An updated long-term homogenized daily temperature data set for Australia

Blair Trewin | Karl Braganza | Robert Fawcett | Simon Grainger | Branislava Jovanovic | David Jones | David Martin | Robert Smalley | Vanessa Webb

Bureau of Meteorology, Melbourne, Vic., Australia

Correspondence
Blair Trewin, Australian Bureau of Meteorology, GPO Box 1289K, Melbourne 3001, Vic., Australia.
Email: blair.trewin@bom.gov.au

Funding information
This work was funded by the Australian Bureau of Meteorology.

Abstract
A new version of the long-term Australian temperature data set, known as ACORN-SAT (Australian Climate Observations Reference Network—Surface Air Temperature), has been developed. ACORN-SAT includes homogenized daily maximum and minimum temperature data from 112 locations across Australia, encompassing the period from 1910 to the present, with 60 of the locations having data for the full 1910–2018 period. Homogenization is achieved using a percentile-matching methodology with a number of improvements beyond practices used in previous versions, including more effective detection and removal of potentially inhomogeneous reference stations and an enhanced breakpoint detection methodology. Explicit corrections have also been introduced for a change in instrument screen size, whilst an assessment has found that the transition from manual to automatic instruments and changes in effective response time of automatic instruments have had a negligible impact on the data. Adjustments associated with documented site moves from in-town to out-of-town locations are predominantly negative, particularly for minimum temperature, with other adjustments showing no strong bias towards either positive or negative values. The new data set shows slightly stronger warming (0.12°C per decade in mean temperature over the 1910–2016 period) than either the previous ACORN-SAT version (0.10°C) or the un-homogenized gridded data (0.08°C), primarily due to more effective treatment of systematic moves of sites out of towns and the removal of a rounding bias in the version 1 methodology.

Dataset
Identifier: https://doi: 10.25941/5d28a5d352de7
Creator: Blair Trewin, Karl Braganza, Robert Fawcett, Simon Grainger, Branislava Jovanovic, David Jones, David Martin, Robert Smalley, Vanessa Webb
Title: Australian Climate Observations Reference Network - Surface Air Temperature (ACORN-SAT)
Publisher: Australian Bureau of Meteorology
Publication date: 2019
Resource type: Dataset and metadata
Version: 2

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. Geoscience Data Journal published by Royal Meteorological Society and John Wiley & Sons Ltd.
1 | INTRODUCTION

Reliable and homogeneous long-term data sets are required to assess observed climate change at the global and regional levels (Aguilar et al., 2003). A homogeneous data set is one where changes in the data reflect changes in the long-term background climate, as opposed to changes in the way the observations have been made or the conditions under which they have been made. The Australian Climate Observations—Surface Air Temperature (ACORN-SAT) data set is designed to allow assessment of long-term climate change in Australia, both at the national and sub-national scales. Its purpose is to allow the assessment both of changes in mean temperatures and in changes in the behaviour of temperature extremes, which are associated with many high-impact climate events in Australia. It seeks the maximum national coverage possible over the last century whilst optimizing long-term stability of the network used, noting the limited availability of data in some of the more sparsely populated parts of Australia.

Numerous such data sets have been developed over time at global (e.g. Morice et al., 2012; Rohde et al., 2013; Menne et al., 2018; Lenssen et al., 2019), continental and national levels. Some of these, including all the global data sets cited above, take the approach of assessing and adjusting for identified inhomogeneities. Others (e.g. ECA&D; Klein Tank et al., 2002) identify homogeneous and inhomogeneous station time series and leave it to data set users to decide how they use the data.

Inhomogeneities in a climate data set may arise for a number of reasons. For temperature, the most common causes of inhomogeneities (Trewin, 2010) include site moves, changes in instruments (especially changes in instrument shelters, such as the introduction of the Stevenson screen as a standard in the late 19th and early 20th century), changes in observation procedures such as observation times and changes in the local site environment (such as the erection of buildings near a site or substantial changes in ground cover or nearby vegetation). Whilst site-specific inhomogeneities have only a modest impact on assessments of temperature change on land at the global level (Jones, 2016), they can be more important over smaller areas. This is both because some changes can systematically affect a large part of a national network in a similar way, creating a signal over an area too small to have a significant impact at the global scale, and because issues specific to a single station have a larger impact when a relatively small area (such as a country or state) is under consideration. In turn, this can affect assessments of regional-scale climate change, which are often of high interest for local climate change assessment and adaptation.

The original Australian homogenized temperature data set was developed, using annual data, by Torok and Nicholls (1996). This formed the basis for monitoring of Australian and state/territory mean annual temperatures from the late 1990s onwards. This was succeeded by the first version of the ACORN-SAT data set (Trewin, 2012, 2013), which was released in 2012. This data set was at the daily timescale, and hence formed the basis for assessment of change and variability in temperature extremes, as well as mean climatic conditions. A separate, unhomogenized data set, the gridded Australian Water Availability Project (AWAP) data set (Jones et al., 2009), designed primarily for spatial analysis and monitoring of climate variability, was released in 2009 and provides a ‘raw’ temperature comparison data set.

Most homogenized data sets in use for operational climate monitoring apply adjustments based on monthly, seasonal or annual data. Trewin and Trevitt (1996) noted that some inhomogeneities can affect different parts of the temperature frequency distribution in different ways. In particular, they found that local-scale influences on minimum temperatures tend to be larger on cold nights than on warm ones, as calm, clear conditions conducive to radiational cooling also maximize the effect of factors such as local topography and the built environment on temperature.

Hence, adjustments which remove inhomogeneities at the monthly or annual timescale may still leave residual inhomogeneities in extremes. Methods which apply differential adjustments to different parts of the monthly or seasonal frequency distribution of daily temperature (hereafter referred to as ‘daily homogenization’)$^1$ can be applied to address this issue. Evaluations carried out in the development of ACORN-SAT version 1 (v1) (Trewin, 2012) found that, whilst daily homogenization produced only modest improvements over monthly methods in metrics across the full data set, such as root-mean-square (RMS) error, it provided a substantial improvement in some indices of extremes, especially for extreme low minimum temperatures. Comparable results have been found in other data sets (e.g. Vincent et al., 2018). These improvements are important, as many of the impacts of climate variability and change occur through extremes (Melillo et al., 2014).

