Australian Government
Bureau of Meteorology

The Australian Climate Observations Reference Network – Surface Air Temperature (ACORN-SAT) version 2

Blair Trewin

October 2018

Bureau Research Report - 032

The Australian Climate Observations Reference Network – Surface Air Temperature (ACORN-SAT) version 2

Blair Trewin

Bureau Research Report No. 032

October 2018

National Library of Australia Cataloguing-in-Publication entry

Author: Blair Trewin

Title: The Australian Climate Observations Reference Network – Surface Air Temperature (ACORN-SAT) version 2

ISBN: 978-1-925315-98-1

Series: Bureau Research Report – BRR032

Enquiries should be addressed to:

Blair Trewin:

Bureau of Meteorology
GPO Box 1289, Melbourne
Victoria 3001, Australia

b.trewin@bom.gov.au:

Copyright and Disclaimer

The Bureau of Meteorology advise that the information contained in this publication comprises general statements based on scientific research. The reader is advised and needs to be aware that such information may be incomplete or unable to be used in any specific situation. No reliance or actions must therefore be made on that information without seeking prior expert professional, scientific and technical advice. To the extent permitted by law and the Bureau of Meteorology (including each of its employees and consultants) excludes all liability to any person for any consequences, including but not limited to all losses, damages, costs, expenses and any other compensation, arising directly or indirectly from using this publication (in part or in whole) and any information or material contained in it.

Contents

List of Figures

List of Tables

PREFACE

An updated version of the Australian Climate Observations Reference Network – Surface Air Temperature (ACORN-SAT) dataset has been developed.

The purpose of this report is to document the updated version (referred to henceforth as version 2, with the 2012 release and subsequent incremental additions to it referred to as version 1). Whilst the fundamental methods used in version 2 are broadly similar to those used in version 1, there have been a number of changes made which increase the robustness and spatial coherence of the dataset. This report documents those changes, whilst providing sufficient information on the method as a whole to allow it to be reproduced without requiring reference to earlier reports. Data quality control methods are not materially different from those used in version 1 and are not described further here: readers are referred to Trewin (2012) for further details.

As part of the process of developing ACORN-SAT version 2, a number of assessments were undertaken to quantify various sources of uncertainty, benchmark various methods, and assess the impact of potential influences on the data at a network-wide level which had either not previously been considered, or only been considered to a limited extent.

1. INTRODUCTION

To be able to monitor climate on long-term timescales, it is necessary to have datasets where the observations reflect changes in the underlying climate, and not changes in the way the observations have been made.

In the case of temperature, there are a number of factors which can significantly influence observations, other than the underlying climate (Trewin, 2010). Some of these include:

- Moves of the observation site;
- Changes in the instruments used (particularly the instrument shelter, but also the thermometer type, etc.);
- Changes in the local site environment (for example, urban development or changes in local vegetation);
- Changes in observation procedures (e.g., changes in observation times)

There are very few long-term temperature series which have not been affected at some point by one or more of these issues. This makes it necessary to adjust the datasets to remove these issues to make a best estimate of what the observations would have been had the site remained constant. This process is known as homogenisation.

There are two main aspects to climate data homogenisation. The first is *detection*, that is, the identification of potential breakpoints in a time series. This is typically done through a combination of the use of metadata (information about observing sites and practices) and statistical methods, the latter being necessary because many inhomogeneities are only documented to a limited extent in metadata, or not documented at all. The second major aspect is *adjustment*, that is, adjusting data to reduce the influence of inhomogeneities and make the dataset more homogeneous. The detection of inhomogeneities has been a well-explored problem for many years (e.g., Peterson et al., 1998; Venema et al., 2012), with Domonkos et al. (2012) providing an extended historical review, but adjustment has only received detailed attention more recently.

Most published homogenised datasets use monthly data, or adjustments to daily data which are based on monthly data in some form. Adjustments of the probability distribution based on daily data are a more recent development as statistical techniques have become more sophisticated. Whilst adjustments based on monthly means will produce a homogeneous series of monthly or annual means, they will not necessarily produce a homogeneous series of extreme values, as some inhomogeneities affect different parts of the temperature frequency distribution in different ways (Trewin and Trevitt, 1996); for example, if an observing site moves from a built-up site in town to a more open site at an airport, this is likely to have a larger impact on minimum temperatures on calm, clear nights (which are typically cold) than on cloudy, windy nights (which are typically warmer). The methods originally developed for version 1 of ACORN-SAT were designed to address this problem and produce a dataset which is more homogeneous for extremes as well as for means.

Developments in the science of climate data homogenisation since 2012 have mostly been incremental, with most published datasets drawing upon existing techniques. The World Meteorological Organization also established a Task Team for Homogenization in 2014, whose work has been primarily focused on monthly data. Some new techniques have been deployed on an experimental scale at small numbers of locations; for example, the use of wavelet analysis for daily temperature adjustment (Li et al., 2014), and the use of radiative transfer models to undertake a physics-based adjustment for sub-daily data in merging observations from two different screen types (Auchmann and Brönnimann, 2012). On a broader scale, Milewska and Vincent (2016) presented a method for applying wind-direction specific adjustments to maximum and minimum temperature data across a range of Canadian stations.

Whilst a major study was carried out to benchmark different methods of inhomogeneity detection and adjustment in monthly data (Venema et al., 2012), to date only smaller studies have been performed for daily temperature data, although a larger study is planned under the auspices of the International Surface Temperatures Initiative (Willett et al., 2014). The most substantial study of this type was carried out by Killick (2016), using a set of synthetic data based upon observations from the United States. The results of this study, as well as those of Milewska and Vincent (2016), both indicated that day-specific adjustments (as opposed to uniform monthly, seasonal or annual adjustments) produced at best modest improvements to data sets. However, the evaluation metrics and/or methods of constructing synthetic benchmark data used in these studies were not designed to capture the behaviour of extreme low minimum temperatures; Trewin (2012) found that the ACORN-SAT methodology produced generally greater improvement in indicators of these (such as minimum temperatures below the 10th percentile) than in mean-based metrics such as root-mean-square (RMS) error. Vincent et al. (2018) did assess the performance of daily quantile-matching methods against monthly and seasonal methods in homogenising indicators of extreme low temperature, and found that quantile-matching methods performed better, although well-correlated neighbours were required.

The ACORN-SAT dataset was originally released in 2012 (Trewin 2012, 2013). It remains the only known continental-scale temperature dataset with a homogenised daily temperature distribution, although a number of other national or regional datasets have been released since 2012 (e.g., Nemec et al., 2013), with China being a particularly active area (Xu et al., 2013; Cao et el., 2016). The version 2 release builds on the methods used in version 1.

Whilst version 2 uses the same set of 112 stations as that used in version 1, it incorporates additional data, both as the result of further digitisation of historical data, and through incorporating new data at locations where the previous site has been replaced since 2012 (most commonly as the result of a former manual site being replaced by an automatic weather station).

2. DATA AND METADATA

2.1 Data

The ACORN-SAT dataset contains daily maximum and minimum temperature for 112 stations in Australia, commencing in 1910 (Figure 1). Data from 60 stations extends for the full period from 1910 to the present, with the others commencing at various times. 110 of the stations have at least 50 years of available data, with the other two (Learmonth and Rabbit Flat, both in remote areas) having between 40 and 50 years.

The network was chosen for version 1 on the basis of length of available record and spatial coverage, with site quality also a consideration, especially in those more densely sampled parts of Australia with a relatively high density of long-term observations. The principal aim was to achieve sufficient spatial coverage over all parts of Australia to represent Australia-wide temperature change and variability, although there are still some areas (especially the interior of Western Australia) which have no suitable observations, and coverage is limited in parts of interior and northern Australia before the 1950s.

No changes have been made to the network in version 2. However, additional data are available for the existing network, as follows:

- The dataset now extends to the end of 2016 (version 1 finished in 2011). At most stations, this merely involves adding the five years of additional data. At ten stations, the site which was operating as of the end of 2011 has closed, and been replaced in the ACORN-SAT dataset by a new site, making it necessary to create a composite record in which data from the pre-2011 site are adjusted to be consistent with the post-2011 site. There are four other sites where parallel observations programs are currently in progress, but in these cases the old site is still open and has been retained for this version of ACORN-SAT.
- Some additional historical data have been digitised, allowing records to be extended back further into the past. The most substantial digitisation has taken place at three locations: Canberra (1913 to 1939), Moree (1912[1] to 1956) and Sale (1910 to 1945). Small quantities of additional data have also been obtained for Dalwallinu (1955 and 1956) and Tarcoola (1961), extending Dalwallinu's record backwards and partially filling a gap in records at Tarcoola.

The 1910 starting date for the ACORN-SAT dataset has been maintained in version 2. This reflects the fact that the Stevenson screen (the standard screen for housing thermometers) was only adopted as a national standard following the creation of the Bureau of Meteorology as a federal organisation in 1908 (Nicholls et al., 1996); whilst the Stevenson screen was in widespread use by 1895 in Queensland, South Australia, and at the few sites operating in the Northern Territory and Tasmania, the standard was not widely adopted in New South Wales and Victoria until 1906-1908. In Western Australia, there is very little temperature data of any kind outside Perth prior to 1907. All ACORN-SAT data are drawn from Stevenson screens; at the one ACORN-SAT site which did not yet have a Stevenson screen by 1910 (Eucla, where it was installed in February 1913), the pre-Stevenson screen data are not included in ACORN-SAT.

A monthly instrumental temperature dataset for southeastern Australia for the period from 1860 has been developed separately (Ashcroft et al., 2012) and complements the information available from ACORN-SAT.

---

[1] The digitisation at Moree extends back to 1910 but the 1910-11 data are not used in ACORN-SAT because of quality issues.

A full list of the ACORN-SAT stations, and the sites used for those records, is given in Appendix A.

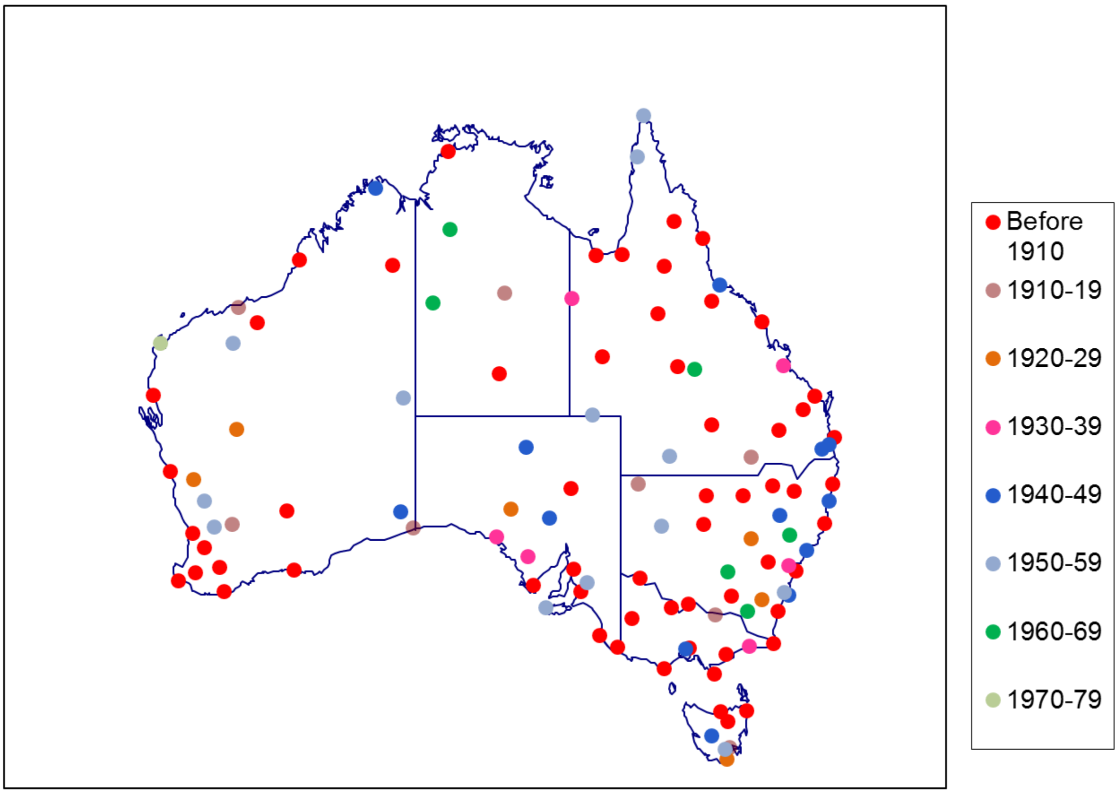

Fig. 1 The ACORN-SAT network by opening date (first available digital daily data).

Large quantities of daily and sub-daily temperature data remain to be digitised in Australia from the pre-1965 period. Whilst the bulk of potentially available daily data have been digitised in the ACORN-SAT network, it is believed that approximately 15 stations still have undigitised daily data which could potentially be added to the dataset. In addition to this, a further 20 stations have been identified which are thought likely to meet the length-of-record and data quality criteria for inclusion in ACORN-SAT and could be considered for future addition to the dataset if their earlier data are digitised.

In addition to the ACORN-SAT stations, other observing stations in Australia play an important role as reference stations for ACORN-SAT homogenisation, and for data quality control. In total, there are more than 1900 stations which have recorded temperature in Australia at some point, of which more than 700 are currently operating. In more recent times, non-Bureau stations have also played some role in quality control, in particular a network established in agricultural areas of Western Australia by the Western Australian Government.

Sub-daily data also play a role in data quality control (as described more fully in Trewin (2012)). The 98 automatic weather stations (AWSs) in the ACORN-SAT network now report data at 1-minute resolution, although most have only done so for 10-15 years, with few stations having 1-minute data prior to 2002. Earlier AWS data are mostly at 30- or 60-minute resolution. The remaining 14 manual stations mostly have observations at 0900 and 1500 local time (one is 0900 only). Prior to the introduction of automatic weather stations, data resolution ranged from 3-hourly through to 0900 and 1500, or 0900 only. Pre-1987 sub-daily data at most sites (other than Meteorological Offices staffed by the Bureau) are only digitised at 0900 and 1500, with other times of day not digitised.

2.2 Metadata

There is a wide range of sources of metadata for the Australian observation network. Most post-1997 station-specific metadata are stored in the Bureau's digital database (SitesDB), with most pre-1997 material generally held on paper files (some of which have been scanned) stored at the National Archives of Australia or in the Bureau of Meteorology's state offices. Metadata become increasingly scarce prior to the 1960s. The available metadata are more completely described in Trewin (2012).

Further metadata have been obtained in the period since 2012, making use of additional documents which have been accessed and a review of existing material, especially around the times of large inhomogeneities found by statistical methods. This has allowed the basis for some adjustments to be determined more fully, and the dates of some inhomogeneities to be determined more precisely than is possible using statistical methods. One example of the use of additional metadata was for a 1969 site move at Launceston Airport. No documentation gave the exact date of the site move, but the date was inferred to within 2-3 days through the discovery of forms for the payment of travel expenses to technical staff sent to install the new site.

In some cases, metadata allow it to be determined that a significant change took place around the time of a statistically detected inhomogeneity, but do not allow the exact date of the change to be determined with precision. For example, at Rutherglen, statistically significant breakpoints were found for minimum temperature in 1966 and 1974. Metadata show that the site moved at least once between 1958 and 1975, but the exact date of the move(s) remains unknown.

3. HOMOGENISATION METHODS

3.1 Detection of inhomogeneities - use of multiple detection methods in parallel

In version 1 of ACORN-SAT, a single statistical method for detection of inhomogeneities was used (Trewin, 2012). This method was based closely on the Pairwise Homogenisation Algorithm (PHA) developed by Menne and Williams (2009), and involves pairwise comparison of data between the candidate station and all sufficiently well-correlated stations in the region, with the Standard Normal Homogeneity Test (SNHT) (Alexandersson, 1986) used to identify significant breakpoints in the difference series. The test was carried out separately on monthly mean anomalies (as a single time series with 12 data points per year), and seasonal mean anomalies, with a breakpoint flagged for further assessment if it was identified in either the monthly series, or (within a window of ± 1 year) in at least two of the four seasons. Further details of the implementation of the PHA in the ACORN-SAT dataset are available in Trewin (2012).

A range of other detection methods have been developed in recent years, many of which were the subject of the COST-HOME intercomparison project (Venema et al., 2012). Three of these methods were selected for use in ACORN-SAT version 2, the selection primarily based on ease of implementation. These methods were:

- HOMER version 2.6, joint detection (Mestre et al., 2013)
- MASH version 3.03 (Szentimrey, 2008).
- RHTests version 4 (Wang et al., 2010).

All of these methods, which use different statistical approaches, have been successfully used across a range of networks since their development. Further details on their implementation are given in Appendix C.

In addition, an alternate version of the PHA method was used, which used annual mean anomalies (in combination with seasonal anomalies) instead of monthly anomalies. For those methods carried out on monthly data, the breakpoint was attributed to the year in which the monthly break was observed.

These five methods (PHA with monthly and annual data, HOMER, MASH, RHTests) were then run in parallel. Initially, a trial was carried out, in which breakpoints were evaluated and the size of the estimated adjustment (using the standard adjustment methodology described in section 3.2) calculated. It was also determined whether or not the estimated adjustment met the standard criteria for an adjustment to be implemented (0.3 °C in the annual mean, 0.3 °C in at least two of the four seasons, or 0.5 °C in one season). These results were then categorised according to the number of methods which had detected the breakpoint in question, within a two-year window. A total of 380 breakpoints were evaluated in this trial, using those parts of the ACORN-SAT dataset which did not involve merging multiple sites, or multiple breakpoints within a 5-year window.

Number of methods	Maximum temperature			Minimum temperature		
	Number of cases	Mean adjustment size (°C)	Percentage of cases meeting criteria	Number of cases	Mean adjustment size (°C)	Percentage of cases meeting criteria
1	57	0.26	48	51	0.26	51
2	69	0.35	80	64	0.38	70
3	32	0.42	87	41	0.48	88
4	14	0.54	100	25	0.64	92
5	9	0.79	89	18	0.79	100

Table 1 Size of estimated adjustments categorised by the number of methods (of the five tested – PHA monthly and annual, HOMER, MASH, RHTests) which identified a potential breakpoint at that location (within a two-year window).

These results show that in the large majority of cases where a potential breakpoint was identified by two or more methods, the breakpoint proved to be sufficiently large to meet the adjustment criteria used for the ACORN-SAT dataset.

In light of these results, a breakpoint was therefore considered a potential candidate breakpoint if it was detected either by the PHA monthly method (the existing operational method), or by at least two of the other four methods (within a two-year window).

