured at supposedly below the freezing point of sea water. This number includes 367 instances of sea surface temperatures being less than -2.6°C, 109 of which were below -3.0°C. The lowest derived SST, at -4.88°C, was reported for a grid cell on Canada’s east coast. These extremely low values are obvious errors.

At the other end of the scale an SST of 32.5°C is about the point at which all incoming solar energy is used in the process of evaporation and sea surface temperatures no longer increase. Of the 424 derived SSTs of greater than 32.5°C a total of 375 were found to occur in the Red Sea or Persian Gulf, which are both recognised locations of excessively high SST and extreme rates of evaporation. These leaves 49 instances of mean monthly SST in excess of 32.5°C outside those regions, the highest of which was 36.01°C just off the east coast of India in April 2017. It is less certain that these extreme maximum temperatures are in error but they warrant closer investigation.

The figures presented above were derived from a “median climatology” developed for this analysis and the HadSST3 “ensemble” dataset. An investigation of the first 10 instances of HadSST3 variants datasets, using the relevant climatologies and SST anomaly files produced similar results, indicating that the above analysis is correct.

While the above data is derived from datasets and climatologies provided by the Hadley Centre and forming part of the composite collection of HadSST3 material, it has uncertainties. An investigation of an unusually high SST 36.01°C (off the east coast of India, April 2007) was derived from a median climatology of 29.15°C and a HadSST3 anomaly of 6.86°C. A check of the HadSST3 data source shows just two SST measurements for this grid cell in the month, both on the first of the month and both measured at the engine cooling system intake. One reading was 28.7°C and the other 30.0°C, which averages 29.35°C, just 0.2°C above the median climatology. It seems unlikely that an adjustment of almost 6.7C was made to the data but that is the most obvious reason for the discrepancy. The investigation of another of 35.72C (near Indonesia, July 2007), with climatology 28.85 and HadSST anomaly of 6.87C) revealed no SST measurements whatsoever for that grid cell and month, the data presumably extrapolated from neighbouring grid cells.

Clarification is required in relation to the above and for now it can only be said that there appears to be errors in the Hadley Centre's handling of this data.

Finding 6 - The technique of creating 100 variants of the HadSST3 dataset and then assembling a 'median' dataset appears to fail to recognize the physical limits of sea surface temperatures.

3.5 Management issues with observation station data

3.5.1 No clear identification of which stations were included

The CRU publishes a dataset containing the metadata and temperature data for 10,295 stations. The metadata for each station contains no explicit indication of whether the data for the station was included in the creation of a CRUTEM4 dataset. In total 2693 (26.2%) of the stations fail to meet the criteria for inclusion that are described in Jones et al (2012), specifically the need for average temperatures for at least 14 years from 1961-1990 and standard deviations (used to identify outliers) derived from at least 15 years of data from 1941-1990. A further three stations have invalid latitudes and longitudes (Lat. -99.9, Long. -199.9 or -999.9) but these are not explicitly mentioned by Jones et al (2012). While the software to create the CRUTEM4 dataset should exclude them, these instances should have been documented. (Most of the instances of latitude -99.9 and longitude -199.9 have flag values for long-term average temperatures or for standard deviations that will see them excluded, but three instances do not.)

According to Morice et al (2012) stations having less than 14 years of data during 1961-90 were included in processing for the HadCRUT4 dataset (cf. CRUTEM4 dataset) if long-term average temperatures for those stations were available from the WMO, but no mention is made of any alternative source for standard deviations

If no criteria were applied then 35 stations were included in HadCRUT4 that are not present in CRUTEM4. The 35 stations are spread across 24 grid cells, only seven of which have less than three other reporting stations. It is unclear why stations whose data is less certain, because no standard deviations are available for the identification of outliers, would be included when the grid cell already contains a substantial number of stations.

Ambiguity and doubt exist around the inclusion and exclusion of data for specific observation stations and this indicates poor data quality.

For the purposes of this analysis, in the absence of any statement from Morice et al (2012) it was assumed that stations without standard deviations were excluded from HadCRUT4 as per the description provided by Jones et al (2012) for CRUTEM4. As indicated above, this assumption might be incorrect.

Finding 7 - The composite set of observation station files published by the CRU includes data from stations that were not included in the processing to create the CRUTEM4 and HadCRUT4 datasets but these excluded stations are not clearly identified.

3.5.2 Inconsistencies and lack of clarity in station and country names

Metadata at the start of each station file shows a unique station number for each station, followed by latitude, longitude, the station name or location, the country and several other fields. Six stations have the name UNKNOW or UNKNOWN, three of which were presumably excluded because they either have invalid latitude and longitude or no long-term average temperatures. Another station has UNKNOWN as a country name.

The formatting of station and country names is inconsistent. Firstly, some names are entirely in upper case and others have only the first character in uppercase. Secondly the fields for station and country names usually blank-filled (i.e. filled with space character) but in 842 cases they are filled with the '-' character.

Alaska and Hawaii appear in the country name field, despite not being countries, and the names of 24 countries have a variety of forms (e.g. "UK" and "UNITED KINGDO", "USA" and "UNITED STATES" and four alternatives for Russia). Some are simply errors (e.g. "VENEZUALA" and "VENEZUELA", "HAWAII" and "HAWAAI", "BANGLADESH" and "BNGLADESH", and "LESOTHO" and "ESOTHO"). "REPUBLIC-OF-M" and "REPUBLIC-OF-K" are listed, resolving to Macedonia and South Korea respectively.

Inconsistent country names handicap the identification of stations specific to a given country and are contrary to the principles of good data quality. If simple meta information about observations stations is this sloppy then how much confidence can we have in the data itself?

Finding 8 - The CRU observation station data contains instances where the station cannot be identified by name.

Finding 9 - The CRU data for observation stations is inconsistent in its formatting of information, inconsistent in its country naming and sometimes clearly incorrect.

3.5.3 Suppliers of observation station data are not clearly identified

The station metadata has fields designated as "Source ID" and "Source file" but both appear to be from older versions of the CRUTEM4 dataset and neither provides the name of the national meteorological service (NMS) that supplied the data. This impedes raising question about the validity of the data and restricts access to upstream data (e.g. mean monthly maximum and minimum temperatures)

Unlike earlier versions of CRUTEM4 (and presumably HadCRUT4), the adjustment of data is now the responsibility of the NMS. Without an explicit indicator there is no certainty as to whether data has been adjusted or if so, when and how. Further, the data for the period 1961 to 1990 might have been adjusted and this would alter the long-term average temperatures and therefore every temperature anomaly for the observation station. Good data collection and handling practices should be transparent. Adjustments to the data should be clearly identified with reasons for the adjustment, the magnitude and direction of the adjustment and details of those who adjusted the data.

It will be shown later than several obvious errors exist in the station data and that there are other instances in which data is plausibly in error. It appears, but is not certain, that the CRU failed to question the data suppliers about these apparent errors.

From data quality perspective these issues indicate the inclusion of irrelevant information and the absence of meaningful information. The presence of obvious errors in the data indicates poor data quality and management by the suppliers of the data and therefore good reason to doubt the accuracy of the data.

Finding 10 - The supplier of each set of station data is not immediately obvious and yet the supplier is responsible for the accuracy and adjustment of the data submitted to the CRU for possible inclusion in the dataset. Any thorough audit of the data would require this information so that data could be validated against the supplier's records and so that questions might be raised with the supplier.

3.5.4 Other relevant station metadata is not supplied

As noted above the station metadata contains no information about data adjustments but this is not the only relevant data that is missing. It will be shown later (Chapter 9) that the World Meteorological Organization (WMO) assigns certain classes to observation stations according to the proximity of factors that might distort the measured temperature (e.g. shading by buildings or vegetation). Some of the classes assigned to poorly located stations have uncertainties due to siting, expressed in degrees Celsius, the worst of which is 5°C. The station class should be identified in the metadata, and if it has changed over time then the dates and relevant classes should be shown, all so that the appropriate mathematic uncertainties can be applied to the data.

Also absent from the station metadata is any certification that the temperature data was measured and processed according to WMO standards, and that the data has been checked and approved. There are instances where the data is unusual for the given month and might well have been derived from the weather over just a few days. There is also at least one instance where a sequence of low temperatures of about -15°C is interrupted by one value of 15°C, and an instance of where the mean temperatures in a certain calendar month are typically about 24°C but were reported twice as 0°C – both almost certainly errors.

Full details of station adjustments, particularly those involving station relations and possible changes of the WMO class, should also be provided if the data is to be properly audited or error margins properly calculated.

Finding 11 - The station metadata fails to provide details about the WMO station class, compliance with WMO standards and temperature adjustments but these are essential for proper audit and for the calculation of error margins.

3.6 Issues with the structure behind HadCRUT4

The fundamental concept of using a system of grid cells with data from observation stations and sea surface temperatures has three distinct flaws.

The first of these is that the processing of observation station data treats each grid cell in isolation from its neighbours. It is conceivable that data from stations in adjacent grid cells are more representative of the target cell. This might occur where a grid cell varies in its geography but observation stations are not distributed across the different regions, whereas stations in adjacent grid cells provide better coverage of those regions. A simple example of this is where one grid cell contains a reporting station near a coastline but none further inland whereas adjacent stations do have stations further inland.

The second is that the temperatures are measured close to the boundary of the medium in which they are measured and another medium with very different thermal properties. In the case of near-surface temperatures, the measurement is in air about 1.5 metres above the ground and in the other the measurement is supposedly taken within the top 500mm of the ocean (or if measured deeper in the water the temperatures are adjusted accordingly). In both cases they are susceptible to influences of the other medium, such as air temperatures being impacted by surface vegetation and by ground moisture, or sea temperatures impacted by surface winds.

The third issue is that the land and sea temperatures, the method of their measurement, and their processing differ considerably (Table 3-1). The month to month variation in land-based temperature data is approximately eight times that of sea surface temperature, and the range of measured data is likewise very different. On this basis, different proportions of land and sea surface data would alter the HadCRUT4 global average temperature anomaly even if the temperature was held constant. In other scientific fields the merging of two very different sets of data would be unacceptable, especially when the range and variation of the data sets differs enormously.

Factor	Temperature on land	Sea surface temperature
Medium in which temperature is measured	Air	Water
Thermal capacity (i.e. energy required to heat unit volume by 1°C)	For air, 0.0003 J/mL/°C	For water, 1.00 J/mL/°C (i.e. 3300 times that of air)
Frequency of measurement	Minimum and maximum once per day	Intervals of 4 to 6 hours
Method	Minimum and maximum thermometers, initially mercury but increasingly electronic	Six methods of water sampling, plus hull mounted sensors, engine intakes and buoys
Range of temperatures	Approx. -70°C to 55°C (range of ~125°C)	Approx. -1.8°C to 32.5°C (range of ~35°C)
Altitude	1.3 to 2 metres above ground, and ground height will vary	Surface to ~3 metre depth, sea level changing with season and air pressure by very few metres.
Method recorded?	No but WMO standards apply	Not always, and assumptions required.
Consistent location?	Semi-permanent, usually fixed for extended periods.	No. Wind-driven ships according to wind. Powered ships follow designated routes but measurements made according to time rather than location.
Thermal layering requiring data adjustment?	No because instrument height fixed	Yes, when measuring at depth in calm conditions. Layering might even be carried over from previous day or even weeks
Susceptible to brief spikes?	Yes. Mercury thermometers respond more slowly than the newer electronic thermometers.	Very minor issue with subsea volcanoes.
Susceptible to wind?	Yes. Hot winds and cold winds will vary temperature.	Yes. Variable ocean overturning, which causes mixing of water layers and increased exposure to the atmosphere (which might mean warming or cooling). This overturning makes data adjustment complex.
Data adjusted?	Yes. For instrument, screening and siting changes, each of which might be constant for 10 or more years.	Yes. Data needs to be brought to a common base, which requires various assumptions and various adjustments.
Impacted by local environment?	Yes. Includes surface moisture, land-use changes, casting of shadows, blocking of winds and urbanisation.	Not in open ocean with homogenous surrounds but some impact when ships in port or near coasts.

Table 3-1 Comparison of measurement and influences on land and sea surface temperatures

Finding 12 - The grid cell system is at times constraining and data might not be representative of the entire cell.

Finding 13 - Temperatures are measured in thermally complex regions near the interfaces of two very different mediums (i.e. land-air and water-air).

Finding 14 - Land and sea temperatures are measured by different methods, according to different schedules, subject to different influences and have vastly different plausible ranges. The merging of such data into a single dataset would be unacceptable in most scientific fields.

3.7 Issues with station data using minimum daily temperature

The observation station data is expressed as the mean monthly temperature, which is the average of the average daily minimum temperature and the average daily maximum temperature. The former is particularly problematic.

On calm, cloud-free mornings the minimum temperature frequently occurs less than five minutes after sunrise, just before the moment at which the incoming solar radiation is equal to the outgoing radiation from the Earth's surface. At this time of day, the sun is only slightly above the horizon and there is a very real prospect of natural and man-made objects casting shadows across the observation station and the ground beneath it. Chapter 8 will show that the World Meteorological Organization considers shading to be important and uses angles to clear sky, i.e. above obstacles causing shadows, as a factor for determining the quality or "class" of observation station.

In urban environments early morning shadows are likely to be issues, but so too is coincident morning peak hour traffic, which for many cities seems to be starting earlier due to increased urban populations. Cities are becoming active earlier, meaning increased private and public transport and earlier start-up of air conditioning systems and other systems that generate heat. McKitrick & Nierenberg (2010) and McKitrick (2013) noted that socio-economic development appeared to influence recorded temperatures and this earlier urban activity might account for their observations.

Finding 15 – Daily minimum temperatures are subject to numerous external, non-meteorological influences and changes in these could easily account for some of the increase in HadCRUT4 temperature anomalies over time.

4 Issues with data coverage

4.1 Introduction

The area of the Earth for which temperature data is available is known as coverage. As we will see, with the HadCRUT4 dataset coverage is more of a concept than an absolute value and the coverage of the dataset has varied over time. At times coverage is so low or so heavily biased towards certain regions that the average temperature anomaly calculated from the data can hardly be regarded as *global* average.

4.2 Coverage is a factor of grid cell size

As chapter 2 mentioned, data coverage, as a percent of the Earth's surface, is based on the size of each grid cell, and that to take into account the reduction in size of the grid cells as latitudes increase towards the poles a weighting factor is applied.

Because coverage is determined by the presence of observations within a grid cell it follows that different grid cell size will alter the coverage. Simply using larger grid cells will increase the notional coverage, and using smaller grid cells will reduce it. This point is easily illustrated by considering a single observation station within cells of different sizes. Adding further stations will likely alter the coverage but it depends on how they are distributed across the grid cell. Another station close to the first could put both of them in the same small grid cell with no increase in coverage, but if that new station is some distance away it might be in another grid cell and the coverage be the total of the two cells. The measurement of sea surface temperatures is done according to a time schedule rather than location but the distribution of the changing locations will have a similar influence on coverage.

This is not easy to illustrate with the entire HadCRUT4 dataset because sea surface measurements are taken at different locations. It can however be shown for the observation stations of the CRUTEM4 dataset because those stations are in (almost) fixed locations. Figure 4.1 shows the annual average CRUTEM4 coverage derived from all stations that reported in each month for three different sizes of grid cell. (NB. This includes data from HadCRUT4 coastal and island grid cells because some stations are located in those grid cells.)

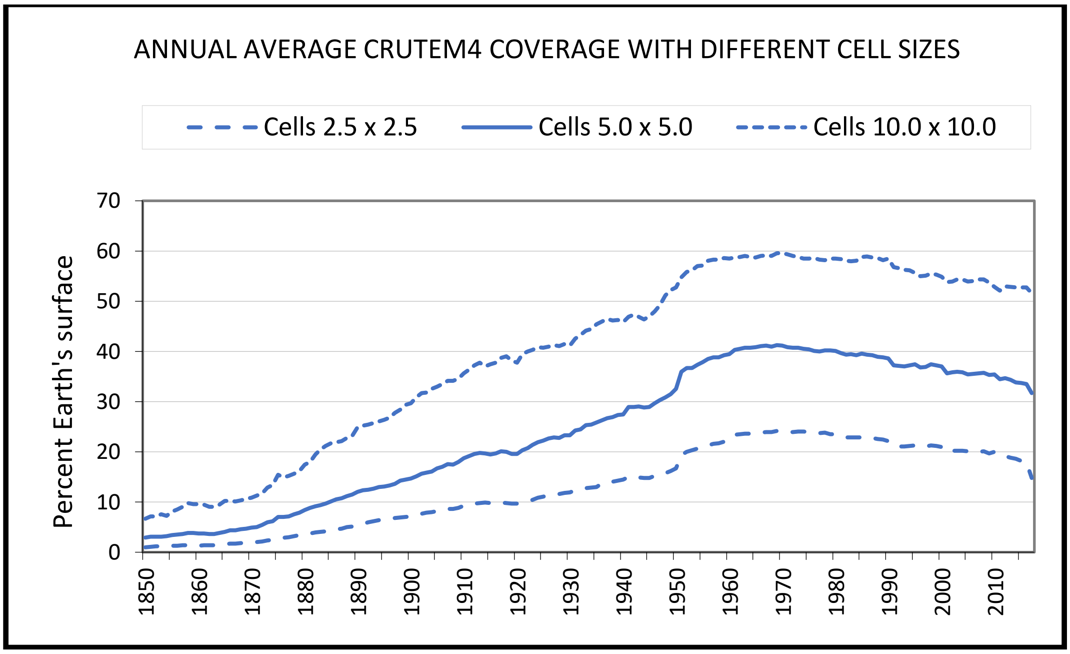


Figure 4.1 CRUTEM4 coverage under different grid cell sizes. The solid line indicates the coverage based on the standard CRUTEM4 and HAdCRUT4 grid cell size. (Sizes are in degrees latitude and longitude)

Finding 16 - The coverage according to HadCRUT data is not absolute but based on grid cell sizes and requires the existence of temperature measurements within the given grid cell and month. Coverage derived from the same data would differ if different cell sizes were used.

4.3 Variation in global and hemispheric coverage over time

Annual average global and hemispheric data coverage, calculated according to the HadCRUT4 method, has varied over time (Figure 4.2). At monthly rather than annual average level, over the entire period of the dataset to date, the global coverage has ranged from 12.2% to 91.3%, northern hemisphere coverage from 10.9% to 95.7% and southern hemisphere coverage from 11.0% to 89.6%. For HadCRUT4 the southern hemisphere monthly coverage fell to 23.8% in November 1918 and to 24.1% in September 1945, the latter being unusual because it meant greater coverage from observation stations than at sea in that hemisphere.

The annual average coverage exceeded 66.6% (i.e. two-thirds) in only three years prior to World War II, in 1935-1937. Southern hemisphere coverage exceeded 50% in 16 of the years prior to World War II but exceeded 55% in just four of them and 60% in one of them. It was not until 1948 that Southern hemisphere annual average coverage again exceeded 50% (at 50.2%) before consistently being above 55%.

Without coverage for a large proportion of the Earth it is not possible to know how the inclusion of temperature from those regions might have altered the HadCRUT4 global and hemispheric average temperature anomalies.

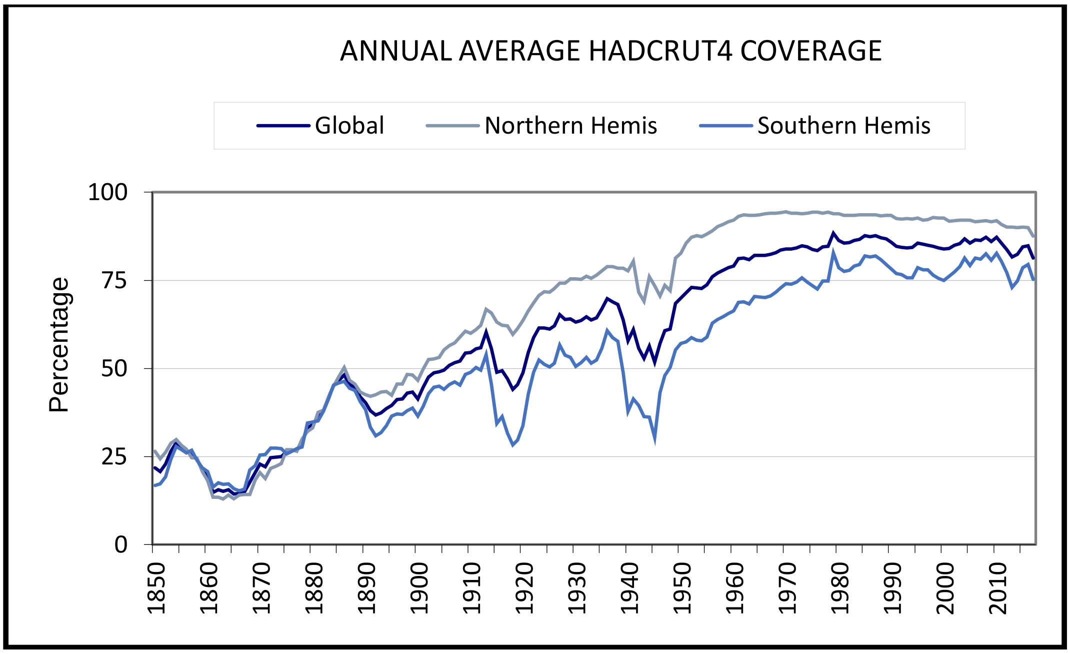


Figure 4.2 Annual average coverage globally and for both hemispheres. Global coverage is expressed as percentage of the Earth's surface and hemispheric coverage as a percentage of the hemisphere surface.

Finding 17 - Global coverage of the HadCRUT4 dataset has varied throughout the data record and falls short of even 50% for most of the first 100 years of the 168-year record. Such low coverage cannot be considered global unless the reporting data was relatively evenly distributed around the Earth, which it is not.

4.4 Spatial bias in coverage – latitude bands in hemispheres

Analysis of coverage by latitude bands and longitude bands (see Appendix 1) shows contributions to hemisphere coverage that are disproportionate regarding the area of the band relative to the total hemisphere surface area.

Finding 18 - Coverage over time has been far from homogenous. Coverage of the northern hemisphere reached homogeneity around 1950 but coverage of the southern hemisphere at the end of 2017 was still not homogenous.

4.5 Spatial bias in coverage - regions

Analysis of specific regions in both hemispheres, defined on the general basis of HadCRUT4 grid cells reporting data in at least 200 of the 360 months from 1850 to 1879 then followed by minor shaping, showed a percentage contribution to the hemisphere coverage far greater than the proportion of the area.

In the Northern Hemisphere a contiguous region covering Europe, the main shipping route south to the equator and west to the USA, plus part the US north-east (See Appendix 1 for mapping co-ordinates) covers just 12.1% of the NH. (Figure 4.3 suggests a higher percentage but this is due to the map's representation of the spherical Earth. It might help to know that 50% of hemisphere coverage is between 30N and the equator.) This region accounted for as much as 77.1% of NH coverage (in Dec. 1862). Annual average contributions to coverage are shown in Figure 4.4. In 82 of the 96 months from January 1861 to December 1868 the contribution to total NH coverage exceeded 60%. The emphasis on European temperatures at this time is important because the continent was in the process of recovering from the Little Ice Age that at times saw London's Thames River freeze over.

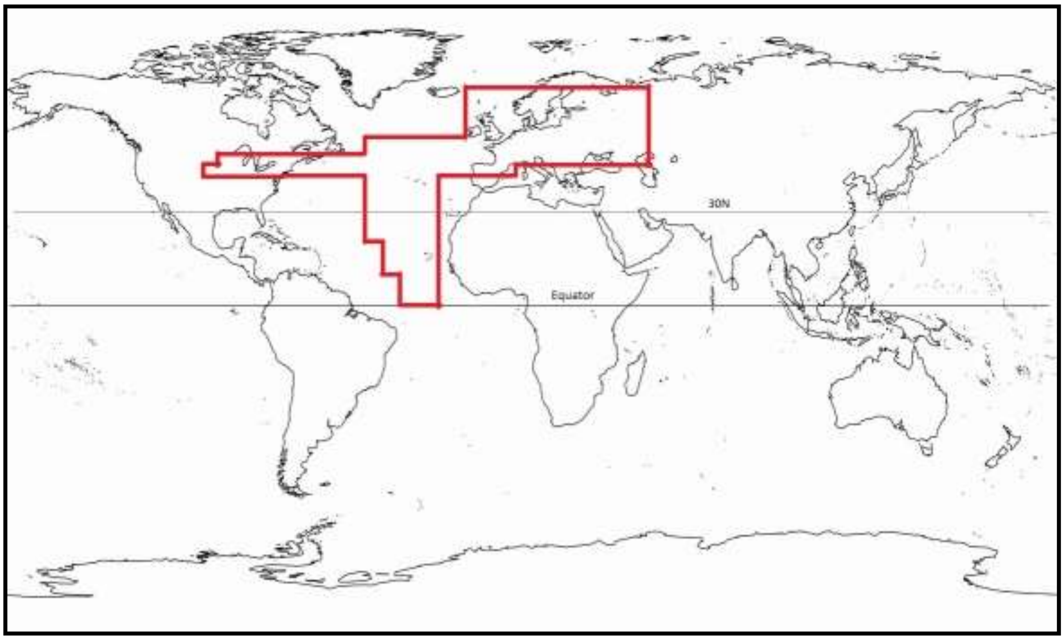


Figure 4.3 Contiguous Northern Hemisphere region containing grid cells with at least 200 months of data in the first 360 months (30 years) of the HadCRUT4 record.

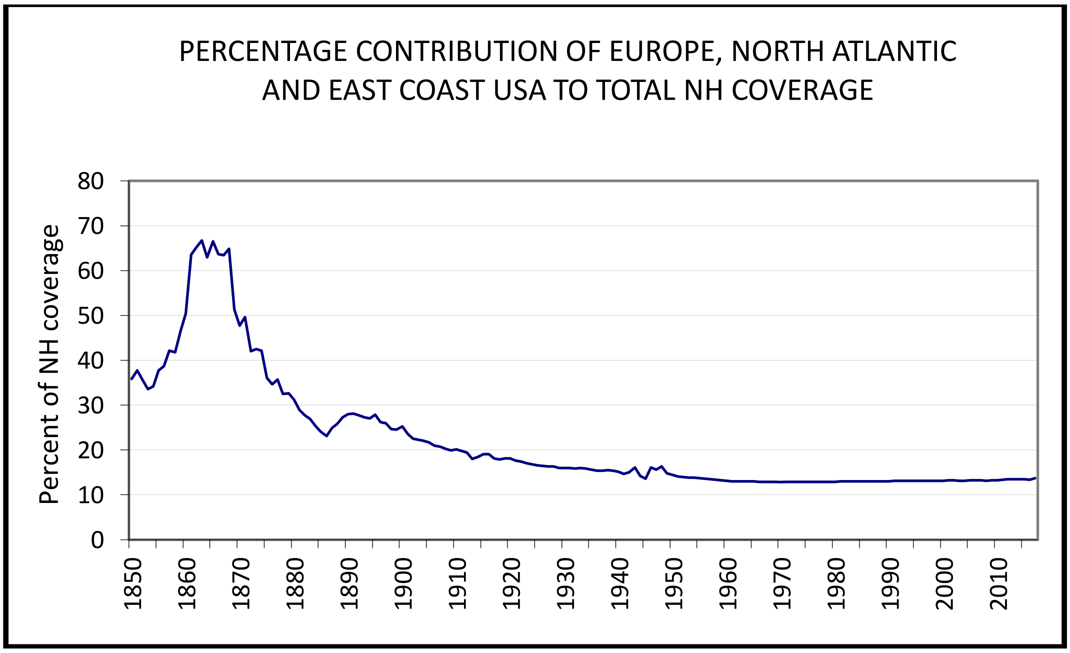


Figure 4.4 Annual average percentage contribution of the region in Figure 4.3 to total Northern Hemisphere coverage.

For the Southern Hemisphere, the region meeting the criteria described above corresponded to shipping routes from Europe into the South Atlantic and across the Indian Ocean, probably Dutch shipping to Indonesia and British shipping to south-east Asia (Figure 4.5). The area covered by the region accounts for 14% of the hemisphere but its contribution to SH coverage exceeded 80% in six months of the 1860s (five of them in 1866) and it exceeded 70% in 45 of the 96 months 1861-68 (Figure 4.6). The annual average contribution to coverage exceeded 60% in seven of those eight years.

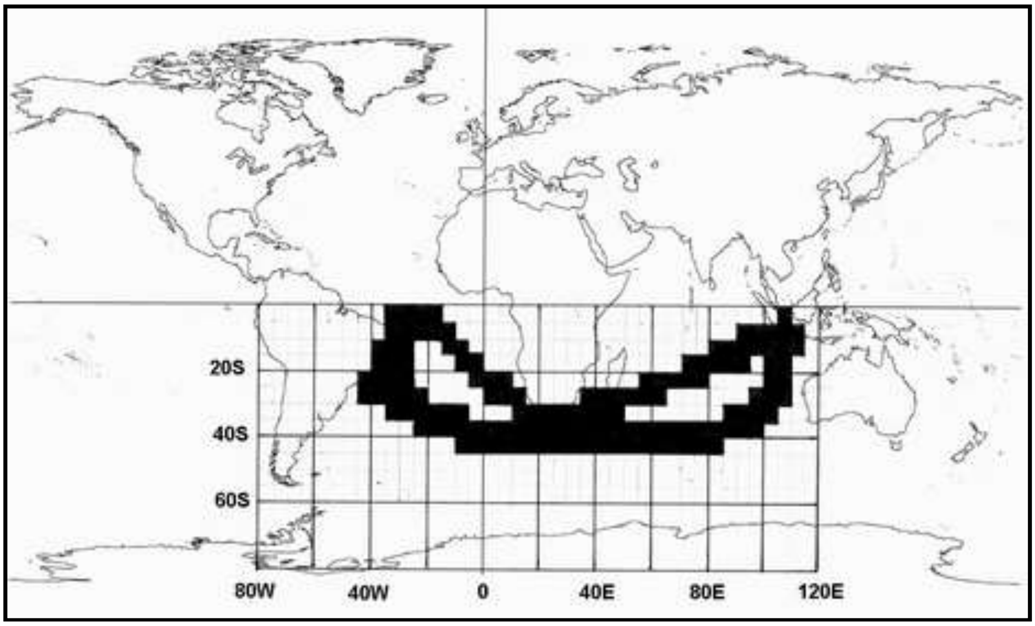


Figure 4.5 Region of the Southern Hemisphere with grid cells reporting data in at least 200 months of the 360 months from 1850 to 1879.

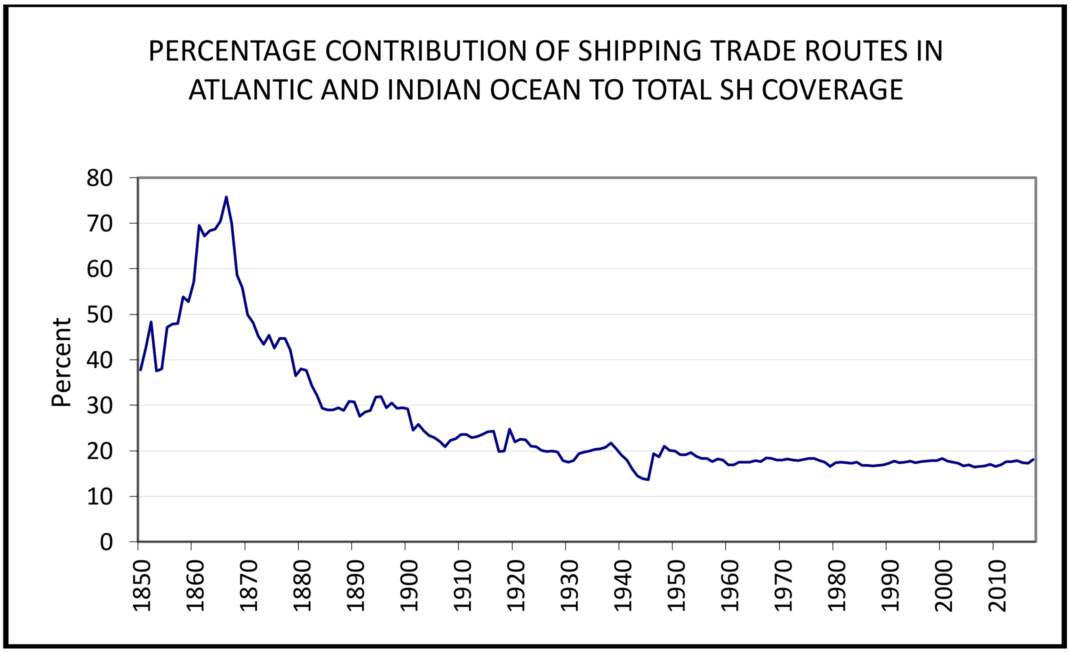


Figure 4.6 Annual average percentage contribution of the region in Figure 4.5 to total Southern Hemisphere coverage.

Taken together the regions described above for each hemisphere account for more than 60% of the total global coverage for most of the 1860s despite only amounting to about 13% of the Earth's surface. For both regions it was not until after 1900 that their contributions to hemispheric coverage fell to approximately correspond with the percentage of surface area that they covered.

In contrast, the coverage of the region from 150°E across the dateline to 75°W (i.e. the Pacific Ocean plus much of North America) and from 90°N to 90°S did not exceed 50% until 1922 (based on annual average coverage) and dipped below that level again during World War II (Figure 4.7). This is despite the region accounting for 40.3% of the Earth's surface.

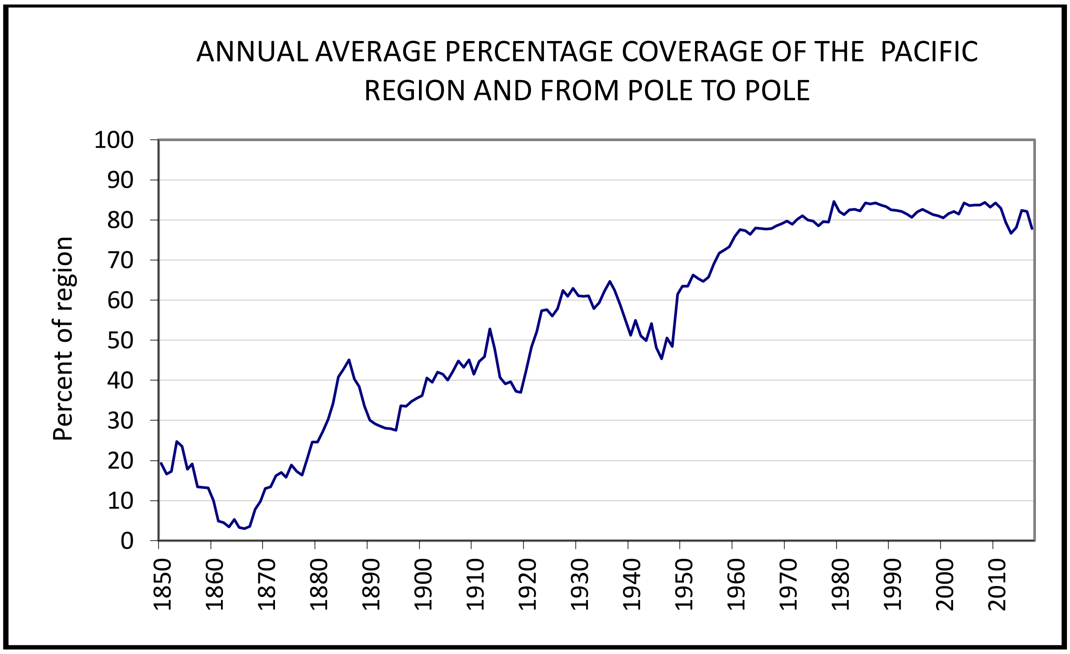


Figure 4.7 Percentage coverage of the Pacific Ocean longitudes (150°E across the dateline to 75°W) from pole to pole. El Nino and La Nina events in this region make it significant for global average temperature.

The situations described above show that the contribution that certain regions have made to hemispheric and global coverage has been greatly disproportionate to their area. This has inevitably resulted in hemispheric and global average temperature anomalies that are biased in favour of the anomalies in those regions.

Finding 19 - In the 1860s and 1870s certain regions of the world accounted for a far greater proportion of total coverage than their physical areas would suggest and coverage in other areas was very poor. So-called global average temperature anomalies at that time cannot be regarded as "global".

4.6 Variation in proportion of coverage from land and sea

Global average temperature anomalies for HadSST3 and CRUTEM4 data are not the same (Figure 4.8) and the month-to-month variation over land is 10 times the variation at sea. Taken together these mean that the HadCRUT4 global average anomalies could well be influenced by the proportions of land and sea coverage, especially at the start and end of HadCRUT4 data when their average temperatures differ.

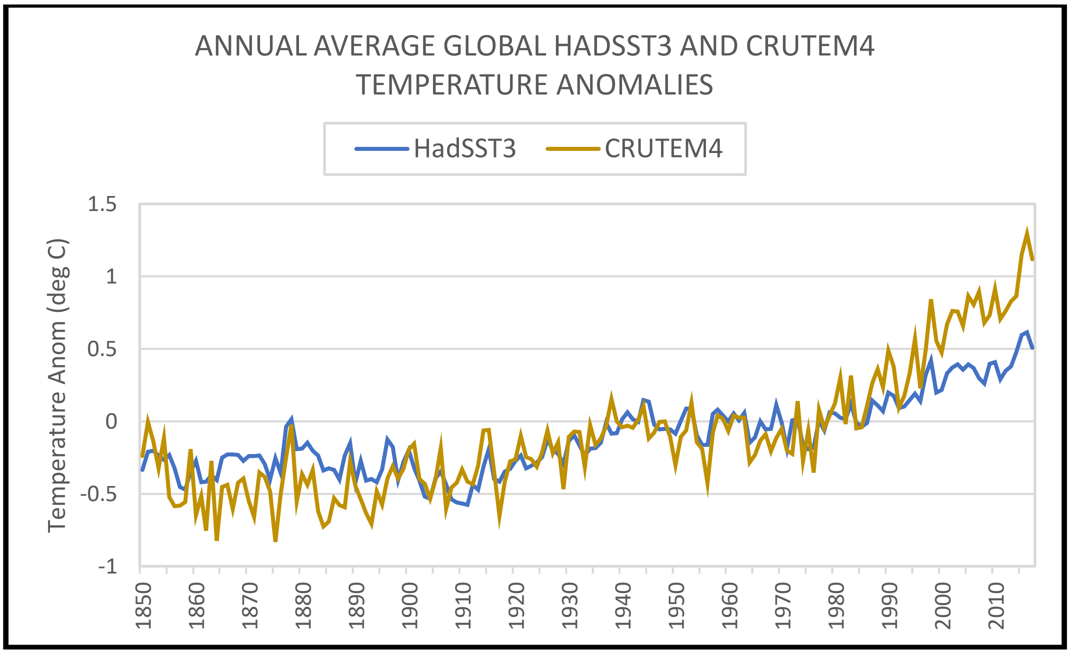


Figure 4.8 HadSST3 and CRUTEM4 Annual average global temperature anomalies

Coverage by sea surface temperature measurements (HadSST3) and observation stations (CRUTEM4) has varied in different manners over time with the latter showing an almost constant increase from 1850 until the 1960s and the former fluctuating over time, particularly during wartime periods (Figure 4.9).