There have been a number of benchmarking studies applying different methods to synthetic data at the monthly scale

---

$^1$Two other methods to which the term ‘daily homogenisation’ is sometimes applied are excluded by this definition: methods where the adjustment for a date is derived by interpolation between adjustments derived from monthly means for the preceding and following months, and methods which use the frequency distribution for a full year, and hence largely reflect seasonal variations in adjustments rather than variations within a month/season.
(e.g. Venema et al., 2012), but the only known benchmarking study of this type for daily data was carried out by Killick (2016). In contrast to the above, this study found limited benefits in the use of daily homogenization methods, although it used a narrower range of evaluation metrics, particularly for extreme values. The results of benchmarking studies using synthetic data will also be sensitive to the nature of the synthetic data—for example, a synthetic breakpoint characterized as a change in the mean and variance of a Gaussian function may not capture the full range of real-world inhomogeneities.

A new version of the ACORN-SAT data set (hereafter known as version 2, or ACORN-SATv2) has been developed. Many major climate data sets have regular major update cycles every few years, incorporating new data and methodological developments since the previous version. ACORN-SATv2 includes a number of methodological improvements from version 1, some of which arise from an external review carried out between 2015 and 2017 (Commonwealth of Australia, 2015, 2017), as well as additional recent and historical data. Some issues which have arisen since the release of version 1 are also assessed. The purpose of this paper was to describe the ACORN-SATv2 data set, and how the methods and outcomes differ from those in v1.

## 2 DATA AND METADATA

There are currently about 700 locations in Australia where temperature is measured, but most have a relatively short period of record. The ACORN-SAT data set, which is designed to assess long-term climate, consists of daily maximum and minimum temperature data\(^2\) from 112 locations\(^3\) distributed across Australia (Figure 1). The first year of the data set is 1910. This is near the start of the period for which there is reasonable national coverage (in particular, very limited data exist in Western Australia prior to 1907). It also marks the start of consistent national instrument exposures, as the Stevenson screen was not widely introduced in New South Wales and Victoria until the Bureau of Meteorology was established as a federal body in 1908 (Nicholls et al., 1996). Jones and Trewin (2002) have previously analysed the adequacy of daily temperature data and shown that a network in the range of 100–200 stations is largely sufficient for monitoring regional temperature change. A 112-station network incorporates the bulk of Australian locations which have data of acceptable quality and homogeneity for a substantial period of time.

60 of the 112 locations have data for the full period from 1910 to the present, and 110 of the 112 locations have at least 50 years of data (the two exceptions, Rabbit Flat and Learmonth, are both in remote areas which would not otherwise have coverage). The only ACORN-SAT location which had non-Stevenson screen data after 1910 (Eucla) is not used prior to the Stevenson screen installation in February 1913.

The locations are distributed across Australia, although coverage was limited in much of central and northern Australia before the 1950s. There are still some significant data voids, such as the western interior and the north-east of the Northern Territory, where temperature observations are either non-existent or only commenced in recent years. A full listing of the locations and years of record at each location is in Table S1. The same 112 locations have been used in v1 and v2.

The ACORN-SAT data set documented in this paper ends in 2016, but is updated operationally using real-time data, with annual reassessments of data quality and homogeneity.

Historically, most daily temperature data in Australia had only been digitized from 1957 onwards. This has progressively been addressed over the last 20 years, with the

---

\(^2\)Mean temperature in Australia is defined as \((\text{maximum temperature} + \text{minimum temperature})/2\) and is hence derived from the maximum and minimum temperature data.

\(^3\)In this context, a ‘location’ may include data from two or more sites in close proximity which have operated over different periods.
majority of ACORN-SAT locations now having fully digitized daily data for their period of record. Compared with v1 of ACORN-SAT, there are five locations with additional digitized historical data in v2.

There has been a transition from manually read, liquid-in-glass maximum and minimum thermometers to automatic weather stations with electronic temperature probes over the last 25 years. 98 of the 112 ACORN-SAT locations now use automated instruments, with a further three in parallel observations programmes between manual and automatic instruments as of mid-2019. The question of potential systematic differences between temperatures measured by manual and automatic instruments is addressed further in section 3.5.

Quality control is an important part of any climate data set (WMO, 2018). Recent operational data collected by the Bureau of Meteorology are subject to routine quality control procedures, including internal consistency and spatial checks. An independent quality control process (Trewin, 2012, 2013) has been carried out to apply—as far as was possible with the available data—a consistent quality control procedure through time, with a comparable level of quality control to historical data to that which is carried out operationally now.

Station metadata are currently maintained in an electronic Bureau database (SitesDB), which includes information such as site locations and photographs, instruments in use and results of tolerance checks on them, and other comments. SitesDB was created in 1997; most pre-1997 metadata are on individual station history files (hard copy or scanned, but not indexed or easily searchable). Station metadata of all kinds become sparse prior to the 1960s. In addition to station-specific metadata, more general metadata (such as network-wide observation procedures) are held in a range of internal documents.

3 | THE DATA HOMOGENIZATION PROCEDURE

The homogenization process (Peterson et al., 1998; Venema et al., 2012) involves two principal steps: the detection of potential inhomogeneities, and the assessment, and adjustment for (where warranted), these inhomogeneities. Detection can be carried out using either metadata or statistical methods, whilst a variety of methods are used for adjustment. The choice of adjustment methods depends on whether there are parallel (overlapping) data available. If there is not, methods using reference series are used, with the exact method depending on the availability of reference stations.

3.1 | Detection of inhomogeneities

Generally accepted best practice for detection of inhomogeneities involves a combination of metadata and statistical methods (Aguilar et al., 2003; Venema et al., 2012). Metadata, which may be specific to individual stations or cover issues which affect a large part of the network, often provide the most definitive evidence that a change has occurred at a given site on or near a certain date. However, metadata are often missing, incomplete or imprecise (e.g. where site photographs or inspection reports some years apart reveal that a move has taken place between those dates, but not the exact date). They may also be useful for confirming a change but not as a primary tool (e.g. a change in local site environment which only becomes apparent when site information is examined in detail following the identification of a shift by statistical methods). Statistical methods are therefore important in comprehensively assessing potential inhomogeneities, both documented and undocumented.