In all cases, statistical detection methods were carried out separately for maximum and minimum temperature, as many inhomogenities affect maximum and minimum temperature differently. Where metadata existed and a potential metadata-defined breakpoint was within two years of a statistically detected breakpoint, the date defined in the metadata took precedence.

This process produced a set of potential breakpoints for further investigation, some detected by statistical methods and some by metadata. Where potential breakpoints were three years or less apart, they were considered according to the procedures outlined below for 'spike' breakpoints (as discussed further in section 3.2.4). In general breakpoints which were found in the first, or last, three years of a site's record were not considered further, as the uncertainty in statistical detection of inhomogeneities is typically large near the end points of a time series; this limit was relaxed to two years for breakpoints supported by metadata.

Breakpoints were attributed to 1 January in the year concerned, except for metadata-supported breakpoints with a more precise date. This has little impact on any data outside the year of the breakpoint. The results of Menne and Williams (2009) indicate that the uncertainty in the timing of a breakpoint detected by the PHA monthly method is typically in the order of a few months.

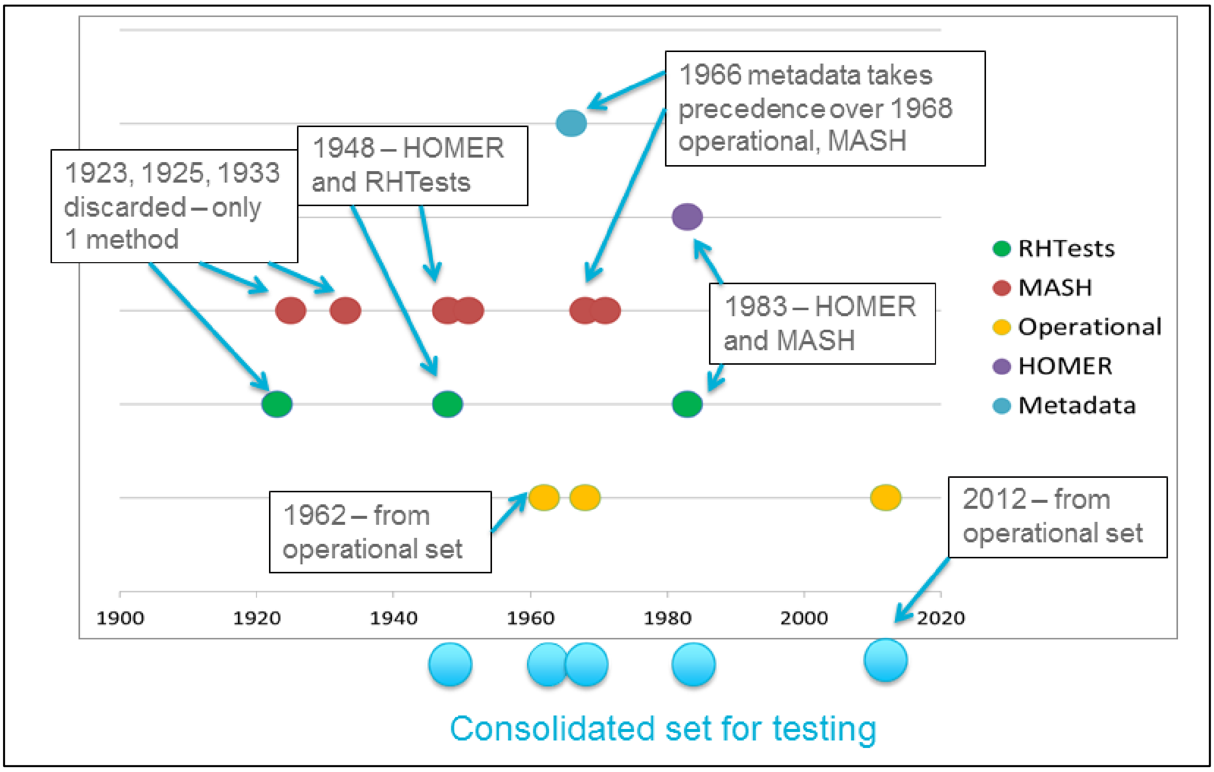

Fig. 2 An example of the consolidation of potential breakpoints for evaluation – Moruya Heads minimum temperature.

3.2 The adjustment procedure in ACORN-SAT

There were four different adjustment procedures used for version 2 of the ACORN-SAT dataset, depending on the circumstances as described further below. The percentile-matching overlap (3.2.1) and non-overlap (3.2.2) methods used are similar to those used in version 1, with more substantial changes to the monthly (3.2.3) and spike (3.2.4) methods.

Adjustments for time of observation changes, or for changes of screen size, were applied regardless of the adjustment size, these being systematic changes which affected a substantial part of the network despite their small size at individual stations.

In all other cases (except for 'spike' adjustments, for which a minimum 0.5 °C annual adjustment was applied), adjustments were applied only if they met at least one of the following criteria:

- At least 0.3 °C in the annual mean;
- At least 0.3 °C in at least two seasons. (In this case, the seasonal adjustments do not necessarily have to be of the same sign; there are cases where adjustments of opposite sign in winter and summer cancel out to a near-zero annual mean adjustment.)
- At least 0.5 °C in one season.

These criteria are based on an evaluation carried out at the time of the development of version 1 of ACORN-SAT (Trewin, 2012), which found that applying adjustments of less than 0.3 °C, on average, did not produce a time series which was a significantly better simulation of a homogeneous time series than would occur with no adjustment at all.

3.2.1 The percentile-matching (PM) algorithm – overlap case

This method was applied where two sites with an overlap of at least 12 months, with at least 50 observations in common for each set of three consecutive months of the year, were being merged into a single record. This typically occurs when a new site is opened but the former site is continued for a period of time as a comparison.

The algorithm involves the following steps:

(a) For each site, calculate daily temperature anomalies from the mean. These anomalies were from a climatology for each day, calculated by linear interpolation from the monthly means (attributing the monthly mean to the middle day of each month), with the monthly means calculated using a 1961-1990 base period if there were at least 12 years of data in the 1961-1990 period, or all years of record otherwise.

(b) Select the period used for data-matching between the sites. (In many cases this was the full period of overlap, but a subset was chosen in some cases where a breakpoint had been identified at one or both of the sites during the overlap period.) For this period, for each of the 12 calendar months, select the daily anomalies for all days that have data for both sites, and which are either in the target month or the month either side (for example, if March is the target month, February-April data are used). For example, if there are 5 years of overlap with complete data, this will result in a comparison set with between 450 and 460 paired observations for each target calendar month.

(c) For the comparison set of daily anomalies, for each site, calculate percentile points for the 5th, 10th, 15th, … , 90th and 95thpercentiles.

(d) Reconvert the anomaly percentile points to temperatures by adding the monthly mean for the target month.

(e) Define a transfer function using the percentile-point pairs as fixed points. For values below the 5th percentile, the temperature difference between the sites is assumed to be the same as the difference between the 5th percentiles (and similarly for values above the 95th percentile), with linear interpolation between the two nearest fixed points being used for values between the 5th and 95th percentiles.

(f) Convert values at the old site to equivalent values at the new site using the transfer function to produce a composite record, with the new site taking precedence where data exist at both sites.

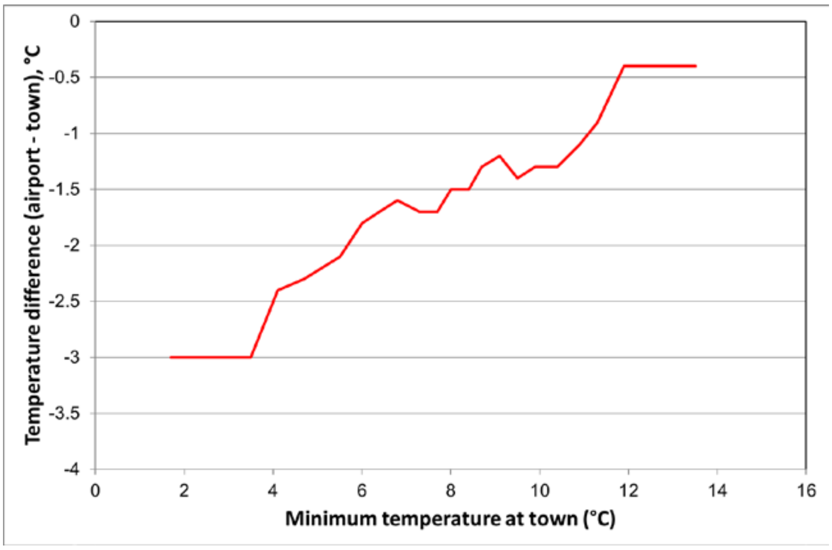

Fig. 3 Example of a transfer function – overlap case (Port Macquarie, July minimum temperatures). The changes in slope in this function indicate the 5th, 10th, …., 95th percentiles, with linear interpolation between these.

A period of parallel observations is only useful in defining the relationship between an old and a new site if the new site during the parallel period is representative of the new site following the parallel period, and the old site during the parallel period is representative of the old site prior to the parallel period. There are some situations where this was not the case, for a variety of reasons; for example, where significant building takes place near the old site before the end of the comparison, or where a new site is opened in an area which is still a construction zone for the first part of the comparison (Figure 4). Some such cases had been identified during the development of version 1, whilst others were identified through metadata, homogeneity testing of the individual site time series, or as a result of the second-round homogeneity testing process (section 3.2.5). In such cases, depending on the circumstances and the available length of parallel observations, either only a subset of the period of parallel observations was used to merge the sites (as described for Port Lincoln in section 8.1), or the parallel observations were not used at all, the adjustment instead being carried out using nearby reference stations as in the non-overlap case described below.

Fig. 4 The new Canberra Airport site (site number 70351) (left) at the time of its opening in November 2008 and (right) in March 2016. The lack of established ground cover during the 2008-2010 period of parallel observations with the old site (70014) makes that period unrepresentative of its behaviour during the post-2010 period; the use of neighbouring sites as references indicates that minimum temperatures at the new site are about 0.5 °C lower than the data from the parallel observation period would suggest.

3.2.2 The percentile-matching algorithm – non-overlap case

The most commonly used adjustment method used was the non-overlap case of the percentile-matching algorithm, which was used in all cases except those with overlap data, those where there were insufficient reference stations, or those where two breakpoints were sufficiently close in time to warrant use of the 'spike' method (see below).

In the non-overlap case, a set of reference stations well-correlated with the candidate station was used. The reference stations were chosen in correlation order with the candidate station, excluding reference stations with known inhomogeneities in the five years before or after the breakpoint being tested. (For these purposes, the correlation between daily temperature anomalies at site pairs was calculated for each of the 12 calendar months, and the correlation index was determined as the median of those 12 monthly correlation values, avoiding distortion arising from the strong seasonality of some correlation length scales, especially for minimum temperature in the tropics, where correlation length scales in the dry season are several times those in the wet season.) Only stations with a correlation index of 0.6 or above were considered; if there were sufficient reference stations meeting this requirement, the 10 best-correlated stations were used as a reference. (The validity and use of the 0.6 correlation threshold is discussed further in Trewin (2013).)

For each reference station, a two-step procedure was used. First, a transfer function was defined, as in the overlap case above (3.2.1), between the candidate station and the reference station, for a period before the breakpoint (the 'first reference period'). A second transfer function was then defined between the reference station and the candidate station for a period after the breakpoint (the 'second reference period'). These two transfer functions were combined to create a single transfer function matching the candidate station before the breakpoint to the candidate station after the breakpoint (Figure 5).

The first reference period was normally the five calendar years before the breakpoint and the second reference period the five calendar years after the breakpoint (for example, for a breakpoint in April 1986, the first reference period was normally 1981 to 1985, and the second reference period 1987 to 1991). A shorter reference period (with a minimum of three years) was used where there was another breakpoint within what would normally be the reference period.

There were also some cases where the period immediately before or after the breakpoint was unrepresentative (a common scenario here is that something occurs at a site – e.g., the construction of a new building nearby – and the site is moved a year or two later). In such cases, the reference periods were sometimes shifted earlier or later to avoid the unrepresentative period. To make such cases more detectable, a diagnostic was included in the adjustment code which indicated mean differences between the candidate station and reference stations in each individual year, enabling years where those differences were anomalous to be identified.

The algorithm operates as follows:

(a) Identify a set of N neighbouring sites with sufficient overlapping data with the candidate location both pre- and post-breakpoint (a minimum of 50 observations for each set of three consecutive months of the year).

(b) For each neighbour separately, define a transfer function between the candidate site pre-breakpoint and the neighbour, using the method for the overlap case as described above. As noted above, the period of comparison was generally the five calendar years prior to, but not including, the year of breakpoint (e.g., if the breakpoint was in 1994, 1989-1993 data were used).

(c) Also for each neighbour separately, define a transfer function between the neighbour and the candidate site post-breakpoint, also using the method for the overlap case described above, with the period of comparison generally the five calendar years following, but not including, the year of breakpoint.

(d) For each value at the old site, create an ensemble of *N* estimated equivalent values at the new site, using the *N* pairs of transfer functions defined in steps (b) and (c) (each of which converts a value at the old site to an equivalent value at a neighbour, then the equivalent value at that neighbour to a value at the new site).

(e) Calculate the 'final' equivalent value as the median of the ensemble of N estimates defined above. The use of the median is to minimise the influence of an individual outlying reference station (e.g. one which had an undetected inhomogeneity of its own).

Figure 5 shows the process of developing a transfer function using an individual reference station, whilst Figure 6 shows the process of consolidating information from a number of reference stations into a single transfer function.

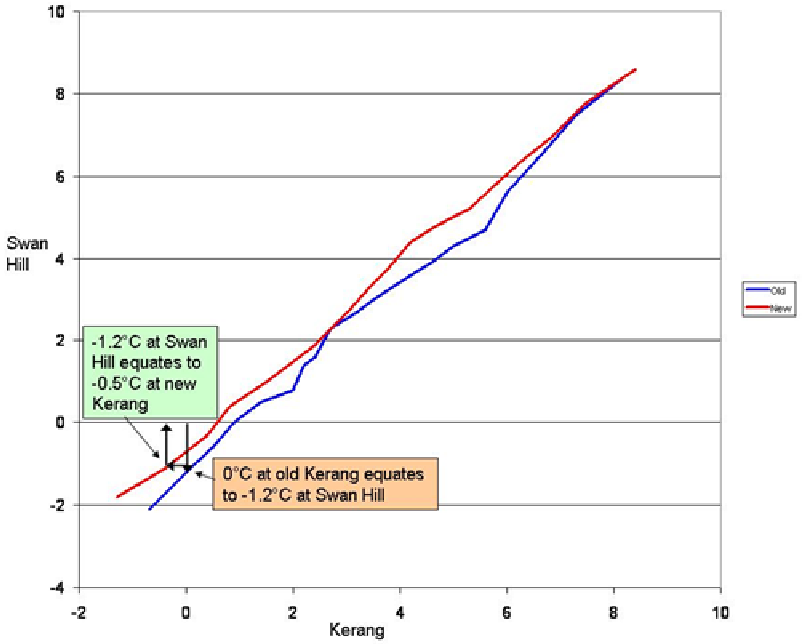

Fig. 5 Example (winter minimum temperatures) of the two step-procedure for a 2000 breakpoint at Kerang. The blue line matches 1996-1999 data at Swan Hill and Kerang, the red line matches 2001-2005 data at Swan Hill and Kerang.

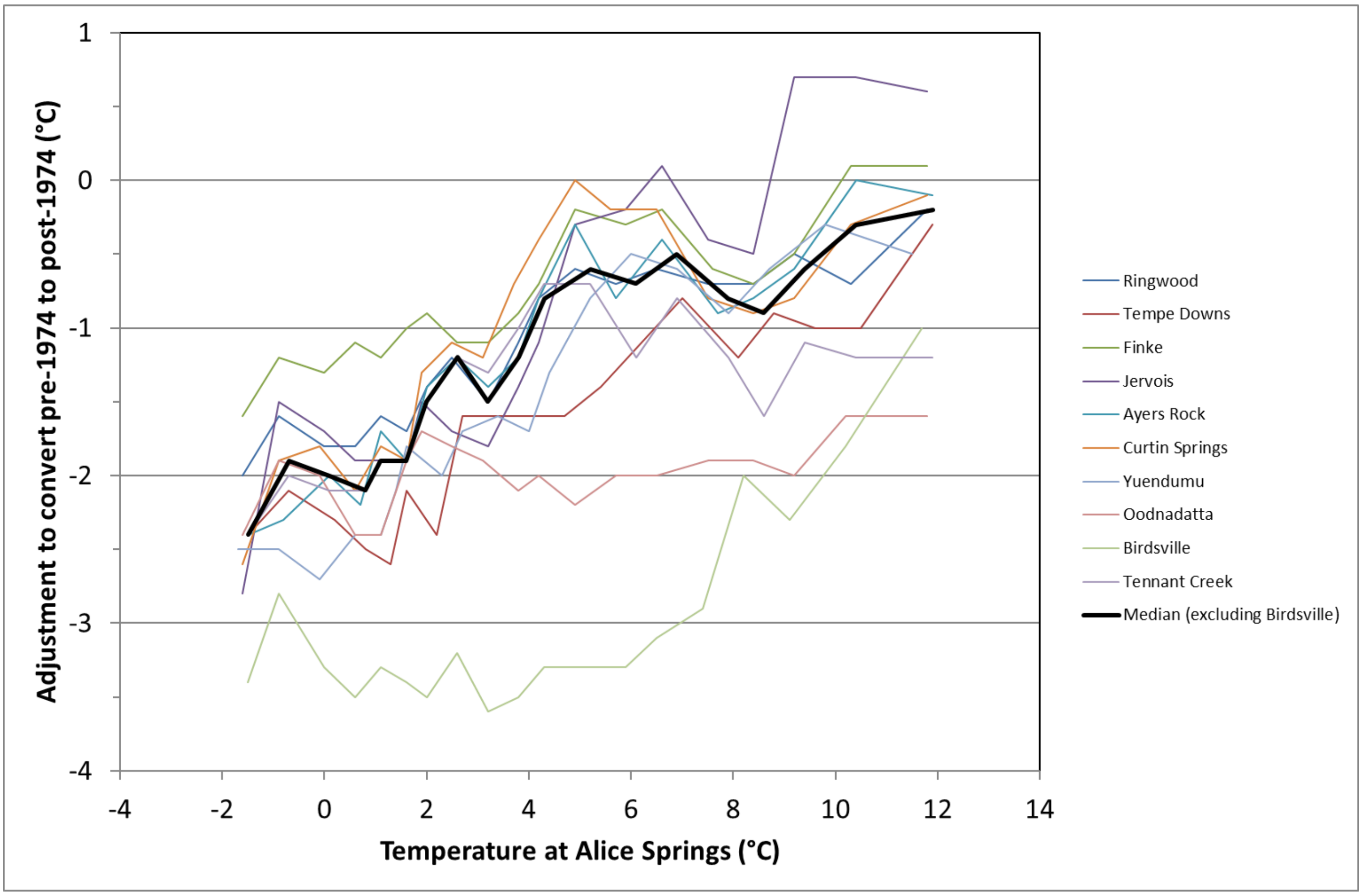

Fig. 6 Example of the consolidation of information from multiple reference stations into a median transfer function – minimum temperature for the July 1974 site move at Alice Springs. Birdsville (lowest line) has a substantial inhomogeneity of its own in 1973 and is hence excluded from the median used for adjustment.