Land accounts for 29% of the Earth's surface and water for 71%. The coverage of CRUTEM4 and HadSST3 data as a percentage the HadCRUT4 coverage has at times been very different to these values. Figure 4.10 shows the annual average coverage of CRUTEM4 and HadSST3 data expressed as a percentage of HadCRUT4 coverage. The two values sometimes sum to greater than 100% because some grid cells have both land and sea data. Annual average sea surface temperature coverage peaked at 90% of the total HadCRUT4 coverage in the mid-1850s and only for seven years in the 1940s did it fall below 70%. In contrast annual average CRUTEM4 coverage as a percentage of HadCRUT4 has fluctuated, peaking at 55.8% in 1945 and falling to 39% in 2017.

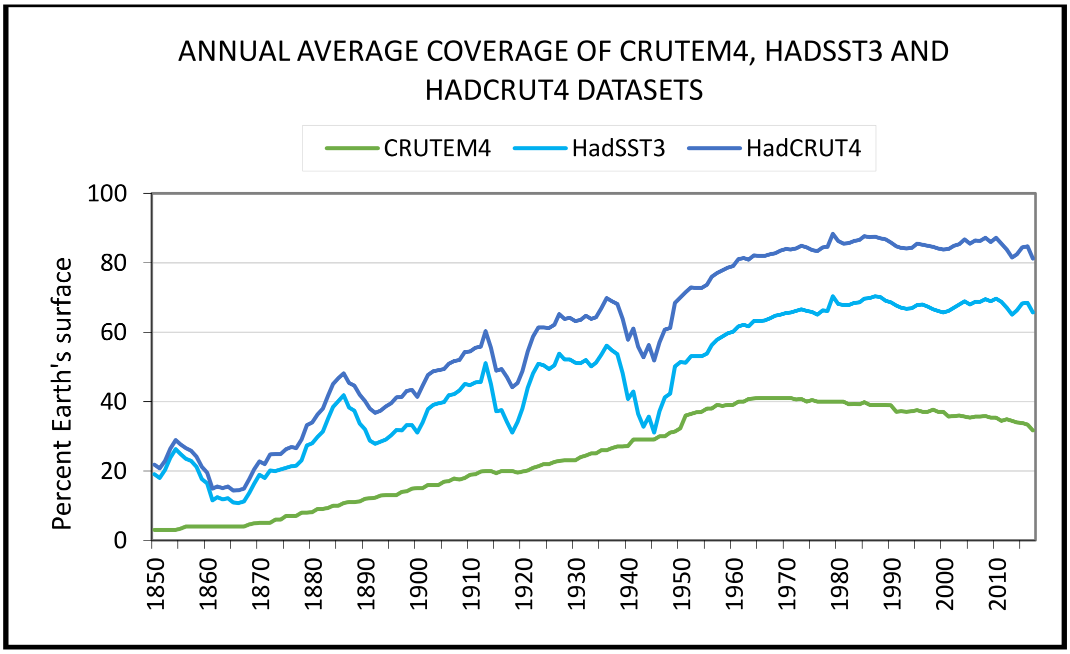


Figure 4.9 Annual average coverage of the three datasets over time

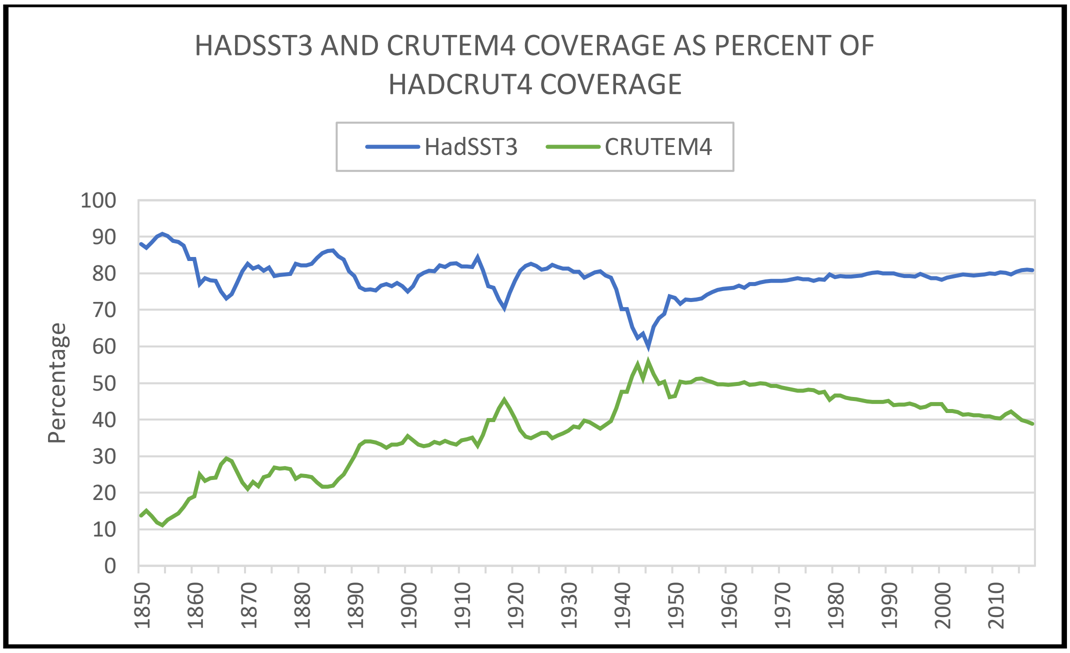


Figure 4.10 Annual average coverage of CRUTEM4 and HadSST3 data expressed as a percentage of the HadCRUT4 coverage at the time. The sum of the two values will sometimes exceed 100% because they might both report data for coastal/island grid cells.

Finding 20 - The proportions of data from land and sea has varied greatly over time and because temperatures over land fluctuate more than sea surface temperatures these changing proportions could account for some of the changes in the HadCRUT4 hemispheric and global average temperature anomalies.

4.7 Coverage and month-to-month variation in average temperature anomaly

Figure 4.11 shows annual average HadCRUT4 coverage and annual average absolute month-to-month variation in HadCRUT4 average global temperature anomaly. The notable feature of Figure 4.11 is the large month-to-month variation in average global temperature anomaly during the 1860s, when coverage was low. It was shown above that at this time much of the data was from a relatively small part of the Earth's surface. Temperature anomalies are very much a function of the weather in the regions from which data was obtained. Spikes in the month-to-month variation can be seen when coverage is reduced during the two World Wars.

The greatest month-to-month variation in the HadCRUT4 dataset global average temperature anomaly was 1.009°C, which occurred in January 1863 when coverage was 14%. In 11 months, all in the 1860s or 1870s, the month-to-month variation exceeded 0.5°C. Since 1961 the annual average coverage has always exceeded 80% and the annual average month-to-month variation has always been less than 0.12°C. In the 684 months (57 years) since 1961 the month-to-month variation exceeded 0.25°C on just 19 occasions and the coverage was below 80% in 18 months, not once below 75%.

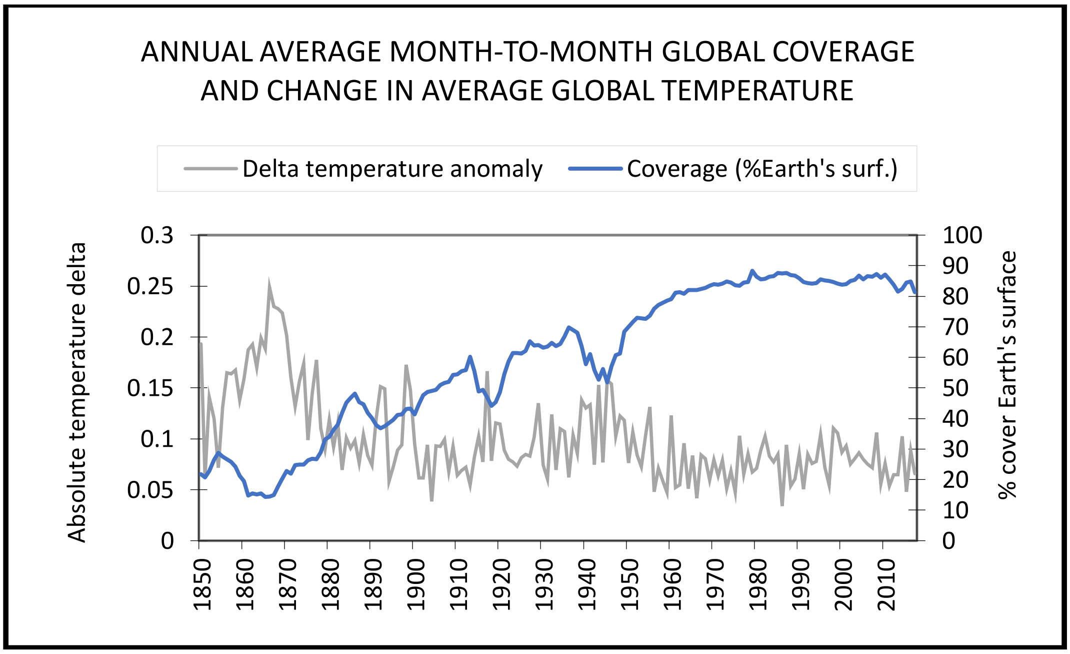


Figure 4.11 Coverage and the month-to-month variation in HadCRUT4 average global temperature anomaly

The relatively large variation in month-to-month HadCRUT4 average temperature anomalies during the first 100 years of the data record, when coverage was generally low, means that error margins are wide and the global average temperature anomalies have little credibility.

Finding 21 - Low coverage in the early years of the record coincides with greater month-to-month variability in average global temperature anomalies, a point repeated during World War I and World War II, indicating a wider error margin during periods of low coverage.

Finding 22 - The HadCRUT4 global average temperature anomalies for the first 100 years of data (i.e. 1850-1949) have little credibility because the error margins caused by low and biased coverage are substantial.

5 Issues with the amount of data

5.1 Introduction

The amount of reported data available for inclusion in the HadCRUT4 dataset is generally related to coverage but is different because changes in the amount of data might not be evenly distributed across grid cells.

5.2 Variable total number of SST observations

The annual average of the total number of SST observations in each month has varied over time by two orders of magnitude. The annual average was less than 10,000 for the first 30 years of the record (1850-1879), less than 100,000 until 1960 and less than 1,000,000 until 2007 (Figure 5.1). According to the HadSST3 dataset as it was in January 2018, the minimum annual average is 1,790 in 1851 and the maximum 1,622,680 in 2010.

While global coverage of SST data has increased from ~20% to ~69%, the increase in total observations has been far greater, indicating that, on average, the data for each HadSST3 grid cell is now derived from far more observations than used at earlier times in the data record. In January 1852 the average number of observations per HadSST3 grid cell was just 4.4, while it did not exceed 8 until June 1854.

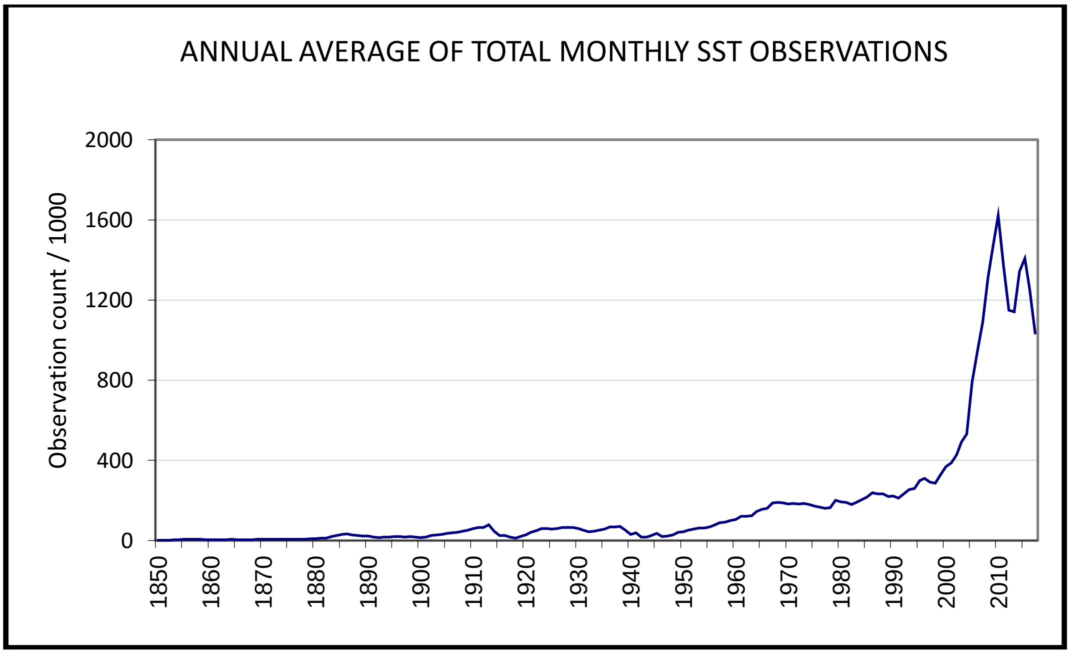


Figure 5.1 Average number of SST observations per month in HadSST3 dataset

Finding 23 - The total number of SST observations was very low in the early years of the record and has increased over time. The decadal average of monthly observations in the 1850s was

4020 and for the 2000s was 781,200, more than 190 times greater. This means that the HadSST3 dataset has been created from very inconsistent numbers of observations.

5.3 Few SST observations for many grid cells

It was mentioned earlier that HadSST3 sea surface temperature anomalies are derived from SST measurements that are first processed using 1° x 1° grid cells and pentads (5-day periods). There are six pentads for each month, except August which has seven. Extrapolation and interpolation as required to produce monthly averages for 5 x 5 grid cells. On this basis 150 SST measurements would be required each month to have just one measurement in each 1° x 1° grid cell in each pentad (175 in August).

Analysis of the HadSST3 observation counts reveals many grid cells with very low numbers of observations. The annual average percentage of HadSS3 grid cells with 15 observations or less accounted for ~50% of reporting grid cells prior to 1950, i.e. for the first 100 years of the data record (Figure 5.2). The percentages of HadSST3 grid cells with certain ranges of observations counts is shown in Supplementary Figure S5.1 (see Appendix 2). Two sub-ranges of 1 to 5 observations and 6 to 15 observations accounted for 20% to 30% of the reporting grid cells during that time, averaging 25.4 and 29.7% respectively across the 100 years, although the former accounted for more than 60% of all reporting grid cells in the early 1850s.

These very low numbers of observations would very likely have been by the passage of a single ship over just a few days and cannot be considered representative of the entire 5° x 5° grid cell across the entire month. (The grid cells that border the equator are about 555 km x 553 km.)

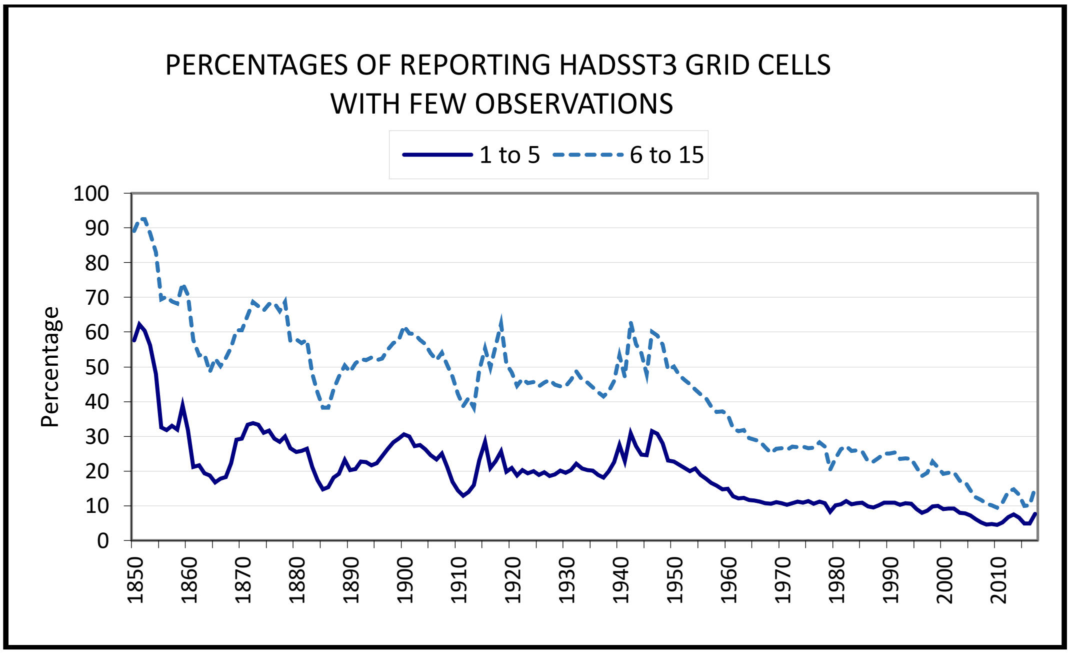


Figure 5.2 Stacked graph of the percentage of reporting HadSST3 grid cells with from 1 to 15 observations in the calendar month.

Table 5-1 shows the total number of instances of extreme HadSST grid cell temperature anomalies and the number and percentage of these that were derived from fewer than five SST observations. The unanswerable question is whether the inclusion of more observations from the same grid cell would have reduced the grid cell's temperature anomaly for the month.

Metric	Number of instances	Less than 5 observations	
		Count	Percent
Grid cell temperature anomalies ≤ -7.5°C	29	25	86%
Grid cell temperature anomalies ≥ 7.5°C	88	79	90%
Grid cell temperature anomalies < -5°C or > +5°C	3014	2055	68%

Table 5-1 Extreme SST anomalies and low numbers of observations in those grid cells during those months. The data shown here suggests a link between the two.

Finding 24 - The high proportion of grid cells with few SST observations casts doubt on the accuracy of any global, hemispheric or regional averages at least until 1950.

Finding 25 - Instances of from 1 to 5 observations for a grid cell across an entire month could all been recorded in just a few days by a single vessel travelling through the region covered by that grid cell and the data is not necessarily representative of the entire cell over the entire month.

Finding 26 - The low number of SST observation for coastal or island grid cells has two problems. Firstly, where no data from observation stations is available the data for the grid cell will be based on few observations. Secondly if observation station data is available the cell value will be a merging of data from one or more stations on land (~24 observations per station if WMO standards are followed) with from just 1 to 5 observations at sea.

5.4 Number of reporting observation stations

The number of reporting observation stations in each month has ranged from 146 (June 1850) to 7,340 (January 1975). The annual average number of reporting stations in each month is shown in Figure 5.3. The sharp decrease at the end of the series is likely due to data for some stations being reported every five years or even every decade. To make some allowance for this, the number of reporting stations in December in both 2009 and 2014 were 5,469 and 4,739 respectively, both well down on the maximum number of reporting stations.

The change in the number of reporting stations involves two possible scenarios: an increase in the number of stations within a grid cell that already contains one or more stations, or stations that now report in grid cells that previously had no reporting stations. The former does not increase the coverage but the latter does.

Comparing Figure 5.3 to the CRUTEM4 coverage in Figure 3.7 shows a generally similar pattern but the coverage does not have the sudden surge around 1890 when almost 900 stations, mainly in the USA, commenced reporting, presumably because many of them shared grid cells with each other.

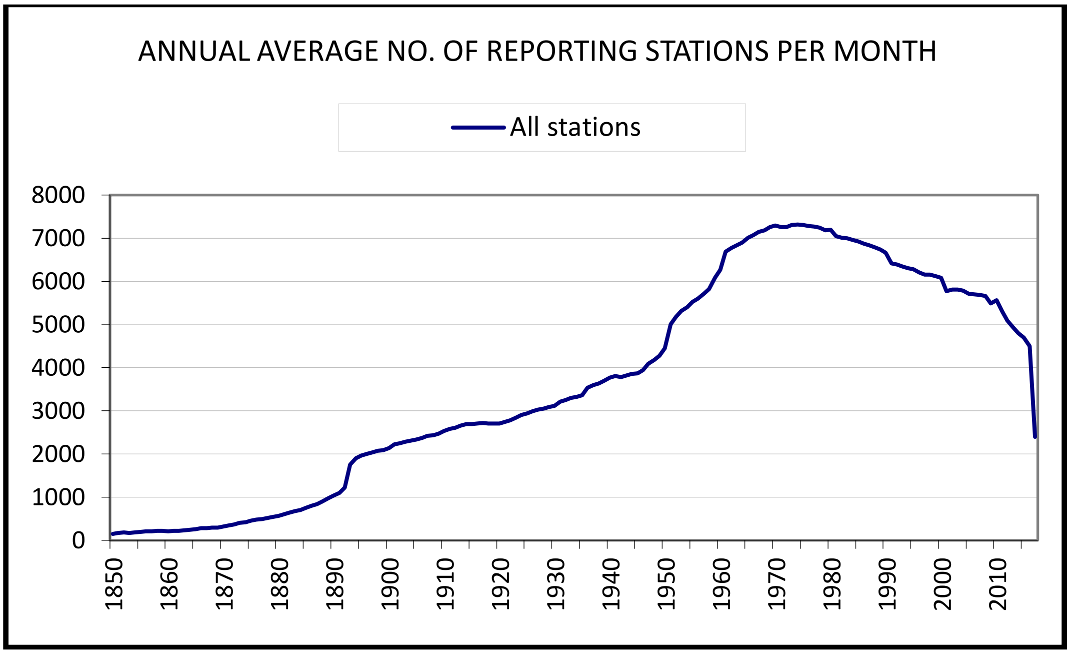


Figure 5.3 Annual average number of reporting observation stations

Finding 27 - The number of reporting observation stations has varied greatly across the record, this not only changing the number of grid cells for which data is reported but changing the number of reporting stations within the same grid cell. Both will almost certainly impact error margins and the accuracy of HadCRUT4 data.

5.5 Number of observation stations in the two hemispheres

The Northern Hemisphere has 67.3% of the Earth's land surface, the Southern Hemisphere just 32.7%. From 1912 to 2010 the number of reporting observation stations in the Northern Hemisphere was between 5 and 10 times the number in the Southern Hemisphere and in earlier years the disparity was even greater. Figure 5.4 shows the average annual number of reporting stations in each hemisphere, with separate vertical axes for each hemisphere because the figures are so different.

In the Southern Hemisphere (SH) a single observation station, at Mia Padang (Indonesia, ID: 961630), provided all of the hemisphere's data from January 1850 to June 1853, when a second station commenced operations. In March 1854 neither station reported temperature data over land (although elsewhere in the SH ships recorded sea surface temperatures).

By the end of the fifth year (i.e. 1854) only three SH observation stations were operating, one of which had only been in service for five months. By the end of first decade just 10 SH observation stations SH were reporting data and Table 5-2 shows the start of "good data" (i.e. supposedly reliable data) of SH stations during the first two decades.

In contrast, 147 observation stations in the Northern Hemisphere reported data at the start of 1850. The records of 179 stations indicate they were operating but 32 failed to report in that month. Of the 179 just 25 of these were not located either in Europe or along the east coast of the USA, and only 18 (10%) were located closer to the equator than latitude 35°N.

ID	Lat	Long	Location	Country	Year	Mon
961630	-0.8	-100.3	MIA PADANG	INDONESIA	1850	1
931190	-37	-174.8	AUCKLAND AERO	NEW ZEALAND	1853	6
619020	-8	14.5	ASCENSION IS.	ST. HELENA (BR)	1854	8
871200	-26.8	65.2	TUCUMAN AERO	ARGENTINA	1855	5
948680	-37.8	-145	MELBOURNE	AUSTRALIA	1855	5
875850	-34.6	58.5	BUENOS AIRES OBS CEN	ARGENTINA	1856	1
688177	-33.9	-18.5	CAPE TOWN /ROYAL OBS	SOUTH AFRICA	1857	1
946720	-34.9	-138.6	ADELAIDE	AUSTRALIA	1857	1
947730	-30.5	-151.7	ARMIDALE (UNI NEW EN	AUSTRALIA	1857	12
947670	-33.9	-151.2	SYDNEY	AUSTRALIA	1859	1
877500	-38.7	62.2	BAHIA BLANCA A	ARGENTINA	1860	1
855740	-33.5	70.7	SANTIAGOWAS_855770	CHILE	1861	1
948140	-35.3	-138.9	STRATHALBYN	AUSTRALIA	1861	1
948060	-35.1	-138.9	MOUNT BARKER	AUSTRALIA	1863	1
933090	-39	-174.1	NEW PLYMOUTH A	NEW ZEALAND	1864	1
938440	-46.4	-168.3	INVERCARGILL A	NEW ZEALAND	1865	1
948420	-38.9	-143.5	CAPE OTWAY	AUSTRALIA	1865	1
967450	-6.2	-106.8	JAKARTA/OBSERVATORY	INDONESIA	1866	1
917890	-21.1	175.2	NUKU'ALOFA	TONGA	1867	1
948690	-35.6	-145	DENILIQUIN	AUSTRALIA	1867	2

Table 5-2 Start months of observation stations in the Southern Hemisphere during the first two decades of the CRUTEM4 and HadCRUT4 datasets. (The metadata for station Pamplemousses, in Mauritius, says that 'good data' started in 1862 but the station's temperature data shows it reported from 1853 to 1861 and then failed to report until 1910.)

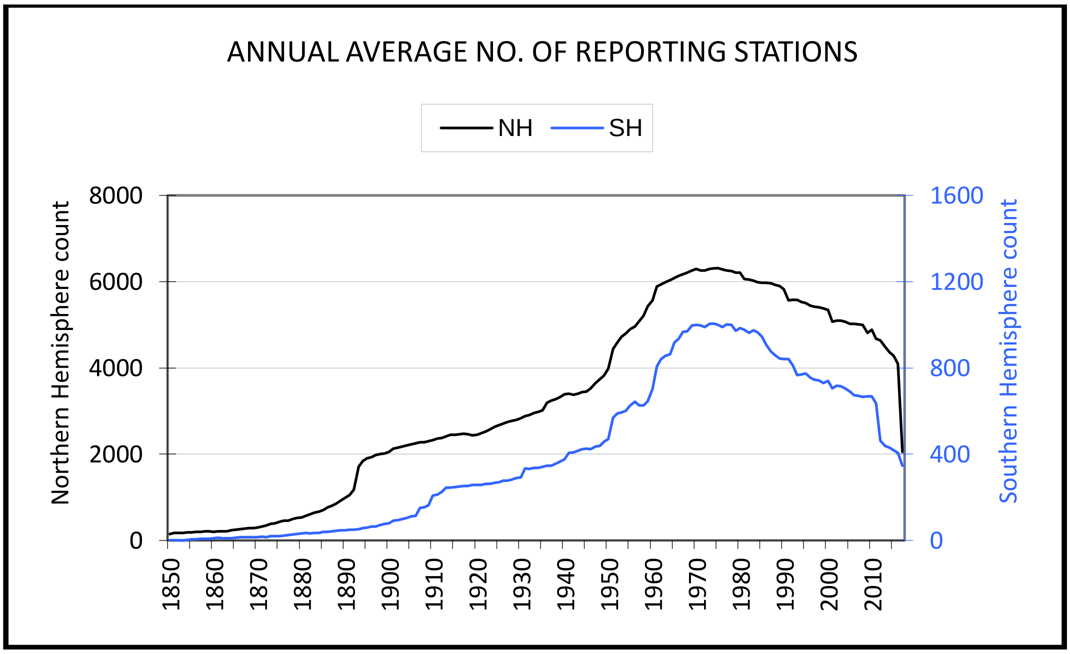


Figure 5.4 Annual average reporting observation stations in each hemisphere

Finding 28 - The very low number of reporting observation stations in the Southern Hemisphere in the first few decades of the record means a very wide error margin in the

CRUTEM4 and HadCRUT4 Southern Hemisphere and global average temperature anomalies during that time.

Finding 29 - In the Northern Hemisphere in the first few decades of the record the majority of data was from Europe and the east coast of North America. Hemispheric and global average temperature anomalies in the early part of the data record are therefore skewed towards those resulting from meteorological conditions in the eastern USA and in Europe.

5.6 CRUTEM4 grid cells with low station/observation counts

Figure 5.5 shows the annual average percentage of grid cells with a given number of reporting observation stations relative to the total number of CRUTEM4 grid cells that contain data for the month. It was not until ~1935 that grid cells with a single observation station accounted for less than 40% of all reporting grid cells, except for three years in the previous ten in which they accounted for between 39 and 40%. Further, it was not until 1951 that there were more than two observation stations in 50% of reporting grid cells.

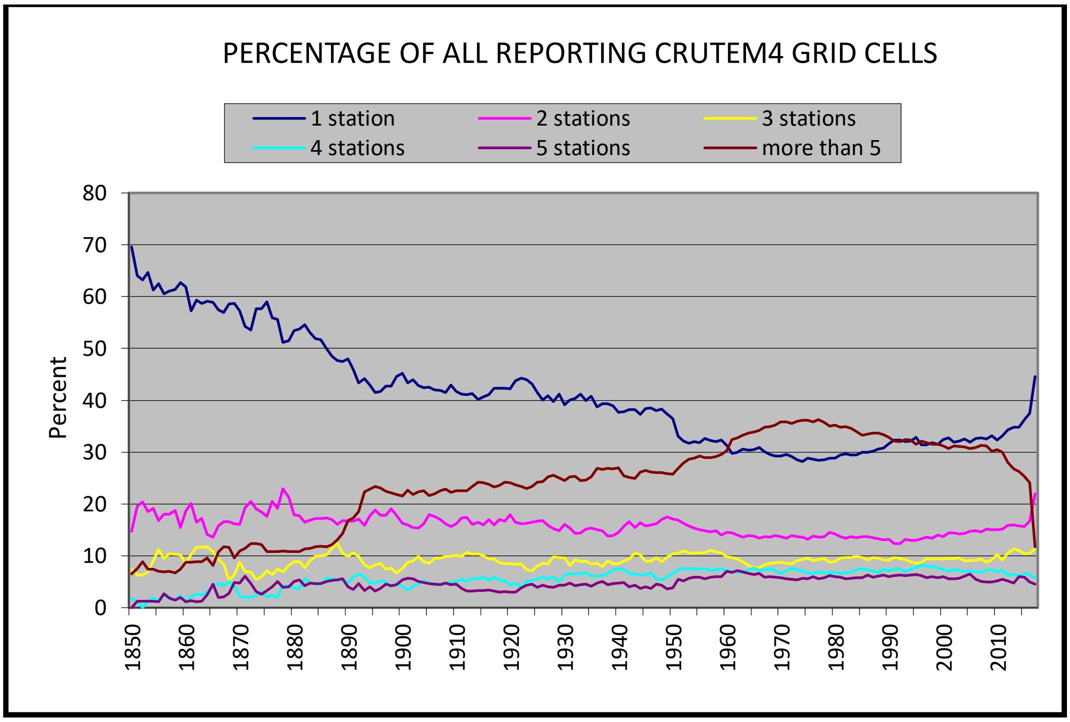


Figure 5.5 The number of grid cells with certain numbers of reporting stations as a percentage of all CRUTEM4 grid cells containing data for that month

Finding 30 - The CRUTEM4 grid cells with a single reporting station has only briefly fallen below 33% (i.e. one-third) of all reporting grid cells in the entire data record. The presence of very few observation stations in a CRUTEM4 or HadCRUT4 grid cell casts doubt as to whether the data is truly representative of the entire cell.

Finding 31 - When few (1 to 4) observation stations report data for a given grid cell any increase or decrease in the number of stations can mean relatively large shifts in the grid cell value because it is the average of the station temperature anomalies.

6 Issues with ' Normals'

6.1 Introduction

As mentioned in Chapter 2, the HadCRUT4 dataset is based on temperature anomalies, which are calculated according to

$$T_{anom} = T_{mon} - T_{base}$$

where T_{anom} is the temperature anomaly, T_{mon} is the mean temperature for a given month and T_{base} is the baseline temperature we are using for that calendar month, which is the long-term average for that month over the period from 1961 to 1990.

For observation station data the baseline temperatures are called Normals. They are calculated for every station before the grid cell value is calculated, which in turn is an average of all station anomalies. (It will be shown shortly that the period over which they are calculated is probably not meteorologically "Normal" in the general sense of the word.)

For the sea surface temperature data, the long-term average temperature is known as the "climatology". Like the rest of the HadSST3 dataset the climatology is based on 1° x 1° grid cells and the 5-day periods called pentads. The sea surface temperature data is always processed at this level and only in the last step are the anomalies converted to 5° x 5° grid cells and months. For both land and sea temperatures the baseline calculations use data from 1961 to 1990.

The accuracy of the Normals and climatologies is vital for the accurate calculation of temperature anomalies across the entire data record for the applicable calendar month.

Because Normals are calculated separately for each observation station and each calendar month this chapter will refer to "station-month combinations", for observation station data, and "cell-month combinations" for HadSST3 sea surface temperature. In both case there are 12 combinations, one for each calendar month.

6.2 The base period, 1961-90, is probably abnormal

The period from 1961 to 1990 was not homogenous and from what is known of the temperature record cannot be considered "normal" in the conventional sense of the word. Figure 6.2 shows this with the accumulated monthly HadCRUT4 global average temperature anomalies for the entire period. The accumulated values themselves are of no importance but the major turning points are. The graph has a general downward trend from January 1961 to January 1977, indicating a dominance of negative temperature anomalies. Next comes a relatively flat period when negative and positive anomalies were approximately balanced. After May 1979 the graph has a general upward trend, indicating a dominance of positive temperature anomalies.

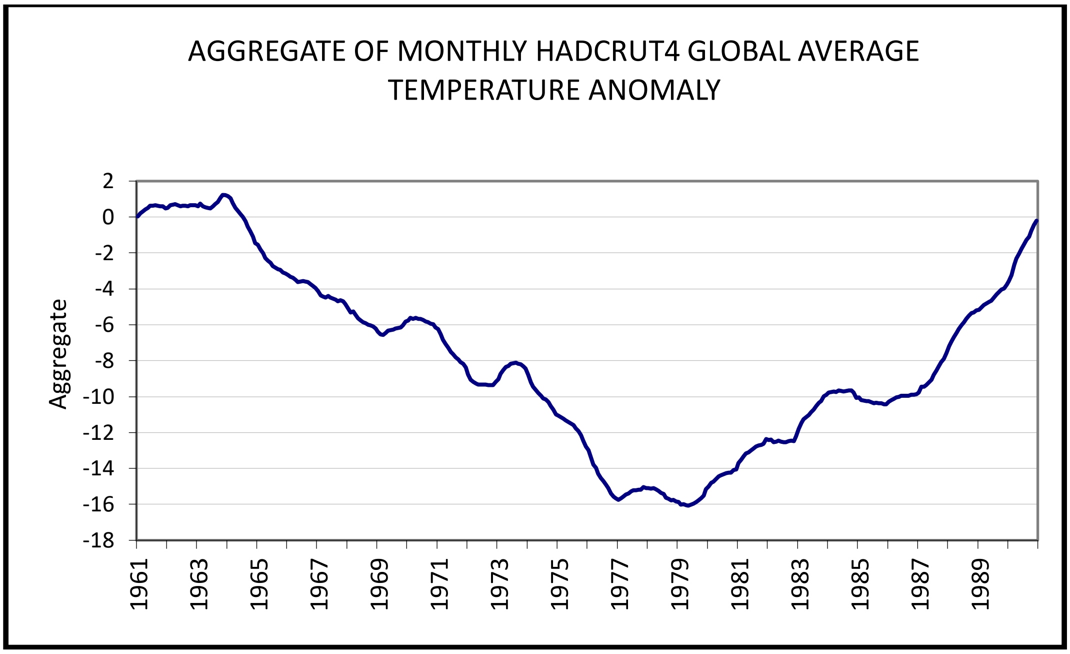


Figure 6.1 Aggregate HadCRUT4 monthly global average temperature

The pattern in the global average anomalies suggests on the balance of averages that if the 30 years of data was incomplete for some observation station or SST grid cell then missing data early in the period would probably omit negative temperature anomalies and late in the period would probably omit positive temperature anomalies. A check of the temperature pattern at every station would be needed to confirm whether this generalisation applied to the specific situation but the global average suggests this is likely.

Consider two stations 500 metres apart that report identical temperatures when they both report. If one is missing some data during even one month of the period from 1961 to 1990 then it is likely to have a different Normal to the other in that month and identical temperatures recorded at the two locations would produce different temperature anomalies. With different Normals the entire data record for that month at the two sites will be different despite identical temperature recordings when both record data.

If the two stations were the only stations in a grid cell then the cell value would be the average of the different anomalies from the two stations. If one of the stations failed to report in some month, either temporarily or permanently because the station was shutdown, the grid cell value would be the anomaly from the other station. The cell value is therefore changing but not for meteorological reasons.

Finding 32 - The period from 1961 to 1990 seems to be meteorologically abnormal with a distinct shift slightly after the mid-point of the period. Any failure to report data in a given month could distort the station Normal or the SST grid cell climatology and therefore distort every temperature anomaly for the calendar month in question.

6.3 Issues with observation station data

6.3.1 Distortion due to the adoption of daylight-saving

Because the "day" for meteorological observations ends at 9:00am local time, the adoption of daylight-saving (aka summer time) can distort an observation station's temperature record. Document WMO 100 (2011) ambiguously instructs meteorological observers "*If daylight-saving time is used for a part of the year, the observations should continue to be made according to the fixed local time; the dates when daylight-saving time commences and ends must be recorded*."

With the shift in local time there is an increased risk that the minimum temperature just after 9:00am (daylight-saving time) will be the minimum temperature until 9:00am the next day, or a risk that a lower minimum temperature will be carried over than would have been the case without daylight-saving. Figure 6.2 shows an example of the latter using hourly data from Australian observation station 'Melbourne (Olympic Park)' for March 27 and 28 in 2018. Daylight-saving was scheduled to end just a few days later, on April 1. Minimum and maximum temperatures are shown, along with those at 9:00am daylight-saving time and 10:00am (i.e. 9:00am if no daylight-saving). The minimum temperature for the 24 hours until 9:00am next day (i.e. March 28) was 11.3°C, that temperature being recorded at moments after 9:00am on March 27, but if daylight-saving was not in operation the minimum would have been 14.2°C.