In ACORN-SATv1, a single statistical detection method was used. This was based closely on the pairwise homogenization algorithm (PHA) developed by Menne and Williams (2009). It was applied to monthly and seasonal data separately at pairs of stations, with a potential breakpoint flagged if it was found in the monthly series or at least two of the four seasons.

Following the development of ACORN-SATv1, a benchmarking study of a large number of detection methods was published, as part of the European HOME project (Venema et al., 2012). This identified a number of different methods with broadly comparable performance (exact rankings depending to some extent on the metrics used for assessment), which in turn indicates that there are multiple different methods available for detection. Squintu et al. (2018) followed the approach of using multiple methods in parallel to increase the robustness of detection. Version 2 also followed that approach, with five different detection methods used in parallel:

- The PHA-based method used in v1.
- A variation of the PHA-based method but using annual rather than monthly data.
- Three other methods chosen largely for ease of implementation, and implemented with monthly data:
  a. HOMER v2.6 with joint detection (Mestre et al., 2013),
  b. MASH version 3.03 (Szentimrey, 2008),
  c. RHTests version 4 (Wang et al., 2010).

Descriptions of, and further details on the implementation of, these methods are contained in Appendix S1. In all cases, the full Australian observing network was considered for use in these methods, subject to data availability.

Following an assessment (Trewin, 2018a) of the effectiveness of these methods in detecting known breakpoints, a breakpoint was considered for further assessment (section 3.2) if it was identified by the PHA method used in v1 (for consistency with v1), or by any two of the other four methods.
within a 2-year window. Breakpoints identified by metadata took precedence over those identified by statistical methods within the 2 years before or after. All breakpoints were attributed to 1 January unless a more precise date could be obtained from metadata records. Maximum and minimum temperatures were considered independently, although metadata-identified potential breakpoints were generally considered for both.

### 3.2 Adjusting for inhomogeneities

Once a potential breakpoint has been identified, the next step is to carry out adjustments, if warranted, to remove the impact of the inhomogeneity and produce a homogeneous record. In all cases, the older data are adjusted to be homogeneous with the most recent data, to allow the data set to be updated with operational data without further adjustment.

In some cases, overlap data will be available for a breakpoint. This occurs when a station move takes place or new instruments are installed, and the old site/instruments are left in place for a period of time after the new site/instruments are established.

In the majority of cases, however, there will not be overlap data of this type. In these cases, adjustment is carried out by comparing the candidate station with a number of other stations in the region which are well-correlated with the candidate station. These are known as reference stations. A requirement for reference stations is that they are themselves homogeneous during the required interval before and after the breakpoint under investigation at the candidate station, to establish that changes in the relationship between the candidate and reference stations result from inhomogeneities at the candidate station and not the reference station. Since, in general, it is not known a priori whether reference stations are homogeneous in that interval (which, in ACORN-SAT, is normally the 5 years before and 5 years after the breakpoint), an adjustment method needs to be robust against previously undetected inhomogeneities at reference stations, including by identifying inhomogeneous reference stations and eliminating them from consideration.

#### 3.2.1 Daily adjustments

Most adjustments in ACORN-SATv2, as in v1, are carried out at the daily timescale, with different adjustments being applied to different parts of the frequency distribution of daily data. The underlying algorithm for daily adjustments, known as the percentile-matching (PM) algorithm, is essentially unchanged from v1. This is applied in two different forms: one where sufficient overlap data exist between an old and a new site, and a second where insufficient (or no) overlap data exist and an independent set of reference stations is required.

### The overlap case

The overlap case 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. Overlap adjustments are applied as the final step in the process after the individual series are homogenized, using the procedures described in later sections.

For the overlap period (or in some cases, a subset of it), transfer functions are developed for maximum and minimum temperatures separately, for each of the 12 months. These transfer functions are developed using deseasonalized data for the candidate month and the preceding and following month (e.g. April, May and June for the May transfer function). They use the 5th, 10th, ..., 95th percentile points of daily temperature during the overlap period at each site to define the transfer function. The difference between the 5th percentile values is applied as an adjustment to all values below the 5th percentile, and that between the 95th percentile values to all values above the 95th percentile. As a separate transfer function is used for each of the 12 months, different adjustments apply to data at the end of 1 month to those at the start of the next. A mathematical description of these functions is contained in Trewin (2013).

An example of such a transfer function is shown in Figure 2, showing a case where a site move from a coastal site to a more sheltered inland site resulted in a cooling of winter (June–August) minimum temperatures, with the cooling being larger on cold nights (which are mostly likely to be calm and clear) than on warm nights (more likely to be cloudy and/or windy). At this location, winter is the driest season and coastal temperature gradients at night are at their maximum. The transfer function is unsmoothed, and fluctuations within the transfer function (as seen in the range between 7°C and 10°C) reflect some level of uncertainty within the transfer function, which in turn is incorporated in the overall uncertainty of the adjustment (section 3.2.5). A consideration here is that an especially important component in the PM methodology is the representation of extremes, and the potential exists for smoothing methods to be influenced at the extremes by a small number of outliers. It was found (Trewin, 2012) that extending the matching points of the transfer function to the 1st, 2nd, 3rd and 4th percentile resulted in a substantially poorer representation of extremes on average, most likely because of sensitivity of the small samples involved to outliers.

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. These cases can be identified in a number of ways, for example through metadata, homogeneity testing at the individual component sites, or the second-round homogenization process (section 3.3). Where such cases were identified, depending on the circumstances, either only a subset of the period of parallel observations was used to merge the sites, 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.

The non-overlap case
The most commonly used adjustment method was the non-overlap case of the percentile-matching algorithm. This was used in all cases except those with overlap data (section 3.2.1.1), those where there were insufficient reference stations meeting the correlation index criteria (section 3.2.2), or those where two breakpoints were sufficiently close in time to warrant use of the ‘spike’ method (section 3.2.2).