Although the use of multiple reference stations provides a high level of robustness against undetected inhomogeneities at individual reference stations (section 5.3), it was still considered appropriate to seek to identify and remove potentially affected reference stations (e.g. Birdsville in Figure 6 above). A diagnostic was included in the code which provided, for each individual reference station, the mean annual adjustment which would have been made had that station been the only reference station. If this value differed excessively from the overall adjustment, the reference station was removed and replaced by the next best-correlated reference; this procedure was repeated until either there were 10 reference stations which did not vary excessively from the overall adjustment, or there were no further potential reference stations available. The criteria for 'excessive' difference were:

- In the initial assessment: differing by more than 0.3 °C (unless at least 50% of the reference stations differed by this amount, in which case the threshold was increased to 0.4 °C).
- In subsequent rounds: differing by more than 0.4 °C.

In some cases, reference stations were excluded by this test because of short-term issues (e.g., temporary vegetation responses to high rainfall in arid areas (section 8.2)) rather than more enduring inhomogeneities. Particularly large numbers of exclusions tended to occur in arid regions for breakpoints around the time of widespread heavy rainfall associated with strong La Niña events, such as those of 1973-74 and 2010-11.

3.2.3 Monthly adjustments

If there were insufficient reference stations with daily data sufficiently well correlated with the candidate station for the PM algorithm to be used, a monthly adjustment method was used instead. This situation principally arises prior to 1957 (when a large number of potential reference stations have data which are only digitised at the monthly timescale), mostly in areas which are data-sparse, or areas which have short correlation length scales (e.g., near some coastlines). Whilst applying a uniform monthly adjustment generally produces inferior results to the percentile-matching method, especially in the simulation of extremes (Trewin, 2012), it is still preferable to making no adjustment in situations where the percentile-matching method cannot be implemented.

In general, only stations with correlations of 0.6 or above were used as reference stations, but if fewer than three such reference stations were available, this limit was relaxed to 0.5. If there were still fewer than three available reference stations, the reference stations were weighted by inverse distance from the candidate station, rather than as per the procedure outlined above. (There was one case, at Darwin in 1937, where only one reference station was available; this adjustment was well supported by metadata.) Reference stations were chosen by the same method as used for daily adjustments above, and the same procedure was followed for the exclusion and replacement of potentially inhomogeneous reference stations.

In version 1 of ACORN-SAT, monthly adjustments were calculated by comparing monthly mean anomalies (using a 1961-1990 baseline) at the candidate station during the first reference period with a weighted mean of anomalies at the reference stations, doing likewise in the second reference period, and then comparing the candidate-reference difference between the two reference periods. The weighting function used was:

$w_s = \exp(-(d/100)^2)$, where d is the distance in kilometres between site s and the candidate site.

It was found, in evaluation of version 1, that the effect of this weighting function was to give excessive weight to the nearest reference station in data-sparse areas (for example, $w_s = 0.368$ for $d = 100$ km, but only 0.105 for $d = 150$ km and 0.018 for $d = 200$ km). In turn, this made the adjustment not robust against potential issues at that reference station.

In version 2, the anomalies at the reference stations were instead combined using a weighted median. Here, the weighting function is defined as:

$w_s = r_s^2/(\Sigma\, r_s^2)$, where r_s is the Pearson correlation coefficient between site s and the candidate station.

(By definition, the values of w_s sum to 1.)

The weighted median is then defined as the station monthly mean temperature anomaly at station k, T_k, where the anomalies at all N reference stations are ranked from lowest to highest, $T_1, T_2, \ldots, T_N$, and k is the lowest value such that $w_1 + w_2 + \ldots + w_k \geq 0.5$. (If the sum of the weights is exactly 0.5, then the weighted median is the mean of T_k and T_{k+1}.)

3.2.4 'Spike' adjustments

The 'spike' adjustment methodology was used in cases where two potential breakpoints were three years or less apart. As a period of less than three years was not considered reliable enough to be a basis for daily adjustment, and to prevent a relatively high-uncertainty adjustment from influencing other parts of the dataset, such periods were adjusted independently, with the data during the 'spike' period being adjusted, using the monthly methodology above, to match the data from before the 'spike' period. Such adjustments were implemented only if the size of the adjustment in the annual mean was 0.5 °C or greater.

The time series was then adjusted on either side of the 'spike' if required (using the daily or monthly method depending on data availability), using a period before the spike as the first reference period and one after the spike as the second reference period.

3.2.5 Overall summary of adjustments

An overall summary of the total number of adjustments is below. The percentile-matching method with daily data and no overlaps accounts for about 80% of all adjustments, with about 10% from the overlap case, 7% from monthly adjustments and 3% from spikes.

Method	Maximum temperature	Minimum temperature	Total
Percentile-matching with overlap	44	47	91
Percentile-matching with no overlap	379	401	780
Monthly	27	43	70
Spike	13	12	25
Total	463	503	966

Table 2 Total number of adjustments using different methods.

3.2.6 Second round of homogenisation

It was found in version 1 that, after the initial stage of homogenisation takes place, best results will be obtained if a second round of homogenisation takes place to identify possible inhomogeneities in the 'homogenised' series.

The most likely scenarios for residual inhomogeneities after the first round of homogenisation are:

- Inhomogeneities occurring at numerous stations at around the same time, either due to a systematic change (e.g., change in observing practices) which affects numerous stations, or by coincidence. (One example of this, identified in version 1, occurred in northwestern New South Wales and southwestern Queensland in the late 1940s.)
- An adjustment calculated using unrepresentative data; for example, a merge using overlap data which is unrepresentative of the period before or after the overlap, or an adjustment which uses a reference period which is not representative of the longer-term behaviour of the site (e.g., because of building or vegetation changes).

To detect potential cases, the annual mean maximum and minimum temperatures of each homogenised time series were tested for homogeneity, using the RHTests method (which is less dependent on having a relatively large number of reference stations as the other methods used in the first round of detection) and using, in general, the homogenised data from the four nearest ACORN-SAT locations as reference stations. In addition, trends over each 40-year sub-period from 1910-1949 to 1970-2009 were calculated, and stations showing trends strongly anomalous when compared with their neighbours were flagged.

Time series flagged through this procedure were subjected to additional sensitivity testing: depending on the location, this could involve comparing the results of an overlap with those obtained through comparison with other nearby stations, testing the sensitivity of the adjustment size to the choice of reference period, or testing possible breakpoints found in the initial homogenised series to determine if a potential adjustment would meet the normal 0.3 °C minimum criterion.

In total, 22 of the 966 adjustments (12 maximum, 10 minimum) applied in version 2 of the ACORN-SAT dataset arose from this second-round procedure.

3.2.7 Observation time adjustments

Changes in observation time are well known to be a potential source of inhomogeneities in temperature records (e.g., Karl et al., 1986), particularly in the United States of America where the bulk of the observation network does not have a standard observation time. In other countries, changes in observation time have sometimes been made across most or all of the network simultaneously (e.g., Canada in 1961 (Vincent et al., 2009) and Norway in 1938 (Nordli, 1997)), making it difficult to use methods involving reference stations for adjustment.

In Australia, an observation time of 0900 local time has been used across the network since 1910, with two significant exceptions:

- Between 1932 and 1963, Bureau-staffed stations (about 30% of the ACORN-SAT network at the time) used a midnight-midnight day instead of the standard 0900-0900 day.
- Some of the early-generation automatic weather stations measured minimum temperature for the 24 hours ending at 0000 UTC (0800-1100 local time, depending on location and season) and maximum temperature for the 24 hours ending at 1200 UTC (2000-2300 local time). This affected a small amount of ACORN-SAT data in the 1990s and early 2000s, principally in South Australia.

The introduction of daylight saving time in the late 1960s or early 1970s in Tasmania, Victoria, New South Wales and South Australia (with brief trials in Western Australia and Queensland) also introduced an effective one-hour shift in the observation time.

The impact of these changes was assessed in detail as part of version 1 (Trewin, 2012). The use of a midnight-midnight day was estimated to have an impact of −0.08 °C on national mean minimum temperature, which was considered large enough to warrant adjustment, given the large number of locations involved. The impact was most significant on the occurrence of extreme high minimum temperatures, as very warm nights are often followed by a rapid cooling during the day, resulting in the high minimum temperature being 'lost'. (As an example, Melbourne had only one minimum temperature of 25 °C or above in the 32 years from 1932 to 1963, compared with 10 in 22 years from 1910 to 1931, and 35 in 53 years from 1964 to 2016.) There was no significant impact on maximum temperature, nor did the 0000/1200 UTC day or the daylight saving shift have any significant impact on either maximum or minimum temperatures. The impact of a midnight-midnight day on minimum temperatures was also much less significant at tropical locations than it was further south.

The adjustment was implemented by carrying out the standard adjustment procedure on 1 January 1964 (unless there was an already identified inhomogeneity adjusted for within two years of that date) at all Bureau-staffed sites outside the tropics, and implementing the adjustment regardless of size (even if it did not reach the normal minimum adjustment criteria).

3.3 Adjustment methodological changes in ACORN-SAT version 2

This section documents significant methodological changes between ACORN-SAT version 1 and version 2. In addition to this, there were a number of minor bug fixes in operational code which are described in a separate document.

3.3.1 Removal of rounding biases

In the original version of ACORN-SAT, estimated adjusted values based on each individual reference station were rounded to one decimal place prior to the calculation of the median of the estimate from all reference stations. The median was then rounded to one decimal place as the final adjusted value for that adjustment. When taking a median of values which have been rounded to one decimal place, the median will, if there is an even number $2n$ of reference stations, be the mean of the n-th and $(n+1)$-th values. In turn, this means that if the first decimal place is odd for one of these two values and even for the other (which, on average, will occur in 50% of cases), the second decimal place of the median will be 5 (e.g. the mean of 19.0 and 19.3 is 19.15).

It was found after the completion of the previous version that the *anint* intrinsic function in the implementation of Fortran used in this work rounds such values up to the nearest whole number, and hence all such values were rounded up when rounded to one decimal place. In the new version, these values were retained as full-precision reals throughout the process and only rounded to one decimal place in the final adjusted time series after all adjustments had been applied.

Since 50% of cases, on average, are affected and the resultant expected bias in such cases is +0.05 °C, the net result of this is an expected rounding bias of +0.025 °C in any adjustment in version 1 which uses an even number of reference stations. As the standard operating procedure was to use 10 reference stations, this applied to the majority of adjustments in version 1 of ACORN-SAT (the exceptions being merges of parallel stations, adjustments which used monthly data, and some adjustments where fewer than 10 reference stations were available). The removal of this bias in version 2 contributes to version 2 having slightly stronger warming trends than version 1, as discussed further in section 7.

3.3.2 Detection of date shifts

There are a number of cases in the historical record where the maximum temperature is shifted by one day. As noted earlier, the standard observation period in Australia for daily maximum and minimum temperature is the 24-hour period ending at 0900 local time, with the minimum being attributed to the day of observation and the maximum to the previous day. However, there are a number of cases where the maximum was actually recorded against the day of observation, creating an effective one-day date shift in the record.

In version 1 of ACORN-SAT, these were typically detected on an *ad hoc* basis as a result of the station concerned failing other quality control tests (for example, consistency between the maximum temperature and fixed-hour temperatures, or excessive spatial variation with other nearby stations). The issue is most easily detected at stations with two (or more) observations per day, as a station with its maximum temperatures shifted by a day for any substantial period of time will trigger a number of cases where the "maximum" (which is actually from the previous day) is less than the temperature at 1500. It is less readily detected at stations with 0900 observations only.

To ensure that there were no undetected cases of date-shifted observations over a substantial period of time, a test was run for each station. For each station, and each individual month, the daily maximum temperature was compared with that at a number of nearby stations, and the correlation between the maximum temperatures at the station pairs was calculated for (a) the data as they were recorded and (b) the data with the candidate station shifted by one day. Stations/months where, for a majority of reference stations, the correlations were higher in the date-shifted case than for the original data were flagged for further investigation.

This process revealed that a large majority of date-shift cases had already been detected in version 1, but a few additional cases were found and corrected for version 2. In most cases, date shifts persist for a few months at most, but at one location (Rutherglen), the maximum temperatures were date-shifted for much of the 1920s and 1930s (as had been detected in version 1). During this period, only 0900 observations were made at Rutherglen.

3.3.3 Adjustment of negative diurnal ranges

In the ACORN-SAT dataset, maximum and minimum temperatures are adjusted independently of each other. This means that situations can occasionally arise where a zero or small diurnal temperature range in the raw data can be adjusted to a negative diurnal range in the adjusted data. (Since the 'diurnal range' effectively combines the minimum from one 24-hour period with the maximum for the next, a zero diurnal range does not mean that the temperature remains constant for 24 hours – rather, it can occur, for example, when the temperature is falling at the 0900 observation time and does not rise further during the day.)

Whilst such a result is methodologically consistent, it is physically unrealistic. A procedure was therefore adopted under which, if a day had a negative diurnal range in the adjusted data, the maximum and minimum temperatures were each corrected to the mean of the original adjusted maximum and adjusted minimum, creating no change in the daily mean.

About 0.03% of all days were affected by this issue, with adjustments of 1 °C or more in only 90 cases (out of about 3.5 million observation days). At the exposed mountaintop site of Cabramurra, about 1% of days were affected, but no other individual station had more than 0.3% of days affected.

3.3.4 Transition from large to small thermometer screens

Both 'large' and 'small' screens have been used in the Australian temperature network (Figure 7).[2] Originally, large screens predominated in the network, but over time there has been a change to small screens at most sites. The highest frequency of such changes was in the 1990s, but some took place as early as 1967. Only four of the 112 ACORN-SAT sites still have large screens.

A field trial carried out at a Bureau test site at Broadmeadows, Victoria (Warne, 1998) in 1995−96 found that the 'small' screen had mean maximum temperatures 0.094 °C higher than those in the 'large' screen, and mean minimum temperatures 0.082 °C lower. These combine to produce a negligible impact on mean temperatures (+0.006 °C), which was considered too small to warrant adjustment in ACORN-SAT version 1. The impact on diurnal temperature range (+0.176 °C), however, was considered sufficiently large to warrant specific treatment in ACORN-SAT version 2, given the large proportion of the network which was affected.

---

[2] 'Large' screens are approximately 71 cm high and 71 x 53 cm horizontal dimensions; 'small' screens (referred to as 'medium' in Warne (1998)) are 43 x 52 x 27 cm.

Fig. 7 A large screen at Gunnedah (left) and a small screen at West Wyalong (right).

The procedure followed was similar to that already implemented in version 1 for changes in observation time as described above. Specifically:

- At stations where there was a known screen change date without a change in site number (except at sites where an inhomogeneity had already been found and adjusted for within 2 years of the screen change date), the data were adjusted for the screen change using the normal adjustment procedures, but with no minimum adjustment size threshold (i.e. the standard 0.3 °C threshold was not used). Only sites with no known screen changes within five years of the candidate station's screen change date were used as reference stations.
- At stations where a screen change occurred at the same time as a documented site change with a change in site number (with or without overlap), the minimum adjustment size criteria were not applied, so any adjustments for such changes which had not been implemented in the first phase because they failed to meet minimum adjustment size criteria were applied in this stage. These adjustments were derived using either overlap data (if available) or reference station data as per above.

In total, 86 stations had documented screen changes with an exact or approximate date.

There are 22 stations which currently have a small screen and either have no documentary evidence of a large screen ever having been in place (noting the limitations of pre-1960s metadata), or where the date of a change cannot be determined with any confidence. There are an additional four stations which still have a large screen. No adjustments for screen size took place at these 26 stations.

For the stations where the screen change was not associated with a previously identified inhomogeneity (42 stations for maximum temperature, 45 stations for minimum temperature), the mean adjustment was +0.04 °C for maximum temperature and −0.06 °C for minimum temperature (combining to a result of +0.10 °C for diurnal temperature range and −0.01 °C for mean temperature). These inhomogeneities, except for the mean temperature shift (negligible in both cases) are of the same sign as those found by Warne (1998) but somewhat smaller in magnitude. The spread of the results between stations was wide (several tenths of a degree), suggesting that the required adjustments are site-specific (one potential influence being the condition of the screen being replaced). Also of interest is a slight cool bias at tropical stations (an environment not sampled in Warne (1998)), with the 11 tropical stations having a mean shift of −0.02 °C for maximum temperature and −0.12 °C for minimum temperature (+0.10 °C for diurnal temperature range, −0.07 °C for mean temperature).

4. EVALUATION OF OTHER POTENTIAL SYSTEMATIC ISSUES

4.1 The transition to automatic weather stations

Australia, like many other countries, has seen a transition to automatic weather stations (AWSs) over the last few decades (WMO, 2017). These use a platinum-resistance probe in place of manually read liquid-in-glass thermometers. Unlike some other countries which changed their thermometer screen design at the same time as they introduced AWSs (Quayle et al., 1991; Brandsma and van der Meulen, 2008) or retained the same screen design but used plastic instead of wood (Perry et al., 2007), Australia retained the same wooden Stevenson screen design used at manual stations.

The first AWS data in the ACORN-SAT dataset are from 1994, although some non-ACORN-SAT stations had used automatic instruments since the 1980s. By 1996, AWSs had been installed at most Bureau Meteorological Offices to operate in parallel with the manual instruments. On 1 November 1996, the AWS became the primary instrument for temperature at those sites which had both types of instruments. AWSs continued to spread throughout the network over the following years, often installed at airports (or similar) to replace manual sites in town centres. By 2018, 98 of the 112 ACORN-SAT stations were automatic, with another two currently having parallel automatic and manual measurements. It is expected that the remaining manual ACORN-SAT sites will transition to automatic measurements over the next few years.

A major change of instrument type, such as a change from manual to automatic instruments, has the potential to introduce a substantial inhomogeneity into the temperature record (WMO, 2017). Some potential causes of such an inhomogeneity include:

- A change in the properties of the instrument itself, such as its response time;
- A change in the thermometer screen;
- A change in observation procedures (e.g., changing to a midnight-midnight observation day which is more practical for automatic instruments than it would be for human observers);
- The automatic station being installed at a different site to the manual station it is replacing.