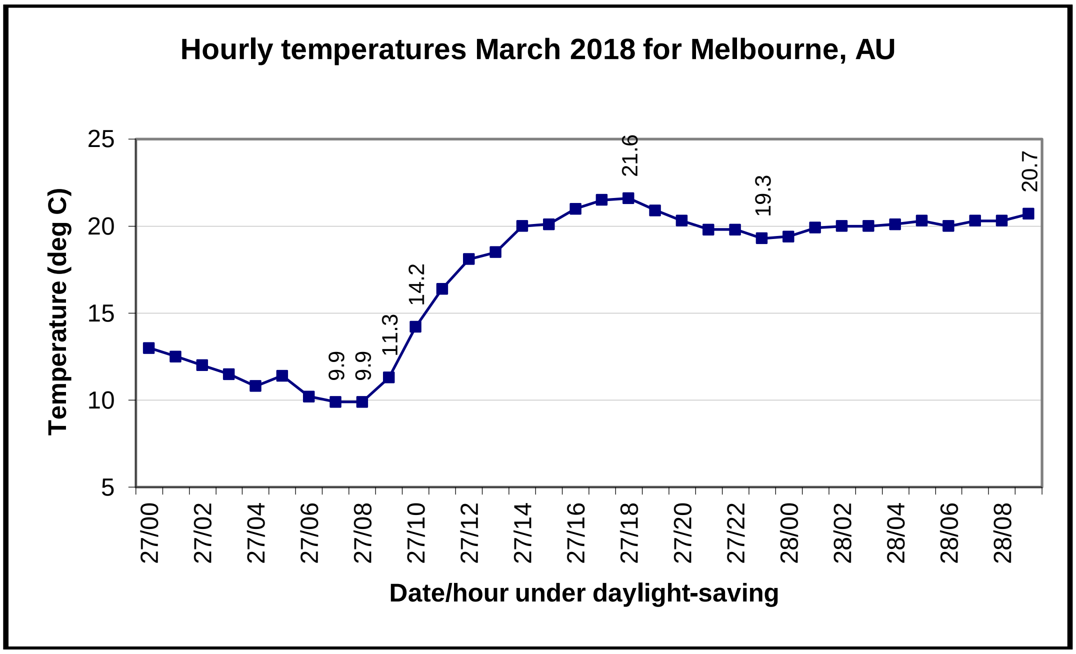


Figure 6.2 Hourly temperatures recorded at Melbourne (Olympic Park) that illustrate a distortion due to the adoption of daylight-saving. The values displayed are for the maximum, minimum, 9:00am and 10:00am recordings.

Carried over minimum temperatures result in lower minimum temperatures being recorded than would have been the case if minimum temperatures were, for example, only from midnight until 9:00am. Daily minimum and maximum temperatures are the basis of mean monthly temperatures so lower minimum temperatures due to daylight-saving mean lower mean monthly temperatures.

Many countries introduced daylight-saving in the 1970s and 1980s. The Normals for observation stations in those countries are likely to be derived from a mixture of monthly mean temperatures recorded with and without daylight-saving. If the entire 30 years of data from 1961 to 190 is not available then the distortion of the Normal could depend on whether daylight-savings was operating at the time. At grid cell level, which ignores country borders, the value could be the average of anomalies calculated from Normals that are inconsistent with respect to daylight-saving. Some countries have never adopted it and some countries adopted it during the period over which Normal are calculated. (Some Australian states have daylight-saving and some do not, moreover some that do have modified the period in which it applies. Few other countries, perhaps none, have these problems.)

Finding 33 - Station data is likely to show lower monthly mean minimum temperatures under daylight-saving than without daylight-saving because 9:00am temperatures are more likely to be the lowest temperature for the next 24 hours.

Finding 34 - The adoption of daylight-saving during the period 1961 to 1990 will likely cause distortion of some station Normals, especially for the months at the start and end of the period of daylight-saving. This has consequences for all temperature anomalies for the calendar month in question and therefore by extension consequences for CRUTEM4 and HadCRUT4 hemispheric and global average temperature anomalies.

6.3.2 More than one method of determining station Normals

The Normal for a given station and calendar month is calculated from the mean monthly temperatures for all instances of that month over the period from 1961 to 1990 but only if 14 or more years of data are available. (The CRUTEM4 documentation describes the minimum as years rather than years for individual calendar months.) When insufficient data is available other methods are used to derive station Normals.

Jones et al (2012) describes how a Normal might be obtained from either data recorded earlier for the same station or from the WMO. Osborn and Jones (2014) have more detail about this, listing the order of preference for estimating the Normal if one is not immediately available from the 1961-1990 data, saying …

> *1. If the station has sufficient data to estimate a Normal for the 1951–1970 period and the grid box from an earlier version of CRUTEM contains data across the longer 1951–1990 period, then we estimate the 1961–1990 Normal for the station using its 1951–1970 Normal adjusted by the difference between the grid-box averages (in the earlier version) for 1961–1990 and 1951–1970.*
>
> *2. If a neighbouring station does have sufficient values to determine its Normal, then we calculate the mean difference between the temperatures recorded at this neighbouring station and the temperature recorded at the current station over a different period when they both have data (e.g. 1951–1970) and assume that this mean difference still holds during the reference period. The Normal for the current station is then calculated as the sum of the Normal for the neighbouring station plus the mean difference between the two stations' temperatures.*

Method 1 is flawed because it can be shown that data used in CRUTEM3 was often inconsistent with data supplied by national meteorological services. A comparison of CRUTEM3 and NMS data for various observation stations in Russia, Norway, Iceland, Switzerland, Australia, New Zealand, Tahiti and the USA revealed differences in mean monthly temperatures and the months for which data was available. The CRU station data for Russian stations Kirensk, Tura and Bratsk differed from the Russian data sometimes in excess of 10°C and data for station Kyra differed by more than 1°C in 67 months of the 360 from January 1961 to December 1990 with just one year in which the monthly mean temperature was identical (see Appendix 3).

Method 2 involves several assumptions, in particular that there is a consistent relationship between the data from the two (or more) locations, which might not be true if the stations differ in their exposure to certain weather conditions.

Methods 1 and 2 described above both come under the general umbrella of using extrapolation to determine station Normals. The station metadata supplied by the CRU for the 113 stations flagged as using extrapolated data fails to detail which of the above two methods were used.

Neither Jones et al (2012) nor Osborn and Jones (2014) describes the method by which the WMO establishes Normals when there is insufficient data for the period 1961-90. As shown above, the global average temperature patterns changed in 1977 and changed again in 1979, so it is doubtful that earlier data would be sufficiently similar.

Finding 35 - Some station Normals are not derived from temperature recordings across the usual period of 1961-1990 and there are good reasons to question the estimated Normals derived by other methods.

6.3.3 Station standard deviations not calculated over same period as Normals

As mentioned in chapter 2, standard deviations are used to provide a threshold beyond which data is regarded as an "outlier" that is probably an error and is rejected from further data processing.

From a statistical perspective it is unusual to calculate a mean value using one set of data and the standard deviation (which indicates the spread of the data) over a different set but that's what is done with the station temperature data. The mean value is calculated from a minimum of 14 values across the period from 1961 to 1990, while the standard deviation is calculated from a minimum of 15 years of data across the period 1941 to 1990. The assumption that patterns in reported temperatures from 1941 to 1990 were similar to those from 1961 to 1990 is flawed and it fails to recognise that average global temperature anomalies increased after 1979.

Some differences between 1941-90 and 1961-90 standard deviations for stations that reported data in the given calendar month a minimum of 21 times during 1961-90 are shown in Table 6-1.

Stn ID	Name	Country	Mon	No years	Mean	61-90 StDev	41-90 StDev	change
152540	PALTINIS	ROMANIA	9	29	10.51	1.7	8.8	-7.1
442827	DAUUNMOD, CENTRAL	MONGOLIA	12	21	-22.1	2.3	4	-1.7
633330	COMBOLCHA	ETHIOPIA	3	30	19.14	0.9	2.2	-1.3
369520	Sarydzhas	KAZAKHSTAN	8	29	14.37	0.8	2	-1.2
306120	BALAGANSK	RUSSIA	12	30	-20.4	2.8	3.9	-1.1
357000	ATYRAU	KAZAKHSTAN	2	30	-7.96	3.6	4.7	-1.1
442410	BARUUNKHARAA	MONGOLIA	3	30	-8.3	3.0	4.1	-1.1
443040	UNDURKHAAN	MONGOLIA	12	30	-19.92	2.7	3.8	-1.1
170900	SIVAS	TURKEY	12	30	-0.46	2.2	3.2	-1.0
343980	FURMANOVO	KAZAKHSTAN	2	30	-11.23	4.0	5	-1.0
345610	VOLGOGRAD	RUSSIA	2	30	-7.48	3.6	4.6	-1.0
486250	IPOH_AERODOME	MALAYSIA	6	30	27.3	0.4	1.4	-1.0
716841	VILLE MARIE	CANADA	3	27	-6.81	2.1	3.1	-1.0
315380	SUTUR	RUSSIA	6	30	15.34	3.2	2.5	0.7
762200	TEMOSACHIC,CH	MEXICO	3	22	10.35	3.9	3.2	0.7
870460	JUJUY A	ARGENTINA	10	30	20.64	1.9	0.8	1.1

Table 6-1 Observations station with a large shift in standard deviation for the given calendar month if the standard deviation was calculated across 1961-1990. A change in standard deviation will shift the threshold for identifying outlying values.

Finding 36 - The use of different periods for the calculation of Normals and standard deviations makes the dubious assumption that the two periods are meteorologically similar and that the recorded temperatures will show similar patterns.

6.3.4 Outliers in observation station data during 1961-90

According to Jones et al (2012) data is considered to be an outlier if it exceeds five standard deviations from the mean. On the basis of the probabilities associated with a normal distribution this would exclude 1 in 1.6 million values, this despite each station being independent and having a maximum only 168 values for any calendar month from 1850 to 2017.

According to the station data file for Apto Uto (Colombia, ID:800890) in April, June and July of 1978, the mean monthly temperatures were 81.5°C, 83.4°C and 83.4°C respectively. These temperatures are clearly in error but they were included in the HadCRUT4 dataset. Other stations within the same grid cell reported data in those months and therefore the grid cell value is the average of all reporting stations, but the influence of the Apto Uto data is clear (Figure 6.3).

The false entries occur during 1961 to 1990 so they are included in both the Normals for that station and for the calculation of the standard deviations. The Normals for that station in most calendar month range between 24.0°C and 24.6°C because it is so close to the equator, but for the three months with bad data the Normals are 27.8°C, 27.9°C and 28.0°C respectively, with standard deviations of 11.9°, 11.8° and 12.0°C when in the other nine months of the year they are less than 0.75°C. The false entries distort the Normals and every anomaly in those same calendar months throughout the entire data record (1947 to 1988, but much missing) have negative temperature anomalies, which will also impact on the grid cell values because values are the average for all stations in the grid cell.

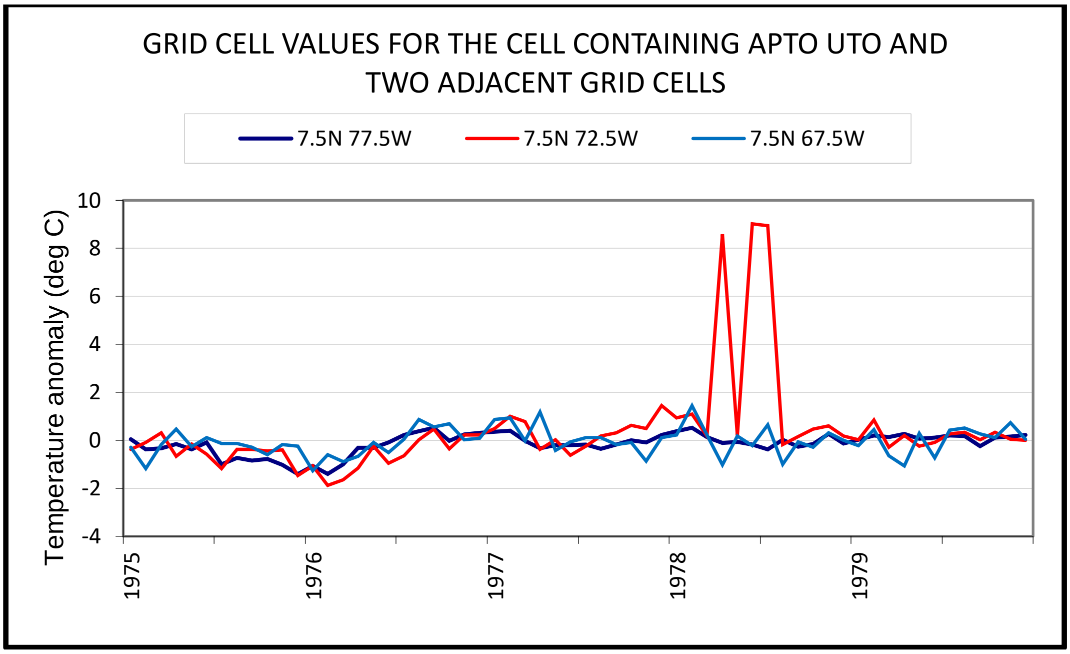


Figure 6.3 Monthly temperature anomalies for the grid cell containing station Apto Uto (in red) and the cells to the east (blue) and west (black) of it during the period 1975-1979

The station data was analysed to determine the change in the Normal for each station-month combination if the maximum mean monthly temperature during 1961-1990 was omitted. Including Apto Uto discussed above, 102 station-month combinations would see the Normal fall by 0.5°C or more, 10 of which would fall by 0.75°C or more. In a separate analysis the minimum of the mean monthly temperatures was omitted and it was found that 184 station-months would increase their Normal by 0.5°C or more, 9 of which were 0.75°C or more. Together these show that Normals can be sensitive to extreme values in the set of data.

Finding 37 - The presence of outliers in the data across the period 1961-1990 for some station-month combinations has distorted the Normals that are calculated from that data. Given that the Normal has implications for the identification of outliers during the rest of the data record from that location and is used in the calculation of every temperature anomaly for that station in that calendar month, the temperature anomalies for the station are likely to be distorted and so too the HadCRUT4 and CRUTEM4 grid cell values.

6.3.5 Failure to observe WMO standards regards minimum data

WMO 100 (2011) provides guidelines about the procedures and practices when determining Normals (i.e. long-term average temperatures). The document says "*As a guide, Normals or period averages should be calculated only when values are available for at least 80 per cent of the years of record, with no more than three consecutive missing years*." (pg4-17)

According to Jones et al (2012) the CRU derives its averages from a minimum of 14 years of data, which is less than 50% of the 30-year period. Despite the processing being performed on a monthly basis (i.e. average for calendar month and corresponding anomaly) the CRU minimum is for the number of years over which data was recorded, meaning that Normals might be determined from less than 14 values.

As noted above, calculating average temperature anomalies for a grid cell from anomalies based on Normals over different numbers of years can lead to non-meteorological distortion.

Finding 38 - The Normals for some observation stations are derived from far less data than the WMO standards specify as a minimum number of entries. The error margins in these cases, both for Normals and the subsequently calculated temperature anomalies, could be quite high.

6.4 Issues with SST data

6.4.1 SST averages not always based on measurements

SST baseline average temperatures used for calculating SST anomalies are created using a system of grid cells of 1° x 1° (latitude x longitude) and the 5-day periods known as pentads. (This is the usual method for dealing with sea surface temperatures right up to the final step which is to extrapolate and interpolate the spatial and temporal data into grid cells of 5° x 5° and time periods of one month.)

According to Kennedy et al (2012b) the SST average temperatures are estimated mathematically, presumably using principles of physics and geometry. If actual SST measurements are available for the grid cell and pentad during 1961-90 they are used to modify the estimated averages.

The assumptions behind the estimates are not clearly defined. The estimates might be incorrect because the region covered by the grid cell include a narrow particular cold or warm current. They might also be incorrect because assumptions about cloud cover and winds were false.

Finding 39 - Average sea surface temperatures for the period 1961-90 are based largely on estimates calculated according to assumptions about cloud cover, ocean surface turnover and other factors.

6.4.2 Some grid cells have little data during 1961-1990

If we consider each cell-month combination (i.e. cell and calendar month) we find that 20875 combinations reported data during the period from 1961 to 1990 and of these 11421 (54.7%) reported data in all 30 years. If we apply the same rather generous minimum of 14 years as for CRUTEM4 observation station data we find 16494 (79.0%) of cell-month combinations met the criteria, which leaves 4381 combinations that failed to meet that lower limit.

Of the those reporting in less than 14 years 1006 reported data in a single year, 512 in just two years, 360 in three years and then between 200 and 225 reported data in from 3 to 13 years. The largest block, 2426 of the 4381, were in the five 5°-latitude bands from 45° to 70°S, with a similar but smaller group north of 75°N. It will be shown later that some cell-month combinations reported no data during 1961-1990 and that the anomalies are based entirely on the estimated sea surface temperature.

The discussion above refers to HadSST3 grid cells of 5° x 5° and monthly data but the climatologies and most of the HadSST3 processing is conducted on 1° x 1° grid cells and using pentads rather than months.

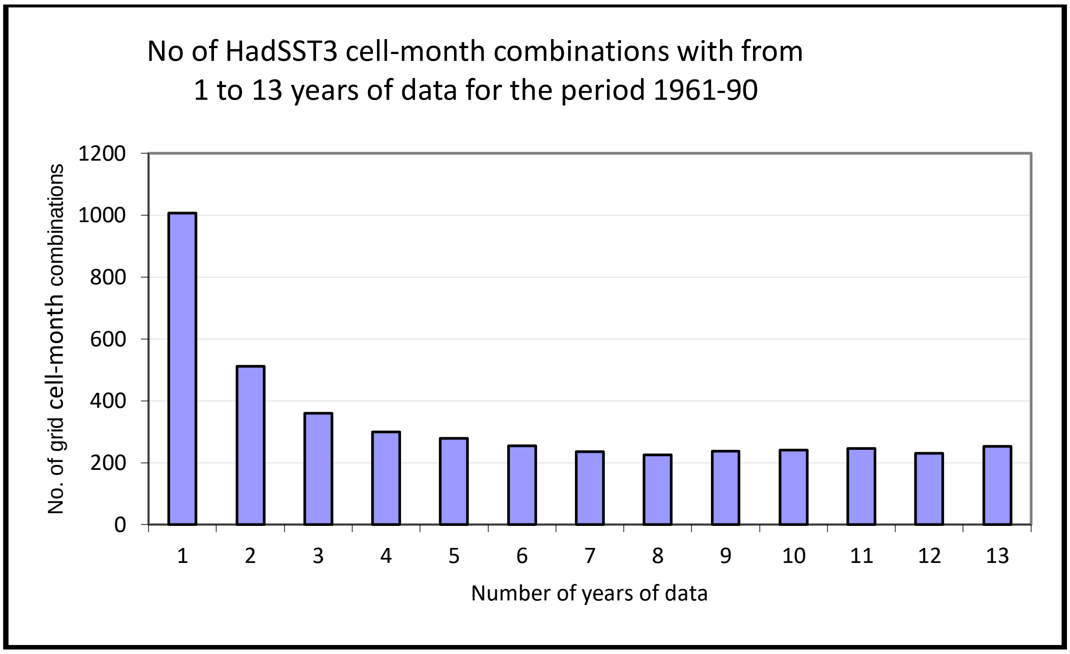


Figure 6.4 Number of HadSST3 cell-month combinations that fail to meet the CRUTEM4 minimum amount of data for observation stations

Finding 40 - Even with a generous minimum of 14 of 30 years, just over 20% of all cell-month combinations that reported in data during that period failed to meet the criteria used by CRUTEM4 for data from observation stations. When we also take into account the fact that climatologies are resolved on a 1° x 1° grid cells and pentads in many cases there was very little data available for adjusting the estimated average SSTs.

6.4.3 Reporting data is one thing but the amount of data is another

Cell-month combinations might have reported data in many or even most years of the period from 1961 to 1990 but that doesn't mean that many measurements were made in each month.

Of the 11241 cell-month combinations with 30 years of data for the period, 8782 (78%) did so with more than five SST measurements for the corresponding calendar month in every year. At the opposite end of the scale there is one cell-month where 25 of the 30 years of data were based on less than six SST measurements in the entire month. In regard to the 4381 cell-months with less than 14 years of data from 1991 to 1990, 942 of them did so with less than six SST measurements every time, including 529 of those that reported data in just one year.

Low numbers of observations mean that estimated average sea surface temperatures are modified according to relatively little data.

To investigate extreme values in the HadSST3 data all instances of grid cell SST anomalies the period 1961-1990 of ≥ 4°C or ≤ -4°C were identified. Of the 940 instances of grid cells in the group with the maximum ≥ 4°C, 417 (44.4%) had a single observation in the month of that maximum and a further 157 (16.7%) had just two observations in that month. Of the 398 grid cells with anomalies of ≤ -4°C 155 (38.9%) had a single observation and 67 (16.8%) had two observations.

Finding 41 - In many cases where data is reported for cell-month combinations during 1961-90 very few observations were made for the given month and extreme values of sea surface temperature anomalies were often based on very few SST observations for the month in which they occurred. No great confidence can be placed in the accuracy of SST average temperatures that are first estimated and then modified according to such small amounts of data.

6.4.4 Range of measured temperatures for grid cell-month combinations

The range of HadSST3 SST anomalies for each grid cell-month combination across the period 1961-1990 gives an indication of how sensitive the data might be to the inclusion of outliers. This was examined by considering the change in average SST anomaly for each cell-month combination when the maximum and minimum SST anomalies were excluded from calculations of the average across that time.

It was found that in 107 instances the exclusion of the maximum anomaly would decrease the average anomaly by 0.3°C or more, 28 of which were decreases of more than 0.4°C. The exclusion of the minimum anomaly over the same period would mean that 55 instances increased the average anomaly by 0.3°C or more, 12 of which were by more than 0.4°C.

Finding 42 - The SST measurements by which the estimated long-term average SSTs are adjusted are at times very variable in some calendar months and average measured SSTs in those calendar months come with wide standard deviations and therefore wide error margins, margins that should be passed along to the modified estimated averages.

6.4.5 Sum of grid cell SST anomalies 1961-1990 SST not equal to zero

If anomalies are calculated relative to an average over a period of time then it follows that the sum of the individual anomalies across that period should be zero. This is not the case with all SST cell-month combinations, which might be understandable when there is not data for all years of 1961-1990 but it occurs even when data is reported in all 30 years of the calendar month in the period.

The analysis of this data involved summing the anomalies for each HadSST3 grid cell across the period and dividing the result by the number of years of data. The result was named the "average discrepancy" and is the adjustment that would need to be applied to the cell-month anomaly in each year of the record in order to make the sum of the anomalies equal zero.

Some examples of extreme average discrepancies are shown in Table 6-2. At one extreme is an average discrepancy of -2.72°C, albeit for only 17 years of data. At the other extreme is a discrepancy of 3.22°C despite reporting data in 24 years of the 30. Positive discrepancies of more than 1°C occur in 63 of 16383 instances of non-zero discrepancies and negative discrepancies of more than 1°C in 49.

Month	Lat.	Long.	No of Years	Avg. Discrep.	Month	Lat.	Long.	No of Years	Avg. Discrep.
1	37.5	52.5	17	-2.719	2	42.5	52.5	19	-1.563
1	37.5	117.5	23	-2.293	12	37.5	-77.5	30	-1.550
1	37.5	47.5	14	-2.136	2	32.5	-87.5	29	-1.513
1	42.5	47.5	18	-2.122	1	37.5	-77.5	30	-1.485
2	37.5	117.5	22	-2.080		Positive discrepancies below			
2	37.5	52.5	18	-1.933	1	67.5	-32.5	27	2.009
1	32.5	-87.5	28	-1.841	7	72.5	-127.5	20	2.082
2	37.5	-77.5	30	-1.841	7	72.5	-22.5	19	2.137
12	37.5	117.5	20	-1.794	8	57.5	-77.5	19	2.146
8	67.5	-37.5	29	-1.751	7	82.5	12.5	19	2.150
12	42.5	47.5	19	-1.698	6	72.5	-17.5	16	2.169
1	42.5	52.5	17	-1.696	8	72.5	77.5	23	2.345
12	37.5	52.5	19	-1.677	7	67.5	-137.5	22	2.564
12	32.5	-87.5	30	-1.664	7	72.5	-132.5	22	2.593
3	37.5	117.5	20	-1.660	7	67.5	-132.5	24	3.220

Table 6-2 Extreme average discrepancies (i.e. sum of anomalies divided by number of years) for grid cells centred according to given latitude and longitude and in the given month.

Average discrepancies when data is reported for all 30 instances of the corresponding calendar month during 1961-1990 range from -1.84°C to 1.13°C, in fact 1776 (15.7%) of the 11325 cell-month combinations that reported 30 years of data have discrepancies of more than 0.1°C in either a positive or negative direction, the average of which was 0.22°C.

Finding 43 - The failure of the SST anomalies for many cell-month combinations to sum to zero casts doubt on the average sea surface temperatures from which those anomalies were calculated.

6.4.6 SST outliers during 1961-1990

Average sea surface temperatures are first estimated and then modified according to SST measurements. Analysis using the same approach as CRUTEM4 data will give some indication of the spread of temperature anomalies and identify those that exceed the limit for outliers. The CRUTEM4 approach uses the mean value calculated from a minimum of 14 years across the period 1961-1990 and the standard deviation calculated from at least 15 years of data over the period 1941-1990. The threshold for outliers is more than five standard deviations above or below the mean. The presence of very abnormal values distorted some CRUTEM4 standard deviations but the principle remains the same.

The results of this analysis were that there are 12 notional outliers in the period from 1961 to 1990 that are more than five standard deviations from the mean, all but three of them below the mean value rather than greater than the mean. The outliers range from -5.65 to +5.34 standard deviations. The probability of 5 standard deviations being exceeded is 1 in ~1.6 million so the probability of these extreme number of standard deviations in a maximum of 30 entries is extremely low.

Finding 44 - In several instances the calculated sea surface temperature anomalies for the period 1941 to 1990 seem improbably low. This might be due average sea surface temperatures being higher than they should be or could be due to very few SST measurements being made in the grid cell during the month and those measurements recording little more than the consequences of weather over just one or two days (see above in this chapter)

7 Other issues with the HadCRUT4 dataset

7.1 Introduction

Issues of coverage, amount of data and the calculation of Normals are the most significant areas of uncertainty, but uncertainties can also be found in several other aspects of the creation of the HadCRUT4 dataset. This chapter will discuss various other inconsistencies, uncertainties and apparent errors in that dataset and the associated two datasets.

7.2 Inconsistencies between the datasets

Datasets HadCRUT4 and HadSST3 use an ensemble approach, under which multiple versions of the datasets are created with different combinations of certain assumptions. The CRUTEM4 data is even artificially manipulated to create similar variation.

By making allowances for the nature of the ensemble datasets there should still be relative consistency between HadCRUT4 and the other two datasets whose data it uses. In particular the HadCRUT4 dataset should not contain data for a grid cell when neither HadSST3 nor CRUTEM4 have corresponding data (and vice versa), and when only one of the other datasets has data for the given grid cell and month then the HadCRUT4 value should be equivalent to it. Finally, when both CRUTEM4 and HadSST3 datasets have data for a given month and grid cell the HadCRUT4 value should not be outside the range of the two values.

(a) HadCRUT4 has data when neither other data set does

The HadCRUT4 grid cell centred at 22.5°N 82.5°E has data for September 2009, when neither the HadSST3 nor CRUTEM4 datasets contain data for that grid cell. The presence of this single example is a decrease from the two instances that could be found in the set of data files for January 2016.

(b) Failures to correspond with single other reporting dataset

The use of ensemble averages required an allowance for differences between datasets, which in this analysis meant that differences of less than 0.15°C were ignored. The analysis revealed 10,014 instances of differences between the HadCRUT4 dataset and the single reporting other dataset. No differences were found when HadSST3 data was present but not CRUTEM4 data. Of the 10,014 instances, 977 were of differences of 0.25°C or more, 301 were of 0.3°C or more and 43 of 0.4°C or more, and of that last group eight have occurred since year 2000. Positive difference s (i.e. HadCRUT4 data being the greater) were found in 6,515 (65%) instances and negative differences in on 3,499, indicating that HadCRUT4 might be exaggerating temperatures from observation stations.

(c) Outside the range defined by other two datasets

When grid cell data is available from both HadSST3 and CRUTEM4 datasets they will define a range in which the HadCRUT4 value should lie. The HadCRUT4 dataset has two instances that are 0.15°C or more outside the range of values defined by the other two datasets. One is where the HadCRUT4 grid cell has the value 0.06°C, HadSST3 has -0.24°C and CRUTEM4 -0.1°C, the other grid cell having 2.28°C, 2.62°C and 2.443°C respectively.

Reduce the threshold for data outside the range to 0.10°C and there are 36 instances of exceeding the range, 29 of which are from the same grid cell (centred at 7.5°N, 2.5°W, which is near the border of Ghana and Ivory Coast), with all of these instances occurring between January 1908 and May 1935.

Finding 45 - The HadCRUT4 dataset is sometimes inconsistent with the associated datasets HadSST3 and CRUTEM4 by either:

- Containing data when there is no corresponding data in either the HadSST3 or CRUTEM4 datasets,
- Differing excessively from the single other dataset reporting at the time (43 are different by more than 0.4°C), or
- When both HadSST3 and CRUTEM4 have data, the HadCRUT4 dataset sometimes has data whose value falls beyond the range defined by the values given in the other two datasets, at times by 0.15°C or more.

7.3 Different CRUTEM4 and HadSST3 global averages

Across the entire period of the datasets the annual average global temperature anomalies for CRUTEM4 have differed from those for HadSST3 (Figure 7.1). They were similar from 1900 to 1980, the difference rarely exceeding 0.2°C but at other times the difference has been greater, peaking at 0.68°C in 2016. Sometimes the average for HadCRUT4 has been closely aligned to the HadSST3 average but at other times it has been closer to the mid-point between the two.

The difference between the CRUTEM4 and HadSST3 annual global averages is shown in Figure 7.2. Explaining the reason for the pattern is one thing but the other is that the HadCRUT4 dataset and therefore its annual global average will be impacted by the relative coverage of the land and sea surface data, and not in a simple linear fashion.

Readers are also reminded that Chapter 2 mentioned the deliberate adjustment of sea surface temperatures to bring the HadSST3 global average SST anomaly into line with the global average anomaly from observation stations in the 1930s and 1940s. This prompts the question of whether the data should have been adjusted as much as it was and the possibility that the oceans take longer to warm and cool than does the land.

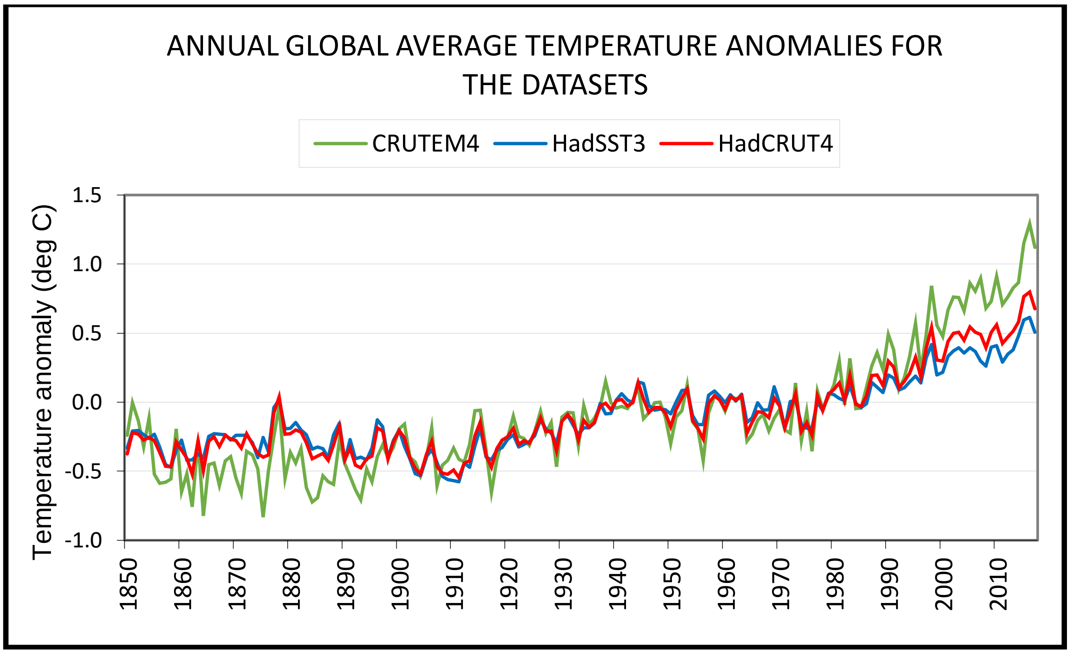


Figure 7.1 Annual average anomalies for CRUTEM4, HadSST3 and HadCRUT4 datasets

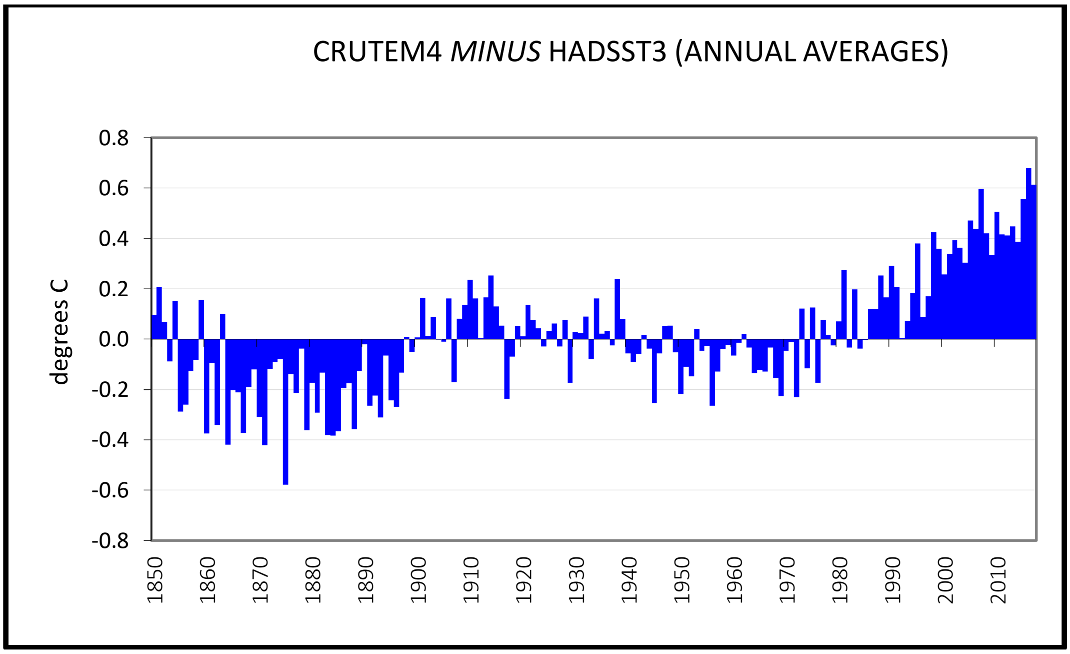


Figure 7.2 Difference between CRUTEM4 and HAdSST3 annual average temperature anomalies.

Finding 46 - Global average temperature anomalies for CRUTEM4 and HadSST3 show quite different patterns. It appears likely that the HadCRUT4 global average temperature anomaly at any time will be impacted by the proportion of coverage of the data of the two other datasets. The unanswered question is whether SST data should have been adjusted as much as it was for the period prior to 1946, because perhaps the oceans warm more slowly.

7.4 Large differences between data from adjacent grid cells

The homogeneity with adjacent grid cells was analysed by comparing the data in each grid cell to the surrounding eight grid cells when a 3 x 3 matrix of cells is used, with the target cell at the centre. The analysis was constrained by having no more than one corner cell without data, which meant that each target cell had either seven or eight neighbours.

In total 1,199 showed differences from the averages of their neighbours of 3°C or more and differences from the mathematically nearest of those neighbours of 2°C or more. Table 7-1 shows the most extreme cases of the difference ('DifToAvg') between the centre cell ('TargCell') from the average of the neighbouring cells ('AvgNbrs') and with large differences to the mathematically nearest of the surrounding seven or eight grid cells('DifToNearest'). The most extreme instance is of a grid cell value of -10.26°C when the average of the eight surrounding cells was 0.76°C, which a difference of 11.02°C from the average and 9.04°C from the mathematically nearest of its neighbours.

Year	Mon	Lat.	Long.	TargCell	AvgNbrs	DifToAvg	DifToNearest
1978	12	32.5	82.5	-10.26	0.764	-11.024	9.04
1932	11	-42.5	-62.5	-8.93	0.813	-9.743	8.53
1995	11	-47.5	-87.5	-6.94	0.827	-7.767	6.61
1879	1	37.5	-102.5	-7.32	-0.096	-7.224	4.00
1959	3	62.5	112.5	-1.99	5.177	-7.167	4.73
1985	5	-2.5	-147.5	-6.9	0.201	-7.101	6.48
1993	10	2.5	-117.7	-6.11	0.416	-6.526	5.63
1972	10	-42.5	52.5	-6.49	0.024	-6.514	5.38
1877	1	37.5	-102.5	-7.06	-0.551	-6.509	4.03
1953	10	-22.5	22.5	-6.39	0.090	-6.48	5.46
2016	11	-22.5	-162.5	-5.48	0.856	-6.336	4.58
1854	12	32.5	-82.5	-6.72	-0.386	-6.334	5.12
1942	3	37.5	-57.5	-6.35	-0.025	-6.325	5.96
1916	5	17.5	-87.5	-6.23	-0.296	-5.934	5.47
2006	12	-17.5	32.5	-3.97	1.569	-5.539	4.39
1853	1	42.5	-42.5	-5.29	0.244	-5.534	4.91
1938	3	-42.5	107.5	-5.98	-0.448	-5.533	4.47
1854	9	27.5	-32.5	-6.14	-0.663	-5.477	3.20
1998	8	-12.5	-27.5	5.31	0.234	5.076	4.19
1897	6	-47.5	-57.5	4.15	-1.025	5.175	4.53
2006	10	-22.5	-102.5	6.42	0.565	5.855	5.58
1936	2	22.5	-122.5	6.59	0.615	5.975	5.05
1912	7	-37.5	-42.5	5.94	-0.106	6.046	4.19
1888	5	52.5	27.5	5.15	-0.996	6.146	4.45
1922	6	32.5	-127.5	6.54	-0.681	7.221	6.63

Table 7-1 Examples of unusual differences between a grid cell and the seven or eight reporting cells that surround it. (Lat. and Long. are for grid cell centre.)

Finding 47 - The HadCRUT4 gridded dataset has some grids cell values that seem implausible when compared to the values in the surrounding grid cells.