In the non-overlap case, a set of reference stations well-correlated with the candidate station was used. In most of Australia, significant correlation exists on daily to monthly timescales over distances of hundreds to thousands of kilometres (Trewin, 2001) owing to the dominance of synoptic features in much of the climate variability, especially over the relatively flat terrain typical of much of inland Australia. Correlation length scales are shorter near the coast, and for minimum temperatures at tropical sites during the wet season.

Reference stations were chosen from the full Australian network, which contains approximately 1900 stations, of which 700-800 have been operating at any given time for most of the last 60 years, with smaller numbers prior to 1957. The reference stations were chosen in correlation order with the candidate station, excluding reference stations with already known inhomogeneities in the 5 years before or after the breakpoint being tested. For this purpose, a correlation index was defined by calculating separately for each of the 12 months as the correlation of daily maximum (minimum) temperatures in that month, then calculating the index as the median of these 12 values. This is done as our interest is in how closely related day-to-day temperature changes are at a specific time of year, whereas taking correlations across the full year would in many cases be dominated by the seasonal cycle of temperature or, where anomalies are used, seasonal cycles in temperature variability. The index is defined separately for maximum and minimum temperatures. Optimally, 10 reference stations were used, to minimize uncertainty in the adjustment, but sensitivity studies (section 3.2.5) found that an acceptable level of uncertainty still existed for as few as three reference stations, if there were not 10 suitable stations available (as sometimes occurs in data-sparse areas or regions with sharp climatic gradients).

Only stations with a correlation index of 0.6 or above, and which had data for a minimum of three of the 5 years before, and three of the 5 years after, the breakpoint being tested, were considered. (It was found in Trewin (2013) that a 0.6 threshold was sufficient for the use of such a reference station to add significant skill to the adjustment). If there were sufficient reference stations meeting this requirement, the 10 best-correlated stations were used as a reference, otherwise all stations meeting the 0.6 correlation index were used (with a minimum of three). An iterative process (section 3.2.3) was used to identify reference stations with potential inhomogeneities during the period of interest (generally from 5 years before to 5 years after the breakpoint being tested) before the final set of reference stations was finalized. Any inhomogeneous reference stations were then removed from the final set of reference stations.
stations from the initial selection were replaced with the next best-correlated (providing they met the 0.6 correlation index criterion).

For each reference station, a two-step procedure was used. First, a transfer function was defined, as in the overlap case above, 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. 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. As in the overlap case, a separate set of transfer functions was defined for each of the 12 months, using data from the candidate month and the preceding and following month.

The final adjusted value for each data point at the candidate station was taken as the median of the estimates derived using each of the individual reference stations (Figure 3). The use of the median is to minimize the influence of an individual outlying reference station (e.g. one which had an undetected inhomogeneity of its own). The overall process is shown in Figure 4.

There were some cases where the period immediately before or after the breakpoint was unrepresentative, or where there was a second inhomogeneity within a few years of the inhomogeneity being considered. A common scenario is that something occurs at a site—for example the construction of a new building nearby—and the site is moved as a consequence a year or two later. In such cases, the reference periods were sometimes shifted earlier or later, or shortened, 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.

3.2.2 Monthly and spike adjustments

In some cases, there are insufficient reference stations with daily data to allow use of daily adjustment methods. As shown in section 3.2.5 below, uncertainty increases sharply if fewer than three reference stations were used, and hence, daily adjustments were not used if there were not at least three potential reference stations with a correlation index > 0.6. This applied most commonly in two situations: in remote areas, and in the pre-1957 period (when many potential reference stations have digitized monthly, but not daily, data) in regions with short correlation length scales. These include sites near the east coast where sea breezes and sharp-land sea contrasts may dominate at some times of the year.

In those situations, a monthly adjustment method was used, which applies a uniform adjustment to the daily data for each of the 12-calendar months. Reference stations were chosen in order of correlation, as for daily adjustments. Only stations with a correlation index > 0.6 were used as reference stations if at least three such stations were available, with the threshold relaxed to 0.5 only if there were fewer than three stations satisfying the 0.6 threshold. There was one case (Darwin in 1937) where only one reference station was available; this adjustment was well supported by metadata.

In ACORN-SATv1, 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 v1 during the development of v2, that the effect of this weighting function was to give excessive weight to the nearest reference station in data-sparse areas (e.g. \( w_s = 0.368 \) 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 issues at that reference station.

In v2, the anomalies at the reference stations were instead combined using a weighted median. Here, the weighting function is defined as follows:

\[ w_s = r_s^2/(\sum r_s^2) \]

where \( r_s \) is the correlation index (as defined earlier) between site \( s \) and the candidate site.

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, \( T_1, T_2, ..., T_N \), and \( k \) is the lowest value such that \( w_1 + w_2 + ... + w_k \geq 0.5 \). (If the sum of the weights is exactly 0.5, then the weighted median is the mean of \( T_k \) and \( T_{k+1} \).)

As for the daily method, the use of a median of multiple reference stations provides some robustness against inhomogeneities in reference stations, undetected inhomogeneities within the adjustment window (generally 5 years before and 5 years after the breakpoint) at reference stations could still affect adjustments. To mitigate this, a new process was used in v2 which involved the running of a diagnostic as part of the daily (non-overlap) and monthly adjustment methods, in which the estimated mean annual adjustment arising from each single anomalous reference station.

Where there were fewer than three reference stations, as occurred in a very small number of cases, a simple inverse-distance weighting was used instead of the procedure above.

A monthly adjustment method was also applied to ‘spike’ adjustments, defined as those where two breakpoints were found (either statistically or through metadata) within 3 years of each other. In such cases, the monthly adjustment method above was applied, using (normally) the 5 years before the start of the ‘spike’ as the first reference period and the ‘spike’ period itself as the second reference period. Once the ‘spike’ had been adjusted to be homogeneous with the period before it, an adjustment (if warranted) was then applied across the ‘spike’ period. This used the years before and after the ‘spike’ as references, and daily adjustments if the criteria for these were met. The overall procedure was adopted to prevent adjustments applied to short periods, which are likely to have a large uncertainty, from influencing adjustments applied to other parts of the data set.