As noted above, in Australia, the thermometer screens did not change. In general, the same observation time was also retained, although some of the earliest AWS data in the 1990s used a 0000 UTC observation day for minimum temperatures and 1200 UTC for maxima (in contrast to the 0900 local time otherwise used as a standard), something which had no significant impact on mean temperatures (see section 3.2.6). However, many automatic weather station installations also involved site moves. These were dealt with in the same way as other known site moves (both in version 1 and version 2 of ACORN-SAT).

Previous analyses carried out prior to the release of ACORN-SAT version 1 (Trewin, 2012) assessed the impact of the change from manual to automated observations, using sites where there were no known changes around the time other than the instrument change, and using only manual stations as reference stations. (In most cases, these were sites where both manual and automatic instruments were in place, which switched from manual to automatic observations as their primary instrument on 1 November 1996). This analysis showed no significant systematic change to either maximum or minimum temperature, with a mean inhomogeneity with AWS introduction of −0.04 °C for maximum temperature (positive at 8 locations, negative at 14) and −0.03 °C for minimum temperature (positive at 13 locations, negative at 13).

Another approach is to consider locations where automatic and manual observations took place at the same time at the same place, or in very close proximity. There are nine such locations with potentially suitable parallel observations:

- Cape Byron (28.64 °S, 153.64 °E) and Marble Bar. At these sites, manual and automatic instruments were in the same screen but their data were archived under different station numbers.
- Point Perpendicular, where the manual and automatic instruments were in different screens three metres apart.
- Mildura, Adelaide, Charleville, Cairns, Launceston Airport, and Giles, which switched from manual to automatic instruments as the primary instrument on 1 November 1996, but with manual observations (in the same screen) continuing to be made as a comparison. (This was also done for a period of time at many other Meteorological Offices, but these are the only such sites to date where the post-1996 comparison data have been digitised.)

On examination of the digitised data, Cairns, Charleville and Adelaide had anomalous data for a period during the parallel observations. At Charleville and Adelaide this appears to be related to a probe which failed a tolerance check and was subsequently replaced, whilst at Cairns it is most likely the result of a faulty manual minimum thermometer (the minimum thermometer was replaced twice in late 1999 and early 2000). These three locations were excluded from further consideration, leaving six locations available.

Results from this comparison are shown in Table 3. Included in this are the result of instrument tolerance checks on the automated probes carried out during the course of the parallel observations, as part of routine maintenance. It may be seen that:

- Differences in mean temperature between manual and automatic instruments are less than 0.25 °C at all six locations. Furthermore, at two of the sites (Mildura and Cape Byron) where the differences are more than 0.1 °C, these differences align closely with the results of tolerance checks on the automated probe, indicating that the differences are primarily accounted for by the probe being slightly out of calibration (something which should affect maximum and minimum temperature equally). The differences between manual-automatic differences and tolerance check results are larger at Giles and Launceston Airport but only at Giles do they exceed 0.1 °C.
- Differences in diurnal temperature range between manual and automatic instruments are small at all six sites, and are not consistent in sign. The largest shift in DTR (+0.13 °C) is at Point Perpendicular, the one location where the manual and automatic instruments are not in the same screen, potentially increasing the uncertainty associated with the comparison.

Location	Dates	Difference (automatic – manual) (°C)				Mean tolerance check result (°C)
		Maximum temperature	Minimum temperature	Mean temperature	Diurnal temperature range	
Mildura	Dec 1997 – May 2000	−0.22	−0.17	−0.19	−0.05	−0.11
Point Perpendicular	May 2001 – Jun 2004	0.08	−0.05	0.01	0.13	0.05
Cape Byron	Nov 2002 – Aug 2007	0.12	0.14	0.13	−0.02	0.16
Marble Bar	Sep 2000 – Dec 2004	0.01	−0.10	−0.05	0.11	0.02

Giles	Nov 1996 – Dec 2001	-0.27	-0.17	-0.22	-0.10	-0.06
Launceston Airport	Nov 1996 – Dec 2001	-0.10	-0.21	-0.15	0.11	-0.06
Mean		-0.06	-0.09	-0.08	0.03	0.00

Table 3 Differences between automatic and manual temperatures at locations with parallel observations, and tolerance check results on the automatic probes during the automatic observations period.

These results indicate that there is no evidence of any significant systematic change in Australian maximum and minimum temperatures arising from the transition from manual to automatic instruments. In particular, any change in diurnal temperature range, as would occur if there were a major change in effective response time between the two instrument types, appears to be minimal, in the order of a few hundredths of a degree at most. (A shorter response time, as discussed further in the next section, would potentially lead to greater sampling of short-term fluctuations in temperature, and hence higher maximum temperatures, lower minimum temperatures, and a larger diurnal temperature range.)

These results indicate that no network-wide adjustment for the transition from manual to automatic instruments is warranted in the ACORN-SAT dataset, only site-specific adjustments in cases where a site change has occurred when the automatic instruments are installed. 41 such adjustments associated with site moves were carried out.

4.2 Potential changes in response times with changes in automatic weather station probes

In the absence of any other influences, an instrument with a faster response time will tend to record higher maximum and lower minimum temperatures than an instrument with a slower response time. This is most clearly manifested as an increase in the mean diurnal range. At most locations (particularly in arid regions), it will also result in a slight increase in mean temperatures, as short-term fluctuations of temperature are generally larger during the day than overnight (Trewin, 2018).

The lack of any substantial signal in diurnal temperature range at the time of the 1996 transition from manual to automatic weather stations suggests that the automatic weather stations used at that time had similar response time characteristics to the manual instruments which they were replacing.

The question remains as to whether any subsequent changes in automatic weather station technology have impacted instrument response times. The key instrument used for temperature measurement in Australian automatic weather stations is a platinum-resistance probe. The most common probe type currently in the Australian network is the Rosemount ST2401, but numerous other types of probes are also found in the ACORN-SAT network (Rosemount with no version number, Temp Control TCBMP01, Wika TR40).

To assess the effective response time of the observations, mean values of the one-minute temperature variation (that is, the difference between the highest and lowest temperature in a 1-minute period) were calculated from the available one-minute data. At 1500 local time in summer (near the typical time of maximum temperature), these mean values are typically between 0.15 °C and 0.30 °C, reaching up to 0.40 °C in some arid regions, but at 1500 in winter they are below 0.10 °C at most southern locations. At 0600 local time, near the usual time of minimum temperature, mean values of the one-minute variation are between 0.05 °C and 0.10 °C all year.

At 17 of the 98 ACORN-SAT locations with automatic weather stations and one-minute data, there was a significant breakpoint during the period of automatic observations in the time series of annual means of the mean one-minute temperature variation. Such a breakpoint indicates a likely change in the response time of the instrument, and will hence affect its sampling of maximum and minimum temperatures. In 16 of the 17 cases, the breakpoint coincided with the documented replacement of a Rosemount probe with no version number (some were replaced by a Rosemount ST2401, some by a Wika TR40 and some by a Temp Control).

Figure 8 shows two examples of this. Alice Springs is the most extreme example; the November 2011 probe replacement there resulted in an increase in mean one-minute temperature fluctuations of approximately 0.16 °C at 1500 and 0.03 °C at 0600. Assuming that the increased variation is distributed symmetrically about the one-minute mean and that the change of probe did not introduce any inhomogeneities into the one-minute mean, this equates to an upward shift of about 0.08 °C for maximum temperature and a downward shift of 0.01-0.02 °C in minimum temperature. In less arid climates the effect is smaller (e.g., for Sydney, around +0.03 °C for maxima and –0.01 °C for minima).

Given the relatively small proportion of the network which is affected, it is estimated that the overall effect on national maximum temperatures is in the order of +0.01 °C, and on minimum temperature, between zero and −0.01 °C. In the context of overall Australian temperature change and variability, these network-wide impacts are negligible, whilst even at the worst-affected stations, the size of the impact falls well below the 0.3 °C minimum threshold normally applied for station-specific adjustments in the ACORN-SAT dataset. No specific adjustment for this change was therefore made in version 2 of ACORN-SAT.

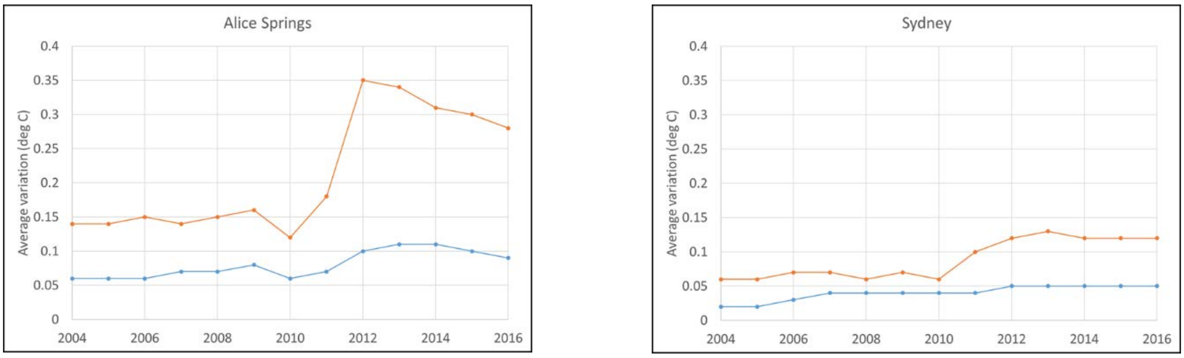

Fig. 8 Mean annual values of one-minute temperature variation at 1500 (orange line) and 0600 (blue line) local time, showing the effect of temperature probe changes at Alice Springs (left, November 2011) and Sydney (right, June 2011).

5. SENSITIVITY AND UNCERTAINTY TESTING

5.1 Parallel station sets used in uncertainty testing

A subset of ACORN-SAT locations with periods of parallel observations between two stations was used to explore various aspects of uncertainty in the analysis, as described in subsequent sections. In total, station pairs at 16 locations were used, with a period of parallel observations between four and nine years. These pairs encompassed a range of circumstances, with some pairs having relatively small temperature differences (less than 0.3 °C) between the paired stations, and some others having relatively large differences (greater than 1 °C in some cases). The use of paired stations allows the construction of a synthetic time series which switches over from the 'old' station to the 'new' station at the start of the parallel observations period; reference stations can then be used to adjust this time series, which can then be compared with the continuation of the 'old' station as a baseline. In some cases, circumstances at a particular station pair (e.g., missing data at a candidate or reference station) prevented it from being used in a specific test.

The station pairs used and further details about these stations are shown in Table A3.

5.2 Sensitivity of adjustments to choice of reference stations

To determine the sensitivity of the size of adjustments to the choice of reference stations, a series of trials was carried out for 14 of the 16 test stations defined in Table A3, separately for maximum and minimum temperature.

For each station/element, the following process was followed:

- A set of potential reference stations was selected, consisting of all stations which had correlations with the candidate station above a certain threshold (defined in Table 4) and had data for at least three years before and after the breakpoint being tested (the start of the parallel observations period at the test station). This typically defined a pool of between 30 and 60 potential reference stations (the varying thresholds were set to achieve a pool of approximately this size).
- An ensemble of 50 separate sets of 10 reference stations were chosen at random from within the pool of potential reference stations. (This compares with the operational procedure for adjustment, which uses as a starting point the 10 best-correlated stations.)
- For each ensemble member, the 10 reference stations in that set were used to adjust the series at the test station, and a range of indicators describing the adjustment were recorded. These indicators were the size of the adjustments in the annual mean and the mean for the four calendar seasons, and the size of the adjustments for the 10th and 90th percentile values in summer and winter.
- The standard deviations of the above indicators were calculated across the 50 ensemble members.

The results of this analysis are shown in Table 4. This shows that at most stations, the uncertainty in annual and seasonal mean adjustments arising from the choice of reference stations is in the range from 0.05 °C to 0.1 °C (slightly higher for maximum temperatures in summer), whilst in most cases, the uncertainty in the adjustment size for the 10th and 90th percentile is in the range from 0.1 °C to 0.2 °C.

The largest uncertainties in the test dataset are for maximum temperatures at Birdsville, particularly large uncertainties for 10th-percentile maximum temperatures in summer. Birdsville is in a data-sparse region. While correlation length scales for maximum temperature in general in the region around Birdsville are long (allowing for a large number of potential reference stations), most likely due to the flat terrain and the lack of coastal influence for many hundreds of kilometres in all directions, individual cool days in summer at Birdsville (which are almost always associated with significant cloud cover, and often rain) may not be well-correlated with cool days at other locations.

The mean of the adjustments from the 50 ensemble members at each test station was also calculated, and compared with the mean difference between the paired parallel stations at that test station. In most cases, these matched reasonably well. However, there were some cases (two for maximum temperature, five for minimum temperature) where the mean adjusted data in the test series differed by more than 0.3 °C in the annual mean from the continuation of the older station in the parallel dataset (Table 4). In these cases, the results suggest that the parallel observations period is not representative of the true relationship between the old and new stations, potentially because of an inhomogeneity during the parallel observations period at either (or both) of the stations in the pair. The case with the largest differences for both maximum and minimum temperature, Port Lincoln, is discussed as a separate case study (section 8.1).

Location	Minimum correlation threshold	Seasonal and annual means (°C)					Winter percentile (°C)		Summer percentile (°C)		Difference (continuation – adjusted) (°C)
		MAM	JJA	SON	DJF	Annual	10th	90th	10th	90th	
Maximum temperature											
Cunderdin	0.8	0.083	0.073	0.063	0.069	0.059	0.076	0.134	0.151	0.125	-0.20
Wandering	0.8	0.100	0.051	0.047	0.133	0.070	0.086	0.075	0.123	0.155	-0.14
Port Lincoln	0.75	0.107	0.083	0.063	0.087	0.070	0.090	0.130	0.134	0.291	0.61
Birdsville	0.7	0.115	0.085	0.146	0.227	0.121	0.111	0.192	0.431	0.118	0.00
Gayndah	0.7	0.109	0.092	0.069	0.060	0.069	0.099	0.121	0.077	0.074	0.01
Brisbane AP	0.6	0.052	0.049	0.043	0.058	0.035	0.078	0.067	0.068	0.084	-0.09
Miles	0.7	0.080	0.072	0.082	0.140	0.073	0.090	0.112	0.169	0.181	0.28
Thargomindah	0.7	0.179	0.105	0.081	0.109	0.097	0.149	0.126	0.166	0.112	0.57
Pt Macquarie	0.6	0.064	0.067	0.098	0.092	0.060	0.120	0.116	0.133	0.139	0.00
Dubbo	0.8	0.126	0.075	0.088	0.078	0.070	0.125	0.099	0.150	0.140	0.22
Deniliquin	0.8	0.120	0.118	0.082	0.118	0.097	0.119	0.213	0.111	0.146	0.13
Nhill	0.8	0.081	0.062	0.067	0.074	0.056	0.079	0.092	0.155	0.131	-0.23
Sale	0.7	0.075	0.069	0.078	0.099	0.059	0.099	0.137	0.116	0.257	0.09
Launceston	0.7	0.069	0.041	0.064	0.079	0.056	0.061	0.060	0.099	0.171	0.05
Mean (14 stations)		0.097	0.074	0.077	0.102	0.071	0.099	0.120	0.149	0.152	
Mean (12 stations, excluding Port Lincoln and Thargomindah)		0.089	0.071	0.078	0.103	0.069	0.096	0.119	0.149	0.144	
Minimum temperature											
Cunderdin	0.7	0.098	0.087	0.069	0.086	0.076	0.122	0.127	0.094	0.150	-0.62
Wandering	0.8	0.069	0.105	0.067	0.082	0.066	0.143	0.071	0.129	0.126	0.04
Port Lincoln	0.6	0.059	0.056	0.048	0.049	0.045	0.079	0.122	0.149	0.052	1.54
Birdsville	0.6	0.131	0.132	0.087	0.144	0.101	0.190	0.151	0.110	0.209	0.18
Gayndah	0.7	0.069	0.074	0.067	0.059	0.054	0.119	0.145	0.067	0.081	0.13
Brisbane AP	0.7	0.064	0.087	0.079	0.068	0.061	0.121	0.097	0.103	0.101	-0.31
Miles	0.7	0.123	0.137	0.164	0.095	0.108	0.187	0.114	0.122	0.084	0.09
Thargomindah	0.6	0.106	0.079	0.083	0.087	0.075	0.144	0.089	0.126	0.172	-0.14
Pt Macquarie	0.6	0.069	0.153	0.083	0.080	0.075	0.210	0.144	0.103	0.080	0.73
Dubbo	0.7	0.161	0.165	0.110	0.085	0.112	0.201	0.121	0.147	0.129	0.36
Deniliquin	0.8	0.081	0.098	0.064	0.067	0.065	0.169	0.094	0.079	0.123	0.04
Nhill	0.7	0.084	0.066	0.061	0.063	0.052	0.090	0.086	0.137	0.146	-0.09
Sale	0.6	0.091	0.087	0.081	0.076	0.063	0.210	0.103	0.116	0.081	-0.23
Launceston	0.7	0.094	0.072	0.070	0.066	0.061	0.164	0.113	0.150	0.145	0.26
Mean (14 stations)		0.093	0.100	0.081	0.079	0.073	0.154	0.113	0.117	0.120	
Mean (9 stations, excluding Cunderdin, Port Lincoln, Brisbane, Port Macquarie and Dubbo)		0.094	0.094	0.083	0.082	0.072	0.157	0.107	0.115	0.130	

Table 4 Standard deviation of estimated adjustments from ensemble members at test stations, and mean temperature difference between mean adjusted dataset at test station and continuation of oldest station. Multi-station means are also calculated excluding those stations where the difference between the continuation and adjusted dataset is more than 0.3°C.

5.3 Sensitivity of adjustment to number of reference stations used

To gain an indication of the sensitivity of adjustment size uncertainty to the number of reference stations used, a series of trials was carried out using Port Macquarie, which was chosen as the size of the expected adjustments was relatively large for both maximum and minimum temperature.

These were carried out using a similar method to that described above to assess adjustment uncertainty due to reference station selection. However, the assessment was carried out for ensembles of 50 randomly chosen sets of *N* stations, where *N* took each value from 1 to 10. (Hence, 500 sets of stations were tested in total for each of maximum and minimum temperature.) As in the analysis in the previous section, the standard deviation of various indicators across the 50 ensemble members was then calculated.

Results of this testing are shown in Figure 9. These show that, while the uncertainty due to reference station selection generally decreases with an increasing number of reference stations, the relationship becomes relatively weak for most indicators for cases with 5 or more reference stations. Conversely, the uncertainty increases sharply when the number of reference stations is three or fewer, and particularly when a single reference station is used. As an example, for maximum temperature, the uncertainty in the annual mean adjustment is 0.060 °C when 10 reference stations are used, 0.078 °C when 5 reference stations are used, but 0.202 °C when a single reference station is used; for minimum temperature, the values are 0.075 °C, 0.098 °C and 0.271 °C respectively.