7.5 Inconsistent source for coastal/island grid cells

As shown in chapter 2, coastal and island (C/I) grid cells account for about 28% of the Earth's surface, which is more than land-only grid cells. The total coverage by C/I cells has changed over time (min: 4.9% of Earth's surface in 1861, max: 26.3% 1980, (see Appendix 4), as has the percentage contribution that they make to the total global coverage in any given month.

Data for C/I grid cells can be anomalies from Observation stations on land, sea surface temperature anomalies or a merging of the two. Figure 7.3 shows the coverage of coastal/island grid cells from different sources over time. Given the ensemble nature of HadCRUT4 and HadSST3 data the analysis compared the HadCRUT4 data ±0.125°C to the other two datasets. It first tested for a match to CRUTEM4 data (land only in Fig 6.3), then to HadSST3 (sea only) and then falling within a range defined by HadSST3 and CRUTEM4 values (merged), the sequence of tests modified if either HadSST3 or CRUTEM4 data was missing. The category 'unknown' was assigned when the HadCRUT4 data could not be matched to any of the above, perhaps because one of CRUTEM4 and HadSST3 was missing and HadCRUT4 value ±0.125°C failed to match the data from the other dataset.

Figure 7.3 shows that sea surface temperature data was the dominant source for coastal/island grid cells until about 1900, with a peak contribution of 73% of the total coverage in 1854. From 1955 to 1981 it contributed less than 10%, except for 1977 (10.1%), with a minimum of 6% (Appendix 4, Figure A4.2).

Examination of coastal and island grid cells in each calendar month showed that an average of 84.4% of cells had a change of data source at some point across the record. Between 630 and 650 coastal and island grid cells reported data at some time in each calendar month and only from 13.3% to 18.6% of those grid cells consistently had the same source for their data.

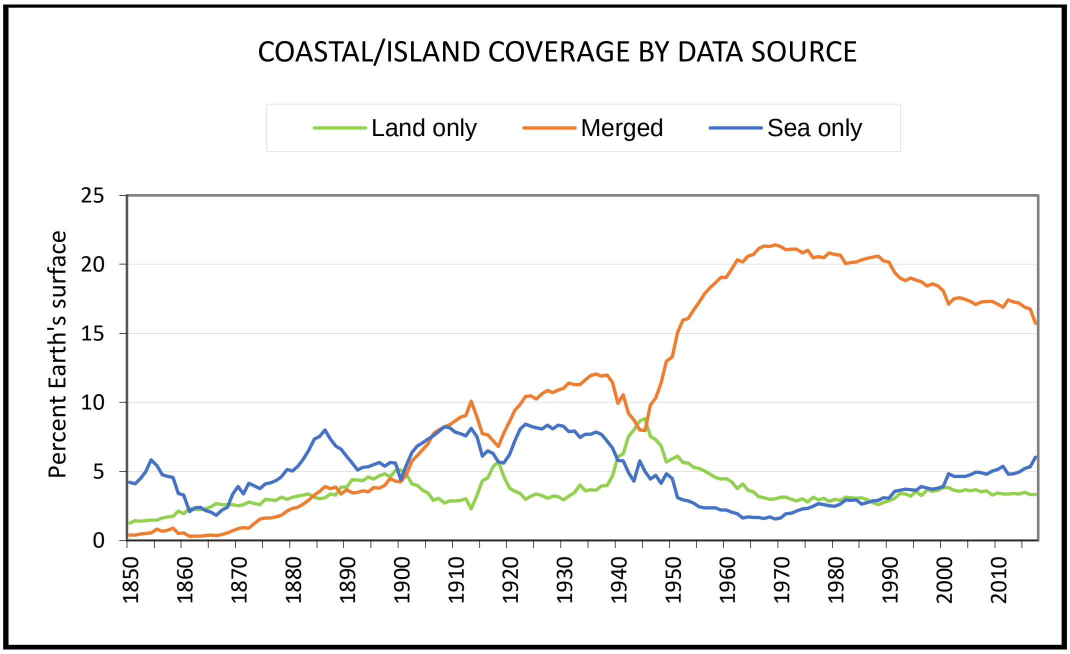


Figure 7.3 The sources of data for coastal and island grid cells, with 'merged' indicating a merging of land and sea temperature data. The very small (1%-2%) of cells whose source was 'unknown' were omitted from this figure (see text for details).

Finding 48 - The data sources for coastal and island grid cells have changed over time in about 85% of such grid cells and the changes of data sources for these grid cells, as well as the change in their contribution to total HadCRUT4 coverage, could well be causing changes in HadCRUT4 global and hemispheric averages.

7.6 Outliers in CRUTEM4 data (i.e. observation stations)

Data is excluded as an 'outlier' (i.e. extreme and possibly erroneous value) if it lies beyond a certain range from the mean value. The limit used in CRUTEM4 and HadCRUT4 for station data is five standard deviations, which is quite extreme. Based on a normal distribution of temperatures for a particular place and calendar month the probability of data exceeding five standard deviations is ~1 in 1.6 million. The HadCRUT4 and CRUTEM4 data from 1850 to 2017 covers only 168 years, so any station and month combination will report no more than 168 values (which is about 0.01% of 1.6 million).

Chapter 6 showed that very extreme values are not removed before calculating Normals or standard deviations. This means that the range of acceptable values is sometimes abnormally and erroneously large.

As chapter 2 showed, some standard deviations for observation stations are very low, particularly in the tropics where the monthly mean temperatures vary little. A standard deviation of 0.2°C would result in a monthly mean temperature being rejected if it was more than 1.0°C from the mean, but variations of that magnitude could probably be caused by genuine meteorological events and should not be assumed to be data errors. An example of this is Penang, Malaysia (ID: 486010) where the September Normal (i.e. 1961-90 average September mean temperature) is 26.4°C, with a standard deviation of 0.2°C. In 12 of the years between 1997 and 2017 the mean monthly temperature reported by this station was in the range 27.5°C to 28.1°C, all of which were outside the threshold of five standard deviations and were therefore rejected.

In total 2,341 of station mean monthly temperatures fall beyond the "five standard deviation" limit but at the same time there are instances of extreme temperatures that fall within that limit and are included in the CRUTEM4 and HADCRUT4 processing. The subsections below illustrate some different types of outliers in the station data.

Extreme temperatures

In the regions beyond 60° latitude North or South extremely cold mean monthly temperatures are to be expected but extremes above 55°C even in tropical regions seem unlikely. Table 7-2 shows some extreme mean monthly temperatures in the CRUTEM4 station data.

ID	Location	Latitude	Year	Month	Temp °C
627510	WAD MEDANI	14.4	2011	4	99.9
852420	ORURO	-18	2011	9	90.0
852010	LA PAZ/ALTO	-16.5	2017	10	88.0
800890	APTO_OTU	7	1978	6	83.4
800890	APTO_OTU	7	1978	7	83.4
800890	APTO_OTU	7	1978	4	81.5
670090	DIEGO-SUAREZ	-12.4	2013	11	67.3
986440	TAGBILARAN	9.6	2012	3	-70.6

Table 7-2 Instances of extreme temperatures in CRUTEM4 station data

Instances of temperature of 0.0°C when the mean temperature is far from zero

The station data contains several instances of 0.0°C as a mean monthly temperature when it falls more than five standard deviations from the Normal for the month (Table 7-3). These might be instances where 0.0°C was wrongly used to indicate missing data. The last entry in the table, for Golden Rock, has two December values of 0.0°C, giving rise to a large standard deviation and therefore the two instances of zero degrees are not rejected as outliers.

ID	Location	Year	Month	Temp °C	Normal °C	Std. Dev °C	No. of Std. Devs
913340	TRUK WSO A	2012	1	0.0	27.4	0.4	-68.5
873050	JACHAL	1994	12	0.0	23.8	1.3	-18.3
341720	SARATOV	1875	7	0.0	22.0	1.7	-12.9
267060	ZELEZNODOROZNY	1997	6	0.0	15.6	1.3	-12.0
879340	RIO GRANDE A	1994	11	0.0	8.2	0.8	-10.3
401830	JERUSALEM CENTER	2012	12	0.0	10.9	1.4	-7.8
315380	SUTUR	1969	6	0.0	15.3	2.5	-6.1
788580	GOLDEN ROCK	1981 & 1984	12	0.0	23.4	7.8	-3.0

Table 7-3 Instances of stations reporting mean monthly temperature of 0.0C when the Normal for that month is very different

Examples of outliers distorting standard deviations and Normals

When extreme values occurred during the period from1941 to 1990 they affect the calculation of the standard deviation for the given observation station in the given calendar month and after 1960 they affect the Normal. Changes in the standard deviation will alter the data that is rejected from the processing for CRUTEM4 and HadCRUT4, and shifts in the Normal will alter every temperature anomaly for the given station in the given month.

Table 7-4 shows some examples of extreme or very unusual temperatures distorting the standard deviation. The columns labelled 'inclusive' and 'exclusive' show the value with the extreme temperature included in the calculation and without it. Where the data applies to some month in the period from 1961 to 1990 the change in the Normal for that month is also shown. With acceptable values being five standard deviations either side of the mean the total span is 10 standard deviations. For the first entry in the table the span was 119°C but falls to 6°C if the temperature of 81.5°C is excluded. Removing that value also reduces the Normal by 3.3°C, which means that every temperature anomaly for that month (April) would be reduced by that amount.

ID	Location	Year	Mon	Temp °C	Normal (i.e. long-term average temperature)		Standard deviation	
					Inclusive	Exclusive	Incl.	Excl.
800890	Apto Uto	1978	4	81.5	27.8	24.5	11.9	0.6
800890	Apto Uto	1978	6	83.4	27.9	24.6	11.8	0.6
800890	Apto Uto	1978	7	83.4	28.0	24.6	12.0	0.6
80150	Oviedo El Christo	1945	6	0.1	16.0	16.0	3.6	1.6
152540	Paltinis	1953	9	-46.4	10.5	10.5	8.8	1.6
417000	Bahawalpur	1947	10	-13.4	26.9	26.9	6.8	1.1
417000	Bahawalpur	1976	11	3.2	20.5	19.5	3.0	1.0
418630	Dinajpur	1964	7	-5.2	27.0	28.9	5.5	0.5
315380	Sutur	1969	6	0.0	15.9	15.3	2.5	1.3
560040	Tuotuohe	1985	11	-28.1	-11.4	-10.8	3.5	1.9
560040	Tuotuohe	1985	12	-26.9	-14.8	-14.2	2.5	1.2
560040	Tuotuohe	1986	1	-28.3	-15.8	-15.4	2.7	1.5
560040	Tuotuohe	1986	2	-20.8	-13.4	-13.1	2.0	1.6
788580	Golden Rock	1981 & 84	12	0.0	23.4	26.0	7.8	0.5

Table 7-4 Calculations of the Normal and standard deviation with and without the extreme values to illustrate the impact of their inclusion in CRUTEM4 and HadCRUT4 processing. (Underlined values show data that falls outside the period over which long-term average temperatures are calculated but within the period for calculating standard deviations.)

Examples of outliers many standard deviations from the mean

Data is rejected if it falls more than five standard deviations from the Normal for all stations. Based on a normal distribution this limit has a probability of ~1 in 1.6 million. Among the 2,341 outliers beyond five standard deviations are 262 that fall beyond eight standard deviations and 117 that fall beyond ten standard deviations. Some extreme values are shown in Table 7-5. These would all have been rejected by CRUTEM4 and HadCRUT4 processing because none occur in the period 1941 to 1990 but their inclusion, like many other extreme mean monthly temperatures, shows a lack of diligence by the organisation that supplied the data and the failure of the Climatic Research Unit to question it.

ID	Location	Year	Month	Temp °C	Mean °C	Std. Dev °C	No. of Std Devs
986440	TAGBILARAN	2012	3	-70.6	26.4	0.6	-161.7
800010	SAN ANDRES/SESQUICEN	2017	8	2.9	27.8	0.2	-124.5
913340	TRUK WSO A	2012	1	0.0	27.4	0.4	-68.5
801390	PUERTO_CARRENO_AGU	2017	8	2.8	26.5	0.4	-59.3
975300	MANOKWARI	2011	1	2.7	26.3	0.4	-59.0
800970	CUCUTA/DAZA A	2017	8	2.9	28.1	0.5	-50.4
800220	CARTAGENA/NUNEZ A	2017	8	2.9	28.0	0.5	-50.2
802340	VANGUARDIA	1993	12	1.0	25.6	0.5	-49.2
802340	VANGUARDIA	2017	8	2.6	24.8	0.5	-44.4
803150	NEIVA/SALAS A	2017	8	3.0	28.7	0.6	-42.8
982230	LAOAG	1992	6	1.0	28.5	0.7	-39.3
644600	SOUANKE	1999	3	2.5	24.6	0.6	-36.8
802340	VANGUARDIA	1994	5	5.3	25.1	0.6	-33.0
802590	CALI/BONILLA A	2017	8	2.4	24.3	0.7	-31.3
802220	BOGOTA/ELDORADO A	2017	8	1.4	13.0	0.4	-29.0
679640	BULAWAYO/GOETZ-OBS.	2013	5	-16.3	16.9	1.2	-27.7
480800	THANDWE	2013	6	43.4	27.0	0.6	27.3
916430	FUNAFUTI A	2016	6	40.2	28.0	0.3	40.7
654420	KUMASI	2010	8	50.7	23.9	0.6	44.7
200260	VIKTORIYA_ISLAND	1994	8	27.3	-0.5	0.6	46.3
627510	WAD MEDANI	2011	4	99.9	31.1	1.1	62.5
200260	VIKTORIYA_ISLAND	1994	7	26.4	0.1	0.4	65.8
670090	DIEGO-SUAREZ	2013	11	67.3	26.5	0.6	68.0
852420	ORURO	2011	9	90.0	10.6	0.9	88.2
852010	LA PAZ/ALTO	2017	10	88.0	10.0	0.8	97.5

Table 7-5 Examples of extreme variation from the Normal (i.e. long-term average temperature).

Finding 49 - The use of a "five standard deviation" limit for the identification of outliers in all circumstances is very generous in some instances and yet very restrictive in locations where mean monthly temperatures for a given calendar month are fairly consistent (such as the coastal or ocean tropics) where that limit could be exceeded by the influence of minor weather events.

Finding 50 - Standard deviations and average long-term temperatures (i.e. Normals) have been distorted by the inclusion of extreme values that should have been removed prior to calculating these key values.

Finding 51 - The inclusion of extreme values in station temperature records calls into question the competence of the national meteorological services that supply the data, and of the Climatic Research Unit for failing to question the presence of those values. Many of these extreme temperatures are obvious errors, some of which might have arisen from having just a few days of temperature recordings and others been errors in transcription. This raises the question of how many less obvious errors exist within the station data.

7.7 Outliers in HadSST3 grid cell values

The HadSST3 temperature anomaly data was analysed by processing in the same manner as the observation station data, i.e. calculating standard deviations based on 1941 to 1990, subject to a minimum of 15 years for the calendar month in question and calculating mean temperatures from a minimum of 14 years of data for the month across the period 1961 to 1990. (The HadSST3 dataset is of temperature anomalies. That only means that a constant value has been subtracted for each calendar month for each grid cell; the distribution about the mean value remains the same.) The data for the grid cell-month combinations that met the two criteria was then examined for entries that are more than five standard deviations from the mean. Table 7-6 shows some entries at the upper and lower ends of the scale.

Year	Mon	Latitude	Long.	Value	Mean	Sdev	SDevs
1851	8	22.5	-67.5	-6.29	-0.002	0.262	-24.028
1858	6	2.5	107.5	-6.82	-0.014	0.273	-24.886
1872	10	2.5	107.5	3.61	-0.007	0.275	13.134
1918	8	22.5	-72.5	-6.91	-0.006	0.289	-23.906
1885	9	7.5	87.5	-5.82	-0.033	0.292	-19.798
1869	10	2.5	-32.5	4.55	-0.049	0.335	13.728
1858	6	7.5	107.5	-6.61	-0.009	0.342	-19.278
1878	10	27.5	-67.5	-6.43	0	0.348	-18.456
1854	5	2.5	87.5	-6.36	-0.097	0.351	-17.859
1851	8	-2.5	62.5	-6.57	0.044	0.357	-18.515
2013	1	-2.5	72.5	-6.20	-0.08	0.359	-17.070
1879	6	22.5	-72.5	-6.89	0.014	0.368	-18.787
1897	9	-27.5	2.5	-6.84	-0.008	0.369	-18.529
1878	12	2.5	-32.5	-6.14	-0.027	0.371	-16.476
1854	9	17.5	-57.5	-6.19	-0.024	0.372	-16.593
1857	11	-2.5	47.5	-6.45	0.024	0.379	-17.081
1866	3	2.5	-37.5	-6.19	-0.054	0.380	-16.151
1859	9	17.5	122.5	-6.26	-0.010	0.383	-16.324
1851	8	-2.5	67.5	-6.88	0.037	0.391	-17.672
1851	3	7.5	92.5	-7.34	-0.050	0.394	-18.489
1859	4	22.5	-82.5	5.16	0.031	0.396	12.945
1858	5	7.5	107.5	-6.78	-0.027	0.401	-16.819
1940	8	17.5	42.5	-7.35	-0.031	0.434	-16.869
2010	12	-62.5	-52.5	6.91	-0.077	0.440	15.865
2010	7	-7.5	-22.5	5.63	0.099	0.446	12.398
1893	7	-27.5	167.5	6.40	-0.030	0.457	14.069
1910	10	-17.5	-2.5	6.87	-0.017	0.502	13.714
1857	10	-37.5	-167.5	6.44	-0.058	0.504	12.893
1910	10	-17.5	-7.5	6.92	-0.031	0.506	13.743
1893	8	-27.5	167.5	6.75	-0.064	0.526	12.959
2017	6	2.5	-22.5	7.31	-0.062	0.605	12.186
2011	10	72.5	-72.5	7.43	-1.038	0.625	13.550

Table 7-6 The most extreme examples of HadSST3 data. For CRUTEM4 data the maximum acceptable range is from -5 standard deviations to +5, but these far exceed that range.

The annual totals of HadSST3 grid cell-month combinations where the data fell outside five standard deviations were also calculated (Figure 7.4) and, not surprisingly, there were very few instances during 1961 to 1990. It is possible that this is linked to the low number of SST measurements discussed in Chapter 5 where it was shown that 68% of HadSST3 anomalies beyond 5.0°C were from less than five observations in the entire 5° x 5° grid cell.

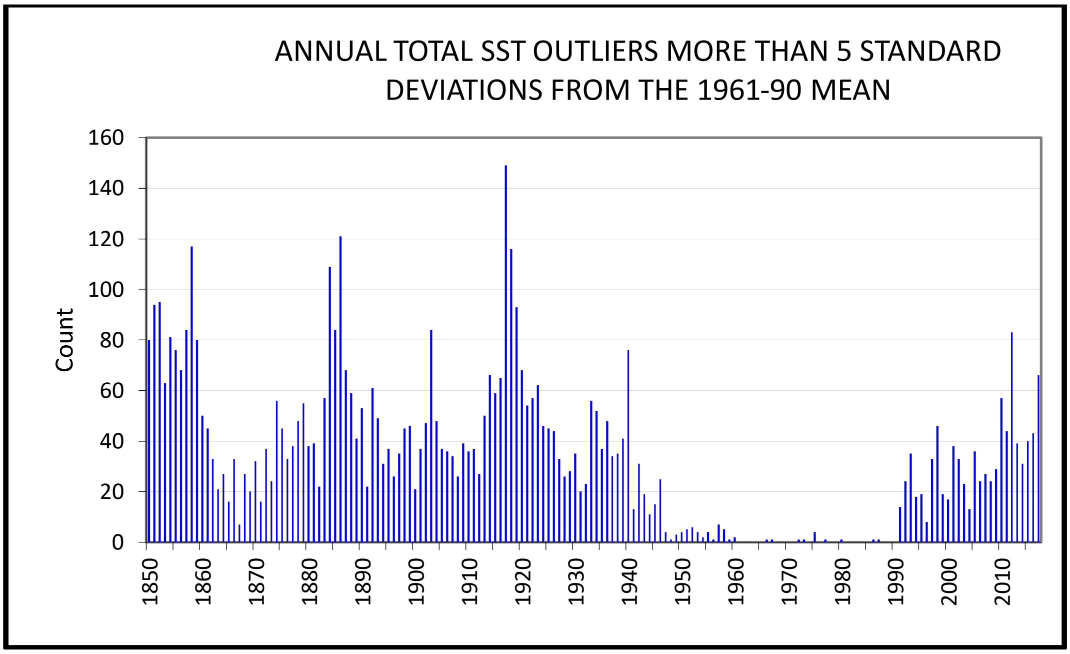


Figure 7.4 Total number of SST entries that were outside five standard deviations from the 1961-90 mean anomaly in each year.

Finding 52 - The presence of outliers in the HadSST3 conflicts with the extremely low probability that such values would be present. If the HadSST3 processing creates such outliers then it should be questioned. The answer might lie with extreme mean monthly temperature anomalies often being linked to very few measurements being made in the corresponding grid cells during the corresponding months.

7.8 Wide ranges of station anomalies in CRUTEM4 grid cells

The grid cell values for the CRUTEM4 dataset are the average of the temperature anomalies for each reporting station in the grid cell. The HadCRUT4 grid cell values for cells that cover land and those of coastal areas whose data might be from land, sea or a merging of the two, are derived by a similar approach but with artificial variance as mentioned earlier.

If the anomalies for stations within the same grid cell vary substantially it might mean errors in the data or, if the data is correct, it shows that the cell value might be skewed depending on which stations reported in the given month.

According to the CRUTEM4 station data there are 1,448 instances of grid cell-month combinations where the temperature anomalies from the reporting stations within the grid cell exceed 8.0°C and of those 459 exceed 10.0°C. Some of these will disappear if outliers are removed from the temperature data, but as was shown in the previous section, some extreme values can be found in the data from which standard deviations are calculated (i.e. for the period 1941-1990) and unusually large standard deviations mean that other extreme values will not be rejected.

Of the 1,448 instances of the range being greater than 8.0°C 660 are from the period 1941 to 1990. Table 7-7 shows 13 instances of grid cells that at some month in the period 1941-1990 have very wide ranges of temperature anomalies for the reporting observation stations in the cell. It appears that in most instances either the minimum or maximum temperature anomaly is in error. In the remaining instances the anomalies are not so extreme but are still quite different to those for other stations in the grid cell. In this table the stations indicated with '*' have not already been mentioned as possibly having data errors but they do have quite different anomalies to the other reporting stations in the grid cell.

In each of these cases the omission of the single suspect data value would shift the average anomaly (i.e. the CRUTEM4 grid cell value).

Year	Mon	Lat.	Long.	No. Stns	Min. Anom.	Max. Anom.	Anom. Range	Probable station in error
1974	1	62.5	92.5	3	-9.9	15.9	25.8	Tugoncani*
1951	2	62.5	102.5	4	-9.2	10.8	20.0	Kislokan*
1953	9	47.5	22.5	17	-56.9	1.9	58.8	Platinis
1945	6	42.5	-7.5	8	-15.9	2.7	18.6	Oviedo El Christo
1966	11	37.5	72.5	17	-3.3	14.8	18.1	Sarytash*
1951	6	32.5	72.5	8	-18.9	0.1	19.0	Sargoda*
1947	10	27.5	72.5	4	-40.3	-1.6	38.7	Bahawalpur
1964	7	27.5	87.5	12	-32.2	0	32.2	Dinajpur
1981	12	17.5	-62.5	5	-23.4	1.5	24.9	Golden Rock
1984	12	17.5	-62.5	3	-23.4	0.3	23.7	Golden Rock
1978	6	7.5	-72.5	6	-0.7	55.5	56.2	Apto Uto
1978	7	7.5	-72.5	6	-0.7	55.4	56.1	Apto Uto
1978	4	7.5	-72.5	6	-1.2	53.7	54.9	Apto Uto

Table 7-7 Instances during 1941-1990 where stations within the same grid cell had an unusually wide range of temperature anomalies. (Stations flagged with '*' have not been previously mentioned as likely errors.)

Finding 53 - Instances of unusual mean monthly temperature anomalies, relative to those from other observation stations in the same grid cell, suggest temperature data errors.

7.9 Observation stations in close proximity to each other

The station metadata gives the latitude and longitude of the station to one decimal place, which at the equator is 11.13km (c.f. grid cell size of about 555 km x 555 km) but will fall with increasing latitudes (i.e. nearer the poles). According to the metadata for only the acceptable stations as described earlier (Chapter 3), there are 93 instances of two stations sharing the same location and a further two instances of three stations sharing the same location. In some instances, the names of the station that share a location are also identical (e.g. "MT KENAAN" with stations IDs 401530 and 401531, "Jokkmokk" with IDs 21420 and 21461). A sampling of various pairs showed differing periods of data overlap, ranging from no overlap to many decades. (See also some examples in Appendix 4.)

A further 381 instances were identified where stations are separated by 0.1° Latitude, 0.1° Longitude or both, including 28 instances of close proximity to the locations with multiple stations. Also, a pair of Russian stations, Severomorsk and Poljarnoe (IDs 220190 and 220191), one at Latitude 69.0°N, the other at 69.2°N and 0.1° Longitude apart (total distance ~20 km), have almost identical temperature data. An example of the close proximity of stations is shown in Figure 7.5.

Any temporal overlap of stations pairs and triples, and of stations in close proximity to each other, will bias CRUTEM4 and HadCRUT4 grid cell values towards the temperature anomalies for that locality.

Figure 7.5 In Austria, Innsbruck Flughafen (airport) and Innsbruck University (Top left?, Top right? Or the university sports fields in green near the airport?) both report data from 1951 to 2012, the distance apart being at most about 5km. (Source: Google maps)

Finding 54 - The inclusion of data from reporting stations in close proximity to each other will bias grid cell average anomalies towards the temperatures recorded in those localities.

7.10 Observation stations incorrectly assigned to grid cells

The observation station at Garissa (Kenya, ID: 637230) has its latitude given in its metadata as 0.5 (i.e. 0.5°N). A latitude of -0.5 (i.e. 0.5°S) is much closer to the town of Garissa and would perhaps refer to an observation station at Garissa airport. The distance between the two locations is only 125 km but 0.5°N places Garissa observation station in one grid cell and 0.5°S places it in another.

In the January 2016 version of the station data Ghanzi (Botswana, ID: 680240) had a latitude of 21.5 rather than -21.5, placing it north of the equator in southern Libya. This error has since been corrected but its presence at all is a cause for concern regards the accuracy of station data.

Another observation station with a doubtful location is Jinja (Uganda, ID: 636820). The station metadata has the location (-1.3, -30), i.e. 1.3°S 30.0°E, which places it in south west Uganda, near Kabale, but the town of Jinja is located on the other side of the equator at 0.5°N, 33.2°E. It is possible that two locations in Uganda have the same name but Google Maps does not indicate that this is the case. One location is north of the equator and one is south, placing them in different grid cells.

Finding 55 - At least one and perhaps two observation stations have metadata that incorrectly indicates the station location, in both cases meaning that the station will be assigned to the incorrect grid cell.

8 Issues with SST data prior to HadCRUT4

8.1 Introduction

The oceans cover 71% of the Earth's surface, which means that sea surface temperature data comprises the major component of the HadCRUT4 dataset. When HadCRUT4 coastal or island grid cells are taken into account and coverage calculated according to the HadCRUT4 system, sea surface temperature anomalies might contribute to as much as cover 82% of the Earth's surface. Clearly if the HadCRUT4 dataset is to be accurate then sea surface temperature data need to be accurate.

Data for the HadSST3 and HadCRUT4 temperature anomaly datasets is obtained from the ICOADS database. In this chapter the focus is on the data, prior to its inclusion in that database, in the database itself and its processing prior to inclusion in the HadSST3 and HadCRUT4 databases.

8.2 Inconsistent methods of measurement

The methods used for observing sea surface temperature have varied over time, covering buckets (leather, rubber, metal, wood, iron and insulated canvas), temperature sensors at depth (hull-mounted or in inlet pipes for engine cooling system), expendable bathythermographs and moored or drifting buoys. The extent of use of each technique has varied over time, the majority of data since about 2002 was obtained by drifting Argo buoys.

Many papers discuss how different methods of measurement report different temperatures, among them Kennedy et al (2011b), Folland and Parker (1995), Matthews (2013) who cites Brooks (1926), Matthews & Matthews (2013), Kent and Taylor (2006), Emery et al (2001), and Rayner et al (2006)

Folland & Parker (1995) and Matthews (2013) also report that national meteorological services provided differing instruments and instructions as to how a bucketful of sea water should be taken, lifted aboard ship and its temperature measured. For example, the instructions about stirring the water in the bucket variously said "little, if at all", "slowly", "slowly, not touching the walls", "continuously", "quickly" and "vigorously". On the subject of how long the thermometer should be left in the water before it was read some said for one minute or less, others said "2 to 3 minutes", "at least 3 minutes", "3 to 5 minutes" and "4 or 5 minutes". The instructions for UK ships in 1938 said that water was to be drawn from the surface, while ships from the USA were told to draw the water from a depth of 3 to 6 feet (i.e.1 to 2 metres).

If the data as a whole is to be useful then most of it needs to be adjusted in order to bring it to the theoretical equivalent of it all being measured by the same technique, which would not only be the same method but according to a specific set of instructions from a national meteorological service. This has two major obstacles. Firstly, many papers that compared techniques, including most of those listed above, give conflicting estimates of the required adjustments. Secondly, the observations recorded in the ICOADS database do not always indicate the method of measurement that was used, the historical data being worse in this regard (i.e. a higher percentage with no information) than more recent records. These are the principal reasons for the 100 different versions of the HadSST3 dataset being created and the general use of an "ensemble" version compromised of the median of the 100 values for each grid cell and month (even though at most only one of those 100 versions is correct).

Finding 56 - The many and very different methods of measuring sea surface temperature and the variations within those methods are very likely to have different levels of accuracy and different error margins but these issues are ignored in the HadSST3 and HadCRUT4 datasets.

Finding 57 - The technique of creating 100 different datasets and using them to derive the HadCRUT4 dataset has uncertain accuracy, mixes the data from (at most) one correct dataset with 99 incorrect variations and simply hopes that the value used for HadCRUT4, the median of the 100 values, is correct. There is poor justification for this ad hoc approach to poor and incompatible data.

8.3 Issues with thermal layering

Many sea surface temperatures are not measured directly but are derived from the temperature measured one or more metres below the surface. The common situation is the measurement of the temperature of the water as it enters the cooling system of a ship. Buoys and hull-mounted sensors can likewise measure below the surface. Even ignoring the notion of whether sea surface temperature means the surface or some centimetres below it, there is still doubt as to whether temperature measured at one depth can accurately be converted into an equivalent temperature at another depth. A fundamental obstacle to this is the frequent presence of thermal layers in ocean waters.

Donlon (2005) describes how five distinct layers can be identified in the first 10 metres of water (Figure 8.1). The top millimetre of water has three layers, which he designates as (a) STT_{int} - the interface of atmosphere and ocean, (b) SST_{skin} - the skin layer, and $SST_{subskin}$ - below the level at which heat is lost by evaporation. These three layers are involved with exchanging heat with the atmosphere as well as being the first water that would enter a bucket dropped into the ocean.

The other two layers are deeper, the first, which is said to be at about one metre depth, he calls SST_{depth} and describes it as "an in-situ measurement near the surface of the ocean that is typically reported simply as SST". The lowest layer, between five and 10 metres, he calls SST_{fnd} and describes it as the "foundation" temperature obtained from below the layer that varies diurnally.

Many papers agree with Conlon's basic schematic including Woodcock (1941), Ewing & MacAlister (1960), Hasse (1963), Fedorov & Ginsburg (1992), Fairall et al (1996), Wells et al (2009), Webster et al (1996), Soloviev & Lukas (1997) and Kawai & Wada (2007).

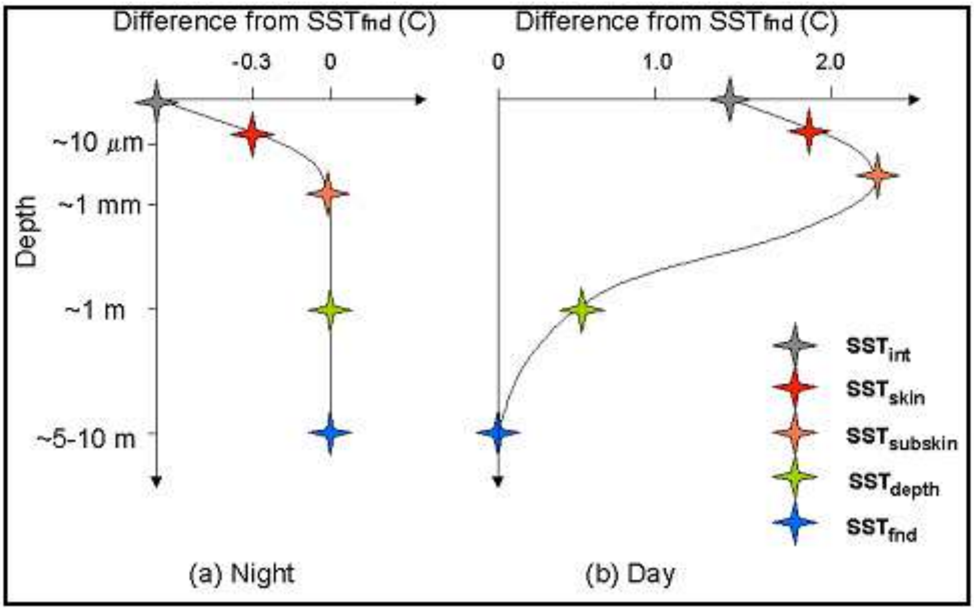


Figure 8.1 (Copied from Conlon et al, 2005) Schematic of the variation in temperature with water depth during the night and day.

The problem is that while the basic situation is well-described, in reality the layering of the water is less clearly defined. The temperature differences between different layers will vary; the depths of the surface skin and the layers below it will vary. As Conlon (2005) shows the pattern varies between day and night. It will also be different under cloudy skies or clear skies and even under clear skies it will depend on the amount of incoming solar radiation. (Farrar et al, 2007, has a figure showing an instance of rather abrupt warming that produced a 2°C increase at one metre depth but negligible change at two metres depth.) The description provided in Conlon (2005) is for calm seas but the seas might be rough and the water so mixed that layers can't be distinguished. At the other extreme is when several days of strong sunlight and calm conditions have caused warming at a greater depth than usual (i.e. below Conlon's "foundation" layer") and not all of that heat is lost overnight. In a similar fashion, factors like the El Nino-Southern Oscillation might cause a sustained increase or decrease in ocean heat that persists for weeks or months and extends below the level of the diurnal heat change pattern.

Conversion of temperatures measured at one depth to equivalent temperatures at another depth would need to take into account a range of factors that might not have been recorded and, in the case of situations over the previous hours or days, quite possibly unknown.

In passing it should be noted that the "standard depth" seems to have switched over time from less than one metre below the surface, to a deeper level consistent with engine cooling water intakes, whose depth in the water varies depending on the ship's load and size.

Finding 58 - The variety of techniques for determining sea surface temperature often requires the adjustment of temperatures taken at different depths to the notional standard depth but correct adjustment requires data related to thermal layering that is unlikely to be available.

8.4 Ships in port when temperatures recorded

Brohan et al (2009) discusses the transcription of temperature observations from the log books of the Royal Navy into the ICOADS database, the source of HadSST3 and HadCRUT4 sea surface temperature data.

Samples of that data were available from Brohan's webpage[5] and data for the 69 ships whose names commenced with 'A', 'B' or 'C', were downloaded for analysis. Of the 253,600 recordings of SST data made from those ships 127,846 recordings (50.4%) were made while the ship was in port (i.e. a non-blank 'port' field in the record) leaving only 125,754 recordings made (49.6%) with the ship at sea. This assumes that the ‘port’ field would not be left blank while the ship was in port but perhaps it was left blank when the ship was in port for an extended period.

In ports the water is shallower, likely fed by local natural and manmade discharges, usually sheltered from currents and often from wind, and there is far less mixing of surface water with deeper water. The water temperature in port cannot be assumed to be the same as the open sea and therefore is unlikely to be representative of any significant part of the grid cell.

For example, the ICOADS database shows that ship "ABERDARE", one of those downloaded as described above, made 32 SST observations between 24th and 30th of September 1939 at latitude 1.4°N longitude 103.8°E. According to ABERDARE’s data this was at "RN Base Singapore". Interrogation of the ICOADS database, covering a box of latitudes 0°N to 2°N and longitudes 103°E to 105°E over the period from 24 to 30 September 1939, revealed 11 ships of the Royal Navy making 445 observations at or very close to the same location over seven days, with just 86 of the observations further away but within the same HadSST3 grid cell, showing that the inclusion of SST data recorded while ships are in port is not an isolated occurrence.

Finding 59 - The source data for sea surface temperatures used in the construction of the HadCRUT4 and HadSST3 datasets, the ICOADS database, appears to contain data recorded when ships were in port, where temperatures are very likely different to those at sea and not representative of the grid cell.

8.5 Macro adjustments to SST data possibly incorrect

Numerous sources (e.g. Folland & Parker, 1995 citing the earlier Folland et al, 1984; Folland et al, 2001; Smith & Reynolds, 2004; Rayner et al, 2006; Kennedy et al, 2011b) report an abrupt shift in average SSTs in December 1941, particularly in the northern part of the Pacific Ocean. Several papers (e.g. Thompson et al, 2008; Kennedy et al, 2011b) also report a similar shift in late 1945, especially in the Southern Hemisphere. These shifts (Figure 8.1) have been attributed to changes in the most common technique of measuring SST, firstly as a shift from buckets (favoured by non-US ships) to engine room intake (favoured by US ships) as the USA entered World War II and then the use of buckets increasing after the end of the war.

[5] See http://brohan.org/hadobs/digitised_obs/docs/

Folland et al (2001) refers to the switch from buckets to engine room intake for SST measurements when describing how SST measurements were adjusted to more closely match global average temperature anomalies derived from models, the assumption being that SST measurements via bucket recorded lower temperatures than measurements via engine room intakes. It states "*The global annual mean bias correction increases steadily from 0.17°C in 1872 to 0.30°C in 1900 and 0.39°C in 1920, remaining around 0.4°C until 1941*."

Folland and Parker (1995) goes on to point out that the first jump, in 1941-2 (Figure 8.2), coincides with the entry of the USA into World War II and "*is likely to have resulted from a realization of the dangers of hauling sea buckets onto deck in wartime conditions when a light would have been needed for both hauling and reading the thermometer at night*".