‘Spike’ adjustments accounted for 3% of the total number of adjustments made in ACORN-SATv2, with other monthly adjustments comprising a further 7%.

### 3.2.3 Identification of inhomogeneous reference stations

Reference stations will only give representative results if they are themselves homogeneous during the period in which they are used. Whilst the use of a median of multiple reference stations provides some robustness against inhomogeneities in reference stations, undetected inhomogeneities within the adjustment window (generally 5 years before and 5 years after the breakpoint) at reference stations could still affect adjustments. To mitigate this, a new process was used in v2 which involved the running of a diagnostic as part of the daily (non-overlap) and monthly adjustment methods, in which the estimated mean annual adjustment arising from each single reference station was calculated. That reference station was then eliminated from the set of reference stations, and replaced by the next best-correlated, if any of the following applied:

- In the initial assessment: differing from the median adjustment from all reference stations 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.

Potential reference stations already known to be possibly inhomogeneous (e.g. other ACORN-SAT locations with an identified breakpoint within 5 years of that under investigation at the candidate stations) were also excluded.
Reference stations were then replaced as required by the next best-correlated station(s), if available, until all reference stations met these criteria.

In some cases, it is likely that reference stations have been flagged as potentially inhomogeneous because of short-term issues rather than longer-term inhomogeneities. In particular, it was found during the development of ACORN-SATv2 that relationships between temperatures at candidate and reference stations could become unstable during unusually wet periods (such as those around the La Niña events of 1973–1974 and 2010–2011) in arid and semi-arid regions. This is most likely because of site-specific vegetation responses to the wet conditions. An example of this occurred for a potential breakpoint at Camooweal in far north-west Queensland in 2012, at the boundary between a very wet period leading up to 2012 and a prolonged drought thereafter. Differences between Camooweal and some other sites in the region (e.g. Trepell, The Monument and Winton) are substantially different (Figure 5) in 2009 and 2011, following the very wet rainy seasons of 2008–2009 and 2010–2011 respectively, to those in 2008 and 2010 which followed much drier summers.

These thresholds were set following an evaluation (Trewin, 2012) which involved 16 pairs of sites with between 4 and 11 years of parallel data. For each of the 16 cases, a series was constructed by combining the ‘old’ station before the start of parallel data with the ‘new’ station from that point onwards, and the later part of that series was adjusted using the PM algorithm. The adjusted data in this series were then compared with the continuation of the ‘old’ station during the parallel observations period, using a variety of metrics (including root-mean-square (RMS) error, and differences in the count of days above the 90th or below the 10th percentile). It was found that, for cases where the two parallel series differed by less than the thresholds above, differences between the adjusted data and the continuation of the ‘old’ station did not differ significantly from those between the unadjusted ‘new’ station and the ‘old’ station, indicating that such small adjustments added no significant skill to the representation of the data set. Above these thresholds, the adjusted data did significantly outperform the unadjusted data.

For spike adjustments, an annual mean adjustment of at least 0.5°C was required.

### 3.2.4 Minimum criteria for adjustment

Adjustments of all kinds (except those for systematic broad-network issues as described in 3.4 below) were applied only if they met a minimum size criterion. As in v1, adjustments were required to meet any one of the following:

- At least 0.3°C in the annual mean;
- At least 0.3°C (not necessarily of the same sign) in at least two of the four seasons;
- At least 0.5°C in at least one season.

### 3.2.5 Quantifying uncertainties in the adjustments

Any adjustment will have a certain level of uncertainty associated with it. In order to quantify aspects of this uncertainty, a series of trials was carried out using 14 of the 16 test sets of stations used in section 3.2.4 above. In each case, to assess the impact of the selection of reference stations on the adjustment, an ensemble of 50 different adjustments was carried out, using as reference stations (instead of the 10 best-correlated stations) a set of 10 reference stations selected...
randomly from the pool of all stations which exceeded a given correlation index threshold. The outcomes of the 50 ensemble members were then compared.

This process found that the average ensemble standard deviation across the 14 test sets was approximately 0.07°C in the annual mean for both maximum and minimum temperatures, with standard deviations for seasonal means between 0.07°C and 0.10°C, and for the values of the seasonal 10th and 90th percentile values between 0.1°C and 0.2°C. More detailed results are available in Trewin (2018a).

A further sensitivity assessment was carried out at one location, Port Macquarie (Figure 6), where the ensemble standard deviation was calculated using an ensemble of 50 sets of \( n \) reference stations, where \( n \) ranged from 1 to 10. This found that the ensemble standard deviation generally increased as the number of reference stations decreased, with the increased sharpening for \( n < 3 \) (and being particularly large where only a single reference station was used). This supports the decision not to use the daily PM algorithm in cases where there are fewer than 3 reference stations meeting the 0.6 correlation index threshold.

3.3 Second round of homogenization

Results from v1 of ACORN-SAT and other homogenization methods (e.g. Squintu et al., 2018) indicate that, after the initial stage of homogenization takes place, best results will be obtained if a second round of homogenization takes place to identify possible inhomogeneities in the 'homogenized' series. This was applied on an ad hoc basis in ACORN-SATv1 but was carried out more systematically in v2.

The most likely scenarios for residual inhomogeneities after the first round of homogenization 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 v1, occurred in north-western New South Wales and south-western 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 homogenized time series were tested for homogeneity. This used the RHTests method (which is less dependent on having a relatively large number of reference stations than the other methods used in the first round of detection) and, in general, the homogenized 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. 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 homogenized series to determine whether a potential adjustment would meet the normal 0.3°C minimum criterion.

In total, 22 of the 968 adjustments (12 maximum, 10 minimum) applied to the raw data in ACORN-SATv2 arose from this second-round procedure.
3.4 Treatment of systematic issues affecting substantial parts of the network

Systematic issues affecting a large part of an observation network can have a substantial impact on large-area averages such as national means, even if the effect at an individual station is too small to be detectable or adjustable under normal procedures. Such issues can also be difficult to detect and adjust for as they may affect a large part of a network simultaneously, limiting the usefulness of reference stations. This may require known issues of this type to be dealt with separately. In some cases, this may require a whole-network adjustment to be applied (e.g. Vincent et al., 2009), especially where it is difficult or impossible to quantify the impact at specific stations.