A likely driver of these results is that when the number of reference stations is small, the results can be susceptible to an inhomogeneity at a reference station. (In these trials, no separate homogeneity testing was carried out on reference stations.) Conversely, the relative stability of the results when five or more reference stations are used indicates that the use of the median outcome from all reference stations gives the adjustment method a relatively high level of robustness against an undetected inhomogeneity at a single reference station.

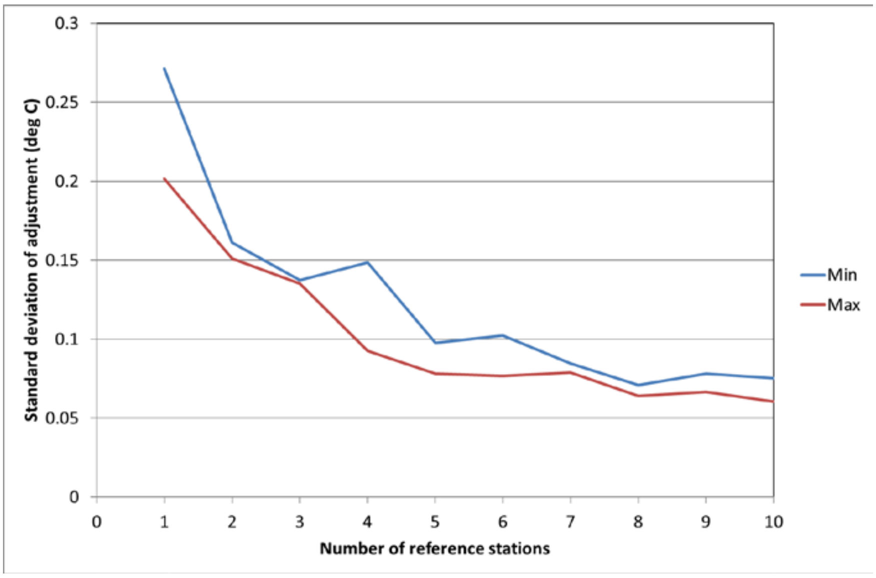

Fig. 9 Reference station-related uncertainty in adjustments as a function of number of reference stations – Port Macquarie example.

5.4 Sensitivity of adjustment to choice of reference period

A series of trials was carried out, using 15 of the 16 test stations defined in Table A3. These tests were designed to determine (a) the additional uncertainty which arises from using a shorter reference period to calculate an adjustment and (b) the effect of displacing the reference periods away from the date of the breakpoint.

The basis for these trials was that, instead of adjustments being carried out using as reference periods the 5 complete years immediately preceding and following the breakpoint being tested, a range of different reference periods was used, as shown in Table 5:

Reference period length (years)	Reference periods used in testing (before and after identified breakpoint)
1	Years 1, 2, 3, 4, 5
2	Years 1-2, 2-3, 3-4, 4-5, 5-6
3	Years 1-3, 2-4, 3-5, 4-6, 5-7
4	Years 1-4, 2-5, 3-6, 4-7, 5-8
5	Years 1-5, 2-6, 3-7, 4-8, 5-9

Table 5 Reference periods used for testing impact of reference period timing and size. The 'years' refer to the number of years displaced from the breakpoint being tested: hence, for example, where the period is 'Years 2-6', the first reference period was from 6 to 2 years before the breakpoint (inclusive) and the second reference period from 2 to 6 years after the breakpoint.

Results from some of these analyses are shown in Table 6. Some key conclusions are:

- When a single year is used as the reference period, the choice of the year used introduces an additional uncertainty into the adjustment in the order of 0.1 °C.
- The relatively large differences (on average, around 0.2 to 0.3 °C) between estimated adjustments based on a 1-year reference period and those based on periods of 3 years or longer suggest that using reference periods shorter than 3 years introduces substantial additional uncertainty into an adjustment.

These results reinforce the approach, used in both version 1 and version 2 of ACORN-SAT, of using five years as the standard reference period and, in normal circumstances, not shortening this to any less than three years.

Assessment of the five five-year periods used indicates that there is no consistent pattern, indicating that there is no reason to adopt a general rule of displacing reference periods away from the breakpoint. However, there may be benefits of displacing in particular cases if evidence suggests that the years immediately before or after the breakpoint are unrepresentative (as discussed in section 3.2.2).

Location	Difference in adjustment size (°C)						Standard deviation of adjustments using single reference years (1 to 5) (°C)	
	Year 1 – (Year 2-5)		Year 2 – (Year 3-6)		Year 2 – (Year 3-5)			
	Max	Min	Max	Min	Max	Min	Max	Min
Kalumburu	-0.75	-0.65	-0.04	-0.10	-0.13	0.06	0.061	0.172
Cunderdin	0.39	0.10	0.30	0.17	0.29	0.31	0.055	0.057
Wandering	0.09	-0.20	-0.20	0.01	-0.13	-0.07	0.079	0.115
Port Lincoln	0.63	-0.10	0.85	-0.50	1.05	-0.43	0.151	0.187
Burketown	-0.73	0.00	0.49	0.36	0.51	0.39	0.077	0.072
Birdsville	-0.23	0.27	-0.35	-0.03	-0.24	0.04	0.054	0.027
Gayndah	0.23	-0.25	-0.07	0.38	-0.05	0.19	0.108	0.118
Miles	-0.36	0.08	0.26	-0.45	0.12	-0.50	0.119	0.037
Thargomindah	-0.35	-0.47	-0.02	-0.51	-0.02	-0.60	0.087	0.171
Pt Macquarie	0.05	-0.10	0.22	0.04	0.19	-0.04	0.084	0.078
Dubbo	-0.23	-0.96	0.34	0.17	0.21	0.08	0.065	0.258
Deniliquin	-0.43	-0.26	-0.31	-0.15	-0.35	-0.09	0.136	0.081
Nhill	0.04	0.26	0.26	-0.45	0.22	-0.39	0.061	0.081
Sale	0.24	-0.03	0.10	-0.16	0.20	-0.12	0.082	0.110
Launceston	-0.13	-0.10	-0.08	0.09	-0.12	0.01	0.170	0.155
Mean (of absolute values)	0.33	0.26	0.26	0.24	0.26	0.22	0.093	0.115

Table 6 Indicators of sensitivity of adjustment size to choice of reference period.

5.5 Comparison of daily and monthly adjustments

To ascertain whether there were any systematic differences between the daily and monthly adjustment procedures, a comparison was carried out between the two. This used stations which met the following criteria:

- No change of station number
- No merges with parallel stations (i.e. all adjustments carried out using reference stations)
- No 'spike' adjustments
- Sufficient reference stations for all adjustments to support the daily method.

A total of 20 stations met these criteria, with a total of 86 adjustments (41 for maximum temperature and 45 for minimum temperature).

For these stations, adjustments were carried out separately using the daily and monthly methods, back to the start of the record. The size of each individual adjustment was compared between the daily and monthly methods. A comparison was also carried out between the adjusted annual means in the first year of the adjusted record.

This comparison showed that there was no evidence of any systematic differences between the daily and monthly methods. The mean adjustment for the daily method was –0.067 °C and for the monthly method –0.069 °C, whilst the mean absolute size of adjustments was 0.465 °C and 0.478 °C respectively. When comparing the annual means in the first year of two adjusted records, for the 16 stations with maximum temperature adjustments, the mean difference (daily – monthly) was 0.01 °C, the mean absolute difference 0.09 °C, with 8 stations showing positive differences and 8 negative; for the 19 stations with minimum temperature adjustments, the mean difference was -0.05 °C, the mean absolute difference 0.12 °C, and 10 stations showed positive differences and 9 negative.

6. MAGNITUDE OF ADJUSTMENTS

In total, there were 966 adjustments applied in version 2 of the ACORN-SAT dataset, 463 for maximum temperature and 503 for minimum temperature. 513 of the adjustments (53%) are supported by metadata in some form (Table 7) and 453 (47%) were only detected by statistical methods. This equates to a mean of 4.1 adjustments per location for maximum temperature and 4.5 adjustments per location for minimum temperature (equating to 4.5 and 4.9 adjustments per 100 years respectively). In all cases, adjustments are applied to adjust historical data to be equivalent to the present.

The number of adjustments is somewhat larger than the 660 applied in version 1, but about half of the increase is accounted for either by the introduction of adjustments for changes of screen size (which accounts for 104 of the 966 adjustments) or by the inclusion of additional data, both recent and historical.

The frequency distribution of adjustments is shown in Figure 10. This shows a two-peaked structure for both maximum and minimum temperature, as would be expected given the general 0.3 °C minimum threshold for the implementation of an adjustment. However, there are more adjustments in the –0.3 °C to +0.3 °C range than in version 1, because of the introduction of an adjustment for screen size which is implemented regardless of magnitude.

Details of the sign and magnitude of the adjustments, depending on the cause, are given in Table 8. Overall, there is little tendency for maximum temperature adjustments towards either positive or negative values, but there is a weak negative tendency in minimum temperature adjustments, with 58%

of minimum adjustments negative and 42% positive. The tendency towards more negative adjustments for minimum temperature is more pronounced for large adjustments; there are 30 negative minimum adjustments of 1 °C or more, but only 13 positive adjustments of that size.

Considering the adjustments by cause, it may be seen that adjustments due to site moves from in-town to out-of-town locations (regardless of the size of the town) show a definite negative tendency, especially for minimum temperatures; the mean minimum temperature adjustment for such a move is –0.65 °C, with 78% of such adjustments being negative. The negative tendency is only weak when the set of all other adjustments (i.e. those not due to a documented move out of town) is considered.

These results indicate that the predominance of negative adjustments for minimum temperature (and, consequently, the stronger warming trend in ACORN-SAT than in the unhomogenised Australian Water Availability Project (AWAP) dataset) is largely driven by the tendency over time for sites to move from in-town to out-of-town locations (and, to a lesser extent, other site moves, some of which involve moves within a town from a poorly exposed to well-exposed location). There is little evidence of any systematic positive or negative tendency in adjustments for breakpoints identified by statistical methods, and hence such breakpoints are unlikely to contribute significantly to the differences between ACORN-SAT and AWAP trends.

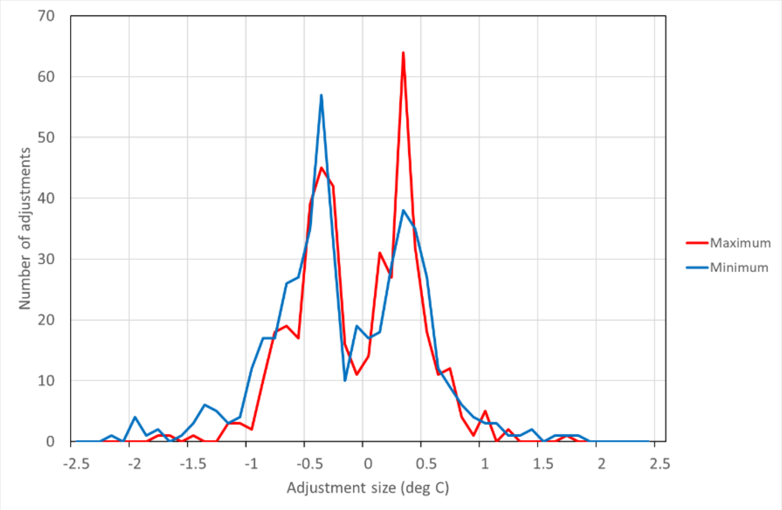

Fig. 10 Frequency distribution of adjustment sizes in ACORN-SAT version 2 (excluding spike adjustments).

Reason for adjustment	Maximum temperature	Minimum temperature
Statistical	209	244
Site moves (all)	176	170
Screen change or condition	79	77
Site condition (without move)	9	5
Observation time change	2	24
Total	450	491

Table 7 ACORN-SAT version 2 adjustments by cause. Totals do not add as some adjustments have multiple causes.

Reason for adjustment	Maximum temperature		Minimum temperature	
	Mean adjustment (°C)	Percentage of positive adjustments	Mean adjustment (°C)	Percentage of positive adjustments
All	-0.03	50	-0.13	42
Statistical	-0.02	50	-0.05	45
All site moves	-0.07	43	-0.30	34
Site moves from in-town to out-of-town	-0.20	37	-0.65	22
All except site moves from in-town to out-of-town	-0.01	50	-0.08	44

Table 8 Mean adjustment size and percentage of adjustments which are positive, by reason for adjustment. Spike adjustments are excluded from this part of the analysis.

7. COMPARISON OF VERSION 1, VERSION 2 AND AWAP

Version 2 of the ACORN-SAT dataset can be compared with version 1, as well as the Bureau's operational whole-network, unhomogenised dataset, AWAP (the Australian Water Availability Project dataset; Jones et al., 2009).

Linear trends of the three datasets are shown in Table 9. All three datasets show substantial warming, particularly after 1960. It had previously been shown (Fawcett et al., 2012) that the version 1 ACORN-SAT dataset warmed more strongly than the unhomogenised AWAP dataset, with the differences primarily occurring prior to 1960. Mean annual temperatures in version 1 warm over the 1910-2016 period at a rate of 0.100 °C/decade, compared with 0.080 °C/decade for AWAP. (These trends are both slightly stronger than reported in Fawcett et al. (2012) because of the inclusion of five additional years of data from 2012 to 2016.) Over the later part of the period the two datasets behave similarly, with mean temperature trends of 0.165 °C/decade and 0.159 °C/decade respectively. Both the differences over the full 1910-2016 period and the similarity for 1960-2016 occur for both maximum and minimum temperatures.

Version 2 shows stronger warming than both version 1 and AWAP, with mean temperature trends of 0.123 °C/decade for 1910-2016. Unlike version 1, it also shows stronger warming than AWAP in the more recent period, with a 1960-2016 trend of 0.200 °C/decade. Whilst minimum temperature warms more quickly than maximum temperature over the full 1910-2016 period, the reverse is true for the more recent 1960-2016 period.

Period	Element	ACORN-SAT version 2	ACORN-SAT version 1	AWAP
1910-2016	Maximum	0.116	0.090	0.068
	Minimum	0.130	0.109	0.091
	Mean	0.123	0.100	0.080
1960-2016	Maximum	0.214	0.174	0.168
	Minimum	0.187	0.157	0.149
	Mean	0.200	0.165	0.159

Table 9 Linear trends (°C/decade) in area-averaged Australian temperatures, using various datasets.

Maps of trends for the 1910-2016 period are shown in Figure 11. As with version 1, version 2 shows warming for both maximum and minimum temperature across virtually all of Australia. The strongest warming trends for mean temperature (mostly between 0.15 and 0.20 °C/decade) occur in central Australia, whilst the weakest trends (between 0.05 and 0.10 °C/decade) are in the northern half of Western Australia (where rainfall has increased substantially since the 1960s), around the southeastern coast from the Eyre Peninsula to far southern New South Wales, and in in northern Tasmania. Maximum temperature trends are strongest in central Australia and weakest in the eastern states (particularly Queensland), whilst minimum temperature trends are strongest in inland Queensland. The only negative trends for any element are for maximum temperature at Moree (potentially influenced by the development of irrigated agriculture in the region, a phenomenon previously documented at Mildura in Trewin (2012)), and for minimum temperature in the east Kimberley. In general, the trends in version 2 are more spatially coherent than those in version 1, especially for minimum temperature.

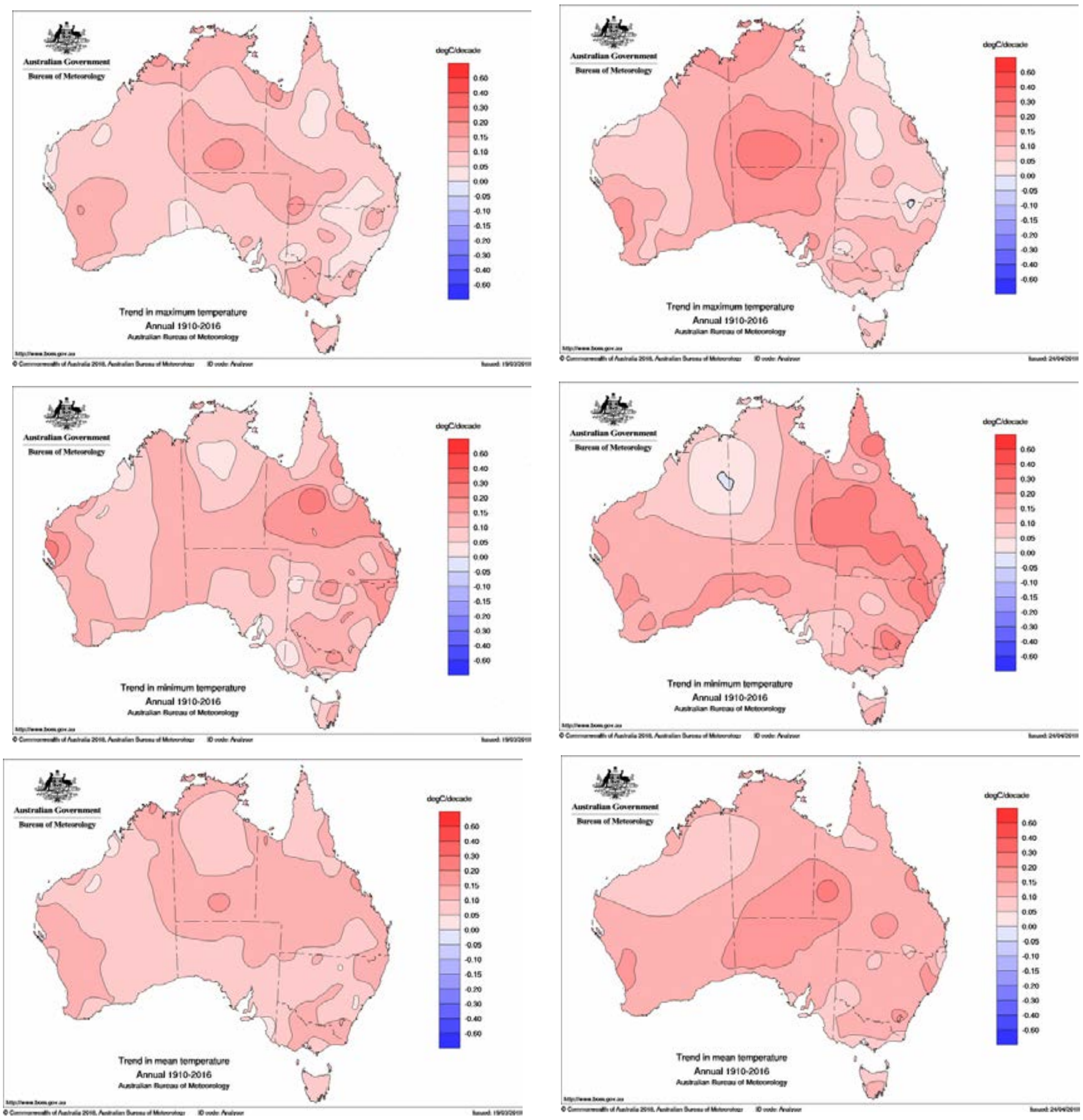

Fig. 11 Temperature trends over Australia in ACORN-SAT version 1 (left) and version 2 (right) for maximum (top), minimum (middle) and mean (bottom) temperature.