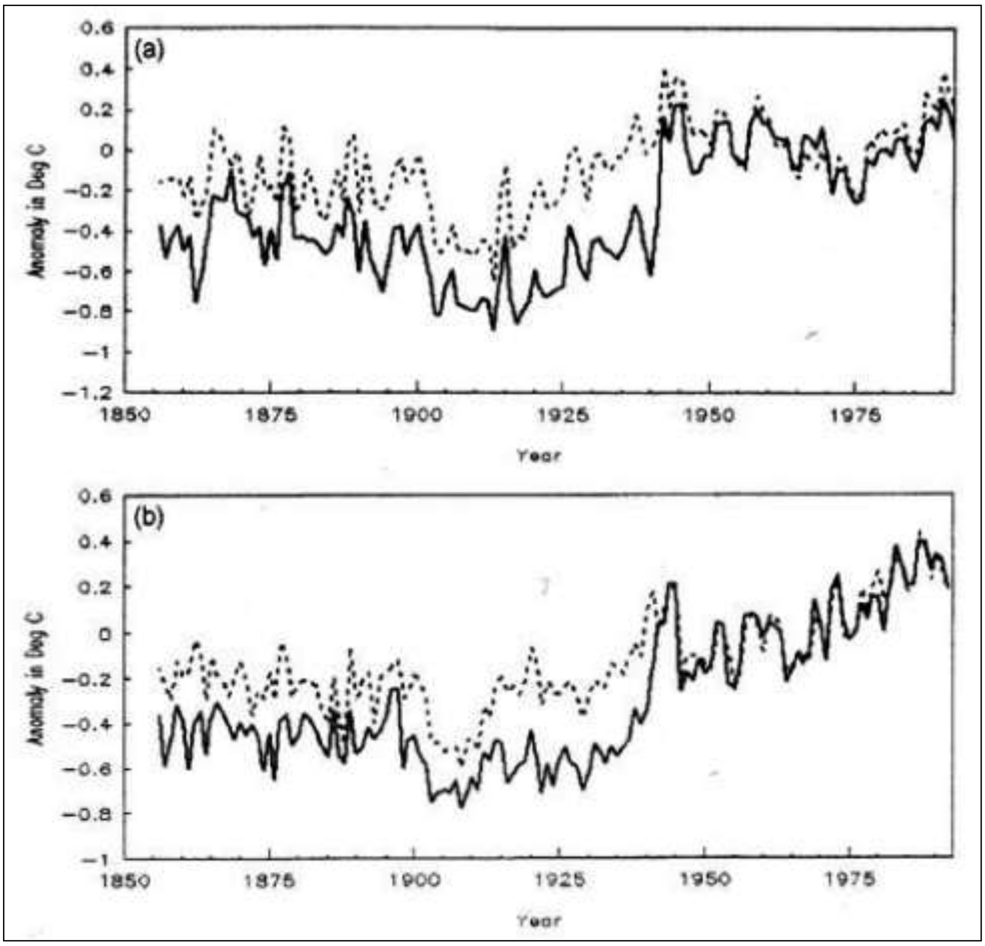


Figure 8.2 Copy of Figure 3 in Folland and Parker (1995) showing the jump in measured SST (solid line) in the mid 1940's and the adjusted - stated as "corrected" - night marine air temperature NMAT (dashed line) for (a) northern (b) southern hemisphere, (1856-1992)

The above papers seem to ignore the large changes in SST coverage that can be attributed to reduced shipping during World War II and its subsequent resumption at the end of the war (Figure 8.3). Coverage in the Southern Hemisphere in March 1939 was 53% but in October that year it had fallen to just 23%. In November 1941 Northern hemisphere coverage was 54% but fell in the next month to 40%. Figure 8.2, from Folland and Parker (1995), indicates a steeper increase in average SST in the Northern Hemisphere than in the Southern Hemisphere. The reduction in coverage, due largely to shipping being limited to key supply routes and to where the war was being fought at the time (e.g. equatorial Pacific Ocean) could easily have caused unexpected global average SST anomalies.

Figures 8.4 and 8.5 show the monthly coverage in 30-degree latitude bands for the Northern and Southern Hemispheres respectively, with sharp decreases in coverage when World War II impacted shipping in the Southern Hemisphere (1939) and Northern Hemisphere (1942). This is particularly true of the first 30 degrees of latitude from the equator, the region that accounts for 50% of the hemisphere coverage.

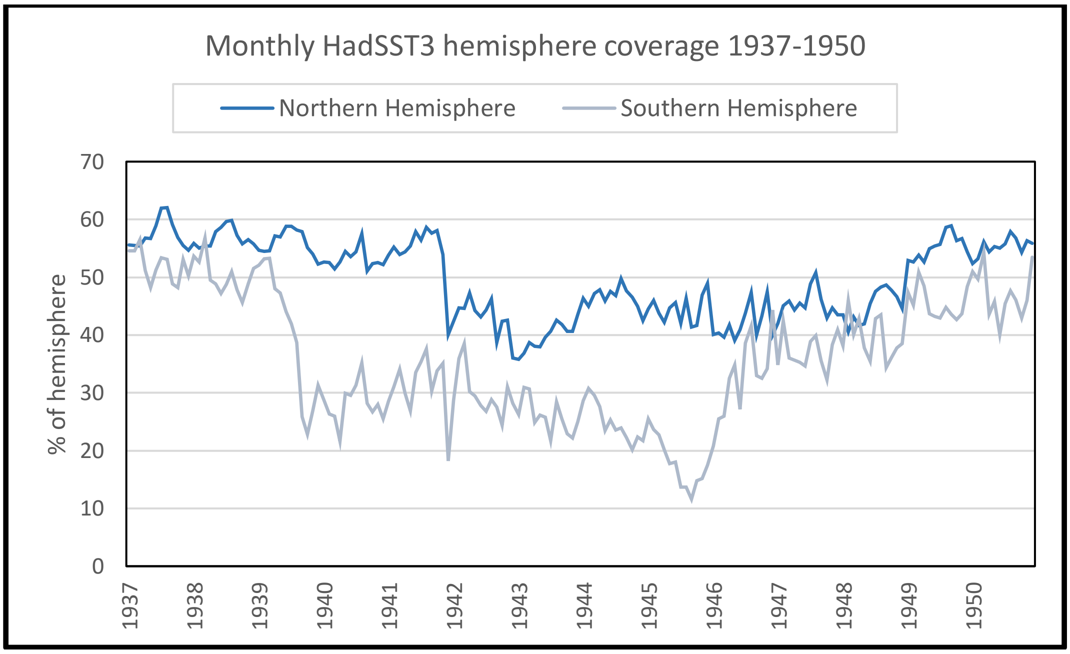


Figure 8.3 HadSST3 coverage from 1937 to 1950 in the two hemispheres, with coverage reducing during World War II, firstly in the Southern Hemisphere (1939) and then in the Northern Hemisphere (1942).

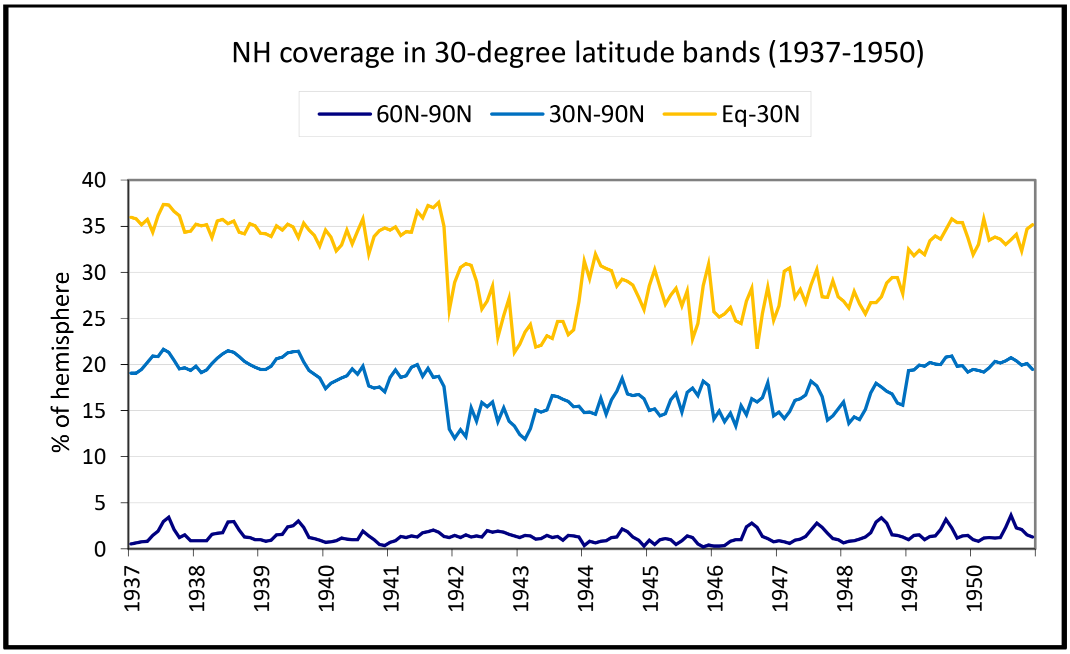


Figure 8.4 HadSST3 coverage in three Northern Hemisphere latitude bands, each of 30 degrees, showing the sharp drop when the USA entered World War II.

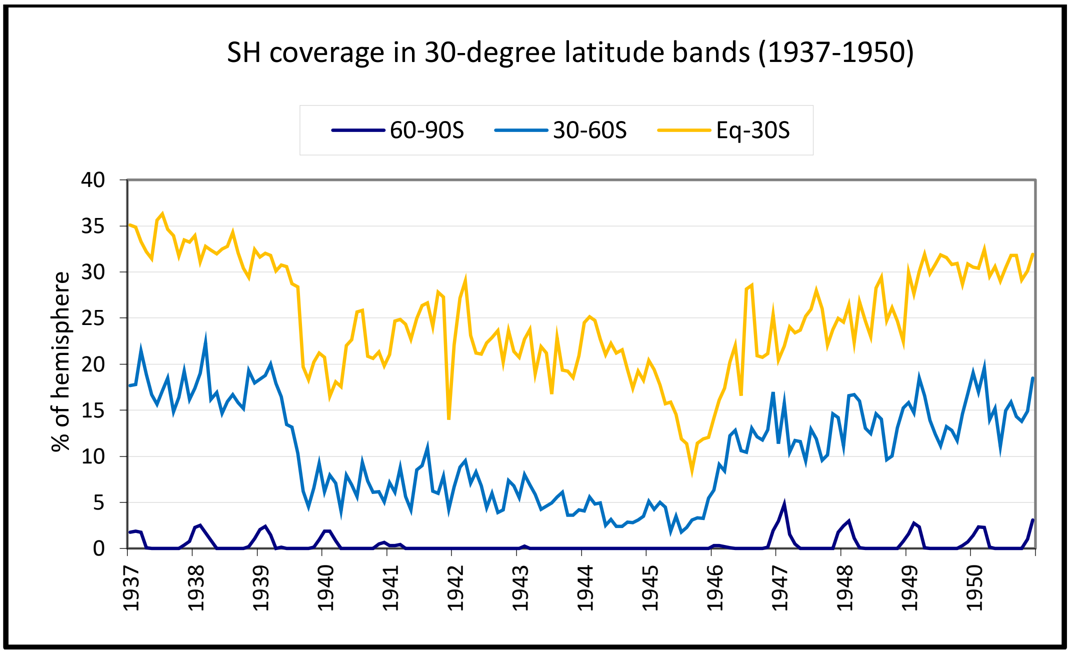


Figure 8.5 HadSST3 coverage in three Southern Hemisphere latitude bands, each of 30 degrees, showing the decrease in coverage during World War II.

Finding 60 - The reduction in SST coverage due to World War II might at least partly account for the widely-reported peak in average hemispheric SST anomaly at about that time. Attributing it all to changes of measurement technique and adjusting sea surface temperatures accordingly is probably unwise.

8.6 Data transcription errors

Data from Brohan et al (2009) was also found to have flaws in its transcription from handwritten logs from Royal Navy ships for the periods 1938-39 and 1941-47 into digital form for inclusion in the ICOADS and HadSST dataset. It appears that at times the automatic digitising (apparently by character recognition software) failed to correctly distinguish between similar digits, such as “1” and “7”, and “4” and “9”. Figure 4 of Brohan et al (2009) provides an illustration of the latter when it shows details from ship HMS Warspite on January 12 1941 (Figure 8.6). Table 8-1 is an extract from the transcribed data after data records in IMMA format had been created (also from Brohan’s website) and Table 8-2 shows the corresponding ICOADS entry.

When Brohan initially made the transcribed data available it contained numerous errors with the latitude and longitude, with such things as some ships apparently moving from one side of the Atlantic to the other in just a few hours! Brohan mentions the correction of this data by careful tracking of each ship's route. It appears that the more difficult task of correcting errors in meteorological data was either not attempted or is incomplete.

Incidentally the ICOADS data in Table 8-2 shows another error in the conversion. A wind direction of NNW equates to 338°, not 348°. It is entirely possible that multiple entries submitted at the same time and referring to the same wind direction likewise have this error.

While errors in just a few transcribed data records that were added to the ICOADS database amount to little in themselves, they raise the question of the accuracy of the transcription of other Royal Navy ship records that were submitted to ICOADS.

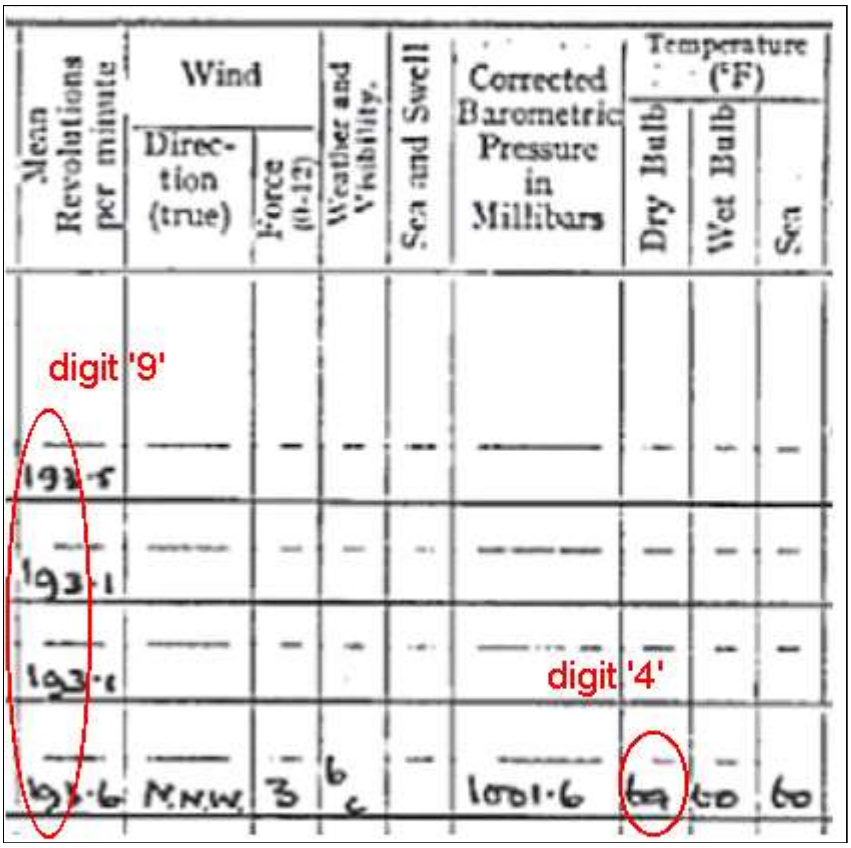


Figure 8.6 Enlarged extract of Figure 4 from Brohan et al (2009), highlighting several instances of the digit '9' and one of the digit '4', the latter shown in the table below to be misinterpreted as a '9'.

Year	Mon	Day	Hour	Lat. * 10	Long. * 10	Wind		SLP*10	Temperature		
						Speed	Direct		Dry Bulb	Wet bulb	Sea
1941	1	11	18	3442	1948	20	SE	10038	64	60	60
1941	1	11	22	3452	2071	24	SE	10025	65	59	60
1941	1	12	2	3462	2194	4	NNW	10016	69	60	60
1941	1	12	6	3472	2317	8	NNW	10023	63	61	61
1941	1	12	10	3410	2427	12	WNW	10053	65	60	60
1941	1	12	14	3349	2537	16	NW	10080	69	61	61

Table 8-1 Extract of transcribed IMMA data from Brohan's web pages. The highlighted data corresponds to Figure 8.6 and shows the error with '4' being transcribed as '9'.

YR	MO	DY	HR	LAT	LON	WindDir	SLP	AT	WBT	SST	Wdir	AT(F)	WBT(F)	SST(F)
1941	1	11	18	34.42	19.48	135	1003.8			15.6	SE			60
1941	1	11	22	34.52	20.71	135	1002.5			15.6	SE			60
1941	1	12	2	34.62	21.94	348	1001.6	20.6	15.6		NNW	69	60	
1941	1	12	6	34.72	23.17	348	1002.3	17.2	16.1		NNW	63	61	
1941	1	12	10	34.10	24.27	293	1005.3	18.3	15.6		WNW	65	60	
1941	1	12	14	33.49	25.37	315	1008.0	20.6	16.1		NW	69	61	

Table 8-2 Part of the ICOADS data corresponding to Table 8-1, with wind direction and Fahrenheit temperatures manually added in rightmost four columns (column heads shaded). The wind direction of 348° does not correspond to NNW; it should be 338°.

Finding 61 - The ICOADS database contains at least one demonstrable error of data transcription and another incorrect conversion. It might contain many more errors that have not been identified.

8.7 Inconsistencies in ICOADS data

As mentioned above, the ship "Warspite" was in the south-east of the Mediterranean Sea on 12 January 1941. Data from the ICOADS database shows that other ships were nearby, but apparently reporting quite different meteorological data. According to data from "Warspite", at 2pm the air temperature was 20.0°C whereas ship "Barham", just 0.07° latitude north and 0.07° longitude west away (approximately 3 km across open sea), recorded the air temperature as 16.1°C.

A few days later, on 14 January 1941, in the same area of the Mediterranean ships "Eagles" and "Barham" reported 6pm air temperatures of 21.1°C and 16.7°C respectively when "Eagles" location was given as 34.85°N 20.83°E and "Barham" at 34.87°N 20.90°E, which puts the vessels even closer than 3 km. Other simultaneous observations made by "Eagle" and "Barham" are shown in Table 8-3, many of which show unexpected large differences. In this table the time is UTC (i.e. Greenwich time), which was probably about two hours behind local time.

Some of the differences in recordings made from ships just a short distance apart[6] are surprising:

- Sea level pressures differ greatly when the ships are very close (e.g. 1005.8 hPa v. 1017.1hPa at 16:00hrs)
- Air temperatures are sometimes more than 3°C different when the ships are close but with wind supposedly from different directions (at 10:00 hrs). Even with wind from the same direction one is 16.7°C while the other 18.3°C (at 16:00 hrs)
- Wet bulb temperatures are consistently different at 16:00 and 18:00 hrs
- The sea surface temperature from ship "Barham" fell by almost 3.5°C between 16:00 and 18:00 hrs but for ship "Eagle" only by 0.5°C, with a difference of 0.6°C at 16:00hrs becoming a difference of 2.3°C at 18:00. (Both SST's at 18:00 hrs appear to be incorrect repeats of wet-bulb temperatures.)

[6] At this latitude and longitude, a difference of 0.1° latitude is ~10 km and 0.1° longitude is ~9.2 km.

Day	Hour	Ship	Latitude (N)	Longitude (E)	Wind direction	Wind speed (m/s)	Sea Level Pressure (hPa)	Air Temp (°C)	Wet Bulb Temp (°C)	Sea Surface Temp (°C)
12	600	BARHAM	34.60	23.20	225	6.7	1004.3	16.7	15.0	17.8
12	600	EAGLE	34.65	23.13	248	6.7	1003.6	18.3	15.6	17.2
12	1000	BARHAM	35.15	23.22	270	6.7	1004.9	15.6	13.3	18.3
12	1000	EAGLE	35.12	23.12	315	6.7	1001.2	18.9	17.2	17.2
12	1400	BARHAM	35.29	23.32	293	6.7	1005.6	16.7	14.4	17.8
12	1400	EAGLE	35.26	23.24	293	6.7	1017.1	16.7	16.1	17.2
12	1600	BARHAM	35.35	23.36	293	6.7	1005.8	16.7	14.4	17.8
12	1600	EAGLE	35.33	23.30	293	6.7	1017.1	18.3	16.7	17.2
12	1800	BARHAM	35.42	23.41	293	4.6	1008.9	15.6	14.4	14.4
12	1800	EAGLE	35.40	23.36	293	4.6	1017.2	17.2	16.7	16.7
12	2200	BARHAM	35.56	23.51	315	2.6	1009.2	16.1	15.0	14.4
12	2200	EAGLE	35.55	23.49	248	2.6	1007.6	16.7	16.1	15.6
13	200	BARHAM	35.69	23.60	315	4.6		17.2	15.6	16.7
13	200	EAGLE	35.69	23.61	293	2.6		15.6	14.4	17.2
13	600	BARHAM	35.83	23.70	293	9.3		12.8	11.1	16.7
13	600	EAGLE	35.83	23.73				15.6	13.9	16.1

Table 8-3 Extract from ICOADS database for two ships in close proximity on 12 and 13 January 1941 and yet often showing considerable differences in recorded meteorological data.

Finding 62 - It appears that the meteorological data in the ICOADS database has not been checked for consistency, either by the organisation that maintains the database or the Hadley Centre, prior to its use in creating the HadSST3 and HadCRUT4 datasets.

8.8 Erroneous locations in ICOADS data

The entries in the ICOADS database often include latitude and longitude co-ordinates but only to one decimal place. This is insufficient to precisely pinpoint the location of the observations and it requires some allowance for rounding. Table 8-4 shows instances of the locations in the ICOADS database from ships making observations along Australia's Great Barrier Reef actually being much further inland than can be explained by rounding to one decimal place. At the first five latitudes shown in Table 8-4 0.1° longitude is equivalent to approximately 10km on the ground.

The examples shown here will still fall within the correct grid cell for the coastal region but it seems likely that there are other cases in the ICOADS database where the incorrect location means the incorrect HadSST3 and HadCRUT4 grid cell.

Latitude	Longitude	Approx. Distance inland	Nearest Coast longitude
-12.9	142.5	80 km	143.3
-13.5	142.5	100 km	143.5
-14.6	143.2	50 km	143.8
-15.6	144.0	125 km	145.3
-15.7	144.3	100 km	145.3
-18.5	145.4	75 km	146.2
-19.8	146.0	80km (SW)	146.5 (at Lat -19.1)
-21.8	148.4	80km	149.4

Table 8-4 Example instances of ICOADS co-ordinates defining inland locations near Australia's Great Barrier Reef rather than locations at sea. (The distances inland are scaled from maps and are approximate.)

Finding 63 - It appears that the ICOADS database has not been checked to ensure that the given locations are in fact at sea and, where possible, are consistent with the path of the relevant ship. The latter might not be easy but as indicated above, it has been done by others.

9 Issues with observation station data prior to HadCRUT4

9.1 Introduction

Errors and uncertainties in the HadCRUT4 and CRUTEM4 databases do not rest with those two datasets alone. Data from observation stations is potentially erroneous and uncertainties from the moment of measurement right to its inclusion in the two datasets. This section will discuss some of those issues, from the recording of individual temperatures through to any processing by the national meteorological services (NMSs) that supplied the data used by the CRU.

Land accounts for 29% of the Earth's surface and water 71% but the grid-based system and the merging of land and sea data for coastal grid cells mean that the contribution of data from over land to the total HadCRUT4 coverage is often more than 29%. Figure 4.10 (chapter 4) shows that the contribution peaked at 56% in 1945. From 1939 to 2012 observation station data consistently contributed over 40% of the total HadCRUT4 coverage, although some of the coverage of land-based data is for coastal grid cells. These levels of contribution give importance to issues with data from observation stations quite.

9.2 Compliance with WMO standards for monthly averages

Neither Jones et al (2012) nor Osborn & Jones (2014) specify the minimum acceptable number of days of data in order that the monthly mean value is acceptable. WMO 100 (2011) says (pg. 4-17) "*It is recommended that a monthly value should not be calculated if more than ten daily values are missing or five or more consecutive daily values are missing*". Later, on the same page we find " '*The Calculation of Monthly and Annual 30-Year Standard Normals' (WMO/TD-No. 341) recommends stricter criteria for calculating averages, with the limits being more than five missing days in total, or more than three consecutive missing days*."

We simply do not know if these WMO standards have been followed by all national meteorological services. Judging by information mentioned earlier this seems unlikely.

Finding 64 - The metadata for each station gives no indication as to whether the national meteorological services complied with WMO recommendations regarding the minimum number of days of data from which monthly mean temperature are calculated, both within the 1961-90 period (which could impact all temperature anomalies throughout the entire record) and outside it.

9.3 WMO station classifications and their error margins

The WMO does not define minimum standards for observation stations but defines a set of station classes based on certain standards (WMO 8, 2017). Table 9-1 (below) shows the major requirements of each class and, for three of the classes the uncertainty estimated by the WMO.

Station classes might vary over time as the station is relocated or the local environment changes. There is no reason to assume that the class of a given station has been unchanged since the station commenced reporting or that the station class did not change during the period from 1961 to 1990. Together these mean that the uncertainty associated with the data from a given station could well have changed over time and consequently the uncertainty with temperature anomalies.

Issues with a single station are one problem but the HADCRUT4 and CRUTEM4 datasets use average temperature anomalies from multiple stations, which means a mixture of uncertainties from each station.

Factor	Class 1	Class 2	Class 3	Class 4	Class 5
Flat open ground sloping less than 1/3 (19 deg.)	Yes	Yes	-	-	Not meeting class 4
Natural representative vegetation, maximum height	10cm	10cm	25cm	-	
Distance to artificial heat or reflective sources (#1)	>100m	>30m	>10m	-	
Distance to expanse of water (#2)	>100m	>30m	>10m	-	
A heat source or expanse of water maximum as (X,Y) meaning X% of radius Y metres surrounding the screen	(10,100) or (5,10-30) or (1,10)	(10,30) or (5,5-10) or (1,5)	(10,10) or (5,5)	(50,10) or (30,3)	
Minimum angle for clear sun (i.e. no shade) (#3)	5 deg	7 deg	7 deg	20 deg	
Additional estimated uncertainty due to siting	**-**	**-**	**1°C**	**2°C**	**5°C**

Table 9-1 WMO station classes

(Notes: **#1** - includes buildings, concrete surfaces, car parks; **#2** - includes lakes, ponds, irrigated areas and with proviso "unless significant of the region"; **#3** - the corresponding tangents are 5 deg: 0.0875, 7 deg: 0.1228 and 20deg: 0.3640, which means heights of 8.75%, 12.28% and 36.40% of the horizontal distance to the shading obstacle.)

WMO 8 (2017) also describes the standards that should be followed for the design and operation of observation stations. For example, it points out that good airflow within an instrument enclosure is essential because without that airflow the difference between the temperature of the outer walls of the enclosure might be "*markedly different from the air temperature*", "*perhaps reaching 2.5K and -0.5K respectively in extreme cases*."

The station metadata provided by the CRU gives no indication of station class or changes to it, nor does it give any indication as to whether the observation stations fully comply with WMO standards. According to WMO standards, the uncertainty from siting alone could be as much as 5°C, plus a further 2.5°C if the instrument enclosure is poorly designed and constructed. The uncertainty for a given station is likely to change over time as the station is relocated and the enclosure repaired or replaced.

Finding 65 - The processing for the HadCRUT4 and CRUTEM4 datasets fails to recognise the uncertainties associated with siting and the types of enclosures on even a static basis, let alone that they might change over time. The uncertainties associated with various forms of siting and various enclosure are individually much greater than the increase in the annual HadCRUT4 global average temperature anomaly since 1850.

9.4 Measurement and sampling error

Brohan et al (2006) and Morice et al (2012) both discuss measurement errors in general terms and say that such errors are dealt with at a grid cell level rather than when dealing with individual stations. In other words, issues with the accuracy of individual station recordings are ignored in favour of some generalisation.

In discussing factors leading up to the adjustment at grid cell level Brohan et al (2006) says that the "*random error in a single thermometer reading is about 0.2C (1 standard deviation)*" and cites Folland et al, (2001) for this information. Folland et al (2001) actually says "*We estimated the two standard error (2σ) measurement error to be 0.4°C in any single daily [land surface-air temperature] observation*", indicating that the error margin is based on two standard deviations rather than one. Brohan et al (2006) goes on to say "*So the error in the monthly average will be <u>at most</u> 0.2/ √60 = 0.03C*" (my underlining) when in fact it will be at least double that figure.

It is worth noting that error margins are derived from the available data and it is assumed to be a good reflection of the data that is missing because all of the data is assumed show a normal distribution pattern about a monthly mean. Both assumptions could well be false.

Frank (2010) suggests, on the basis of a number of factors, that the error margin for a station anomaly is 0.46°C. He points out that Brohan et al (2006) and Folland et al (2001) used assumptions of station noise on the basis that no surveys of stations had been performed at a scale approaching global. In an empirical attempt to determine temperature uncertainty he shows that the average noise uncertainty for annual mean and 30-year mean temperatures is 0.2°C. Given that a temperature anomaly is derived from the two values, each with their own independent noise-uncertainty error margin, the uncertainties must be combined in quadrature (i.e. square root of the sum of the squares). This produces $\sqrt{(0.2°C^2 + 0.2°C^2)}$, which is 0.283°C. Frank (2010) adds a lower limit instrument uncertainty of 0.359°C and combines it with the 0.283°C in the same quadrature fashion to conclude that the error margin for a single temperature anomaly is 0.46°C.

Finding 66 - The CRUTEM4 and HadCRUT4 temperature anomaly datasets are created from temperature data that was possibly in error from the moment that it was recorded, but the magnitude of the error margin at either a station level or as a grid cell average is unclear (besides which the sources that supply data vary every month).

9.5 Mercury thermometers v. fast-reacting electronic thermometers

WMO 8 (2017) notes (pg. 65) "*the temperature of the air continually fluctuates up to one or two degrees within a few seconds*". It goes on to list and discuss seven possible errors with all liquid-in-glass thermometers (pg. 70), four with electrical resistance thermometers (pg. 80) and seven with thermocouples (pg. 80).

Later WMO 8 (2017) recognises the problem that electronic thermometers react faster that mercury-in-glass thermometers and how this might record the short-term fluctuations in temperature that glass thermometers might not. It says (pg. 532), "*Electrical thermometers usually have a short time-constant and, when sampled by fast electronic circuits, their output* *reflects high-frequency low amplitude fluctuations of the local temperature. This problem can be avoided by using sensors with a long time-constant, by artificially damping the response with a suitable circuit to increase the time constant of the output signal, or by averaging digitally the sampled outputs in the CPU*."

Finding 67 - The CRUTEM4 station metadata fails to mention if the station is automated and if so, when the change from manual instruments occurred and whether the station is fully compliant with WMO standards regards recording temperature data with fast-acting electronic sensors in a manner comparable to that for slower-acting mercury-in-glass thermometers.

9.6 Issues with shifting observation times

On the matter of the time of observations WMO 100 (2011) states "If conditions dictate that only one observation a day is possible, this observation should be taken between 0700 and 0900 local standard time. ... In selecting the schedule for climatological observations, times at or near the normal occurrence of daily minimum and maximum temperatures should be avoided."

For the sake of convenience, we will call the observation that marks the end of a 24-hour period and the start of the next the “principal observation”. The common practice is that this observation is at 9:00am, which is unfortunate because it risks the temperature just after 9:00am one morning being the minimum temperature for the next 24-hours.

In this case the “time of day” bias could well cause lower minimum temperatures to be reported on consecutive days than would have been the case if the temperatures had been recorded at for example, 12:00 midday, when there is less risk of temperatures being the minimum or maximum for the next 24 hours.

Some stations originally recorded the data at other times (e.g. 3:00pm), which had their own risks of carried-over minimum or, more likely, maximum temperatures. The data for those stations has been adjusted, using statistics and assumptions, to estimate the minimum and maximum temperatures had they been recorded at 9:00am. The accuracy of such estimates is impossible to determine.

Finding 68 - The temperature data for many stations is likely to be biased low as a consequence of the recommended time of the principal observation. In the case of stations where the principal observation time was changed to between 0700 and 0900 the data has been artificially altered but no indication is shown in the station metadata. Because of these adjustments the HadCRUT4 and CRUTEM4 datasets seem likely to be merging data that has different levels of confidence.

9.7 Issues with daylight-saving

In chapter six, section 6.3, it was mentioned that the data might be distorted downwards because 9:00am when daylight-savings is in operation is equivalent to 8:00am when it isn't. Figure 6.3 showed an example of a carried-over minimum temperature that would have been almost 3.0°C higher if daylight-saving not been operating and it discussed the consequences for the calculation of station Normals.

Using that same data, showing 11.3°C at 9:00am daylight-saving time (i.e. 8:00am normal local) and 14.2°C at 10:00am daylight-saving time (i.e. 9:00am normal time), consider the situation if a temperature of 13.7°C had been recorded at 6:00am the next morning. With daylight-saving in operation the minimum temperature for the 24 hours would still be 11.3°C but if daylight-saving wasn't in operation the minimum temperature would be the 13.7°C recorded at 6:00am.

This situation shows that the use of daylight-saving increases the probability that a temperature at 9:00am (local time, daylight-savings) might be the minimum for the next 24 hours. Despite this, document WMO 100 (2011) says *"If daylight-saving time is used for a part of the year, the observations should continue to be made according to the fixed local time …"*

Ultimately the data that is submitted for inclusion in the HadCRUT4 dataset might, for the same calendar month, be derived from a mixture of temperatures some of which were recorded under daylight-saving and some that were not. Further, some HadCRUT4 grid cells cover a mixture of locations, some of which have adopted daylight-saving and others that do not. It is possible that a 9:00am temperature at one station will be its minimum for the next 24 hours whereas another station recording identical temperatures but located just a few kilometres away, where daylight-saving does not apply, will have a quite different minimum by 9:00am the next day.

As mentioned earlier, many countries adopted daylight-saving in the 1970s and 1980s. Some countries adopted daylight-saving and later abandoned it, meaning that some temperature data is more likely to being biased towards lower mean temperatures and more recent data is not.

Recent media reports indicate that several European countries are reconsidering daylight-saving. If the practice is abandoned then the bias towards lower values during the months when daylight-saving operates would be removed, which would mean that an increase in average monthly minimum temperatures is likely. If the media reports of daylight-saving time applying all year (i.e. no change of clocks) then the bias towards lower recorded temperatures will increase.

With the benefit of hindsight, the recording of temperatures at midday when daylight-saving is not applied and 1:00pm when it is would have minimised "time of day" bias and removed the influence of daylight-saving on temperatures, but historical data is not recorded to this pattern and any adjustment would add its own uncertainty.

Finding 69 - Given that local minimum and maximum temperatures are recorded as at 9:00am clock time, the adoption of daylight-saving time and continuing to make temperature observations at 9:00am local time (i.e. 8:00am without daylight-saving) is an inconsistency that is likely to result in lower monthly average minimum temperatures than when daylight-saving is not used.

Finding 70 - Certain HadCRUT4 and CRUTEM4 grid cells will be susceptible to having data from some observation stations from regions that adopt daylight-saving and some that do not.

9.8 General issues with temperature data adjustments

Temperature data is adjusted (aka homogenised) in an attempt to compensate for non-meteorological changes in the recorded data. For land-based observation stations these changes might be caused by the replacement of instruments or the screening around the instruments (typically a Stephenson screen), changes to the local environment or the relocation of the observation station. The aim of homogenisation is to bring historical data into line (theoretically at least) with all of the observations being made at the current location with the current screening and current instruments. A minor exception to this is when a new station has recently started operation in parallel with an old station but the adjustment hasn't yet been calculated.

The WMO has much to say on the topic, particularly WMO/TD 1186 (2003) which has a list of 14 methods of identifying steps in the data that require adjustment and sometimes a related technique of making that adjustment. It also notes that one problem with homogenisation is that "... on most occasions the magnitude of the inhomogeneities is the same or even smaller than that of true climate-related variations", meaning that it can be difficult to determine natural variation from those with a non-meteorological cause.

The methods listed by WMO/TD 1186 (2003) often have variations. One example is the Standard Normal Homogeniety Test (Alexandersson, 1986; Alexandersson and Moberg, 1997; Menne and Williams, 2009) which looks for changes in the relationship between temperature data at one site and temperature data from one or more "reference" sites. But which relationship and what type of comparison?

Some variations use all neighbouring stations as reference stations but others use only neighbouring stations with strong correlations to the target site (e.g. Hausfather et al, 2013). Some compare the data from the target station to the data each individual from reference station while others merge the data from those neighbouring stations into a composite reference sequence (e.g. Tuomenvirta, 2002). Some determine the relationship using the actual recorded temperatures but use the difference between successive recorded temperatures at each site (e.g. Peterson & Easterling, 1994) or it might be based on the normalised difference (e.g. Alexandersson & Moberg, 1997; Toreti et al, 2010) or even on temperature percentiles (Trewin, 2013). Typically any adjustment is made using some method associated with the method used to identify the problem.

The methods described by WMO/TD 1186 (2003) are of two basic types - one comparing data from a target station to a single other station and the other comparing that data to data from, or based on, multiple other "reference stations". The first type is typically used when an existing station and its intended replacement are run in parallel for a few years to collect data that will determine how the old data should be adjusted. The second type is when there is no new station but changes seem to have occurred at the old station.

Some significant assumptions are made in both methods:

a) That weather systems influence all of the relevant stations in a consistent manner (which usually means that it was large enough to do so and that for weather system X the effect on recorded data from any given station is always Y)
b) Related to the above but encompassing more issues, there is general similarity of data from target and reference stations and a general one-to-one correspondence in data from different sites (e.g. stations with substantially similar exposure to weather conditions)
c) That all stations are sited well and have no issues with vegetation, landforms casting shadows, bodies of water or anything else in the local environment that might distort temperatures (as per the WMO station classes discussed earlier in this chapter).
d) That any data adjustments made to the data from the comparison station(s) were accurate and correct
e) That any comparison stations whose data has not been adjusted have not themselves suffered from inhomogeneities

As mentioned near the start of this chapter, the WMO advises that some classes of observation station should have certain error margins. The correct adjustment of temperature data should therefore take these into account, probably with error margins applied in quadrature to successive adjustments of the same data (e.g. two adjustments where error margins are each 2°C would produce an error margin of $\sqrt{(2^2 + 2^2)} = \sqrt{8} = 2.8$°C).

Two examples will illustrate some of these issues. Temperature data prior to 1961 for Orbost (Victoria, Australia) was adjusted by comparing it to neighbouring sites, some more than 200 km away, in various geographic settings. Figure 9.1 shows wind roses at four locations near some of the comparison sites. The central circle, the hub of the spokes, indicates the percentage of time that conditions were calm and the radial lines indicate wind speed (thickness of radial) and for each direction the percentage of time (distance outward from the centre, concentric circles at intervals of 10%) the wind was from that direction and at that strength. The wind roses indicate quite different patterns of prevailing winds at these sites, which suggests that the relationship between sites for given weather systems is likely to be inconsistent (points a and b above).