Two such issues in the ACORN-SAT network are observation times and changed screen sizes. In both cases, whilst the issue itself affects a large part of the network, its impact on the data is specific to each individual station, and hence, the adjustments were carried out as part of the overall sequence of station-specific adjustments.

3.4.1 Observation times

The standard observation time for daily maximum and minimum temperatures is 0900 local time4 (including any local use of daylight saving time), which varies from 2200 UTC to 0100 UTC depending on location and time of year. This has been in general use across the Australian network since 1964, except for a few early-generation automatic weather stations (mostly in South Australia) which used days ending at 0000 UTC for minimum temperature and 1200 UTC for maximum temperature for their first few years of operation. The use (or non-use) of daylight saving time also introduces some effective one-hour shifts of observation time. Prior to 1964, Bureau-staffed stations such as airport and capital city meteorological offices (and some other locations, such as lighthouses) instead used an observation day ending at 0000 local time (Trewin, 2012).

Observation time is known to have the potential to influence daily maximum and minimum temperatures, both in countries where there is no fixed standard for observation time such as the United States (Karl et al., 1986) and where a change has been made across most or all of a national network simultaneously, such as Canada (Vincent et al., 2009) or Norway (Nordli, 1997). In the ACORN-SATv1, observation time issues were assessed in detail (Trewin, 2012). It was found that the use of daylight saving time, or of a 0000/1200 UTC day, had no significant impact. However, the impact of use of a 0000 local time day was sufficiently large at stations outside the tropics that it warranted a specific adjustment at affected stations. As in v1, this was implemented by applying adjustments to minimum temperature on 1 January 1964, whether or not that adjustment met the normal 0.3°C threshold. Only unaffected stations were used as reference stations.

3.4.2 Screen sizes

Two different Stevenson screen sizes have been used in the network over time (Warne, 1998): ‘large’ (approximately 71 × 71 × 53 cm) and ‘small’ (43 × 52 × 27 cm). 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 and some as late as 2012. Only four of the 112 ACORN-SAT sites still have large screens.

In ACORN-SATv1, no specific adjustment was applied for changes in screen size, based on the outcome of a field study at Broadmeadows, near Melbourne (Warne, 1998), which found a negligible impact on mean temperature, with the small screen having maximum temperatures 0.094°C higher and minimum temperatures 0.082°C lower. However, in ACORN-SATv2, it was considered that the size of the impact on diurnal temperature range which was reported in the Warne (1998) study was sufficiently large to warrant specific treatment.

This was implemented in a similar way to the observation time adjustment. An adjustment was made for both maximum and minimum temperatures on the date of the screen change (unless an adjustment had already been made within 2 years of that date), without the usual minimum adjustment size criteria being applied. Only stations with no known screen changes within 5 years of the date were used as reference stations. In total, there were documented screen changes at 86 of the 112 locations, with four still having large screens and the remaining 22 having no documented evidence of ever having large screens.

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 size. The spread of the results between stations was wide (several tenths

4Local time ranges from 8 hours (Western Australia all year) to 11 hours (New South Wales, Victoria, Tasmania and the ACT in summer) ahead of UTC.
of a degree), suggesting that the required adjustments are site-specific (one potential influence being the condition of the screen being replaced).

3.4.3 | Urbanization

One potential influence on temperature data which could make it unrepresentative of the large-scale climatic conditions is urbanization, which can lead to anomalous warming, particularly of minimum temperatures. Whilst some of the temperature impacts will arise from specific changes near an observation site (e.g. the construction of a new building or paved surface nearby) which are detected as inhomogeneities and adjusted accordingly, this does not eliminate the possibility that there may be residual urban effects which generate anomalous warming trends at sites in urban locations.

To assess this, locations were initially divided into three categories:

- Urban: locations where the current site is clearly within a city or town with a population of more than 10,000 (according to the 2016 Census).
- Potential urban: locations where either (a) the current site is in a non-urbanized part (e.g. a large park or an airport) within a city or town with a population of more than 10,000 or (b) the current site is in a non-urban location but previous sites have been urban (e.g. where a site has moved from a town centre to an airport outside the town boundary).
- Non-urban: locations not associated with a town or city with a population of more than 10,000, or associated with such a centre but clearly outside its boundary.

The sites in the ‘potential urban’ category were then assessed by comparing warming trends in their minimum temperatures with those at non-urban ACORN-SAT locations in the same region. Those sites which showed evidence of anomalous warming were given a final classification of ‘urban’. This process was first applied in v1 and continued in v2.

As a result of this process, eight of the 112 locations were classified as urban: four from the first round of classification (Sydney, Melbourne, Adelaide and Hobart) and four ‘potential urban’ sites which were found to have anomalous warming trends (Laverton, Richmond (NSW), Rockhampton and Townsville). These eight sites were classified as urban: four from the first round of classification (Sydney, Melbourne, Adelaide and Hobart) and four ‘potential urban’ sites which were found to have anomalous warming trends (Laverton, Richmond (NSW), Rockhampton and Townsville). These eight sites were retained in the ACORN-SAT data set as it is considered important to have high-quality long-term time series for the assessment of urban climates. However, they are not included in many downstream products, including the gridded data used for the calculation of national and regional spatial means and long-term trends in these (section 4.1). There was no change in the sites classified as urban between v1 and v2.

It is interesting to note that of the four ‘potential urban’ sites which had anomalous warming trends, three are either located in growth corridors in the outer suburbs of major cities (Laverton and Richmond) or near a city which is rapidly growing in general (Townsville). In particular, the anomalous warming signal at Laverton, located on a (now mostly decommissioned) Air Force base, develops from the late 2000s onwards, coinciding with the construction of a new suburb about 1 km west of the site. It is also of interest that the Sydney record shows no evidence of anomalous warming relative to the region, indicating that the urban influence at that location (which is in a part of the city which was already heavily built-up by the late 19th century) was already fully developed by 1910.