The differences over time between version 1 and version 2 are shown in more detail in Figure 12. It may be seen that the relationship between the two versions is relatively stable prior to 1960, but that after 1960, version 2 progressively warms relative to version 1. The maximum temperature difference stabilises after 2000 but minimum temperatures in version 2 continue to warm (relative to version 1) until the most recent period.

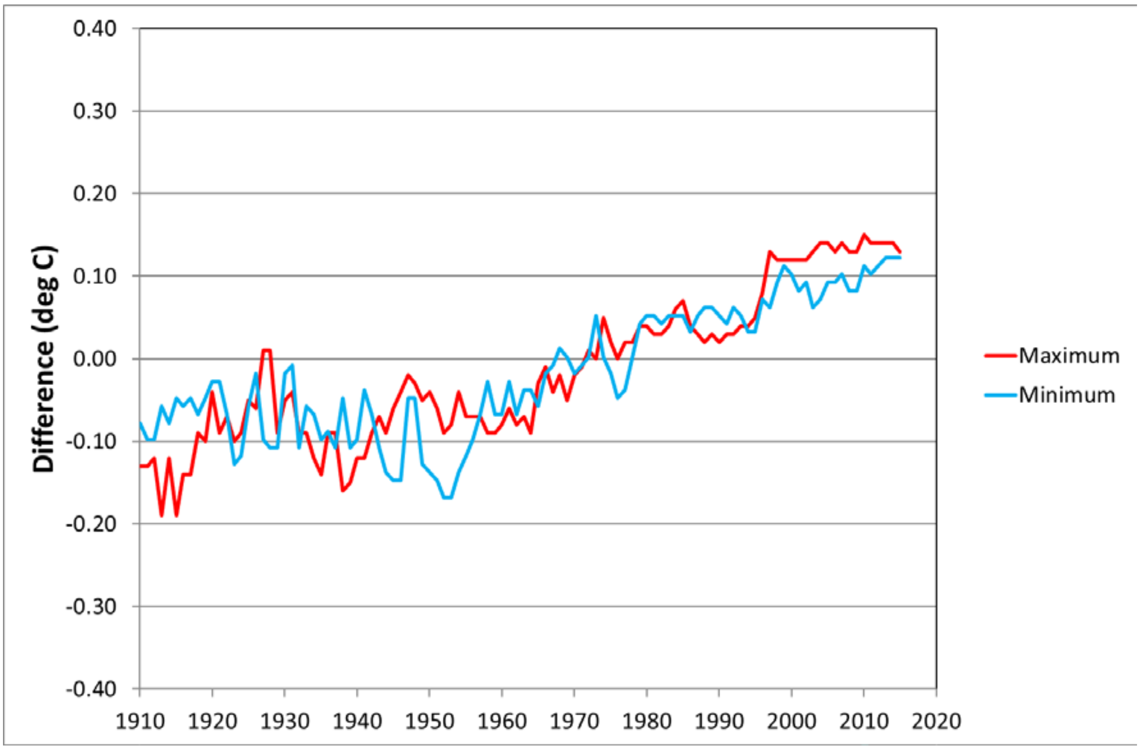

Fig. 12 Difference in mean annual area-averaged Australian temperature anomalies (base period 1961-1990) between version 2 and version 1 of ACORN-SAT (version 2 – version 1).

The stronger warming trends in the homogenised datasets, relative to AWAP, largely reflects the tendency over time for sites to move from in-town to out-of-town locations, something which largely accounts for the negative bias in adjustments, particularly for minimum temperature (section 6). There are a number of reasons why version 2 would be expected to show stronger warming trends than version 1. The most prominent are:

- Removal of the rounding bias in adjustments. As noted in section 3.3.1, the expected impact of this is 0.025 °C per adjustment (for those adjustments which use percentile matching without overlap, about 80% of all adjustments), which, given the mean number of adjustments per location, would equate to 0.08-0.09 °C per station. However, a number of long-term stations in remote areas have a particularly large number of adjustments (for example, Alice Springs has 4 for maximum temperature and 11 for minimum temperature, although only 7 of the latter use percentile matching, with the others using monthly adjustments as outlined in section 3.2.3); as remote stations in data-sparse areas have a greater influence on the gridded analysis than stations in more data-dense areas, it is likely that the expected impact on a national spatial average would be slightly higher.
- More effective treatment of the transition from manual stations (often in towns) to automatic stations (usually at airports or similar locations). This predominantly occurred

from the mid-1990s to the mid-2000s (Figure 13). Most such transitions involved a substantial period of overlapping data, which was used in version 1, but it was found in the course of version 2 that some of these overlap periods were unrepresentative; in these cases, other stations in the region were used as references instead.

- Adjustment of more recent data. Version 1 contains no additional adjustments after 2009, with unadjusted data being appended to the dataset, whilst version 2 includes adjustments up to 2013. This encompasses a renewed period of movement of stations out of towns, particularly in Western Australia and New South Wales; several of these moves involved substantial cooling of minimum temperatures which was not adjusted for in version 1. Separate to this, there have also been recent negative adjustments at some key stations, including maximum temperatures at Halls Creek and Kalgoorlie, and minimum temperatures at Alice Springs.

These outcomes also suggest that the similarity in post-1960 trends in version 1 and the AWAP dataset are the result of two approximately offsetting biases, with the rounding bias in version 1 adjustments offsetting the effect of not fully accounting for site moves associated with automatic weather station introduction.

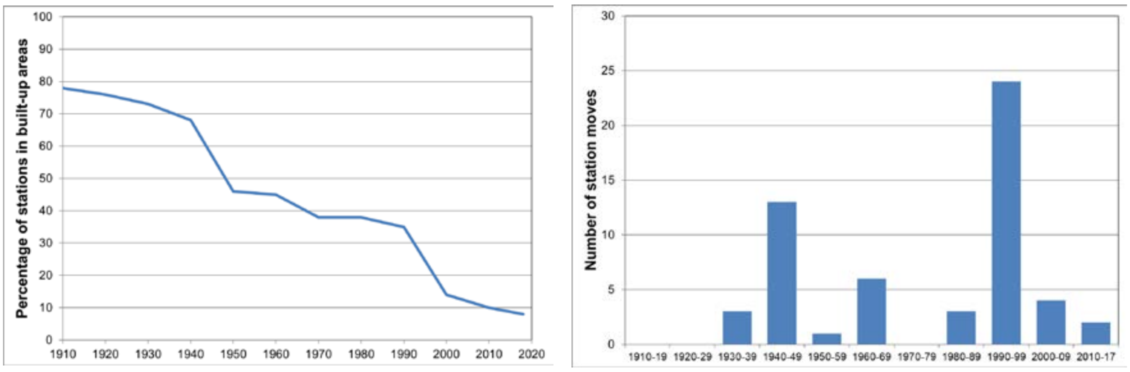

Fig. 13 Percentage of stations within urban areas (of any size) over time (left) and number of site moves from in-town to out-of-town locations by year (right).

8. CASE STUDIES

8.1 Unrepresentativeness of overlap periods

As noted early, on occasions, a period of overlapping observations between two nearby sites which are to be merged into a single time series may not be fully representative – either the old site is no longer representative of its condition before the overlap period began, or the new site is not representative of its ultimate condition post-overlap. In such situations, the relationship between the old and new sites during the overlap period may not be representative of the longer-term relationship between the sites, making the overlap data potentially unsuitable for use in adjustment.

An example of this occurs at Port Lincoln (Figure 14). Observations had been carried out at a number of sites in the central town area, with town observations eventually ceasing in 2002. Meanwhile, an automatic weather station had opened at the airport, approximately 14 kilometres north of the town, in 1992.

The town site became progressively more enclosed in its last few years, with several new trees and bushes growing within 20 metres of the instruments between the 1997 and 2001 inspections. Meanwhile, the initial installation at the airport was an early-generation automatic weather station, which behaved somewhat erratically in its first two years before improved software was deployed.

The effect of this is that the town site progressively warmed relative to the airport in its last few years, whilst the automatic station was too erratic to be suitable for a comparison in its early period. In this case, the 1992-1994 and 1999-2002 parts of the overlap period were excluded and the merge between the two records was carried out using the 1995-1998 period only.

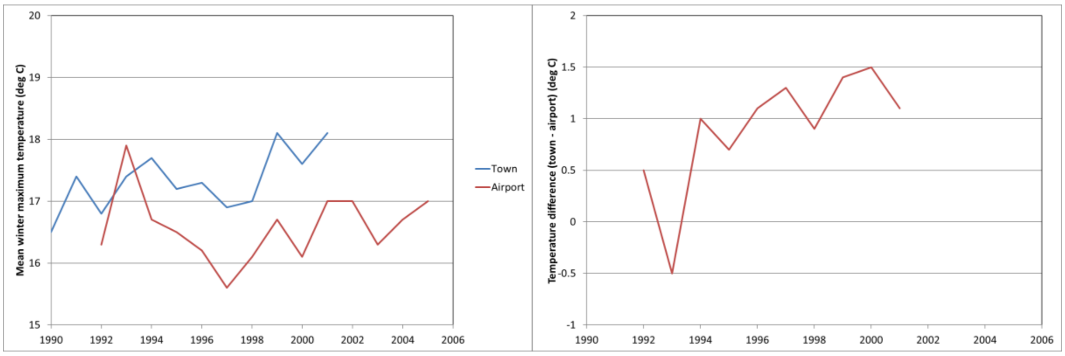

Fig. 14 (left) Mean winter maximum temperatures at the two Port Lincoln sites. (right) Difference in mean winter maximum temperatures between the two sites. This shows the erratic behaviour of the airport site before 1994, and the progressive warming of the town site in the later years of the overlap.

8.2 Variable inter-site relationships during high-rainfall periods in arid regions

In the course of carrying out homogenisation, it became apparent that there were frequently large differences between the behaviour of different reference stations in arid regions, particularly around very wet periods. This led to relatively large numbers of potential reference stations being rejected as potentially inhomogeneous around the time of the breakpoint being investigated.

An illustrative example of this was a potential 2012 breakpoint for maximum temperature at Camooweal, in far northwestern Queensland. 2012 also marked the boundary between a period of very high rainfall (in addition to the 2010-11 La Niña, the 2008-09 wet season was also extremely wet in the region (Bureau of Meteorology, 2010)) and the multi-year drought which affected much of western Queensland from 2012 to 2016. Of 18 potential reference stations which met the correlation criteria, 10 were rejected as potentially inhomogeneous, amongst the highest proportions for any breakpoint considered.

Figure 15 shows the differences in mean annual maximum temperature between Camooweal and seven other sites in the region. The variable inter-site relationships at the annual timescale during the high-rainfall phase are clearly apparent. At Trepell, The Monument and Winton, the sites are much warmer (relative to Camooweal) in 2009 and 2011, following the very wet wet seasons of 2008-09 and 2010-11, than they are in 2010 and particularly in 2008, following a very dry wet season in 2007-08 (2009-10 was near or slightly above average for rainfall). Tennant Creek shows similar behaviour to a slightly lesser extent, although without a strong response in 2011, whilst Cloncurry and Century Mine show more consistent differences with Camooweal. Relationships of temperature between Camooweal and the other sites (and hence with each other) are generally much more consistent during the drought period from 2013 to 2015.

Whilst the causes of this variability have not been rigorously determined, a potential cause is site-specific responses to high rainfall, through changes in soil moisture and/or changes in vegetation. In arid and semi-arid regions of Australia, it is normal for vegetation to remain dormant through prolonged drought periods and then respond rapidly during periods of abnormally high rainfall. It would be expected that the extent (if any) to which this affects any specific site would depend on the local site environment (for example, some sites may become surrounded by green vegetation during a wet period with an associated cooling effect on maximum temperatures, others may remain surrounded by bare ground), and hence that inter-site temperature relationships would potentially become unstable during very wet periods.

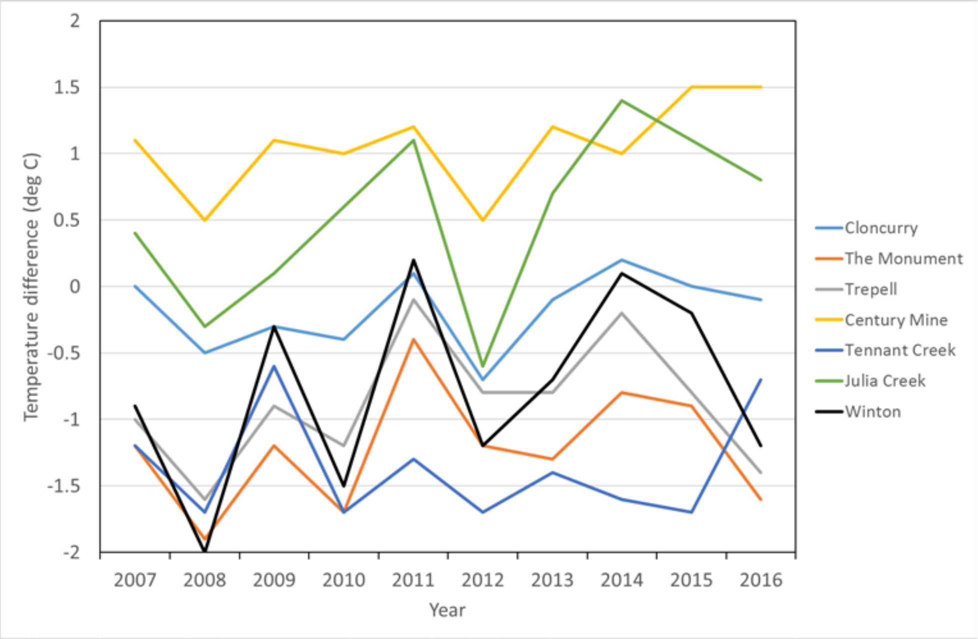

Fig. 15 Mean annual maximum temperature differences (site – Camooweal) between Camooweal and seven other sites in the region from 2007 to 2016. The anomalous behaviour of 2009 and 2011 (following very wet summers) at some sites is evident.

9. CONCLUSION

Version 2 of the ACORN-SAT dataset continues the development of an Australian homogenised temperature record which began with version 1. Compared with version 1, it has increased robustness and greater spatial coherence, especially for minimum temperature, and is expected to support improved analyses of long-term changes in mean and extreme temperatures for Australia.

Evaluations of the uncertainties associated with homogeneity adjustments were generally supportive of the methodological choices used in the original ACORN-SAT dataset, with new adjustments included for factors such as screen size changes, which were found to have no significant impact on mean temperatures but do influence the diurnal temperature range. No evidence was found of any significant systematic impact arising from the change from manual to automatic weather stations (other than that associated with site moves which often accompanied the changes).

Version 2 of ACORN-SAT has 1910-2016 trends in Australian temperature about 0.02 °C/decade larger than those found in version 1, with the differences primarily in the second half of the period (whereas the differences between both ACORN-SAT versions and the unhomogenised AWAP dataset are primarily found prior to 1960). A major contributor to this difference is the removal of a rounding issue which affected version 1, whilst improved accounting for the widespread 1990s/2000s moves of sites out of towns, and the incorporation of recent data from new sites, are also substantial contributors.

10. REFERENCES

Alexandersson, H. 1986. A homogeneity test applied to precipitation data. *J. Climatol.*, 6, 661-675.

Ashcroft, L., Karoly, D. and Gergis, J. 2012. Temperature variations of southeastern Australia, 1860-2011. *Aust. Met. Oceanogr. J.*, 62, 227-245.

Auchmann, R. and Brönnimann, S. 2012. A physics-based correction model for homogenizing sub-daily temperature series. *J. Geophys. Res.*, 117, D17119.

Brandsma, T. and van der Meulen, J.R. 2008. Thermometer screen intercomparison in De Bilt (the Netherlands) – Part II: description and modelling of mean temperature differences and extremes. *Int. J. Climatol.*, 28, 389-400.

Bureau of Meteorology. 2010. *Annual climate summary 2009*, Bureau of Meteorology, Melbourne. Available at http://www.bom.gov.au/climate/annual_sum/2009/AnClimSum09_HR1.1.pdf.

Cao, L., Zhu, Y., Tang, G., Yuan, F. and Yan, Z. 2016. Climatic warming in China according to a homogenized data set from 2419 stations. *Int. J. Climatol.*, 36, 4384-4392.

Domonkos, P., Venema, V., Auer, I., Mestre, O., and Brunetti, M. 2012. The historical pathway towards more accurate homogenisation. *Adv. Sci. Res.*, 8, 45-52.

Fawcett, R.J.B., Trewin, B.C., Braganza, K., Smalley, R.J., Jovanovic, B. and Jones, D.A. 2012. On the sensitivity of Australian temperature trends and variability to analysis methods and observation networks. *CAWCR Technical Report 50*, Centre for Australian Weather and Climate Research, Melbourne.

Jones, D.A., Wang, W. and Fawcett, R. 2009. High-quality spatial climate data-sets for Australia. *Aust. Met. Oceanogr. J.*, 58, 233-248.

Karl, T.R., Williams, C.N., Young, P.J. and Wendland, W.M. 1986. A model to estimate the time of observation bias associated with monthly mean maximum, minimum, and mean temperature for the United States. *J. Clim. Appl. Met.*, 25, 145-160.

Killick, R.E. 2016. Benchmarking the performance of homogenisation algorithms on daily temperature data. Ph.D. thesis, University of Exeter, United Kingdom.

Li, Z., Yan, Z., Cao, L. and Jones, P. 2014. Adjusting inhomogeneous daily temperature variability using wavelet analysis. *Int. J. Climatol.*, 34, 1196-1207.

Menne, M.J. and Williams, C.N. 2009. Homogenization of temperature series via pairwise comparisons. *J. Clim.*, 22, 1700-1717.

Mestre, O. and coauthors. 2013. HOMER: HOMogenisation software in R – methods and applications. *Időjárás*, 117, 47-67.

Milewska, E.J. and Vincent, L.A. 2016. Preserving continuity of long-term daily maximum and minimum temperature observations with automation of reference climate stations using overlapping data and meteorological conditions. *Atmosphere-Ocean*, 54, 32-47.

Nemec, J., Gruber, C., Chimani, B. and Auer, I. 2013. Trends in extreme temperature indices in Austria based on a new homogenised dataset. *Int. J. Climatol.*, 33, 1538-1550.