The Orbost historical temperature data was also adjusted after the observation station was automated and relocated. The old and new stations ran in parallel from 2000 to 2007 before the old station was closed. Figure 9.2 shows the range of minimum temperatures below 9.0°C recorded at the new station (Australian ID: 84145), on the y-axis, for different minimum temperatures at the old station (ID: 84030), on the x-axis. The range of temperatures at the new site often exceeds 2.0°C. There is no one-to-one correspondence between the data from the two sites and the average might be incorrect in a given situation by 1.0°C or more.

The other important aspect to adjustments is that usually all prior data is adjusted and that historical data might have been adjusted multiple times for different situations.

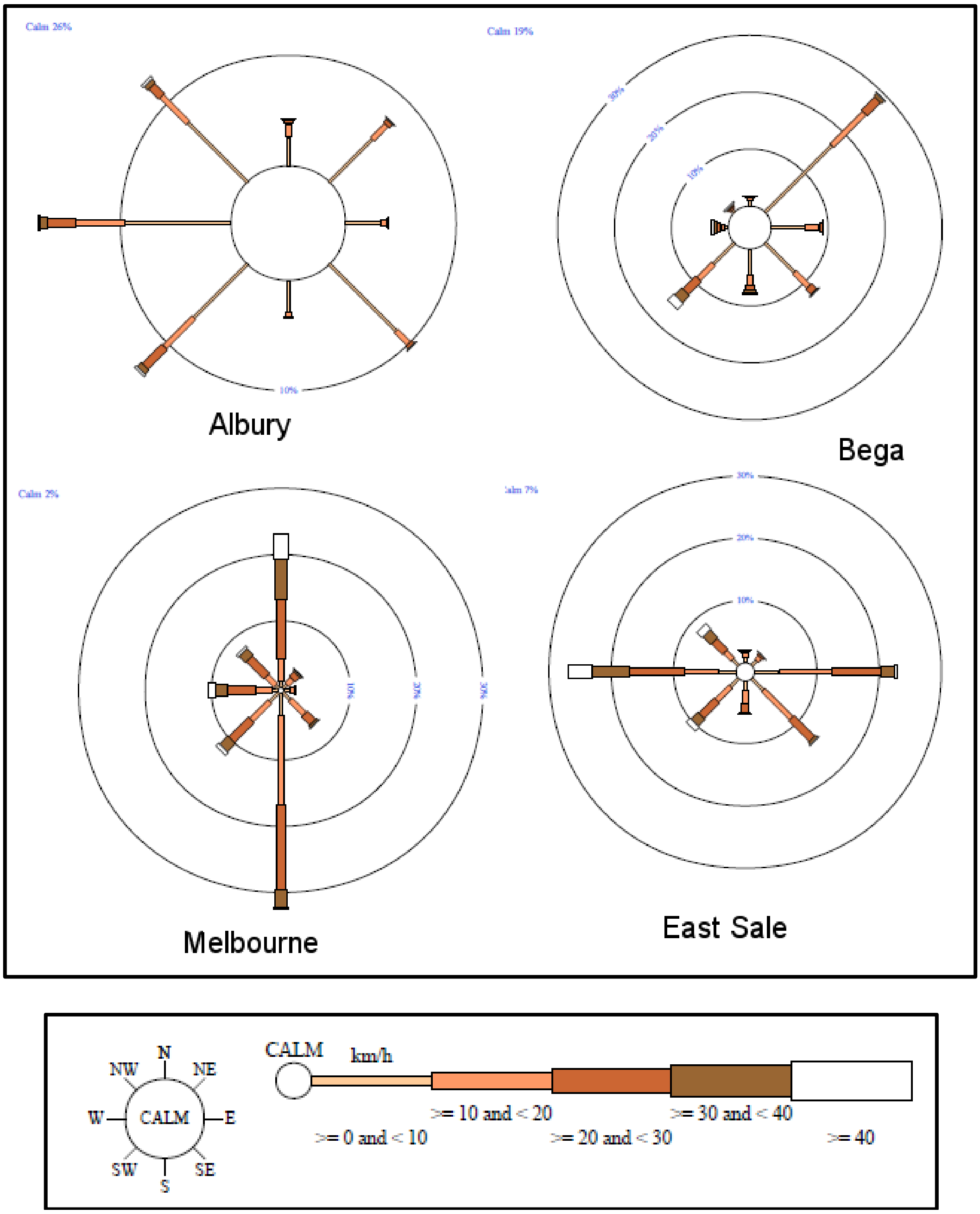


Figure 9.1 The long-term 3pm wind roses at four locations near sites whose temperature data was used to adjust temperature data from a fifth site. Calm conditions occur 26% of the time at Albury, 19% at Bega, 2% at Melbourne and 7% at East Sale. (Data from Australia's Bureau of Meteorology)

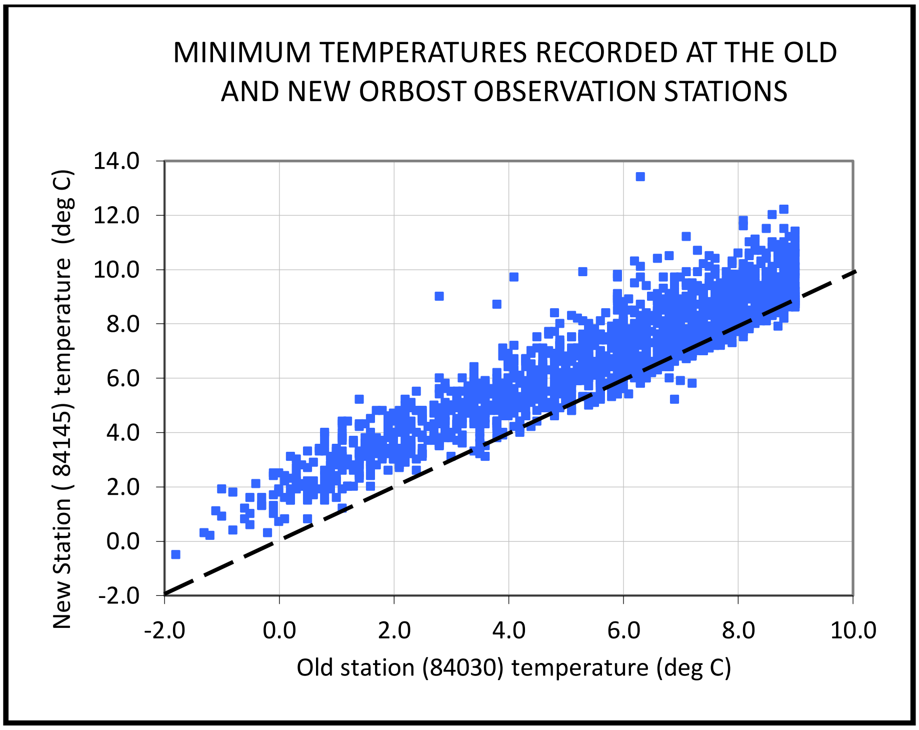


Figure 9.2 Minimum temperatures at the new Orbost observation station (y-axis) measured for each temperature below 9.0°C recorded at the old Orbost station (x-axis). The broken line indicates equal temperature at both sites.

Finding 71 - Much of the observation station data used in HadCRUT4 is likely to have been adjusted but there is no record of the method used or the amount of the adjustment, which makes the data impossible to independently audit.

Finding 72 - The station temperature data used in HadCRUT4 and CRUTEM4 has likely been adjusted one or more times and any errors in the adjustment likely been compounded, as the error margin should have been.

9.9 Issues with adjusting for gradual distortion

The gradual distortion of temperature measurements is a common problem with observation station temperature records. The distortion might be due to the growth of vegetation causing shading, the blocking of certain winds or to urbanisation near the observation station. As noted above, adjustments to temperature data are cumulative, which would mean that if data is incorrectly adjusted each time and in the same direction (e.g. old data too low) the cumulative adjustments will result in the oldest data (i.e. most adjusted data) being the most incorrect and greatly distorting the temperature record. This section discusses a possible cause of incorrect data adjustments for situations where urbanisation is an issue.

Hansen et al (2001) mentioned a concept called “undisturbed temperature” that referred to temperature in an undisturbed (i.e. constant) local environment. The paper didn’t fully explore the situation and its implications but it provides a useful perspective on external influences on measured temperature.

"True Temperature" can be defined as the temperature at a given location with its fixed geography and baseline physical environment. The temperature is "true" in the sense that it is consistent with the location, the environment and meteorological conditions.

"True Temperature" will vary throughout the calendar year because of the changes to the position of sunrise and sunset, the changes in the angle of incidence of insolation, cyclic changes in the local environment and of course changes in meteorological conditions. The key characteristic with True Temperature is that the local environment does not change other than through a natural cycle that has the same conditions at the same point in that cycle (e.g. leaves on trees, shadows cast).

Urbanisation is usually an increasing influence, not consistent in a static or cyclic sense, so over time the recorded temperature will slowly move away from True Temperature. The adjustment of temperature data needs to take into account this divergence from True Temperature, but it often appears that it fails to do so.

Figure 9.3 shows a station initially located at Site 1 and recording True Temperature 1 before becoming urbanised (line segment B to C), then being relocated to Site 2 where it records True Temperature 2. The x-axis of the Figure indicates time but has no scale because the exact times are unknown. The straight-line segment B to C is a simplification because the effects of urbanisation are irregular and depend on the influence (magnitude and direction) and its proximity to the site. At times the effect might even be negative, such as with the removal of a wind block.

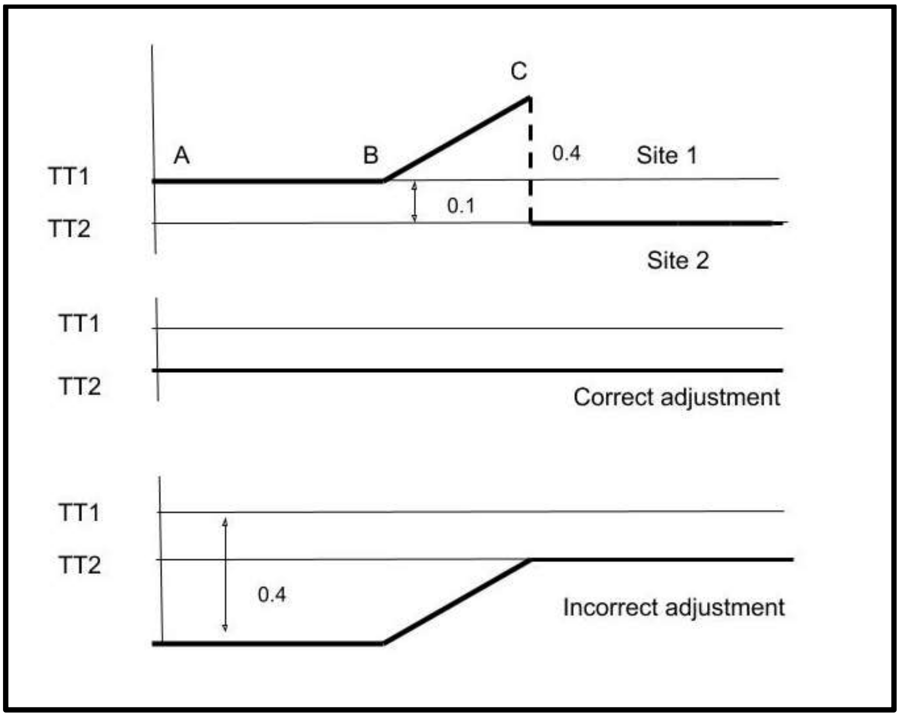


Figure 9.3 Schematic of the influence of urbanisation on an observation station and the correct and incorrect methods of temperature data adjustment

The difference in True Temperature at the two sites is 0.1°C but at C, the point in time when the station is relocated to the new site, the difference is 0.4°C. The correct adjustment, to convert the data recorded at Site 1 into the equivalent at Site 2, would be to remove the urbanisation influence at Site 1 and then decrease all the data recorded at Site 1 by 0.1°C. Failing to consider that Site 1 did record correct temperatures when it started operation, and perhaps for many years after that, causes the inaccurate adjustment of temperatures.

The correct adjustment is not easy because it requires knowledge of when the divergence from True Temperature occurred, the point in time noted as 'B' on the Figure and the difference in True Temperature at the two sites despite the fact that it relies on conditions at Site 1 at some time in the past. If the adjustment was to be very accurate the pattern in divergence from B to C would need to be known because, as mentioned earlier, the straight line shown in Figure 9.3 is an approximation.

When an observation station starts operating it records True Temperature and would continue to do so if no gradual, external, non-meteorological influences distorted the measurements. Any temperature adjustment should therefore, at the most, taper back from the calculated value at the end of the station's reporting to zero adjustment for urbanisation when the station began operating. Documentary evidence might indicate that urbanisation did not start until later, in which case the data adjustment should cease at that later date. The difference in True Temperature at the two sites is more difficult to determine but is necessary if the adjustment is to be accurate.

It is incorrect but seemingly common to apply a constant adjustment to all data from the first site according the calculated difference between it and the second site when the data at the first site has been distorted by gradual changes in the local environment. It defies logic that a location that was abandoned because it was contaminated by external influences was equally contaminated when the observation station was first established there and throughout the entire time that the station was at that site, but this is what a constant data adjustment implies. In the case of Figure 9.3 the decreasing all of the data recorded at Site 1 by a constant 0.4°C leaves the earlier data recorded at Site 1 0.3°C below its correct value

As mentioned earlier, it is not only data from the previous site that is adjusted when moving to a new site. All of the data prior to that the station being at that site is also adjusted to theoretically make it equivalent to being recorded at the newest site.

The impact of a sequence of flawed constant adjustments is illustrated in Figure 9.4. In this simplified example the observation station has been moved three times, each site having the same True Temperature and each adjustment being made after a period of 20 years when the recorded temperature was 0.5C above True Temperature. To further simplify matter the True Temperature at every location is identical. This situation might be thought of as urbanisation influencing temperatures at an observation station, that station being moved multiple times but urbanisation repeatedly reaching it.

The top portion of the Figure represents the sequence of relocations. True Temperature is recorded for a certain period and then urbanisation means that the recorded temperature is increasingly above True Temperature. The correct data adjustments are shown immediately below that, with each relocation requiring that all historic data is adjusted in order that the record be consistent with recording all data from the latest location. The bottom portion of the Figure shows the impact of the flawed constant adjustments to the data, with the second adjustment decreasing even further the data that has already been wrongly adjusted, and the third decreasing it again. In this case the earliest data is decreased 1.5C below the correct value.

The end-to-end difference is 1.5C, consistent with the flawed magnitude of the decrease in the earliest data from the first location. The trend will be around 1.2C/60 years (or 1.4C/century) because the data doesn't increase for the entire period. The real temperature at the site would need to have a falling trend of the same size in order to cancel out this trend that results from flawed data adjustment. Conditions in the real world will be more complex but the principle remains the same – a constant adjustment to data that has been subjected to gradually increasing urbanisation will create a rising temperature trend that does not exist in the true temperature.

The situation of falsely adjusting data by a constant amount is detailed further in Appendix 5, part 1, where two examples with different patterns of influence on measured temperatures are discussed.

Any flawed adjustments will be carried forward into regional or national records. They will be more evident when urbanised relocated stations are a significant proportion of that data. A likely example of this is to be found in New Zealand's principal national temperature record. This record and the findings of a paper discussing the data's incorrect adjustment can be found in Appendix 5, part 2.

This discussion has raised two important points. Firstly, where the temperatures recorded at an observation station have been subjected to gradual non-meteorological influences and the station eventually relocated, the adjustment of all previous data by a constant amount, the average difference in temperatures at the two sites, is flawed. It implicitly assumes that old location always suffered from exactly the same influence as it did when the station was relocated, and while it will adjust the recorded temperature data it will not remove the non-meteorological trend. The effect of adjustments is cumulative because all previous data is adjusted each time. This means that in the case of gradually increasing urbanisation causing multiple station relocations over time, the temperatures recorded earlier will be excessively decreased multiple times.

Secondly, with the situation described above, the false adjustment of temperatures since 1961 will decrease the Normals (i.e. long-term averages) calculated over the period from 1961 to 1990. This will not alter the historic temperature anomalies for the station because all data that is adjusted will be adjusted by a constant amount and the relationship pf temperatures to Normals will remain the same. The anomalies for the recent, unadjusted data will increase because the Normal has decreased. A second post-1961 adjustment will decrease the Normals even further, adjust the previously unadjusted data in line with the again-decreased Normals (which means the excess anomalies will be unchanged) and the temperature anomalies for the most recent, unadjusted data will be calculated using Normals that have decreased even further. Data adjustments since 1961, and even since 1990, are very likely because of station relocations caused by either a switch to automated weather stations or an increased awareness of the need for observation stations to be as free as possible from non-meteorological influences.

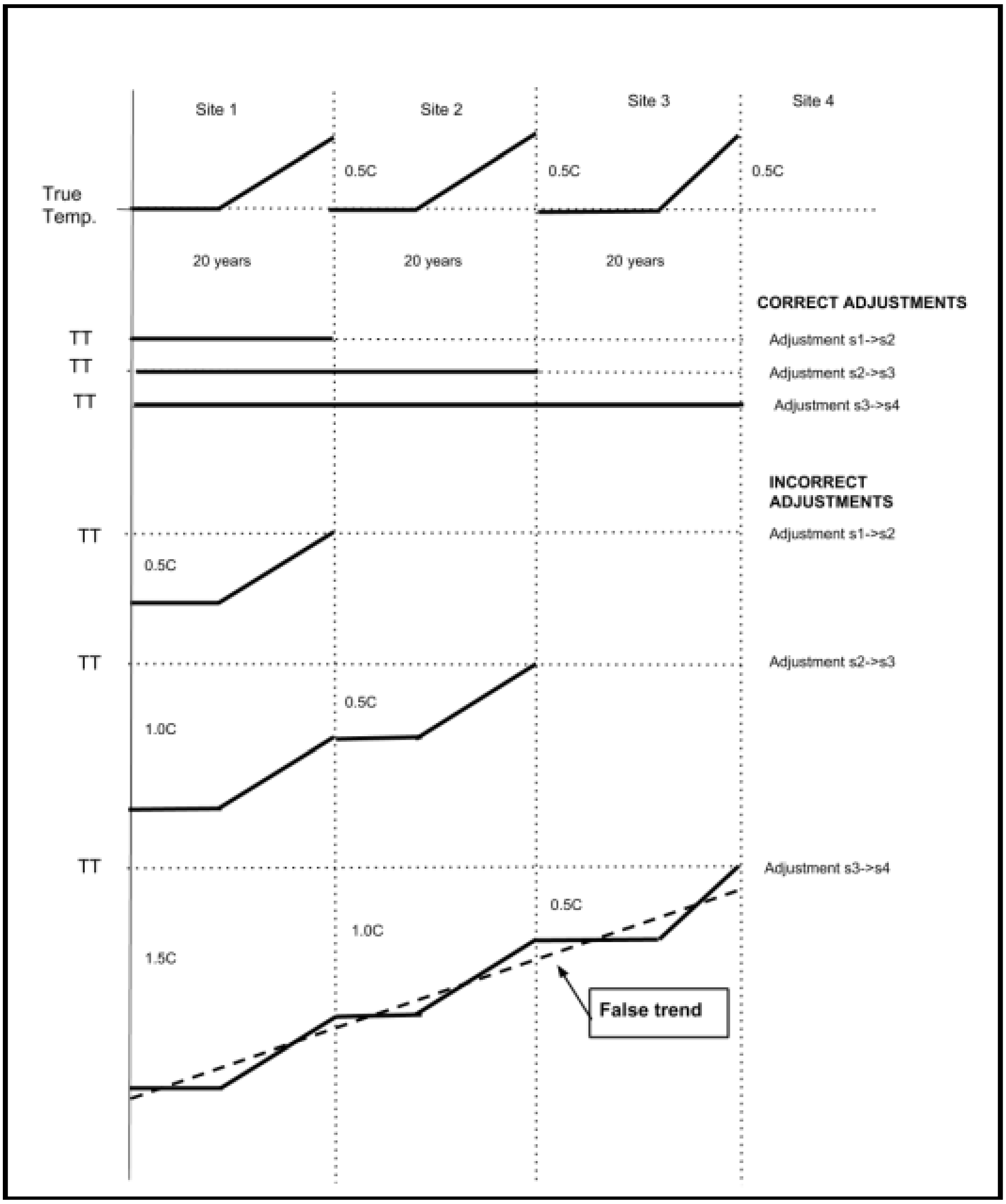


Figure 9.4 A simplified illustration of the distortion of data and generation of false trends caused by adjusting gradually distorted temperature records by a constant amount

Finding 73 - Where non-meteorological influences on recorded temperatures have gradually changed over time, the adjusting of all previous temperatures by a constant amount falsely assumes a constant distortion of all prior data and creates a false temperature trend. For station suffering increasing urbanisation this will typically mean that earlier data is excessively adjusted downwards, producing a false warming trend

Finding 74 – Adjustments of the type described above in Finding 73 that occurred after 1961 will falsely adjust the Normals (i.e. long-term average temperatures) from which temperature anomalies are calculated. The temperature anomalies for all data adjusted by that constant amount will be unchanged but temperature anomalies for recent unadjusted data will be calculated using altered Normals. In the case of increasing urbanisation, the decreased

Normals will mean a false increase in temperature anomalies, leading to exaggerated warming trends in the HadCRUT4 dataset.

9.10 Data not adjusted after station closures

The descriptions of data homogenisation from the CRU, WMO and others indicate that adjustments are usually made for changes to instruments, screening and the location of the site. The last factor is particularly linked to urbanisation of the local environment, which distorts the temperature. There appears to be no mention of adjustment to data when a clearly urbanised observation station ceases operation.

Jones (2012) reported that some countries only update the station data at the end of each decade, meaning that the published station data might not be accurate after 2009. By the end of that year 1338 CRUTEM4 stations had already ceased reporting data. There are another 1460 stations that had not reported data since the end of 2015, suggesting that as many as 2798 (36.8% of the 7599 seemingly acceptable stations) might have closed.

Without detailed investigation it is impossible to know which of these closed stations might have suffered from uncorrected urbanisation. The point can however be illustrated by the list of closed United Kingdom observation stations including Cambridge, Glasgow, Edinburgh and York, all of which have substantial urban areas.

Finding 75 - It seems very likely that when urban stations are closed rather than relocated their data is not adjusted for urbanisation and any warming due to non-meteorological influences will have remained in the HadCRUT4 record.

10 Conclusions

This report has identified more than 70 issues of concern regarding the HadCRUT4 dataset. They encompass the dataset's creation from the measurement of temperature on land and at sea, through the adjustment of data, the methodology used in the dataset, the processing to create the dataset, to the credibility of statements such as "global average".

The dataset has been shown to include erroneous temperature data, both on land and at sea, and much data of doubtful accuracy. The source data from observation stations and the ICOADS database have been shown to contain errors. It seems that neither the Climatic Research Unit or the Hadley Centre audit the data that they use. Worse, the Climatic Research Unit incorporates erroneous data into its calculations of station Normals and the standard deviations that are used to determine the range of data will be included in the processing that creates the HadCRUT4 and CRUTEM4 datasets.

A general summary of the findings is as follows:

(a) Failure to properly check source data for errors, query and correct it or reject it, particularly prior to calculating long-term average temperatures and standard deviations for observation station data. Examples:
  - Obvious errors in observation station metadata and temperature data.
  - Errors in the ICOADS database.

(b) At times grid cell values, hemispheric averages and global averages have been derived from so little data that they cannot be considered reliable. Examples:
  - A significant proportion of SST grid cell values are based on less than six temperature measurements in the month in question.
  - Many grid cells over land have a single reporting observation station.
  - Just one observation station reported data for the Southern Hemisphere in the first two years and five months of the data record and only two more stations at the end of the fifth year of the record.
  - Coverage (by HadCRUT4 methodology) was less than 50% of the Earth's surface for most of the first 100 years of data.
  - So-called "global" average temperature anomalies have at times been heavily biased towards certain areas of the world and at other times there is a lack of coverage in specific regions.

(c) Both SST and observation station data has been adjusted according to assumptions that might not be true. In the case of observation station data, the adjustments are cumulative and therefore errors might be compounded. Examples:
  - Adjustments to bring SST data into line with measuring by the same method each time involve heroic assumptions because the necessary information about the conditions was not recorded. They are likely to be flawed.
  - Adjustment to station data for gradual inhomogeneities is likely to fail to take into account the gradually-changing environment around the old observation station, in particular the situation when the station commenced operation, which could be very different to the situation when it ceased operation.
  - The bulk adjustments to SST data are questionable in their own right on the basis of coverage but SST data has been adjusted to approximately correspond to

temperatures over land, so if data from observation stations has been wrongly adjusted then those SST adjustments will also be flawed.

(d) Other issues. These are a variety of issues of which some examples are:

- Poor choice of the base period for calculating averages because of significant shifts in weather patterns.
- The sum of the SST anomalies when data for all 30 years from 1961 to 1990 is present is not always zero.
- An unknown amount of SST was recorded when ships were in port, which are very different environments to the open sea.
- Inclusion of extreme values of SST anomalies and of mean monthly temperatures on land.
- There are instances of large differences between the values for adjacent grid cells that suggest errors.
- Temperature anomalies for observation stations in the same grid cell can vary widely.
- It is unclear whether the measuring, adjusting and reporting of data from observation stations conforms to WMO standards.
- The close proximity of some stations to each other leads to bias towards meteorological conditions in their specific region of their grid cells.
- Data quality is poor: identification, documentation, identification of sources, relevant data is missing, data file formats inconsistent.
- The techniques used for sea surface temperature data appear to generate implausible temperatures.

The underlying situation is one of multiple inconsistencies.

- Temperature data measured on land has numerous differences to temperature data measured at sea including medium and nature (e.g. min. and max. temperature compared to time intervals),
- Coverage of the reported data,
- Amount of data (i.e. reporting stations or grid cells),
- Instruments, screening (on land) and techniques for measuring temperature,
- Different depths of measurements at sea and different heights on land,
- Physical location where measurements are made,
- Physical surroundings (i.e. local environment),
- Some data recorded during daylight-saving and some not,
- Variable influence of non-meteorological factors (e.g. urbanisation),
- Methods of data adjustment,
- Source of data for coastal/island grid cells (from land, from sea or merging of the two),
- Periods over which Normals and standard deviations are calculated (obs. stations only),
- Inconsistencies in the ICOADS database (e.g. differences between data from nearby ships),

The data prior to 1950 (i.e. the first 100 years of the data record) is particularly unsatisfactory for the calculation of global or hemispheric averages due to:

(a) Poor coverage and the bias towards certain regions.

(b) The high percentage of SST grid cells with data based on from 1 to 5 observations in entire months.
(c) The method of measuring sea surface temperature was less likely to be recorded.
(d) The low number of observation stations.
(e) The likelihood that observation station data prior to 1950 having been adjusted multiple times and errors in those adjustment been compounded.

In the opinion of this author, the data before 1950 has negligible real value and cannot be relied upon to be accurate. The data from individual stations might be satisfactory but only if local environments are unchanged and with no manual adjustments to the temperature data. The many issues with the 1850-1949 data make it meaningless to attempt any comparison between it and later data especially in derived values such as averages and the trends in those averages.

Even after 1950 the data is plagued by several issues, such as the adjustment of SST data, the ad hoc adoption of daylight-saving time and the introduction of electronic thermometers in observation stations, which required the adjustment of earlier temperature data.

The data from 1961 t0 1990 is also hardly a reliable basis from which to calculate long-term average temperatures. Sea surface temperatures are first estimated and then modified by temperature measurements that might be very few in number, and in some cases it seems that those estimates were not modified at all. On land there is a frequent problem with stations failing to report a mean monthly temperature in a given month. Without accurate long-term average temperatures derived from consistent sets of temperature data, the subsequent temperate anomalies cannot be sensibly combined to produce credible and reliable global averages.

It is also the opinion of this author that the HadCRUT4 data since 1950 is likewise not fit for purpose. It might be suitable for single-station studies or even small regional studies but only after being deemed satisfactory regards the issues raised in this report. It is not suitable for the derivation of global or hemispheric averages, not even with wide error margins that can only be guessed at because there are too many points in the data collection and processing that are uncertain and inconsistent.

The conclusion above and the findings of this report in general have several major implications:

i. The HadCRUT4 global annual average temperature anomaly supposedly increased from -0.176°C in 1950 to 0.677°C in 2017, an increase of 0.753°C. When all of the errors discussed in this report are corrected and appropriate error margins somehow determined for other factors, that 0.753°C is likely to decrease and the error margins increase. The new error margins might remove the statistical certainty that any global warming whatsoever has occurred since 1950. Even if warming was found to have occurred but only at 50% of the trend indicated by the HadCRUT4's current figures it might radically change public and political attitudes.

ii. The failings in the HadCRUT4 data cast serious doubt on the output of climate models. The common practice with models is to adjust them until they produce output as close to historical temperatures as possible. Adjusting them to match flawed data can only result in flawed models and models that are incorrect cannot be relied upon to (a) produce accurate estimates mankind's influence on temperature, (b) produce accurate predictions of future temperatures or (c) retrospectively derive accurate historical average global temperatures.

iii. It was shown in chapter one that the Intergovernmental Panel on Climate Change (IPCC) failed to audit the temperature data used in its 2013 report and used in the climate modelling that it commissioned so that its output could be cited. Nothing was discovered to suggest that the temperature data used in earlier IPCC reports was audited either. The findings of this report cast doubt not only on key claims in IPCC climate assessment reports but also the claims of the United Nations Framework Convention on Climate Change (UNFCCC), which the IPCC is tasked with supporting. In particular the Paris Climate Agreement is undermined because not only are current average global temperatures uncertain but the accuracy of supposed pre-industrial temperatures is called into question by this report showing the many flaws in the historical temperature data.

iv. Many of the issues identified in this report relate to shortfalls in amount of data, shortfalls in coverage and the issues with the measurement and adjusting of temperatures. These issues will inevitably apply to any alternative near-surface temperature datasets, such as those from NASA's GISS, because all such datasets inevitably rely on the same data. On this basis no near-surface temperature dataset is likely to be accurate and reliable.

The very large uncertainties about the HadCRUT4 dataset mean that it cannot be accepted as an accurate temperature record. It follows therefore that reports from the IPCC and other bodies that rely on this data cannot be regarded as accurate. Government commitments, policies and spending on climate issues based on such reports appear to have been made without proper due diligence. With the benefit of hindsight, proper audit of the HadCRUT4 dataset (or its predecessor versions) would have been wise before spending billions of dollars addressing supposed issues and trying to re-structure the world's energy market.

References

Alexandersson, H. (1986) A homogeneity test applied to precipitation data. *J. Climatol.*, 6, pp661-675.

Alexandersson, H. and A. Moberg (1997) Homogenization of Swedish temperature data. Part I: Homogeneity test for linear trends. *Int. J Climatol*., 17, pp25–34.

Brohan, P., R. Allan, J.E. Freeman, A.M. Waple, D. Wheeler, C. Wilkinson and S. Woodruff, (2009) Marine Observations of Old Weather. *Bull. Amer. Meteor. Soc.*, 90, pp219–230. doi: http://dx.doi.org/10.1175/2008BAMS2522.1

Brohan, P., J.J. Kennedy, I. Harris, S.F.B. Tett and P.D. Jones (2006) Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850, *J. Geophys. Res*. 111: D12106. doi:10.1029/2005JD006548

Brooks, C.F. (1926) Observing Water-Surface Temperatures at Sea, *Mon. Wea. Rev*., 54, pp241–253, doi:10.1175/1520-0493(1926)54<241:OWTAS>2.0.CO;2

Cowtan, K. and R.G. Way (2014) Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends Q.J.R. Meteorol. Soc. 140: pp1935-1944 July 2014 B DOI:10.1002/qj.2297

Donlon, C. J. and the GHRSST-PP Science Team (2005) The Recommended GHRSST-PP Data Processing Specification GDS (Version 1 revision 1.6). *GHRSST-PP Report Number 17*. The GHRSST-PP International Project Office, Exeter, U.K., 245 pp. (available at: ftp://podaac.jpl.nasa.gov/allData/ghrsst/docs/GDS-v1.0-rev1.6.pdf *)*

Emery, W.J., D.J. Baldwin, P. Schlüssel and R.W. Reynolds, (2001) Accuracy of in situ sea surface temperatures used to calibrate infrared satellite measurements *J. Geophys. Res.* v106 no. C2, pp2387-2405

Ewing, G. and E.D. MacAlister (1960) On the thermal boundary layer of the ocean *Science* 131 pp1374-1376

Fairall, C.W., E.F. Bradley, D.P. Rogers, J.B. Edson and G.S. Young (1996) Bulk parameterization of air-sea fluxes for Tropical Ocean-Global Atmosphere Coupled Ocean-Atmosphere Response Experiment. *J. Geophys. Res.*, 101, pp3747–3764.

Farrar, J. T., C.J. Zappa, R.A. Weller and A.T. Jessup (2007) Sea surface temperature signatures of oceanic internal waves in low winds, *J. Geophys. Res.*, 112, C06014, doi:10.1029/2006JC003947.

Fedorov, K.N. and A.I. Ginsburg (1992) The near-surface layer of the ocean, Koninldije Wohrmann, Utrecht, Netherlands, pp63-108

Folland, C.K. and F.E. Kates, (1984) Changes in decadally averaged sea-surface temperature over the world 1861-1980: pp. 721-727 in "Milankovitch & climate", pr 2, Eds. A. L. Berger et al., D. Reidel

Folland, C.K. and D.E. Parker (1995) Correction of instrumental biases in Historical sea surface temperature data *Q.J.R. Meteorolol. Soc.*, 121, pp. 319-367

Folland, C. K., N.A. Rayner, S. J. Brown, T.M. Smith, S.S.P. Shen, D.E. Parker, I. Macadam, P.D. Jones, R.N. Jones, N. Nichols and D.M.H. Sexton (2001) Global temperature change and its uncertainties since 1861, *Geophys. Res. Lett*., 28(13), pp2621–2624

Frank, P. (2010) Uncertainty in the Global Average Surface Air Temperature Index: A representative lower limit, *Energy & Environment*, vol 21, No. 8, pp969-989

Freeman, E., S.D. Woodruff, S.J. Worley, S.J. Lubker, E.C. Kent, W.E. Angel, D.I. Berry, P. Brohan, R. Eastman, L. Gates, W. Gloeden, Z. Ji, J. Lawrimore, N.A. Rayner, G. Rosenhagen and S.R. Smith (2017) ICOADS Release 3.0: a major update to the historical marine climate record, *Int. J Climatol*., 37, pp2211-2232 (pub. Online 27 June 2016)

Hansen, J., R. Ruedy, M. Sato, M. Imhoff, W. Lawrence, D. Easterling, T. Peterson and T. Karl (2001) A closer look at United States and global surface temperature change, *J.. Geophys. Res.* v106, pp 23947-23963

Hasse, L. (1963) On the cooling of the sea surface by evaporation and heat exchange *Tellus* 15 pp363-366

Hausfather, Z., M.J. Menne, C.N. Williams, T. Masters, R. Broberg and D. Jones (2013) Quantifying the effect of urbanization on U.S. Historical Climatology Network temperature records, *J. Geophys. Res*:*Atmos*, v118, pp481–494, doi:10.1029/2012JD018509

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

Kawai, Y and A Wada (2007) Diurnal Sea Surface Temperature Variation and its Impact on the Atmosphere and Ocean: A Review, *Journal of Oceanography*, v63, pp721-744

Kennedy J.J., N.A. Rayner, R.O. Smith, M. Saunby and D.E. Parker (2011b) Reassessing biases and other uncertainties in sea-surface temperature observations measured in situ since 1850 part 2: biases and homogenisation. *J. Geophys. Res.* 116, D14104, doi:10.1029/2010JD015220

Kent, E.C. and P.K. Taylor (2006) Toward Estimating Climatic Trends in SST. Part I: Methods of Measurement *Journal of Atmospheric and Oceanic Technology* v23 pp464-475

Matthews, J.B.R. (2013) Comparing historical and modern methods of sea surface temperature measurement - Part 1: Review of methods, field comparisons and dataset adjustments, *Ocean Sci*., 9, pp683-694

Matthews, J.B.R. and J.B. Matthews (2013) Comparing historical and modern methods of sea surface temperature measurement - Part 2: Field comparison in the central tropical Pacific, *Ocean Sci*., 9, pp695-711

McKitrick, R. (2013), Encompassing tests of socioeconomic signals in surface climate data *Climatic Change*, June 2013

McKitrick, R. (2010) A critical review of global surface temperature data products, SSRN Working paper 1653928, (75 pages), abstract online at http://ssrn.com/abstract=1653928

McKitrick, R. and N.Nierenberg (2010) Socioeconomic patterns in Climate Data, *Journal of Economic and Social Measurement*, 35:3,4, pp149-175

Menne, M. J. and C. N. Williams Jr. (2009) Homogenization of temperature series via pairwise comparisons, *J. Clim*., 22, pp1700–1717, doi:10.1175/2008JCLI2263.1.

Morice, C.P., J.J. Kennedy, N.A. Rayner and P.D. Jones (2012) Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set *JGR:Atmospheres* Volume 117, Issue D8 27 April 2012, doi:10.1029/2011JD017187

Osborn T.J. and P. Jones (2014) The CRUTEM4 land-surface air temperature data set: construction, previous versions and dissemination via Google Earth *Earth Syst. Sci. Data*, 6, pp61–68, 2014 www.earth-syst-sci-data.net/6/61/2014/ doi:10.5194/essd-6-61-2014

Peterson, T.C. and D.R. Easterling (1994) Creation of homogeneous composite climatological reference series. *Int. J. Climatol*., 14, pp671-679.

Rayner,N.A., P. Brohan, D.E. Parker, C.K. Foland, J.J. Kennedy, M. Vanicek, T.J. Ansell and S.F.B. Tett (2006) Improved Analyses of Changes and Uncertainties in Sea Surface Temperature Measured In Situ since the Mid-Nineteenth Century: The HadSST2 Dataset *J. Clim.* v19(3) pp446-469

Smith, T.M. and R.W. Reynolds (2004) Improved Extended Reconstruction of SST (1854-1997), *J. Clim.* v 17, pp2466-2477

Soloviev, A. and R. Lukas (1997) Observation of large diurnal warming events in the near-surface layer of the western Equatorial Pacific warm pool, *Deep Sea Research I*, v 44 pp1055-1076

Sylte, Gudrun Urd (24 May 2010) "Den aller kaldaste havstraumen". forskning.no (in Norwegian) (https://forskning.no/havforskning-klima-vaer-og-vind/2010/05/den-aller-kaldaste-havstraumen). Retrieved 24 May 2010.