3.5 | Evaluation of other systematic issues

In addition to the systematic issues raised in section 3.4 above, there are a number of other potential systematic issues which are analogous to those which have led to network-wide biases in some other countries. As such, it is necessary to evaluate these issues to provide a level of assurance that no specific adjustments are required for them in ACORN-SAT. Two such issues, considered in more detail below, are the transition to automatic weather stations, and changes in temperature probes which potentially alter instrument response time characteristics. A third such issue, the 1972 transition from imperial to metric measurements in Australia and consequent network-wide replacement of thermometers, was evaluated in Trewin (2012) and is not discussed further here.

3.5.1 | 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-resistant probe in place of manually read liquid-in-glass thermometers. Australia has retained the same wooden Stevenson screen design (except for a few stations outside the ACORN-SAT network) as that used at manual stations. This differs from 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).

AWSs began to be widely used in the Australian network from the late 1980s, although the first AWS data in the

5There were a few earlier AWSs, mostly on remote offshore islands.
ACORN-SAT data set are from 1994. A major transition took place on 1 November 1996, when, at sites (primarily Bureau-staffed Meteorological Offices) which had both manual and automatic instruments, the automatic instruments became the primary instrument for daily maximum and minimum temperatures in the Australian climate database on which ACORN-SAT is based. AWSs continued to spread through the network over the following years; in many cases, these were installed at airports (or other out-of-town locations) to replace manual sites in towns.

A major change in instrument type, such as a change from manual to automatic instruments, can 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). This had no significant impact on mean temperatures (section 3.4.1). However, many automatic weather station installations also involved site moves, often from town locations to airports. These were dealt with in the same way as other known site moves (both in v1 and v2 of ACORN-SAT).

Previous assessments (Trewin, 2012), using nearby stations which had retained manual observations to assess the impact of the 1996 changeover to using automatic instruments as primary observations, had found no significant systematic change to either maximum or minimum temperatures.

To further explore this question, an assessment was carried out using data from ten locations where automatic and manual observations took place at the same time at the same place, or in very close proximity. At nine of these, manual and automatic instruments were in the same screen, whilst at the tenth, they were in different screens 3 m apart. The parallel observations revealed anomalous data for part of the comparison period (most likely associated with faulty or out-of-tolerance instruments) at three of the ten locations, leaving seven available for analysis (Table 1).

The comparison shows that the differences in mean temperature between the manual and automatic instruments are often less than 0.25°C at all seven sites and align closely with the results of tolerance checks at those stations. This indicates that the differences which do exist are primarily the result of differences between individual instruments rather than a more systematic effect. Differences in diurnal temperature range are also small at all seven sites and are not consistent in sign, suggesting that there was no significant change in effective response time between the manual and automatic instruments. (If the effective response time of the automatic instruments was faster, this would be expected to lead to higher maximum and lower minimum temperatures, and hence a larger diurnal temperature range). There is also little seasonality in the results; averaged across the seven

| Location                  | Dates                      | Difference (automatic – manual) (°C) | Mean tolerance check result (°C) |
|---------------------------|----------------------------|-------------------------------------|----------------------------------|
|                           | Maximum temperature        | Minimum temperature                 | Mean temperature                 | Diurnal temperature range | Mean tolerance check result |
| 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        | −0.26                               | −0.17                            | −0.22                        | −0.09                        | −0.06                        |
| Launceston Airport        | Nov 1996 – Dec 2001        | −0.10                               | −0.21                            | −0.15                        | 0.11                         | −0.06                        |
| Brisbane Airport          | Jan 2012 – Dec 2013, Jun-Dec 2015 | −0.04                           | −0.06                            | −0.05                        | −0.02                        | −0.06                        |
| Mean                      | −0.06                      | −0.09                               | −0.07                            | 0.02                         | −0.01                        |
stations, none of the four seasons (not shown) deviate from the annual mean difference by more than 0.03°C for either maximum or minimum temperature.

These results indicate that no network-wide adjustment is warranted for the transition from manual to automatic measurements; only site-specific adjustments in cases where a site change was also involved.

### 3.5.2 Response time changes with different automatic probes

It was also considered whether any subsequent changes in automated temperature probes used after AWS installation could have different response time characteristics. The key instrument used for temperature measurement in Australian automatic weather stations is a platinum-resistant 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). A faster response time would result in higher maximum and lower minimum temperatures (and hence an amplified diurnal temperature range) as shorter-period fluctuations in temperature are sampled.

The effective response time of the observations was assessed by considering the mean value of the 1-minute temperature variation (i.e. the difference between the highest and lowest value within each minute) from the available 1-minute data at 0600 and 1500 local time, representing the typical time of minimum and maximum temperatures, respectively. This value is generally highest during the day in dry conditions, with monthly values generally in the range from 0.15°C to 0.30°C at 1500 local time in summer, reaching up to 0.40°C at some arid locations (Trewin, 2018b). Values are generally below 0.10°C at 1500 in winter at most southern locations, and all year in most locations near 0600 local time.

At 17 of the 98 ACORN-SAT locations with AWS 1-minute data, there was a significant breakpoint at some stage in the mean 1-minute temperature variation during the time of AWS measurements. Such a breakpoint indicates a likely change in the effective instrument response time and hence the 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). Among the most extreme examples of this is Alice Springs (Figure 7), where a November 2011 probe replacement resulted in an increase in mean 1-minute temperature variation of approximately 0.16°C at 1500 and 0.03°C at 0600. Assuming that the increased variation is distributed symmetrically about the 1-minute mean, this equates to a shift of approximately +0.08°C for maximum temperature and −0.015°C for minimum temperature, well below the normal threshold for adjustment. In less arid climates, the effect is smaller. Averaged

**TABLE 2** Count of ACORN-SAT version 2 adjustments by cause. Totals do not equal the sum of counts as some adjustments have multiple causes

| Reason for adjustment          | Maximum temperature | Minimum temperature |
|-------------------------------|---------------------|---------------------|
| Statistical                   | 220                 | 256                 |
| Known site moves (all)        | 176                 | 170                 |
| Screen change or condition    | 79                  | 77                  |
| Site condition (without move) | 9                   | 5                   |
| Observation time change       | 2                   | 24                  |
| Total                         | 463                 | 505                 |
across the network (including the 81 unaffected AWS stations and the 14 manual stations, where there is no impact), the impact of probe changes is estimated as being in the order of +0.01°C for maximum temperature and between zero and −0.01°C for minimum temperature. No adjustment for this was therefore made in ACORN-SAT given the very small size of the effect.