Nicholls, N., Tapp, R., Burrows, K. and Richards, D. 1996. Historical thermometer exposures in Australia. *Int. J. Climatol.*, 16, 705-710.

Nordli, P.Ø. 1997. Adjustments of Norwegian monthly means of daily minimum temperature. *KLIMA Report 6/97*, Norwegian Meteorological Institute, Oslo.

Perry, M.C., Prior, M.J. and Parker, D.E. 2007. An assessment of the suitability of a plastic thermometer screen for climatic data collection. *Int. J. Climatol.*, 27, 267-276.

Peterson, T.C. and coauthors. 1998. Homogeneity adjustments of in situ atmospheric climate data:a review. *Int. J. Climatol.*, 18, 1493-1517.

Quayle, R.G., Easterling, D.R., Karl, T.R. and Hughes, P.Y. 1991. Effects of recent thermometer changes in the cooperative station network. *Bull. Amer. Met. Soc.*, 72, 1718-1723.

Szentimrey, T. 2008. Development of MASH homogenization procedure for daily data. *Proceedings, 5th seminar for homogenization and quality control in climatological databases, Budapest, Hungary, 2006, WCDMP- No.71*, 123-130.

Trewin, B.C. and Trevitt, A.C.F. 1996. The development of composite temperature records. *Int. J. Climatol.*, 16, 1227-1242.

Trewin, B.C. 2010. Exposure, instrumentation, and observing practice effects on land temperature measurements. *WIRES Climate Change*, 1, 490-506.

Trewin, B.C. 2012. Techniques involved in developing the Australian Climate Observations Reference Network – Surface Air Temperature (ACORN-SAT) dataset. *CAWCR Technical Report 49*, Centre for Australian Weather and Climate Research, Melbourne.

Trewin, B.C. 2013. A daily homogenized temperature data set for Australia. *Int. J. Climatol.*, 33, 1510-1529.

Trewin, B.C. 2018. Short-period fluctuations in temperature at Australian locations. *25th AMOS National Conference and 12th International Conference for Southern Hemisphere Meteorology and Oceanography, Sydney, 5-9 February 2018.*

Venema, V. and coauthors. 2012. Benchmarking homogenization algorithms for monthly data. *Clim. Of the Past*, 8, 89-115.

Vincent, L.A., Milewska, E.J., Hopkinson, R. and Malone, L. 2009. Bias in minimum temperature introduced by a redefinition of the climatological day at the Canadian synoptic stations. *J. Appl. Met. Climatol*., 48, 2160-2168.

Vincent, L.A., Milewska, E.J., Wang, X.L. and Hartwell, M.M. 2018. Uncertainty in homogenized daily temperatures and derived indices of extremes illustrated using parallel observations in Canada. *Int. J. Climatol.*, 38, 692-707.

Wang, X.L., Chen, H., Wu, Y., Feng, Y. and Pu, Q. 2010. New techniques for detection and adjustment of shifts in daily precipitation data series. *J. Appl. Meteor. Climatol.*, 49, 2416-2436.

Warne, J. 1998. A preliminary investigation of temperature screen design and their impacts on temperature measurements. *Instrument Test Report 649*, Bureau of Meteorology, Melbourne.

Willett, K. and coauthors. 2014. A framework for benchmarking of homogenisation algorithm performance on the global scale. *Geosci. Instrum, Method. Data Syst.*, 3, 187-200.

World Meteorological Organisation (WMO). 2017. Challenges in the transition from conventional to automatic meteorological observing networks for long-term climate records. *WMO No. 1202*, World Meteorological Organization, Geneva.

Xu, W., Li, Q., Wang, X.L., Yang, S., Cao, L. and Feng, Y. 2013. Homogenization of Chinese daily surface air temperatures and analysis of trends in the extreme temperature indices. *J. Geophys. Res. Atmos.*, 118, 9708-9720.

APPENDIX 1 – the ACORN-SAT observation network and other networks used in analysis

Site number	Name	Latitude (deg S)	Longitude (deg E)	Elevation (m)	First year of data	Population of urban centre
001019	Kalumburu	14.30	126.65	23	1941	412
002012	Halls Creek	18.23	127.66	422	1910	1546
003003	Broome	17.95	122.23	7	1910	13984
004032	Port Hedland	20.37	118.63	6	1912	13828
004106	Marble Bar	21.18	119.75	182	1910	174
005007	Learmonth	22.24	114.10	5	1975	
005026	Wittenoom	22.24	118.34	463	1951	
006011	Carnarvon	24.89	113.67	4	1910	4426
007045	Meekatharra	26.61	118.54	517	1926	573
008296	Morawa	29.20	116.02	271	1925	507
008297	Dalwallinu	30.28	116.67	325	1955	699
008315	Geraldton	28.80	114.70	30	1910	31982
009021	Perth	31.93	115.98	15	1910	1907833
009518	Cape Leeuwin	34.37	115.14	13	1910	1109*
009617	Bridgetown	33.95	116.13	179	1910	1448
009789	Esperance	33.83	121.89	25	1910	10421
009999	Albany	34.94	117.82	68	1910	29373
010092	Merredin	31.48	118.28	315	1912	2636
010286	Cunderdin	31.62	117.22	217	1950	774
010916	Katanning	33.69	117.61	320	1910	3702
010917	Wandering	32.67	116.67	275	1910	
011003	Eucla	31.68	128.88	93	1913	53
011052	Forrest	30.85	128.11	159	1946	
012038	Kalgoorlie-Boulder	30.78	121.45	365	1910	29875
013017	Giles	25.03	128.30	598	1956	
014015	Darwin	12.42	130.89	30	1910	123574
014825	Victoria River	16.40	131.01	89	*1965*	
015135	Downs	19.64	134.18	376	1910	2991
015590	Tennant Creek	23.80	133.89	546	1910	23726
015666	Alice Springs	20.18	130.01	340	1969	
016001	Rabbit Flat	31.16	136.81	167	1949	146
016098	Woomera	30.71	134.58	123	1921	
017043	Tarcoola	27.56	135.45	117	1940	204
017126	Oodnadatta	29.66	138.07	50	1910	101
018012	Marree	32.13	133.70	15	1939	2157
018044	Ceduna	33.13	135.56	57	1930	549*
018192	Kyancutta	34.60	135.88	9	1910	14064
021133	Port Lincoln	33.77	138.22	109	1910	390
022823	Snowtown	35.75	136.60	158	*1962*	
023090	Cape Borda	34.92	138.62	48	1910	1277431
023373	Adelaide	34.48	139.01	275	*1957*	5691
026021	Nuriootpa	37.75	140.77	63	1910	26148
026026	Mount Gambier	37.16	139.76	3	1910	998
027045	Robe	12.68	141.92	18	1959	3899
027058	Weipa	10.58	142.29	4	*1950*	531
028004	Horn Island	16.00	144.08	204	1910	
029063	Palmerville	17.69	141.07	18	1910	1210
029077	Normanton	17.75	139.54	6	1910	162
030045	Burketown	20.73	143.14	211	1910	517
	Richmond (Qld)					

Site number	Name	Latitude (deg S)	Longitude (deg E)	Elevation (m)	First year of data	Population of urban centre
030124	Georgetown	18.30	143.53	302	1910	301
031011	Cairns	16.87	145.75	2	1910	144730
032040	Townsville	19.25	146.77	5	1940	168729
033119	Mackay	21.12	149.22	30	1910	75710
034084	Charters Towers	20.05	146.27	290	1910	8120
036007	Barcaldine	23.55	145.29	267	*1962*	1287
036031	Longreach	23.44	144.28	192	1910	2738
037010	Camooweal	19.92	138.12	231	1939	159
038003	Boulia	22.91	139.90	162	1910	273
038026	Birdsville	25.90	139.35	47	1954	140
039066	Gayndah	25.62	151.62	111	1910	1712
039083	Rockhampton	23.38	150.48	10	*1939*	61214
039128	Bundaberg	24.91	152.32	31	1910	50148
040004	Amberley	27.63	152.71	24	1941	193733*
040043	Cape Moreton	27.03	153.47	100	1910	243*
040842	Brisbane Airport	27.39	153.12	5	1949	2192720
042112	Miles	26.66	150.18	305	1910	1133
043109	St George	28.05	148.59	199	1913	2395
044021	Charleville	26.41	146.26	302	1910	3077
045025	Thargomindah	27.99	143.81	131	*1957*	270
046012	Wilcannia	31.52	143.39	94	*1957*	549
046126	Tibooburra	29.44	142.06	176	1910	134
048027	Cobar	31.48	145.83	260	1910	3748
048245	Bourke	30.04	145.95	107	1910	1824
050017	West Wyalong	33.94	147.20	257	*1965*	2749
052088	Walgett	30.04	148.12	133	1910	1546
053115	Moree	29.49	149.85	213	1912	7383
055024	Gunnedah	31.03	150.27	307	1948	7984
056242	Inverell	29.78	151.11	582	1910	9547
058012	Yamba	29.43	153.36	27	1910	6043
059151	Coffs Harbour	30.32	153.12	3	1943	48225
060139	Port Macquarie	31.43	152.87	4	1910	44814
061078	Williamtown	32.79	151.84	9	1942	885
061363	Scone	32.03	150.83	221	*1965*	4956
063005	Bathurst	33.43	149.56	713	1910	33587
065070	Dubbo	32.22	148.58	284	1921	34339
066062	Sydney	33.86	151.21	39	1910	4446805
067105	Richmond (NSW)	33.60	150.78	19	1939	4446805*
068072	Nowra	34.95	150.54	109	*1955*	30853
068151	Point Perpendicular	35.09	150.80	85	*1946*	
069018	Moruya Heads	35.91	150.15	17	1910	976
070351	Canberra	35.31	149.20	577	1913	432369
072150	Wagga Wagga	35.16	147.46	212	1910	48263
072161	Cabramurra	35.94	148.38	1482	*1962*	37
074258	Deniliquin	35.56	144.95	94	1910	6833
076031	Mildura	34.24	142.09	50	1910	33444
078015	Nhill	36.31	141.65	139	1910	1749

Site number	Name	Latitude (deg S)	Longitude (deg E)	Elevation (m)	First year of data	Population of urban centre
080023	Kerang	35.72	143.92	78	1910	3633
082039	Rutherglen	36.10	146.51	175	1912	2109
084016	Gabo Island	37.57	149.92	15	1910	
084145	Orbost	37.69	148.47	63	1938	2014
085072	Sale	38.12	147.13	5	1910	13511
085096	Wilsons Promontory	39.13	146.42	95	1910	
086338	Melbourne	37.83	144.98	31	1910	4323072
087031	Laverton	37.86	144.76	20	1945	4323072*
090015	Cape Otway	38.86	143.51	82	1910	
091293	Low Head	41.06	146.79	28	1910	4257*
091311	Launceston	41.55	147.21	167	1910	75329
092045	larapuna (Eddystone Point)	40.99	148.35	20	1910	
094010	Cape Bruny	43.49	147.15	55	1923	
094029	Hobart	42.89	147.33	51	1918	204010
094220	Grove	42.99	147.07	65	1952	1840*
096003	Butlers Gorge	42.28	146.28	667	1944	

Table 10 ACORN-SAT locations. Site number, location and elevation for site in operation in December 2016. Population from 2016 Census (not shown for sites more than 10 km from any urban centre with population above 100; * indicates population is that of a larger urban centre within 20 km of the named town). If first year is shown in italics, further undigitised data are believed to exist.

Location	Sites
Kalumburu	1021 Kalumburu Mission (1941-2005), 1019 Kalumburu (1998-)
Halls Creek	2011 Halls Creek (1910-1949), 2012 Halls Creek (1950-), 2071 Halls Creek Comparison* (1996-2001)
Broome	3002 Broome Post Office (PO) (1910-1941), 3003 Broome Airport (AP) (1942-), 3089 Broome Comparison* (1995-2000)
Port Hedland	4002 Port Hedland PO (1912-1948), 4032 Port Hedland AP (1948-),
Marble Bar	4020 Marble Bar (1910-2003), 4106 Marble Bar (2000-)
Learmonth	5007 Learmonth (1975-)
Wittenoom	5026 Wittenoom (1952-)
Carnarvon	6062 Carnarvon PO (1910-1950), 6011 Carnarvon AP (1945-)
Meekatharra	7046 Meekatharra PO (1926-1950), 7045 Meekatharra AP (1950-)
Dalwallinu	8039 Dalwallinu PO (1957-2012), 8297 Dalwallinu (2007-)
Geraldton	8050 Geraldton Town (1910-1941), 8051 Geraldton AP (1942-2014), 8315 Geraldton AP (2011-)
Morawa	8093 Morawa PO (1925-2005), 8296 Morawa AP (1997-)
Perth	9034 Perth Regional Office (1910-1962), 9021 Perth AP (1944-)
Bridgetown	9510 Bridgetown (1910-2012), 9617 Bridgetown (2008-)
Cape Leeuwin	9518 Cape Leeuwin (1910-)
Albany	9500 Albany (1910-1965), 9741 Albany AP (1965-2014), 9999 Albany AP (2012-)
Esperance	9541 Esperance PO (1910-1969), 9789 Esperance Met Office (MO) (1969-)
Merredin	10093 Merredin Research Stn (1912-1985), 10092 Merredin (1975-)
Cunderdin	10035 Cunderdin (1950-1996), 10286 Cunderdin AP (1997-)
Katanning	10579 Katanning (1910-2012), 10916 Katanning (1999-)
Wandering	10648 Wandering Shire (1910-2003), 10917 Wandering (1998-)
Eucla	11003 Eucla (1913-)
Forrest	11004 Forrest MO (1946-1995), 11052 Forrest (1993-)
Kalgoorlie-Boulder	12039 Kalgoorlie PO (1910-1953), 12038 Kalgoorlie-Boulder AP (1944-)
Giles	13017 Giles (1956-)
Darwin	14016 Darwin PO (1910-1942), 14015 Darwin AP (1941-), 14040 Darwin AP Comparison* (2001-2007)
Victoria River Downs	14825 Victoria River Downs (1965-)
Tennant Creek	15087 Tennant Creek PO (1910-1970), 15135 Tennant Creek MO (1969-)
Alice Springs	15540 Alice Springs PO (1910-1943), 15590 Alice Springs AP (1944-)
Rabbit Flat	15548 Rabbit Flat (1969-1996), 15666 Rabbit Flat (1997-)
Woomera	16001 Woomera (1949-)
Tarcoola	16044 Tarcoola (1921-1999), 16098 Tarcoola (1997-)
Marree	17031 Marree (1939-2011), 17024 Farina (1910-1939), 17126 Marree AP (2006-)
Oodnadatta	17043 Oodnadatta AP (1940-1985, 1994-), 17114 Oodnadatta Police (1985-1991)
Ceduna	18012 Ceduna AP (1939-)
Kyancutta	18044 Kyancutta (1930-)
Port Lincoln	18070 Port Lincoln PO (1910-1998), 18192 Port Lincoln AP (1995-)
Snowtown	21046 Snowtown PO (1910-2001), 21133 Rayville Park (1998-)
Cape Borda	22801 Cape Borda (1962-2007), 22823 Cape Borda (2002-)
Adelaide	23000 Adelaide West Terrace (1910-1979), 23090 Adelaide Kent Town (1977-)
Nuriootpa	23321 Nuriootpa (1957-1999), 23373 Nuriootpa (1996-)
Mount Gambier	26020 Mount Gambier PO (1910-1947), 26021 Mount Gambier AP (1942-)
Robe	26026 Robe (1910-)

Location	Sites
Weipa	27042 Weipa (1959-1994), 27045 Weipa AP (1992-)
Horn Island	27022 Thursday Island MO (1952-1993), 27021 Thursday Island (1992-1996), 27058 Horn Island (1995-)
Palmerville	28004 Palmerville (1910-)
Normanton	29041 Normanton PO (1910-2001), 29063 Normanton AP (2001-)
Burketown	29004 Burketown PO (1910-2001), 29077 Burketown AP (2002-)
Georgetown	30018 Georgetown PO (1910-2004), 30124 Georgetown AP (2004-)
Richmond (Qld)	30045 Richmond PO (1910-)
Cairns	31010 Cairns PO (1910-1947), 31011 Cairns AP (1942-)
Townsville	32040 Townsville AP (1940-), 32178 Townsville Comparison* (1994-2000)
Mackay	33046 Mackay PO (1910-1950), 33047 Te Kowai (1947-1965), 33119 Mackay MO (1959-), 33297 Mackay MO Comparison* (1995-2001)
Charters Towers	34002 Charters Towers PO (1910-1992), 34084 Charters Towers AP (1992-)
Barcaldine	36007 Barcaldine (1962-)
Longreach	36030 Longreach PO (1910-1967), 36031 Longreach MO (1967-), 36167 Longreach MO Comparison* (1996-1997)
Camooweal	37010 Camooweal (1939-)
Boulia	38003 Boulia (1910-)
Birdsville	38002 Birdsville Police (1954-2005), 38026 Birdsville AP (2000-)
Gayndah	39039 Gayndah PO (1910-2009), 39066 Gayndah AP (2003-)
Rockhampton	39083 Rockhampton AP (1939-)
Bundaberg	39015 Bundaberg PO (1910-1990), 39128 Bundaberg AP (1990-)
Amberley	40004 Amberley (1942-)
Cape Moreton	40043 Cape Moreton (1910-)
Brisbane Airport	40223 Brisbane AP (1949-2000), 40842 Brisbane AP (1996-)
Miles	42023 Miles (1910-2005), 42112 Miles (1997-)
St George	43034 St George PO (1913-1997), 43109 St George AP (1997-)
Charleville	44022 Charleville PO (1910-1948), 44021 Charleville AP (1949-), 44221 Charleville AP Comparison* (2003-2006)
Thargomindah	45017 Thargomindah PO (1957-1999), 45025 Thargomindah AP (1999-)
Tibooburra	46037 Tibooburra PO (1910-2012), 46126 Tibooburra AP (2007-)
Wilcannia	46043 Wilcannia (1957-2015), 46012 Wilcannia AP (2011-)
Cobar	48030 Cobar PO (1910-1965), 48027 Cobar MO (1962-), 48244 Cobar MO Comparison# (1997-2000)
Bourke	48013 Bourke PO (1910-1996), 48239 Bourke AP (1994-1998), 48245 Bourke AP (1999-)
Walgett	52026 Walgett PO (1910-1993), 52088 Walgett AP (1993-)
Moree	53027 Moree PO (1957-1964), 53048 Moree MO (1965-1998), 53115 Moree MO (1995-)
Gunnedah	55024 Gunnedah Research Station (1948-)
Inverell	56017 Inverell PO (1910-1997), 56242 Inverell (1995-)
Yamba	58012 Yamba (1910-)
Coffs Harbour	59040 Coffs Harbour (1943-2015), 59151 Coffs Harbour (2013-)
Port Macquarie	60026 Port Macquarie (1910-2003), 60139 Port Macquarie AP (2000-)
Williamtown	61078 Williamtown (1942-)
Scone	61089 Scone Soil Conservation (1965-1996), 61363 Scone AP (1995-)
Bathurst	63005 Bathurst Agricultural Research (1910-), 63305 Bathurst Comparison# (1996-1998)
Dubbo	65012 Dubbo (1921-1997), 65070 Dubbo AP (1995-)
Sydney	66062 Sydney (1910-)
Richmond (NSW)	67033 Richmond (1939-1994), 67105 Richmond (1993-)
Nowra	68076 Nowra (1955-2000), 68072 Nowra (2000-)
Point Perpendicular	68034 Point Perpendicular (1946-2004), 68151 Point Perpendicular (2001-)