Toreti,A., F.G. Kuglitsch, E. Xoplaki, J. Luterbacher and H. Wanner (2010) A Novel Method for the Homogenization of Daily Temperature Series and Its Relevance for Climate Change Analysis *J. Clim* v23, pp 5325-5331

Trewin, B. (2013) A daily homogenized temperature data set for Australia *Int. J. Climatol*. 33: pp 1510–1529 (2013), DOI: 10.1002/joc.3530

Tuomenvirta, H (2002) Homogeneity Testing and Adjustment of Climatic Time Series in Finland *Geophysica* (2002), 38(1-2), pp15-41

Webster, P.J., C.A. Clayson and J.A. Curry (1996) Clouds, Radiation, and the Diurnal Cycle of Sea Surface Temperature in the Tropical Western Pacific, J. Clim., v9 pp1712-1730

Wells, A.J., C. Cenedese, J.T. Farrar and C.J. Zappa (2009) Variations in Ocean Surface Temperature due to Near-Surface Flow: Straining the Cool Skin Layer, *Journal of Physical Oceanography*, v39, pp2685-2710

WMO 8 (2017) Guide to Meteorological Instruments and Methods of Observation *World Meteorological Organization*, Geneva (2014 edition updated in 2017)

WMO 100 (2011) Guide to Climatological Practices *World Meteorological Organization*, Geneva

WMO/TD 1186 (2003) Guidelines on Climate Metadata and Homogenization *World Meteorological Organization*, Geneva

Woodcock, A.H. (1941) Surface cooling and streaming in shallow and salt water *J.Mar.Res* 4 p153

Appendix 1 - Supplementary Information chapter 4

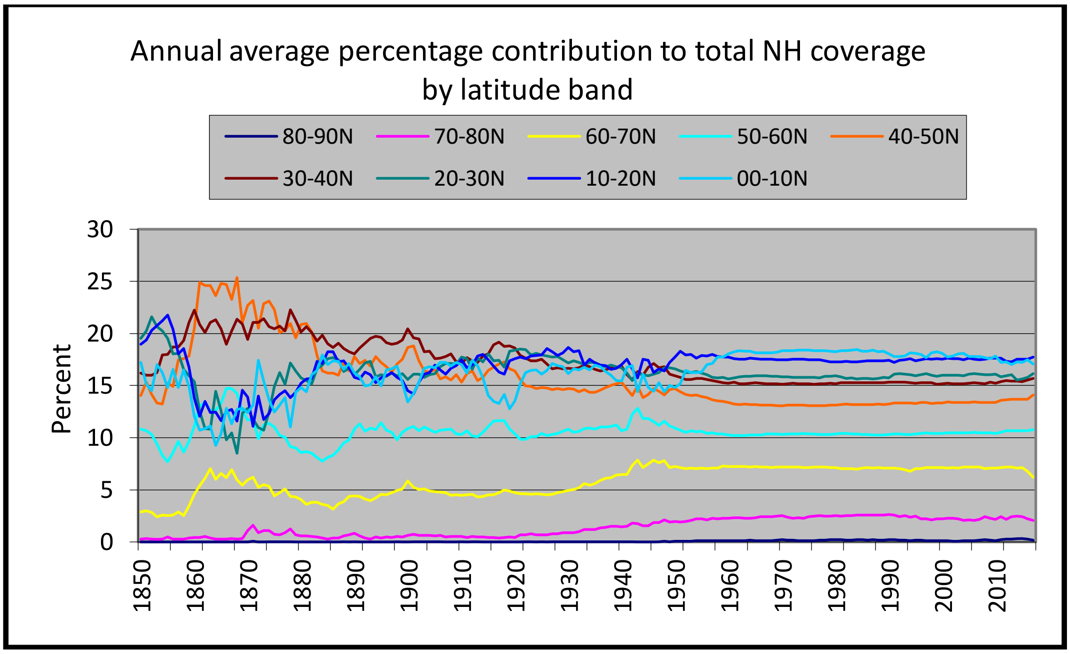


Figure A1.1 Percentage contribution to NH coverage by 10-degree latitude bands.

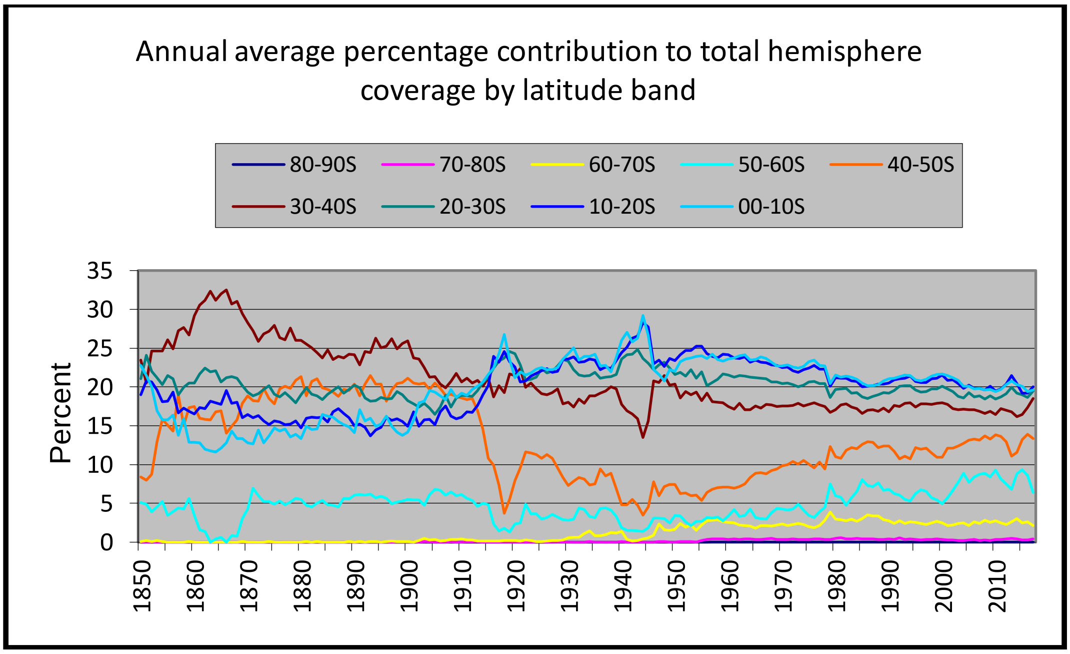


Figure A1.2 Percentage contribution to NH coverage by 10-degree latitude bands

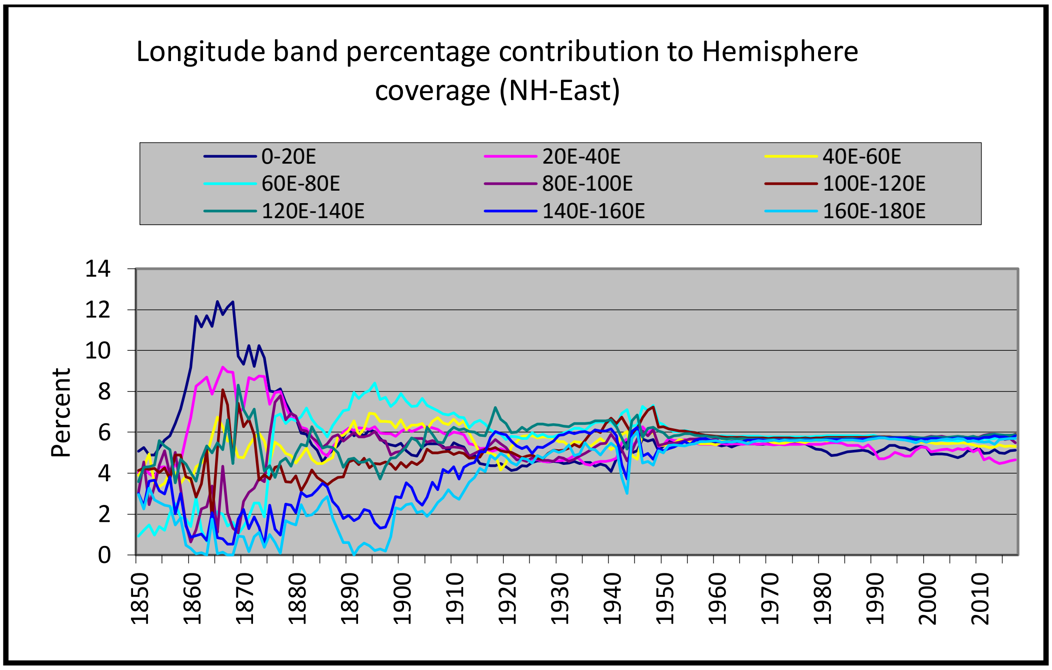


Figure A1.3(a) Contribution to total hemisphere coverage by longitude bands (NH, East)

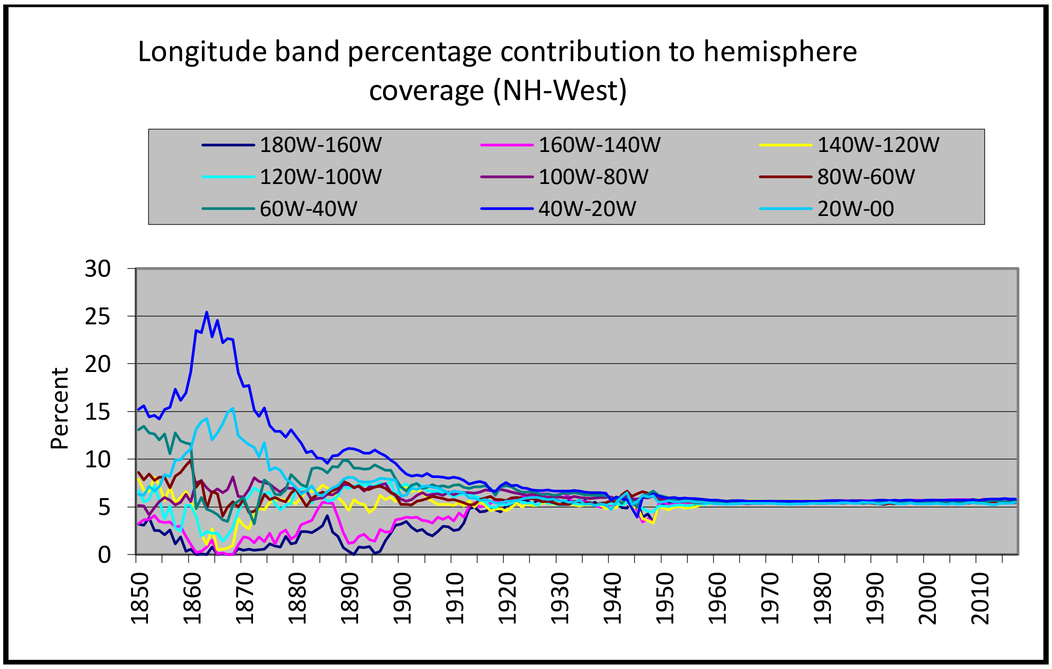


Figure A1.3(b) Contribution to total hemisphere coverage by longitude bands (NH, West)

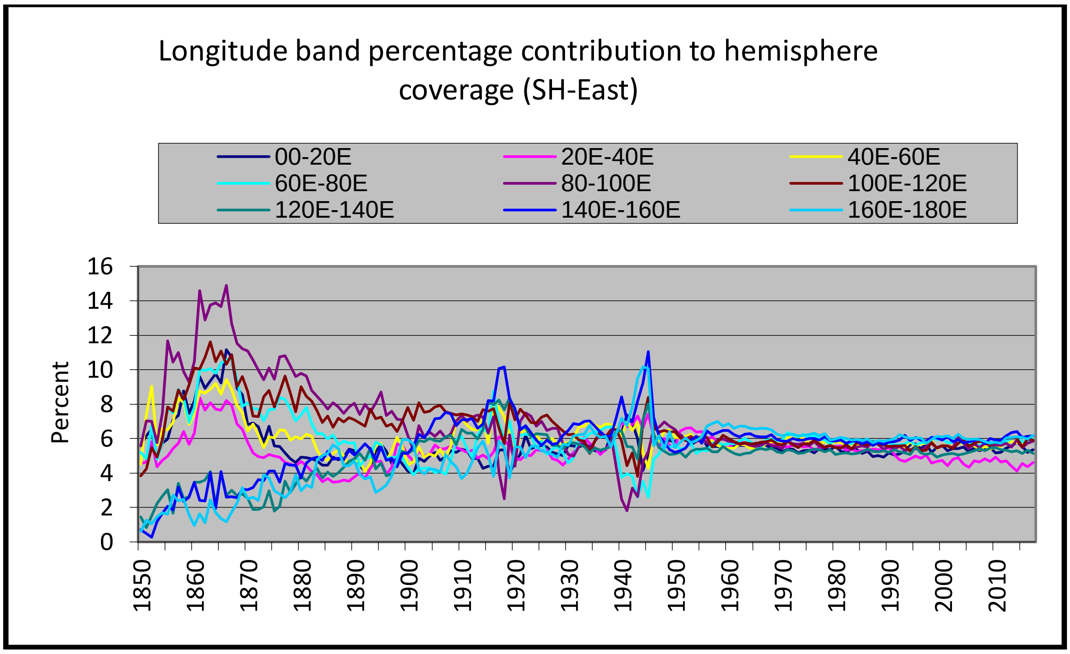


Figure A1.4(a) Contribution to total hemisphere coverage by longitude bands (SH, East)

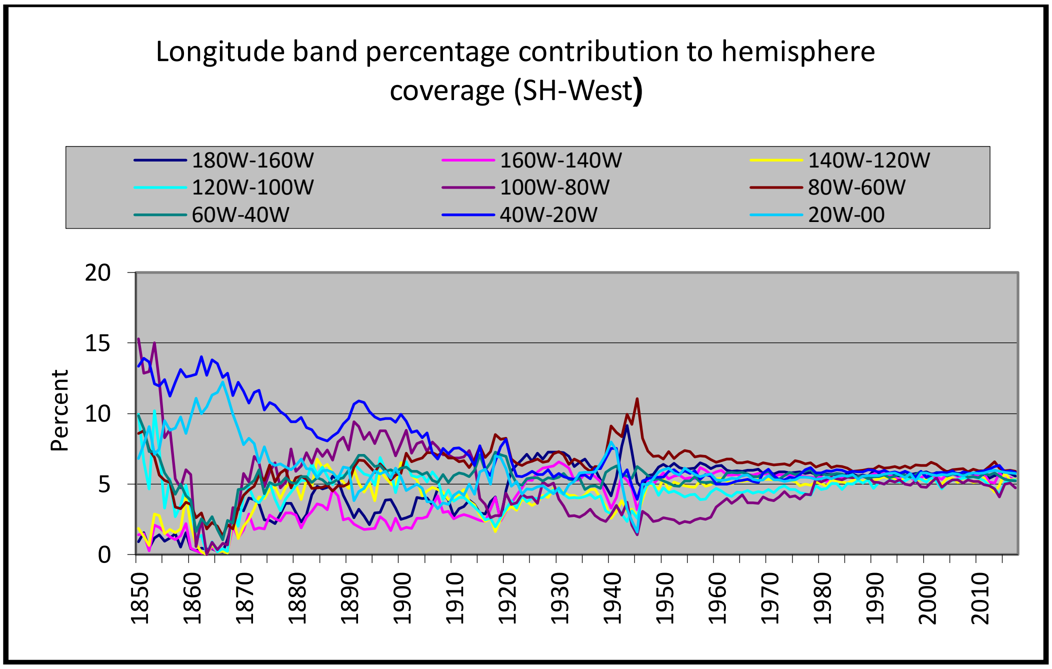


Figure A1.4(b) Contribution to total hemisphere coverage by longitude bands (SH, West)

Corner No.	Lat. Long.	Corner No.	Lat. Long.
1	65N 55E	11	20N 45W
2	65N 10W	12	20N 40W
3	50N 10W	13	10N 40W
4	50N 45W	14	10N 35W
5	45N 45W	15	0N 35W
6	45N 95W	16	0N 20W
7	40N 95W	17	35N 20W
8	40N 100W	18	35N 5E
9	35N 100W	19	40N 5E
10	35N 45W	20	40N 55E

Table A1-1 Mapping locations expressed in latitude and longitude for each turning point in the Northern Hemisphere region identified in Chapter 4, Figure 4.3, anticlockwise from top right of region.

Appendix 2 - Supplementary Information chapter 5

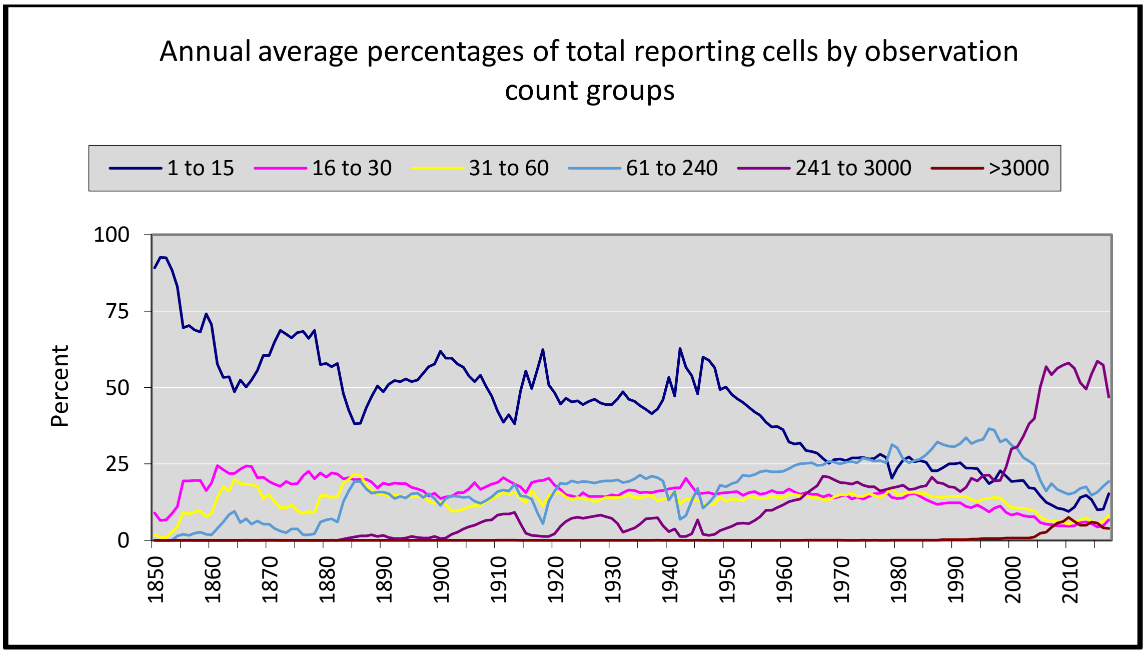


Figure A2.1 Grouped observation counts and their percentage of cells of all that reported in the month (annual averages)

Appendix 3 - Supplementary Information chapter 6

From an analysis conducted by the author in 2012, comparing HadCRUT3 data for locations in Russia to data from North Eurasian Climate Centre (NECC, http://neacc.meteoinfo.ru/)

				HadCRUT3	NECC	difference
Station	**Stn ID**	**Month**	**Year**	**deg C**	**deg C**	**deg C**
Tura	245070	Dec	1945	-25.3	-38.1	12.8
Vitim	300540	Jan	2003	-39.7	-28.1	-11.6
Tura	245070	Feb	1978	-15.0	-26.3	11.3
Vitim	300540	Mar	2002	-18.5	-7.7	-10.8
Bratsk	303090	Feb	1981	-28.6	-18.4	-10.2
Erbogatchen	248170	Feb	1993	-29.6	-19.6	-10.0
Kirensk	302300	Jan	1963	-17.0	-27.0	10.0
Kirensk	302300	Mar	1973	-5.9	-15.9	10.0
Mogotcha	306730	Jan	1969	-24.3	-34.3	10.0
Irkutsk	307100	Jan	1961	-27.9	-18.0	-9.9
Suntar	247380	Dec	1973	-21.0	-30.8	9.8
Vitim	300540	Jan	1993	-29.1	-19.3	-9.8
Tura	245070	Jan	1945	-24.3	-34.1	9.8
Irkutsk	307100	Nov	1996	-12.8	-3.3	-9.5
Bratsk	303090	Jan	1981	-30.5	-21.0	-9.5
Vitim	300540	Feb	2001	-36.5	-27.2	-9.3
Boguchany	292820	Apr	1972	10.6	1.6	9.0
Tura	245070	Dec	1938	-27.0	-36.0	9.0
Bajkit	238910	Jan	1947	-30.5	-38.9	8.4
Minusinsk	298660	Feb	1978	-8.4	-16.7	8.3
Eniseysk	292630	Feb	1978	-11.9	-19.8	7.9
Bajkit	238910	Dec	1949	-29.9	-37.6	7.7
Suntar	247380	Feb	1980	-35.4	-27.7	-7.7
Boguchany	292820	Feb	1978	-13.3	-21.0	7.7
Minusinsk	298660	Nov	1896	3.9	-3.7	7.6
Vanavara	249080	Dec	1949	-28.9	-36.4	7.5
Boguchany	292820	Apr	1983	-3.6	3.9	-7.5
Minusinsk	298660	Jun	1997	10.1	17.1	-7.0
Kirensk	302300	Dec	1926	-31.3	-24.3	-7.0
Kirensk	302300	Feb	1923	-32.4	-25.6	-6.8

Table A3-1 Sorted table of greatest differences between HadCRUT3 station data and data supplied by the North Eurasian Climate Centre for various eastern Russian stations. Several stations have multiple entries.

Appendix 4 - Supplementary Information chapter 7

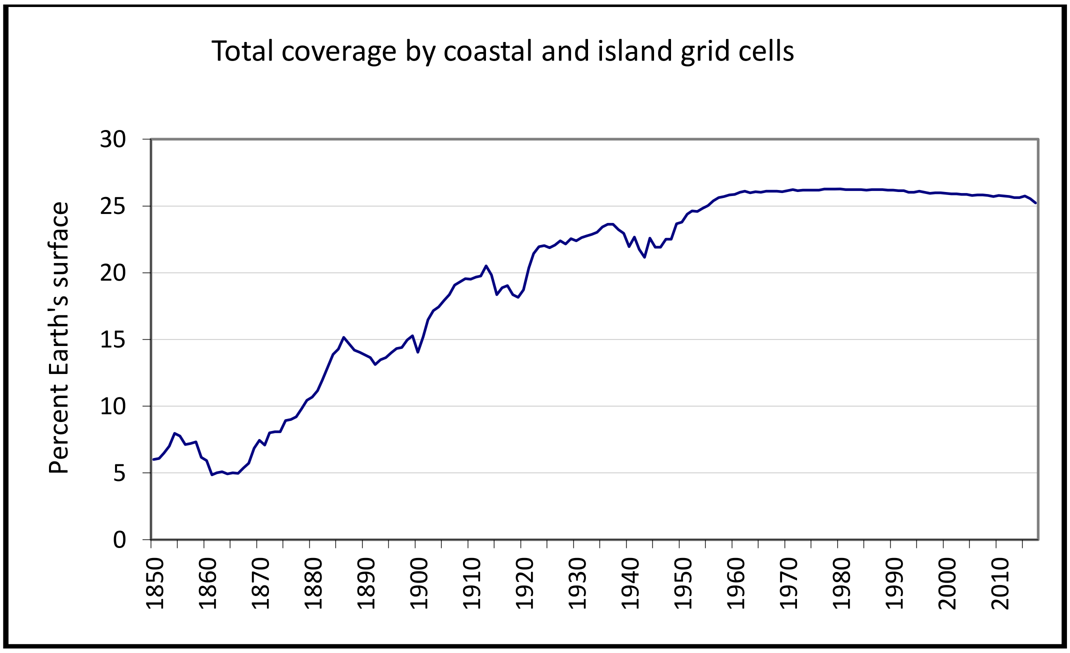


Figure A4.1 Coverage of coastal and island grid cells

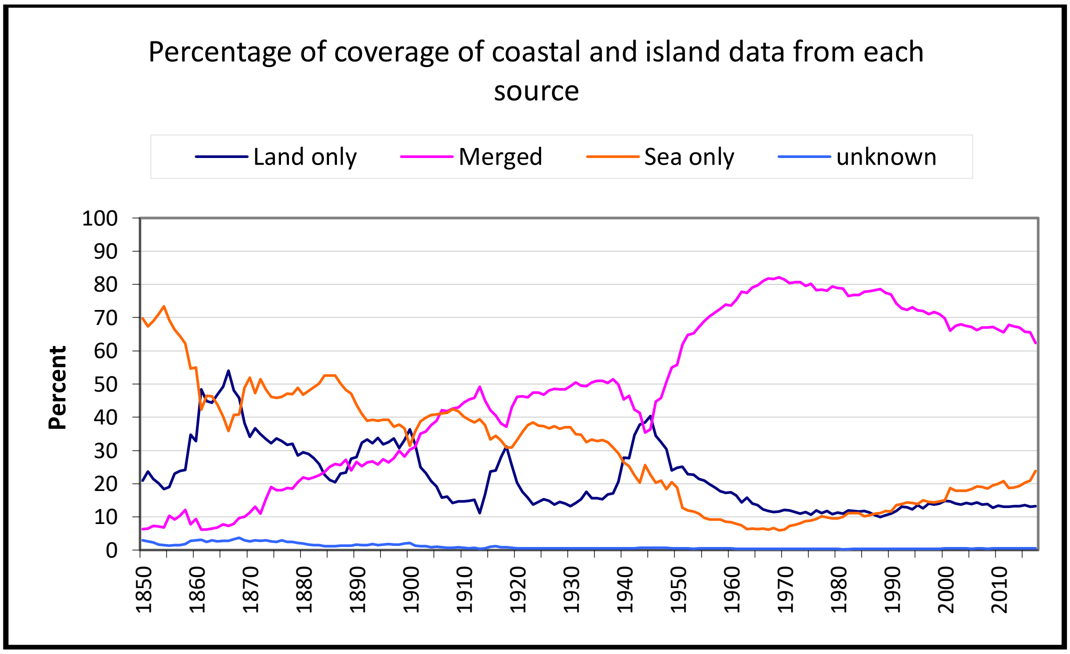


Figure A4.2 Percentage of the coverage from the different sources of the grid cell values

ID	Lat	Long	Location	Country	Start	End
351330	51.5	59.9	ADAMOVKA	RUSSIA	1937	1989
351331	51.5	59.9	ADAMOVKA	RUSSIAN FEDER	1936	1988
43500	67.8	32.3	APUTITEEQ	GREENLAND	1958	1987
43510	67.8	32.3	APUTITEEQ	GREENLAND [DE	1958	1979
623920	27.2	31.2	ASYUT	EGYPT	1902	1990
623930	27.1	31	ASYUT	EGYPT	1902	2010
170200	41.6	32.3	BARTIN	TURKEY	1965	1990
176730	41.6	32.3	BARTIN	TURKEY	1965	1990
104060	51.8	6.5	BOCHOLT	GERMANY	1947	2006
106060	51.8	6.5	BOCHOLT	GERMANY	1947	2006
128390	47.4	19.3	Budapest Lorinc Ai	HUNGARY	1780	2008
128430	47.4	19.2	Budapest Lorinc Ai	HUNGARY	1780	2017
986450	10.3	123.9	CEBU	PHILIPPINES	1891	2012
987113	10.3	123.9	CEBU	PHILIPPINES	1951	1975
854700	27.3	70.4	COPIAPO	CHILE	1951	2016
854701	27.1	70.3	COPIAPO	CHILE	1967	2005
109340	47.7	9.5	FRIEDRICHSHAFEN	GERMANY	1861	1980
109350	47.6	9.5	FRIEDRICHSHAFEN	GERMANY	1861	2014
486252	4.4	101	HOSPITAL BATU GAJAH	MALAYSIA	1967	2000
486257	4.4	101	HOSPITAL BATU GAJAH	MALAYSIA	1967	2000
486654	2.7	102.2	HOSPITAL KUALA PILAH	MALAYSIA	1967	2000
486657	2.7	102.2	HOSPITAL KUALA PILAH	MALAYSIA	1967	2000
427540	22.7	75.8	INDORE	INDIA	1878	2017
427550	22.7	75.7	INDORE	INDIA	1901	1988
21420	66.6	19.6	Jokkmokk	SWEDEN	1860	2017
21461	66.6	19.6	JOKKMOKK	SWEDEN	1951	2016
703813	58.3	134.4	JUNEAU_DWTN	UNITED STATES	1890	1984
703814	58.3	134.4	JUNEAU_DWTN	UNITED STATES	1883	2015
171350	39.9	33.5	KIRIKKALE	TURKEY	1963	1990
177320	39.9	33.5	KIRIKKALE	TURKEY	1963	1990
11240	64.2	12.5	KJOBLI I SNASA	NORWAY	1939	2017
11241	64.2	12.5	KJOBLI I SNASA	NORWAY	1939	2012
401530	33	35.5	MT. KENAAN	ISRAEL	1950	2017
401531	33	35.5	MT. KENAAN	ISRAEL	1950	2012
77470	42.7	2.9	PERPIGNAN	FRANCE	1836	2017
80804	42.7	2.9	PERPIGNAN	FRANCE	1950	2010

Table A4-1 Examples of duplicate stations or near-duplicate station locations, showing station ID, latitude, longitude, name, country and period for which data is available.

Appendix 5 - Supplementary Information chapter 9

PART 1 - Further notes on the adjustment of temperature data

Section 9.9 (of Chapter 9) showed that the adjustment of temperature data by a constant amount was flawed when dealing with situations where the non-meteorological influence on temperature was gradual. In particular it highlighted the issue with urbanisation increasing over time and stations being relocated to correct for that urbanisation, but the adjustment of temperature by a constant amount implicitly assuming that the influence of urbanisation throughout the entire period of the station being located there was a great as it was when the station was relocated. This deserves a more detailed explanation because it appears to be a common practice that will have caused distortion to the HadCRUT4 temperature dataset.

We start by looking only at the influence of urbanisation, assuming for the sake of simplicity that the influence of urbanisation is constantly increasing at a rate of 4.0°C/century. The broken blue line shows the temperature increase due to urbanisation if the observation station had remained at is initial location. In this case it is assumed that at each new location for the observation station the recorded temperature is 1.0°C cooler than it was at the end of the period of using the old location. In the real world the pattern of changes would be less consistent but this simplification can illustrate the problem.

Figure A5.1 deals only with the pattern of urbanisation. It indicates the increasing influence of urbanisation had the station remained at the first location (broken blue line), the influence of urbanisation on the recorded temperature at each successive site (saw-tooth orange line) and the impact of adjusting all previous data after each relocation (broken green line).

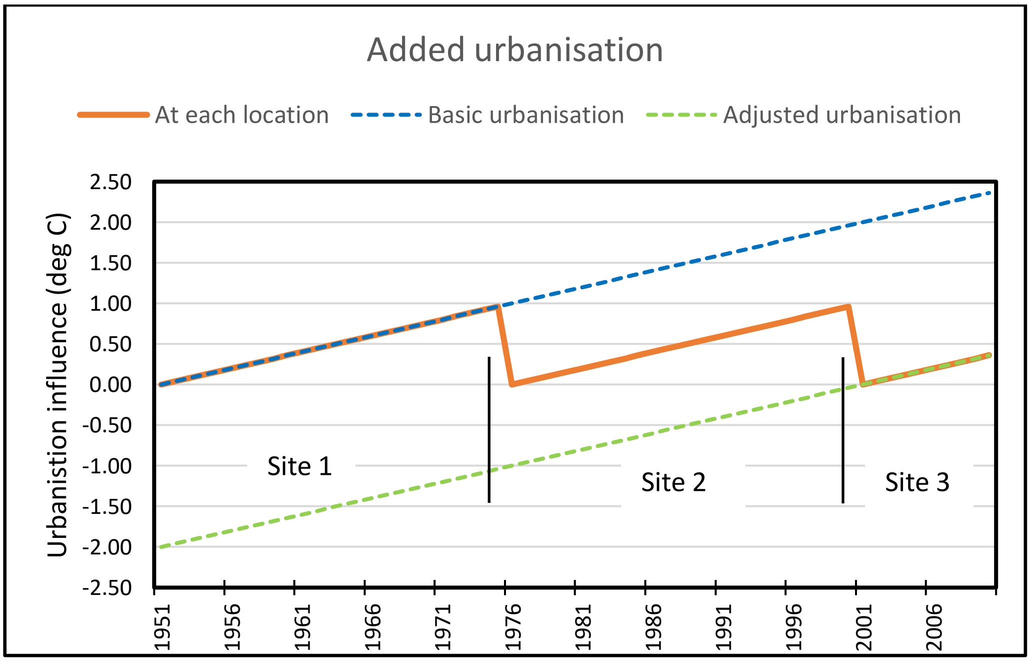


Figure A5.1 – Constant urbanisation influence of 4.0°C/century and adjustment of all previous data each time the station is relocated. The blue indicates the impact on temperature at the first site had it continued to operate, the orange line the impact on recorded temperature at each new site and the green line the impact of adjusting temperatures by a constant 1°C after each relocation

To convert this into a more realistic situation the urbanisation is added to a base temperature of 20°C and 60 random values ranging from -1.0 to +1.0 create a weather-like pattern. (The actual data used here is described in an end-note to this section.) The 60 values simulate natural "weather" variation in the mean monthly temperatures for the same calendar month (e.g. October) from 1951 to 2010 for some observation station.

Two sets of adjusted temperature data are created from these 60 values. The first set is of perfectly adjusted temperatures, with adjustments of diminishing magnitude as we move back to the start of operations at each site. This is the correct adjustment. The second set simply applies constant data adjustments that are the difference between the temperatures at the old and new stations at the time of each relocation. For the second set, the pattern of adjustment for urbanisation is consistent with the broken green line of Figure A5.1

Figure A5.2 plots the two pseudo-temperature patterns, with the continuous blue line indicating the first set and broken red line the second set. The adjustments for the second set might produce reasonable recorded temperatures at each site but the trend for the influence of urbanisation remains in the data and falsely indicates a substantial warming trend.

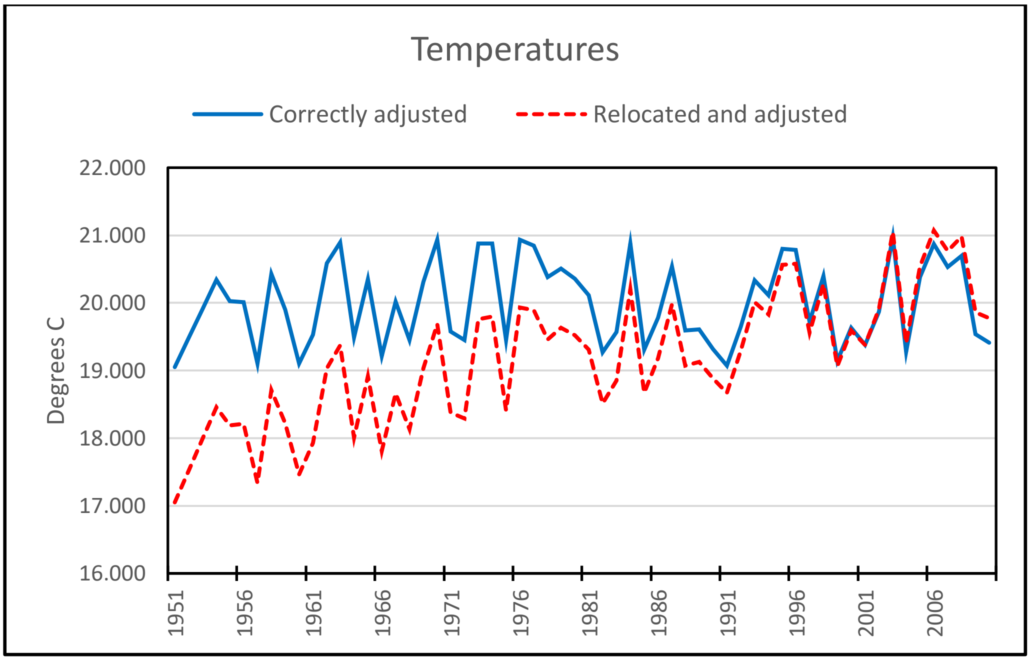


Figure A5.2 Temperatures according to data that has been correctly adjusted for the constant urbanisation shown earlier and data that has been incorrectly adjusted by applying a constant shift to all previous values after each change of location.

Like other global temperature datasets the HadCRUT4 dataset is based on temperature anomalies, which are calculated by subtracting a given temperature from some base. In the case of the HadCRUT4 data the base is the 1961-1990 average temperature for the given location, so if the average is 23C and last month was 24°C, the anomaly for the month was +1.0°C but if last month was 22°C the anomaly is -1.0°C.

Figure A5.3 plots the HadCRUT4 temperature anomalies for the data shown in Figure A5.2. For the second set of data the temperature adjustments have decreased the recorded temperatures but have also decreased the 1961-90 average temperature by the same amount, meaning that the anomalies are unchanged and contain the same influence of urbanisation that they always had.

Figure A5.3 shows a data trend of 0.003°C/year (or 0.3°C/century) for the first set of adjusted data, with correct adjustment for urbanisation, the very small warming trend in the data caused by the sequence of the random data that simulates normal variation. For the second set of data, which is adjusted by a constant amount after each relocation, the trend is 0.043°C/year (or 4.3°C/century), which is the urbanisation trend plus the small trend in the random data.

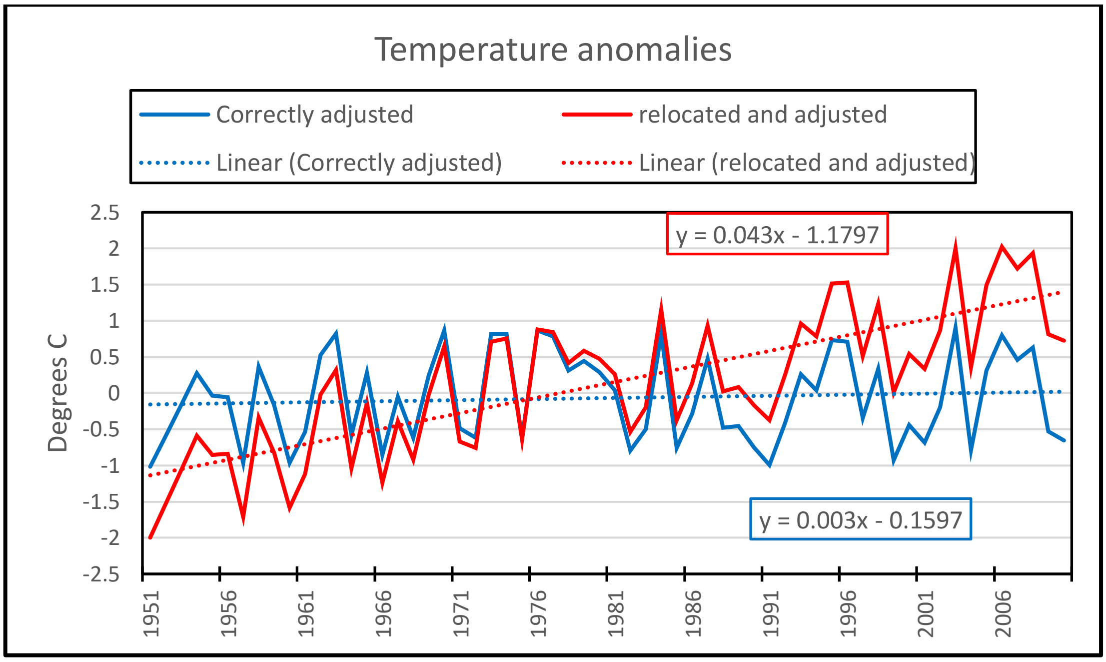


Figure A5.3 The sets of temperature anomalies produced by the temperatures shown in Figure A5.2. The incorrectly adjusted data is too low at the start and too high at the end, producing a false trend that corresponds to the urbanisation

In effect, the relocation of the observation station and the associated adjustment of the temperature data by a constant amount to supposedly remove the effects of urbanisation has done nothing other than reduce the temperatures reported in the past. The adjustment shown here has had no impact whatsoever on the trend or the temperature anomalies that are calculated from that data.

As an example, section 9.9 of chapter 9 discussed an instance where the station was relocated and urbanisation had no influence on the recorded temperature for many years before eventually having an effect. This can be thought of as relocating the station to a point well beyond an urban area and over time urban expansion reaching and passing that point.

Figure A5.4 is similar to graphs in section 9.9 and shows influence of urbanisation at three locations, with no impact on temperatures for 15 years and then rising at 0.1°C/year for the next 10 years (blue line). This simplification assumes a 1.0°C decrease in measured temperature at each new location. Also show is the effect of adjusting all prior temperatures by a constant 1.0C after each move line (gold line).

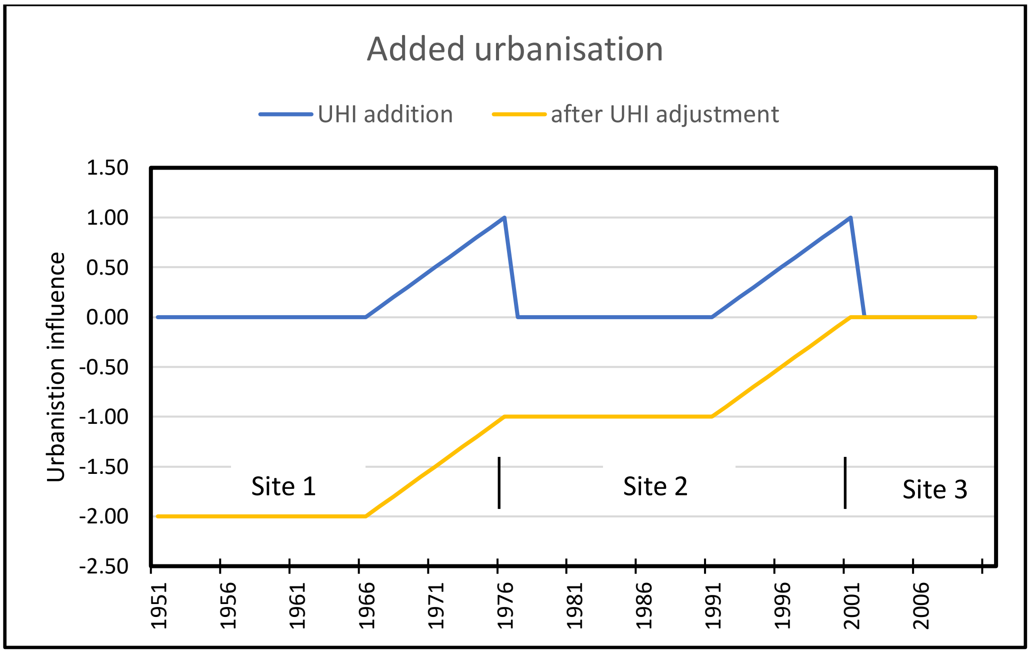


Figure A5.4 Urbanisation relative to stations that are unaffected for 15 years and then are influenced sharply. In this simplified case each relocation of the observation station means a drop of 1.0C compared to the last data from the previous location.

The same data as earlier was used to simulate the variation due to weather and again in one instance we assume that in one case the adjustment for urbanisation is perfect and completely removes it and in the other case that the temperature data is adjusted by a constant amount. The two temperature patterns are shown in Figure A5.5. As well as decreasing all historical data except the most recent (which is in accordance with the Figure above), the flawed adjustment has decreased the average temperature for the period from 1961 to 1990. The average over that period for the flawed adjustment is 18.7°C but is 20.07°C for the correctly adjusted data.

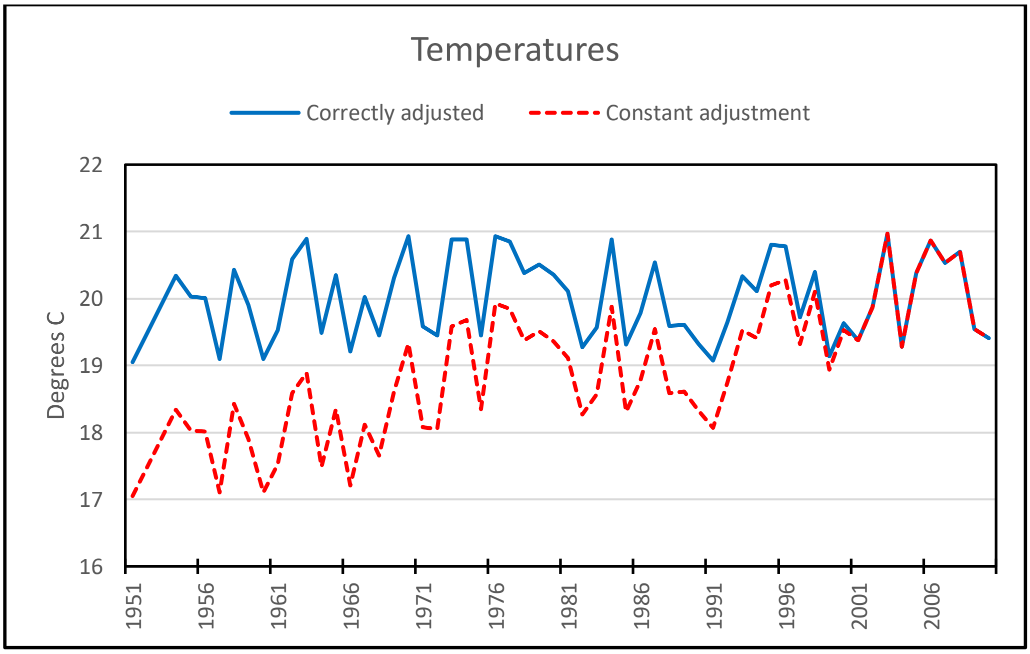


Figure A5.5 – Temperatures that result from the correct and incorrect adjustment for the urbanisation effects shown in Figure A5.4

Because the 1961-1990 averages are different we can expect the temperature anomalies to be different but the added complication is that the falsely adjusted data has a trend that the correctly adjusted data doesn't have (Figure A5.6). In the earlier example the influence of urbanisation increased in linear fashion and the data trend corresponded to that increase. In this current example the trend in the anomalies corresponds to the total trend of pattern of urbanisation, i.e. flat trend for 15 years then sharply rising influence. The trends in the temperature anomalies for incorrectly adjusted data in two examples are similar but this is only because of the influence of urbanisation used in each example.

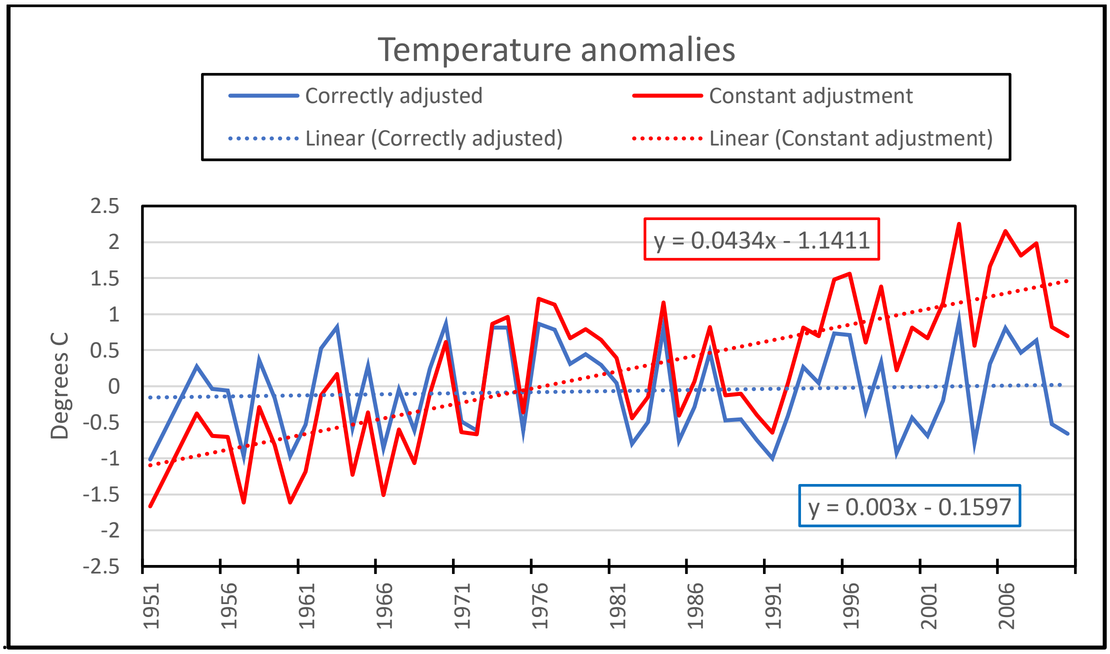


Figure A5.6 Temperature anomalies generated from the temperature anomalies shown in Figure A5.5. The data adjusted using a constant shift in the values produces a false trend because it failed to remove the effect of urbanisation

The patterns shown in the trends of the anomalies for incorrectly adjusted data will depend on the actual data, the data adjustment and when that adjustment occurred but every adjustment from 1962 onwards will impact recorded temperatures during the period from 1961 to 1990 and will therefore affect all anomalies for the station in question.

The effect of adjustments is also cumulative, so it is quite likely that data early in the record has been falsely adjusted multiple times. Decreased 1961-1990 average temperatures means a change in temperature anomalies, with historical temperatures producing excessively low temperature anomalies and recent temperatures producing excessively high anomalies. These incorrect anomalies will be carried through into the HadCRUT4 and CRUTEM4 datasets where they will corrupt the global average temperature anomalies, in all likelihood creating a rising trend of greater magnitude than the true data. These false adjustments might also account for the different trends in global average temperature anomalies on land and at sea.

In more general terms, expressions such as "This month's mean temperature was 2 degrees above the long-term average temperature" are based on the 1961-1990 average. Given that the 1961-1990 average decreases with every new flawed data adjustment, after another relocation and data adjustment the same recorded temperature might be reported as "2.5 degrees above average".

Determining the impact of these flawed adjustments on the HadCRUT4 dataset will require examining the adjustment practices of every national meteorological service, but it appears that both Australian and US services use this flawed technique. The World Meteorological Organization (WMO) document on this (WMO TD 1186, 2003) has a single sentence that mentions correction for gradual non-meteorological influences in its 20-page discussion of data adjustment so perhaps it's not surprising that data would be incorrectly adjusted.

Two final points need to be made. Adjustment by a constant amount is quite reasonable when there is a step change in the temperature data, such as a new instrument that records slightly higher or lower than the old instrument. It should not however be applied to gradual changes.

The second point is that the above scenarios are simplifications. There is often a difference between temperatures recorded at two locations because the exposure to weather at the new location and when the old location was in its *initial state* will very likely vary. This change can quite reasonably be regarded as a step change. It will be difficult to determine the difference in recorded temperature at the new site and the initial state of the old site because, as discussed above, the non-meteorological influences at the old site might have changed over time. Regardless of the difficulty of determining that difference it is absolutely necessary if we are to regard as accurate any temperature data that has been adjusted for gradual non-meteorological influences.

Data used above:

Base temperatures: 20°C in both cases
Urbanization: Scenario 1 – constant at 4.0°/century starting after the first year
Scenario 2 – At each location, no urbanization affecting temperatures for 16 years then 2.0°C/decade

Randomised "weather data" (random number between -1.0 and +1.0)

year		year		year		year	
1951	-0.95	1966	-0.79	1981	0.11	1996	0.78
1952	-0.52	1967	0.02	1982	-0.73	1997	-0.28
1953	-0.09	1968	-0.55	1983	-0.43	1998	0.40
1954	0.34	1969	0.31	1984	0.88	1999	-0.86
1955	0.03	1970	0.93	1985	-0.69	2000	-0.37
1956	0.01	1971	-0.42	1986	-0.22	2001	-0.62
1957	-0.90	1972	-0.55	1987	0.54	2002	-0.13
1958	0.43	1973	0.88	1988	-0.41	2003	0.97
1959	-0.10	1974	0.88	1989	-0.39	2004	-0.72
1960	-0.90	1975	-0.55	1990	-0.68	2005	0.38
1961	-0.47	1976	0.93	1991	-0.93	2006	0.87
1962	0.59	1977	0.85	1992	-0.35	2007	0.53
1963	0.89	1978	0.38	1993	0.33	2008	0.70
1964	-0.51	1979	0.51	1994	0.11	2009	-0.46
1965	0.35	1980	0.36	1995	0.80	2010	-0.59

PART 2 - Likely incorrect data adjustment in national temperature record

The New Zealand Institute of Water and Atmospheric Research (NIWA) publishes a "7-station" composite and an "11-station" composite for the record of New Zealand temperatures over time. Mullan et al (2010) is the primary documentation for the "7-station" composite series, which is created from the data from three observation stations in large urban environments (Auckland, Wellington and Dunedin), one station in a medium-sized urban area (Nelson), one station in a small urban area (Hokitika) and two stations in rural regions (Masterton and Lincoln). All stations except the two in rural locations are situated in towns and cities on the coast, but exposure to the coast is from a different direction in each case.

Site relocations have been rare in the large urban environments for some time. The Auckland observation station relocated once in the 48 years from 1951 to1998, Wellington's location was unchanged from 1928 to 2005 and Dunedin's in the same location for 38 years (1960-1997), meaning they were all susceptible to increasing urbanisation throughout these periods.

In the 101 years described by the documentation (i.e. 1909-2009) 32 data adjustments were made, sometimes based on parallel local observations and at other times according to comparison with other of the seven stations. In the order of the seven stations listed above in the introduction 3, 3, 5, 4, 5, 4 and 8 adjustments were made. The adjustments at each site were sequential, which means that adjustments were carried forward and the earliest data from any individual stations was adjusted 3, 4, 5 or 8 times.

A graph of the annual averages of the raw and adjusted composite data is shown in Figure A5.7 and Figure A5.8 shows the difference between the original and adjusted data. Data prior to 1930 is included for completeness with NIWA's graphs but prior to 1922 fewer than seven stations reported data, as was the case in 1945-7, 1949, 1953-4, 1962-5 some of which coincided with abrupt shifts in the differences.

The discussion of a possible reason for flawed adjustment presented in Chapter 8 can account for the diminishing difference between raw and adjusted data as time passes. The "small noise" in those differences probably due to the number of reporting stations and the temperature "noise" in each year, the latter perhaps with conditions different to those when old and new stations were run in parallel. In one case at least old and new sites were not run in parallel. A site in the corner of a cemetery was required at short notice and the station had to be relocated, the adjustments having to be determined from comparisons to neighbouring stations.

It is interesting that Hessell (1980), an employee of the New Zealand Meteorological Service, finds serious problems with urbanisation at many urban sites and growth of local vegetation to be a problem at many rural sites, especially those run by the New Zealand Forest Service. It states "There are no sites at all which have been unaffected since 1930 by site changes less than 10m, increased sheltering, urbanisation or screen changes." and goes on to say "It is concluded that the warming trends in New Zealand previously claimed are in doubt".

Hessell (1980) predates NIWA's "7-station" composite record but de Freitas et al (2014) does not when it applies a different adjustment technique to that used by NIWA although developed by the same individual (Salinger, 1980; Rhoades & Salinger, 1993). NIWA's adjusted 7-station composite, which uses the 1980/81 homogenisation technique and 35 adjustments, has a warming trend of 1°C/century. de Freitas et al (2014) used the 1993 technique plus two well-documented site changes that did not appear in NIWA's list of adjustments, found an average 0.32°C/century warming for the five urban locations and 0.2°C/century for the two rural stations. Which technique is more accurate might be open to debate but not the fact that different methods of homogenisation produce different temperature estimates.

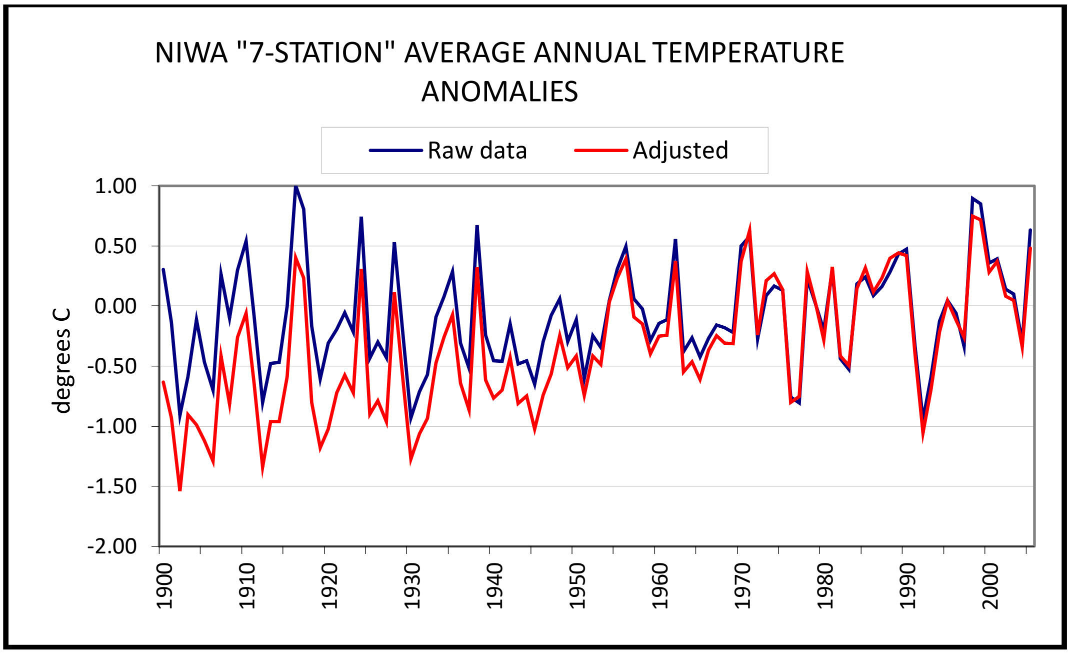


Figure A5.7 Average annual temperature anomalies for the raw and adjusted data used in NIWA's "7-station" composite for New Zealand

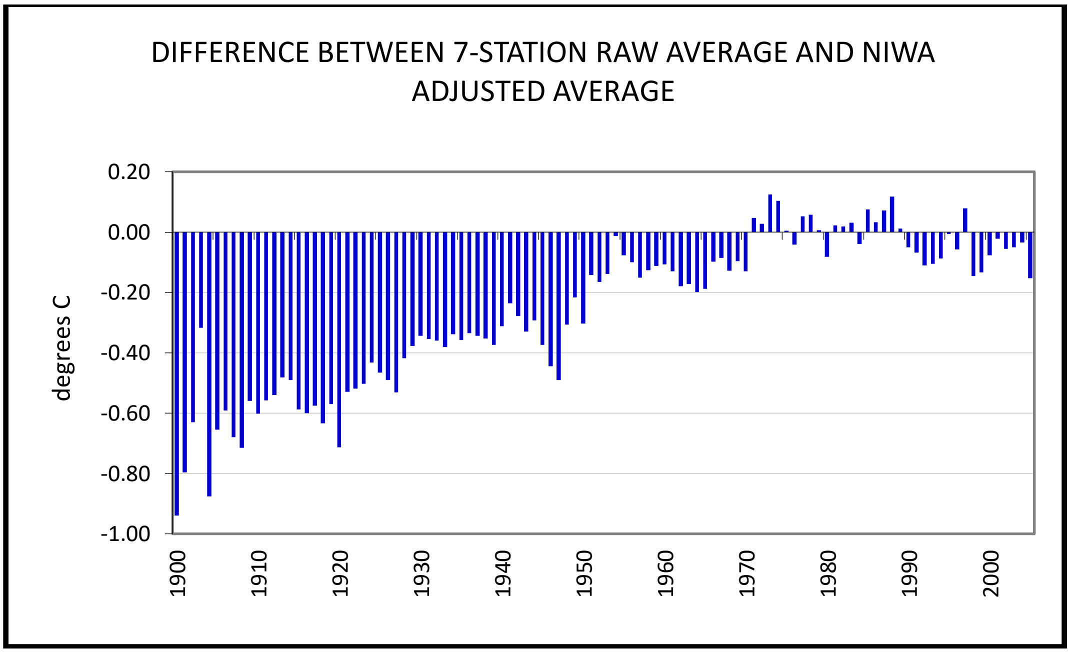


Figure A5.8 - Difference between the two sets of data shown in the previous graph

References for Appendix 5 – part 2

de Freitas, C.R., M.O.Dedekind and B.E.Brill (2014) A reanalysis of long-term surface air temperature trends in New Zealand, *Environmental Modelling and Assessment,* 20(4), 399-410 DOI:10.1007/s10666-014-9429-z

Hessell, J.W.D (1980) Apparent trends in mean temperatures in New Zealand since 1930, *New Zealand Journal of Science*, v23, p1-9

Mullan, A.B; Stuart, S.J; Hadfield, M.G; Smith, M.J (2010). "Report on the Review of NIWA's 'Seven-Station' Temperature Series" *NIWA Information Series No. 78*. p175, online as https://www.niwa.co.nz/sites/default/files/import/attachments/Report-on-the-Review-of-NIWAas-Seven-Station-Temperature-Series_v3.pdf

Rhoades, D. A., & Salinger,M. J. (1993). Adjustment of temperature and rainfall records for site changes. *Int J Climatol*, 13, pp899–913

Salinger,M. J. (1980) The New Zealand temperature series. *Climate Monitor*, 9(4), pp112–118.

Appendix 6 - Consolidated Findings

From chapter 3

1 - The HadCRUT, HadSST and CRUTEM datasets lack clear identification as to when they were created (or published) and the period of data that they contain. The file names should clearly identify when they were created and information describing the contents be available and particularly note the changes since the previous release.

2 - The three datasets are related so it would be logical to format them in identical fashion for ease of processing. A single piece of software could read and process each file if the formats were the same. In terms of data quality this would address the issue of consistency.

3 - The date and time of the download from ICOADS that was used to create the HadSST dataset is essential information for a thorough audit of the dataset but this is not provided. In terms of data quality, the description of the data is incomplete.

4 - The SST data has not been properly audited prior to the creation of the HadSST3 and HadCRUT4 datasets. The carrying forward of errors from a data supplier is contrary to good data quality practices.

5 - Clear descriptions sufficient for others to reproduce them, and therefore verify their accurate construction, are not available for the 100 different variants of the HadSST3 dataset.

6 - The technique of creating 100 variants of the HadSST3 dataset and then assembling a 'median' dataset appears to fail to recognize the physical limits of sea surface temperatures.

7 - The composite set of observation station files published by the CRU includes data from stations that were not included in the processing to create the CRUTEM4 and HadCRUT4 datasets but these excluded stations are not clearly identified.

8 - The CRU observation station data contains instances where the station cannot be identified by name.

9 - The CRU data for observation stations is inconsistent in its formatting of information, inconsistent in its country naming and sometimes clearly incorrect.

10 - The supplier of each set of station data is not immediately obvious and yet the supplier is responsible for the accuracy and adjustment of the data submitted to the CRU for possible inclusion in the dataset. Any thorough audit of the data would require this information so that data could be validated against the supplier's records and so that questions might be raised with the supplier.

11 - The station metadata fails to provide details about the WMO station class, compliance with WMO standards and temperature adjustments but these are essential for proper audit and for the calculation of error margins.

12 - The grid cell system is at times constraining and data might not be representative of the entire cell.

13 - Temperatures are measured in thermally complex regions near the interfaces of two very different mediums (i.e. land-air and water-air).

14 - Land and sea temperatures are measured by different methods, according to different schedules, subject to different influences and have vastly different plausible ranges. The merging of such data into a single dataset would be unacceptable in most scientific fields.

15 - Daily minimum temperatures are subject to numerous external, non-meteorological influences and changes in these could easily account for some of the increase in HadCRUt4 temperature anomalies over time.

From chapter 4

16 - The coverage according to HadCRUT data is not absolute but based on grid cell sizes and requires the existence of temperature measurements within the given grid cell and month. Coverage derived from the same data would differ if different cell sizes were used.

17 - Global coverage of the HadCRUT4 dataset has varied throughout the data record and falls short of even 50% for most of the first 100 years of the 168-year record. Such low coverage cannot be considered global unless the reporting data was relatively evenly distributed around the Earth, which it is not.

18 - Coverage over time has been far from homogenous. Coverage of the northern hemisphere reached homogeneity around 1950 but coverage of the southern hemisphere at the end of 2017 was still not homogenous.

19 - In the 1860s and 1870s certain regions of the world accounted for a far greater proportion of total coverage than their physical areas would suggest and coverage in other areas was very poor. So-called global average temperature anomalies at that time cannot be regarded as "global".

20 - The proportions of data from land and sea has varied greatly over time and because temperatures over land fluctuate more than sea surface temperatures these changing proportions could account for some of the changes in the HadCRUT4 hemispheric and global average temperature anomalies.

21` - Low coverage in the early years of the record coincides with greater month-to-month variability in average global temperature anomalies, a point repeated during World War I and World War II, indicating a wider error margin during periods of low coverage.

22 - The HadCRUT4 global average temperature anomalies for the first 100 years of data (i.e. 1850-1949) have little credibility because the error margins caused by low and biased coverage are substantial.

From chapter 5

23 - The total number of SST observations was very low in the early years of the record and has increased by over 800% since 1850. This means that the HadSST3 dataset has been created from very inconsistent numbers of observations.

24 - The high proportion of grid cells with few SST observations casts doubt on the accuracy of any global, hemispheric or regional averages at least until 1950.

25 - Instances of from 1 to 5 observations for a grid cell across an entire month could all been recorded in just a few days by a single vessel travelling through the region

covered by that grid cell and the data is not necessarily representative of the entire cell over the entire month.

26 - The low number of SST observation for coastal or island grid cells has two problems. Firstly, where no data from observation stations is available the data for the grid cell will be based on few observations. Secondly if observation station data is available the cell value will be a merging of data from one or more stations on land (~24 observations per station if WMO standards are followed) with from just 1 to 5 observations at sea.

27 - The number of reporting observation stations has varied greatly across the record, this not only changing the number of grid cells for which data is reported but changing the number of reporting stations within the same grid cell. Both will almost certainly impact error margins and the accuracy of HadCRUT4 data.

28 - The very low number of reporting observation stations in the Southern Hemisphere in the first few decades of the record means a very wide error margin in the CRUTEM4 and HadCRUT4 Southern Hemisphere and global average temperature anomalies during that time.

29 - In the Northern Hemisphere in the first few decades of the record the majority of data was from Europe and the east coast of North America. Hemispheric and global average temperature anomalies in the early part of the data record are skewed towards those resulting from meteorological conditions in the eastern USA and in Europe.

30 - The CRUTEM4 grid cells with a single reporting station has only briefly fallen below 33% (i.e. one-third) of all reporting grid cells in the entire data record. The presence of very few observation stations in a CRUTEM4 or HadCRUT4 grid cell casts doubt as to whether the data is truly representative of the entire cell.

31 - When few (1 to 4) observation stations report data for a given grid cell any increase or decrease in the number of stations can mean relatively large shifts in the grid cell value because it is the average of the station temperature anomalies.

From chapter 6

32 - The period from 1961 to 1990 seems to be meteorologically abnormal with a distinct shift slightly after the mid-point of the period. Any failure to report data in a given month could distort the station Normal or the SST grid cell climatology and therefore distort every temperature anomaly for the calendar month in question.

33 - Station data is likely to show lower monthly mean minimum temperatures under daylight-saving than without daylight-saving because 9:00am temperatures are more likely to be the lowest temperature for the next 24 hours.

34 - The adoption of daylight-saving during the period 1961 to 1990 will likely cause distortion of some station Normals, especially for the months at the start and end of the period of daylight-saving. This has consequences for all temperature anomalies for the calendar month in question and therefore by extension consequences for CRUTEM4 and HadCRUT4 hemispheric and global average temperature anomalies.

35 - Some station Normals are not derived from temperature recordings across the usual period of 1961-1990 and there are good reasons to question the estimated Normals derived by other methods.

36 - The use of different periods for the calculation of Normals and standard deviations makes the dubious assumption that the two periods are meteorologically similar and that the recorded temperatures will show similar patterns.

37 - The presence of outliers in the data across the period 1961-1990 for some station-month combinations has distorted the Normals that are calculated from that data. Given that the Normal has implications for the identification of outliers during the rest of the data record from that location and is used in the calculation of every temperature anomaly for that station in that calendar month, the temperature anomalies for the station are likely to be distorted and so too the HadCRUT4 and CRUTEM4 grid cell values.

38 - The Normals for some observation stations are derived from far less data than the WMO standards specify as a minimum number of entries. The error margins in these cases, both for Normals and the subsequently calculated temperature anomalies could be quite high.

39 - Average sea surface temperatures for the period 1961-90 are based largely on estimates calculated according to assumptions about cloud cover, ocean surface turnover and other factors.

40 - Even with a generous minimum of 14 of 30 years, just over 20% of all cell-month combinations that reported in data during that period failed to meet the criteria used by CRUTEM4 for data from observation stations. When we also take into account the fact that climatologies are resolved on a 1° x 1° grid cells and pentads in many cases there was very little data available for adjusting the estimated average SSTs.

41 - In many cases where data is reported for cell-month combinations during 1961-90 very few observations were made for the given month and extreme values of sea surface temperature anomalies were often based on very few SST observations for the month in which they occurred. No great confidence can be placed in the accuracy of SST average temperatures that are first estimated and then modified according to such small amounts of data.

42 - The SST measurements by which the estimated long-term average SSTs are adjusted are at times very variable in some calendar months and average measured SSTs in those calendar months come with wide standard deviations and therefore wide error margins, margins that should be passed along to the modified estimated averages.

43 - The failure of the SST anomalies for many cell-month combinations to sum to zero casts doubt on the average sea surface temperatures from which those anomalies were calculated.

44 - In several instances the calculated sea surface temperature anomalies for the period 1941 to 1990 seem improbably low. This might be due average sea surface temperatures being higher than they should be or could be due to very few SST measurements being made in the grid cell during the month and those measurements recording little more than the consequences of weather over just one or two days (see above in this chapter)

From chapter 7

45 - The HadCRUT4 dataset is sometimes inconsistent with the associated datasets HadSST3 and CRUTEM4 by either:

- Containing data when there is no corresponding data in either the HadSST3 or C RUTEM4 datasets,

- Differing excessively from the single other dataset reporting at the time (43 are different by more than 0.4°C), or

- When both HadSST3 and CRUTEM4 have data, the HadCRUT4 dataset sometimes has data whose value falls beyond the range defined by the values given in the other two datasets, at times by 0.15°C or more.

46 - Global average temperature anomalies for CRUTEM4 and HadSST3 show quite different patterns. It appears likely that the HadCRUT4 global average temperature anomaly at any time will be impacted by the proportion of coverage of the data of the two other datasets. The unanswered question is whether SST data should have been adjusted as much as it was for the period prior to 1946, because perhaps the oceans warm more slowly.

47 - The HadCRUT4 gridded dataset has some grids cell values that seem implausible when compared to the values in the surrounding grid cells.

48 - The data sources for coastal and island grid cells have changed over time in about 85% of such grid cells and the changes of data sources for these grid cells, as well as the change in their contribution to total HadCRUT4 coverage, could well be causing changes in HadCRUT4 global and hemispheric averages.

49 - The use of a “five standard deviation” limit for the identification of outliers in all circumstances is very generous in some instances and yet very restrictive in locations where mean monthly temperatures for a given calendar month are fairly consistent (such as the coastal or ocean tropics) where that limit could be exceeded by the influence of minor weather events.

50 - Standard deviations and average long-term temperatures (i.e. Normals) have been distorted by the inclusion of extreme values that should have been removed prior to calculating these key values.

51 - The inclusion of extreme values in station temperature records calls into question the competence of the national meteorological services that supply the data, and of the Climatic Research Unit for failing to question the presence of those values. Many of these extreme temperatures are obvious errors, some of which might have arisen from having just a few days of temperature recordings and others been errors in transcription. This raises the question of how many less obvious errors exist within the station data.

52 - The presence of outliers in the HadSST3 conflicts with the extremely low probability that such values would be present. If the HadSST3 processing creates such outliers then it should be questioned. The answer might lie with extreme mean monthly temperature anomalies often being linked to very few measurements being made in the corresponding grid cells during the corresponding months.

53 - Instances of unusual mean monthly temperature anomalies, relative to those from other observation stations in the same grid cell, suggest temperature data errors.

54 - The inclusion of data from reporting stations in close proximity to each other will bias grid cell average anomalies towards the temperatures recorded in those localities.

55 - At least one and perhaps two observation stations have metadata that incorrectly indicates the station location, in both cases meaning that the station will be assigned to the incorrect grid cell.

From chapter 8

56 - The many and very different methods of measuring sea surface temperature and the variations within those methods are very likely to have different levels of accuracy and different error margins but these issues are ignored in the HadSST3 and HadCRUT4 datasets.

57 - The technique of creating 100 different datasets and using them to derive the HadCRUT4 dataset has uncertain accuracy, mixes the data from (at most) one correct dataset with 99 incorrect variations and simply hopes that the value used for HadCRUT4, the median of the 100 values, is correct. There is poor justification for this ad hoc approach to poor and incompatible data.

58 - The variety of techniques for determining sea surface temperature often requires the adjustment of temperatures taken at different depths to the notional standard depth but correct adjustment requires data related to thermal layering that is unlikely to be available.

59 - The source data for sea surface temperatures used in the construction of the HadCRUT4 and HadSST3 datasets, the ICOADS database, appears to contain data recorded when ships were in port, where temperatures are very likely different to those at sea and not representative of the grid cell.

60 - The reduction in SST coverage due to World War II might at least partly account for the widely-reported peak in average hemispheric SST anomaly at about that time. Attributing it all to changes of measurement technique and adjusting sea surface temperatures accordingly is probably unwise.

61 - The ICOADS database contains at least one demonstrable error of data transcription and another incorrect conversion. It might contain many more errors that have not been identified.

62 - It appears that the meteorological data in the ICOADS database has not been checked for consistency, either by the organisation that maintains the database or the Hadley Centre, prior to its use in creating the HadSST3 and HadCRUT4 datasets.

63 - It appears that the ICOADS database has not been checked to ensure that the given locations are in fact at sea and, where possible, are consistent with the path of the relevant ship. The latter might not be easy but as indicated above, it has been done by others.

From chapter 9

64 - The metadata for each station gives no indication as to whether the national meteorological services complied with WMO recommendations regarding the minimum number of days of data from which monthly mean temperature are calculated, both within the 1961-90 period (which could impact all temperature anomalies throughout the entire record) and outside it.

65 - The processing for the HadCRUT4 and CRUTEM4 datasets fails to recognise the uncertainties associated with siting and the types of enclosures on even a static basis, let alone that they might change over time. The uncertainties associated with various forms of siting and various enclosure are individually much greater than the increase in the annual HadCRUT4 global average temperature anomaly since 1850.

66 - The CRUTEM4 and HadCRUT4 temperature anomaly datasets are created from temperature data that was possibly in error from the moment that it was recorded, but the magnitude of the error margin at either a station level or as a grid cell average is unclear (besides which the data sources and supply vary every month).

67 - The CRUTEM4 station metadata fails to mention if the station is automated and if so, when the change from manual instruments occurred and whether the station is fully compliant with WMO standards regards recording temperature data with fast-acting electronic sensors in a manner comparable to that for slower-acting mercury-in-glass thermometers.

68 - The temperature data for many stations is likely to be biased low as a consequence of the recommended time of the principal observation. In the case of stations where the principal observation time was changed to between 0700 and 0900 the data has been artificially altered but no indication is shown in the station metadata. Because of these adjustments the HadCRUT4 and CRUTEM4 datasets seem likely to be merging data that has different levels of confidence.

69 - Given that local minimum and maximum temperatures are recorded as at 9:00am clock time, the adoption of daylight-saving time and continuing to make temperature observations at 9:00am local time (i.e. 8:00am without daylight-saving) is an inconsistency that is likely to result in lower monthly average minimum temperatures than when daylight-saving is not used.

70 - Certain HadCRUT4 and CRUTEM4 grid cells will be susceptible to having data from some observation stations from regions that adopt daylight-saving and some that do not.

71 - Much of the observation station data used in HadCRUT4 is likely to have been adjusted but there is no record of the method used or the amount of the adjustment, which makes the data impossible to independently audit.

72 - The station temperature data used in HadCRUT4 and CRUTEM4 has likely been adjusted one or more times and any errors in the adjustment likely been compounded, as the error margin should have been.

73 - Where non-meteorological influences on recorded temperatures have gradually changed over time, the adjusting of all previous temperatures by a constant amount falsely assumes a constant distortion of all prior data and creates a false temperature trend. For station suffering increasing urbanisation this will typically mean that earlier data is excessively adjusted downwards, producing a false warming trend.

74 - Adjustments of the type described above in Finding 73 that occurred after 1961 will falsely adjust the Normals (i.e. long-term average temperatures) from which temperature anomalies are calculated. The temperature anomalies for all data adjusted by that constant amount will be unchanged but temperature anomalies for recent unadjusted data will be calculated using altered Normals. In the case of increasing urbanisation, the decreased Normals will mean a false increase in temperature anomalies, leading to exaggerated warming trends in the HadCRUT4 dataset.

75 - It seems very likely that when urban stations are closed rather than relocated their data is not adjusted for urbanisation and any warming due to non-meteorological influences will have remained in the HadCRUT4 record.

***