Location	Sites
Moruya Heads	69018 Moruya Heads (1910-)
Canberra	70014 Canberra AP (1939-2010), 70351 Canberra AP (2010-), 70338 Canberra AP Comparison* (1995-1997)
Wagga Wagga	72151 Kooringal (1910-1946), 72150 Wagga Wagga AP (1943-)
West Wyalong	73054 Wyalong (1965-2007), 50017 West Wyalong AP (2005-)
Deniliquin	74128 Deniliquin (1910-2003), 74258 Deniliquin AP (1997-)
Mildura	76077 Mildura PO (1910-1949), 76031 Mildura AP (1946-)
Nhill	78031 Nhill (1910-2008), 78015 Nhill AP (2003-)
Kerang	80023 Kerang (1910-)
Rutherglen	82039 Rutherglen (1912-)
Gabo Island	84016 Gabo Island (1910-)
Orbost	84030 Orbost (1938-2011), 84145 Orbost (2006-)
Sale	85072 East Sale (1945-), 85298 East Sale Comparison* (1996-2005), 85133 Sale (1910-1945)
Wilsons Promontory	85096 Wilsons Promontory (1910-)
Melbourne	86071 Melbourne (1910-2015), 86338 Melbourne (2013-)
Laverton	87031 Laverton (1945-)
Cape Otway	90015 Cape Otway (1910-)
Low Head	91057 Low Head (1910-2001), 91293 Low Head (1998-)
Launceston	91049 Launceston Pumping Station (1910-1942), 91104 Launceston AP (1939-2006), 91311 Launceston AP (2004-)
larapuna (Eddystone Point	92045 larapuna (Eddystone Point) (1910-)
Cape Bruny	94010 Cape Bruny (1923-)
Hobart	94029 Hobart (1918-)
Grove	94069 Grove (1952-2010), 94220 Grove (2004-)
Butlers Gorge	96003 Butlers Gorge (1944-1993, 2008-), 96071 Lake St Clair (1990-2013)

Table 11 Sites used for ACORN-SAT locations. (* - old site switched to comparison number while number switched to new site; # - comparison site switched to old number after end of comparison)

Location	Site numbers (old/new)	Period of parallel observations	Length of parallel observations (yrs)	Mean temperature difference (new – old, °C) and classification	
				Maximum	Minimum
Kalumburu	1021/1019	1998-2005	6.5	-0.05 (small)	-1.14 (large)
Cunderdin	10035/10286	1996-2007	10.5	0.09 (small)	-0.55 (medium)
Wandering	10648/10917	1998-2003	4.2	-0.34 (medium)	-0.65 (large)
Port Lincoln	18070/18192	1992-2002	9.9	-0.76 (large)	-1.18 (large)
Burketown	29004/29077	2001-2009	7.7	0.91 (large)	-0.34 (medium)
Birdsville	38002/38026	2000-2005	4.7	-0.23 (medium)	0.20 (small)
Gayndah	39039/39066	2003-2009	6.4	-0.28 (small)	-0.25 (small)
Brisbane AP	40223/40842	1994-2000	5.8	-0.23 (small)	0.23 (medium)
Miles	42023/42112	1997-2005	7.6	-0.34 (medium)	-0.05 (small)
Thargomindah	45017/45025	1999-2005	5.7	-0.51 (medium)	0.74 (large)
Port Macquarie	60026/60139	1995-2003	7.6	0.62 (large)	-1.43 (large)
Dubbo	65012/65070	1993-1999	6.9	-0.41 (medium)	-0.16 (small)
Deniliquin	74128/74258	1997-2003	6.1	-0.14 (small)	-0.49 (medium)
Nhill	78031/78015	2003-2008	5.5	0.34 (medium)	0.92 (large)
Sale	85298/85072	1996-2008	9.2	-0.16 (small)	-0.71 (large)
Launceston	91104/91311	2004-2009	4.9	0.69 (large)	-0.10 (small)

Table 12 Station pairs with parallel observations used for technique and uncertainty evaluation.

APPENDIX 2. Summary of changes between version 1 and version 2 of ACORN-SAT

Where the change is discussed in detail in the text, the relevant section(s) is/are shown.

Change	**Details**	**Impact**
1.Recoding of homogenisation software into Python	Both the inhomogeneity detection and homogenitiety correction code has been re-coded in Python. This has allowed an independent validation of code and associated techniques.	No impact on dataset in itself (though noting items 5 and 9 below). Improved user confidence in the accuracy of the data and the methods.
2. Change in output data formats	The ACORN-SAT station time series are now provided as CSV files on the recommendation of the Technical Advisory Forum (TAF).	Increased accessibility of the dataset and ease for use in Excel and other tools. No impact on the dataset.
3. Provision of effective station weights	Monthly effective stations weights are now provided which given an estimate of the relative importance of each station ("footprint") in the all Australia temperature. Multiplication of the station weight by the station value provides an estimate of the all Australia temperature for days and months.	Easier interpretation of areal averages by users. No impact on the dataset.
4. Inclusion of new temperature data and metadata (sections 2.1, 2.2)	Version 2 includes updated data from all stations to near the time of release (i.e., to the end of 2016). It will also include newly digitised historical data from Canberra (1913-1939), Sale (1910-1945) and Moree (1910-1956), smaller quantities (1-2 years) from Tarcoola and Dalwallinu, and newly adjusted data at locations where old manual sites have closed since 2012 and have been replaced by Automatic Weather Stations. Some additional metadata have also been found during the process, in particular through a secondary search of station history files for potential explanations for	Not expected to impact on previously released data. Will lead to a slight improvement in the portrayal of historical climate variations in eastern Australia, and will bring ACORN-SAT data up to date. The additional metadata has allowed the date of changes to be more precisely specified at a number of locations. As moves of stations from in-town to out-of-town locations most often leads to a negative inhomogeneity in temperatures, especially for minima, accounting more accurately for recent site moves leads to an increase in trends in the updated dataset.

	large inhomogeneities found by statistical methods.	

5. Change to a symmetrical homogenisation window for non-overlapping station cases	Recoding into PYTHON of the ACORN-SAT code (for public release) identified a minor error in the homogenisation window for the non-overlap station case. Ideally, the overlap window for the calculation of differences used in homogenisation should be symmetrical, and this was meant to be set at a maximum of 5 years for all stations. The identified error set the first overlap period (before breakpoint) to 5 years, but allowed the second overlap period (after breakpoint) to be as long as 7 years. The recoding fixes the maximum length of overlap to be 5 years for both the first and second overlap periods. All else remains the same.	This was assessed as introducing no systematic bias, but has been corrected. The new data may differ by as much as 0.1 °C at individual stations for particular months.
6. Introduction of a floor of 0 °C for the homogenised data diurnal temperature range (section 3.3.3)	The independent adjustment of the maximum and minimum temperature data allows negative (non-physical) diurnal temperature ranges to occur on a very occasional basis (about 0.03% of cases). This is entirely an artefact of the adjustment process, and is a known limitation of independent data adjustment. All negative occurrences of diurnal temperature range have been removed with the maximum temperature adjusted upwards and the minimum temperature downwards by an equal amount to achieve a 0 °C diurnal temperature range.	The impact on mean temperature and associated trends is 0.0 °C by design. The diurnal temperature range is very slightly increased, with a mean change anticipated to be less than 0.01 °C across the network.

7. Calculation of normal at short-period stations	The ACORN-SAT code includes a process for calculating normal values at each month for candidate and reference stations, and for adjusting them to a 1961-1990 equivalent at stations with insufficient data within the 1961-1990 window. These normal values are then used to calculate anomaly values which are compared between stations (the reason for using anomalies rather than raw temperatures in this part of the analysis is so that the transition seasons are handled appropriately, and so, for example, in an autumn comparison, a situation is avoided whereby all the warm days in the distribution are from March, and all the cold days from May). An error was identified and corrected in the process of correcting minimum temperatures at short-period stations to a 1961-1990 equivalent. This particularly affects cases where parallel (comparison) observations were carried out but the primary station number remained unchanged (e.g. at Townsville, where parallel observations took place between 1994 and 2000, the primary site number (32040) switched over from the old to new site at the start of the comparison, with the old site continuing under the new site number 32178).	In general, no measurable impact is expected on mean temperatures as the incorrectly-adjusted normal values are subtracted from, and then added back to, the raw data. There will be improvements to the identification of, and appropriate adjustment of, extreme days during the transition seasons, but impacts on the overall dataset are expected to be marginal. The issue which affected the parallel observations case had no significant impact at the annual mean level, but did lead to some distortions between months.

8. Use of multiple detection methods in parallel (section 3.1)	In version 1, a single method (pairwise difference, using monthly and seasonal data simultaneously) was used for the statistical detection of potential inhomogeneities. In version 2, five methods were run in parallel – the pairwise difference algorithm (PDA) using monthly data, PDA using annual data, and three methods developed by other international groups (HOMER, MASH and RHTestsV4). Breakpoints flagged by the monthly PDA method, or by any two of the other four methods (within a 2-year window) were identified for further investigation and, if significant, adjustment. (As in version 1, metadata-identified potential breakpoints took priority if within 2 years of a statistical breakpoint). Testing found that a breakpoint identified by 1 method had a ~50% probability of being significant, rising to near 100% if identified by 4 or 5 methods.	It is expected that this process will add to the robustness of the dataset and reduce the probability of undetected inhomogenities. The new process would not be expected to introduce any net bias into adjustments or long-term trends.

9. Identification of rounding bug in version 1 code and move to full precision (section 3.3.1)	During the process of recoding the version 1 adjustment code in Python, a rounding bug was found. The version 1 code used the Fortran *anint* function, but it was later found that this function, as implemented, automatically rounds values ending in 0.5 down to the nearest integer. This affects numerous adjustment transfer functions, as the consolidated transfer function value for any given candidate station value is the median of estimates derived from each reference station, and as the standard number of reference stations is 10 (an even number), it is relatively common for the median value of 10 numbers, each with 0.1 precision, to be a number ending in .x5 (which would then be rounded down to the nearest 0.1). This issue affects all adjustments made using daily data and reference stations (the majority). It does not affect adjustments made by merging two stations with overlapping data, or those made using monthly data. In version 2, data are retained at full precision throughout the process, removing any need for rounding during internal calculations.	The expected impact of the rounding bug in version 1 is −0.025°C per adjustment. Given the total number of adjustments, the overall result of this is assessed to be a negative bias in the 1910-2016 trend in the version 1 data of approximately −0.09°C. The removal of this bug is therefore expected to increase post-1910 trends in maximum, minimum and mean temperature by approximately 0.09°C.

10. Improved method for monthly adjustments (section 3.2.3)	In a relatively small number of cases, adjustments are carried out using monthly rather than daily data. This occurs in the following situations: • Where there are insufficient reference stations (fewer than 3), correlated at better than 0.6 with the candidate station, with daily data in the 3-year periods before and after the breakpoint. Except at the most remote locations, this occurs only prior to 1965. (In many cases there will be additional reference stations with monthly digitised data only). • Where two breakpoints are separated by 3 years or less (“spikes”), as daily data would not be expected to give sufficiently robust results over such a short window. A (rarely invoked) provision in version 1, under which monthly adjustments were used if the 10th best correlated station with monthly data was better correlated with the candidate than the 3rd best correlated station with daily data, was not used in version 2, as it was complex to implement and offered only very limited benefits. In version 1, the monthly adjustment procedure involved a distance-weighted	This change would not be expected to produce any systematic bias in one direction or the other. However, as the adjustments affected are disproportionately in data-sparse areas in the earlier years of the data set, changes in individual affected adjustments (in either direction) may have a noticeable impact on state or national means due to the large effective weighting of the stations concerned.

	mean of differences with reference stations before and after the breakpoint. It was found that in data-sparse areas, this algorithm was dominated by the nearest reference station, and was hence vulnerable to distortions from possible undetected issues at that reference station. In version 2, this was replaced by a correlation-weighted median, which is expected to give more robust outcomes.	
11. More useful diagnostics for identifying and eliminating inhomogeneous reference stations (including comparison stations) and unrepresentative reference periods (section 3.2.2)	A routine set of diagnostics has been introduced into the adjustment code, indicating: (a) For each reference station, the size of the adjustment which would be carried out if that station were the only reference station. (b) A comparison of temperatures at the candidate station and reference stations in each individual year for the 5 years before and after the breakpoint, or, in the case of parallel observations, a year-by-year assessment of the relationship between the two sites. Step (a) is intended to identify potentially inhomogeneous reference stations. If the estimated adjustment attributable to an individual reference station differs from that derived from the median of all reference stations by	It is expected that these procedures will lead to a more robust dataset at the station level, although they should not have any systematic impact in themselves on trends. However, in practice, the replacement of unrepresentative parallel observation periods at a number of sites with the use of other reference stations has led to the cooling impact of the 1990s/2000s transition to automatic weather stations being more faithfully reflected in the data, resulting in a slight increase in warming trends once corrected for.

	more than 0.3°C[3], the reference station is considered to be inhomogeneous, and the adjustment process is repeated with that station replaced by the next best-correlated station. This process is repeated until either a set of 10 reference stations has been found meeting the homogeneity criteria, or there are no further potential reference stations with a sufficiently high correlation. Step (b) is intended to identify cases where the period immediately before or after a breakpoint is unrepresentative. Where this occurs, the adjustment is repeated but using a different reference period (e.g. if the breakpoint is in 1949 but 1947 and 1948 are unrepresentative, 1942-46 is used as the reference period instead of 1944-48). It is also intended to identify cases where a period of parallel observations is unrepresentative (e.g. where the old site becomes more built-up or overgrown during the parallel observations period) and the use of other stations as a reference is more appropriate.	

---

[3] A threshold of 0.4°C is used in second or subsequent rounds, or if 50% or more of reference stations show differences of more than 0.3°C.

12. More rigorous second-round homogenisation procedure (section 3.2.5)	A second-round homogenisation process is carried out in order to identify cases where an inhomogeneity has been missed in the first round, or where an adjustment has been incorrectly applied. Whilst the most common causes of such cases in version 1 (inhomogeneous reference stations and unrepresentative reference periods) were eliminated in the first round in version 2 by use of the methods described above, some anomalies remained which required the use of a second-round process. In version 1, the principal tool for the second-round homogenisation was the comparison of trends for the 1910-2011 period, and sub-periods therein, at each station with other ACORN-SAT stations in the region, and the further investigation of stations with anomalously large or small trends. This process was still carried out in version 2, but in addition to this, the (first-round) homogenised data from each station was passed through the RHTestsV4 detection procedure, with four nearby ACORN-SAT homogenised series used as reference series. Any inhomogenities found in the “homogenised” data by this process were then referred for further investigation. 111 possible breakpoints were found by the second-round procedure, and whilst 47 were found not to warrant further action (mostly associated with local short-term climate anomalies), in 64	No net impact is expected on national mean trends, but initial analysis indicates greater spatial coherence of outcomes in version 2 compared with version 1.

	cases changes were made to the first round homogenisation. This most frequently involved changes to reference periods or reference stations, but in 22 cases new breakpoints missed in the first round were added, and in 6 cases breakpoints provisionally identified in the first round were removed.	
13. Handling of large to small screen transition (section 3.3.4)	A substantial number of sites transitioned from large to small screens, the majority of them during the 1990s. Whilst field trials in the late 1990s, reinforced by observed data, indicate that this change produced no significant inhomogeneity in mean temperature, the opposing impacts on maximum and minimum temperature (typically 0.05-0.10°C in each) result in an impact on diurnal temperature range which is considered large enough to require correction, given that a substantial proportion of the network is affected. An explicit adjustment will therefore be carried out for those sites where there is not already an identified breakpoint adjusted for in the data (in some cases the transition from a large to small screen took place at the time of a site move, and hence the effect of the screen change is already incorporated in the overall adjustment for the site move).	No net impact is expected on national mean trends, but an impact is expected on other variables; at a national level, in the order of −0.05°C in maximum temperature, +0.05°C in minimum temperature, and −0.10°C in diurnal temperature range.

14. Other minor issues	An error was identified in the percentile-matching code in dealing with the case where the 4th and 5th percentiles were identical at one station in a pair but not at the other. Improved methods of detecting observations which have been date-shifted were implemented (see section 3.3.2). As discussed in detail in Trewin (2012), a very small number of station records were classified as unsuitable for homogenisation of extreme temperatures, as the relationship between temperatures at the coastal and inland sites forming the composite is unstable at very high or very low temperatures. There are only four such locations affected within ACORN-SAT: Albany and Port Macquarie for extreme high maximum temperatures, Eucla for extreme high minimum temperatures, and Horn Island for extreme low minimum temperatures. In these examples the adjustment method used produces a realistic time series for mean temperatures but a highly unrealistic one for extremes. These locations are excluded from downstream analyses that involve extremes, but are still considered reliable for analyses involving means, as only a very small proportion of values (less than 1%) are considered unreliable and they have little impact on means. The revised dataset will look at a way to flag these unreliable values, although an	The percentile-matching code error will have a small impact (less than 1°C) on a very small number of individual daily values. Correcting date shifts will have negligible impact at the monthly and annual level but will improve the analysis on individual days.

	initial assessment suggests that some of the most extreme cases identified in version 1 are less extreme in version 2.	

APPENDIX 3 – further details on implementation of detection methods

A3.1. HOMER version 2.6, joint detection

All stations with monthly data covering the span of the candidate station, and correlated with the candidate station at $r > 0.6$, were used as reference stations.

As the software used for the implementation of HOMER requires serially complete data, missing monthly values at candidate and reference stations were infilled by interpolation from other stations using inverse-distance ($1/d$) weighting.

A3.2. MASH version 3.03

The best-correlated neighbouring stations (up to a maximum of 10) were used as reference stations.

A3.3. RHTests version 4

Reference series were constructed using monthly anomalies, using a weighted mean from all stations correlated with the candidate at $r > 0.6$, weighted by r^2.