### 3.6 Number and size of adjustments

In total, 968 adjustments were applied to the original data in ACORN-SATv2 (Table 2), 51% of which were supported by metadata in some form and 49% of which were detected by statistical methods without supporting metadata. This represents a mean of 4.1 adjustments per location for maximum temperature and 4.5 for minimum temperature (equating to 4.5 and 4.9 per 100 years, respectively). About half of the difference between this and the 660 adjustments applied in v1 is accounted for either by the introduction of a new adjustment for screen size, or the addition of new or newly digitized data. 91 of the 968 adjustments used overlap data, with the remainder using reference series in some form.

The size of adjustments shows a bimodal structure (Figure 8) with peaks near +0.4°C and −0.4°C. This is to be expected, as in general adjustments between +0.3°C and −0.3°C are not implemented, except for those which are for time of observation or screen changes, or meet seasonal criteria.

Overall, there is no strong tendency towards positive or negative adjustments for maximum temperature (Table 3), but negative adjustments predominate for minimum temperature, with 58% of adjustments negative and 42% positive. Large adjustments of minimum temperature are largely negative, with 30 negative adjustments >1°C but only 13 which are positive. Minimum temperature adjustments associated with a documented site move from an in-town to out-of-town location are strongly negative, with 78% of adjustments negative and a mean magnitude of −0.65°C. However, there is only a weak negative tendency for adjustments not associated with site moves out of town, implying that the predominance of negative adjustments in

| Reason for adjustment                      | Maximum temperature | Minimum temperature |
|--------------------------------------------|---------------------|---------------------|
|                                            | Mean adjustment (°C) | Percentage of positive adjustments |
| All                                        | −0.03               | 50                  |
| Statistical                                | −0.02               | 50                  |
| All site moves                             | −0.07               | 43                  |
| Site moves from in-town to out-of-town     | −0.20               | 37                  |
| All except site moves from in-town to out-of-town | −0.01               | 50                  |

TABLE 3 Mean adjustment size and percentage of adjustments which are positive, by reason for adjustment. Spike adjustments are excluded from this part of the analysis.
the overall ACORN-SAT version 2 minimum temperature record is largely attributable to sites moving from in-town to out-of-town locations, with other adjustments having no clear positive or negative tendency.

4 | COMPARISON OF ACORN-SAT VERSION 2 WITH OTHER AUSTRALIAN TEMPERATURE DATA SETS AND LONGER-TERM TRENDS

4.1 | Comparison with other data sets

The data from ACORN-SATv2 can be compared with other Australian temperature data sets, such as the unhomogenized AWAP data set, and ACORN-SAT v1. These spatial averages are calculated from gridded data, using the 104 non-urban stations.

Trends for the 1910–2016 and 1960–2016 period in the data sets are shown in Table 4. It had previously been shown (Fawcett et al., 2012) that ACORN-SATv1 and AWAP showed similar warming trends over the post-1960 period, but that ACORN-SAT was cooler than AWAP prior to 1960 (and especially prior to 1930), particularly for minimum temperature. These continue to hold in the results reported in Table 4. However, unlike ACORN-SATv1, ACORN-SATv2 warms more strongly than AWAP over the post-1960 period. It is worth noting that Menne et al. (2018), using a completely independent method, present results from the Global Historical Climatology Network data set which show that Australia warms more strongly in the most recent version of that data set (version 4) than in earlier versions.

A more detailed comparison of ACORN-SATv1 and v2 is shown in Figure 9. The relationship between the two is relatively stable prior to about 1960, but after 1960 v2 warms progressively relative to v1. This levels off after 2000 for maximum temperature, but continues to near the present day for minimum temperature. 2013 is the warmest year on record for Australia in both ACORN-SATv1 and v2.

The most substantial systematic change in the observation network has been the tendency over time for sites to move from in-town locations (most often associated with post offices) to out-of-town locations (such as airports). About 80% of the ACORN-SAT stations which were operating in 1910 were in towns, but by 2016 this proportion had dropped to <10% (Figure 10). Such moves were particularly frequent in the 1940s, associated with the opening of numerous airport meteorological offices during and immediately after the Second World War. A second wave of site moves occurred in the 1990s and early 2000s, when the initial roll-out of AWSs took place (and fewer post offices made observations following the corporatization of Australia Post in 1989). As the majority of site moves from in-town to out-of-town locations result in non-climatic cooling (section 3.6), a larger warming trend in homogenized than non-homogenized data is consistent with what would be expected in a network with a systematic tendency to move to out of town locations over time.

There are a number of factors which potentially contribute to the stronger warming seen in ACORN-SATv2 relative to v1. Predominant among these are.

- It was found following a review of the v1 code that the use of the Fortran intrinsic anint function resulted in values where the second decimal place was 5 (e.g. 0.15) being rounded up when rounded to the nearest 0.1 degree (where the values were pre-multiplied by 10 to allow integer conversion). As final adjustments from the ‘standard’ procedure are calculated as the median of adjustments derived from all reference stations (normally 10), this would be expected to affect 50% of cases where the median of an even number of reference stations is being calculated, with an expected bias in adjustment size of +0.05°C. Consequently, this creates an expected bias across all ‘standard’ reference station adjustments of +0.025°C. In version 2, this was addressed by retaining full precision in all calculations until after all adjustments are completed. Given the number of affected adjustments per station, this would be expected on a raw averaging to account for a difference of approximately 0.09°C between v1 and v2 trends over the full data set.
- V2 employed more rigorous methods both for determining whether reference stations used in adjustments were themselves locally homogeneous over the period where

| 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 |
they were used as reference stations, and for determining whether overlap periods between new and old stations were representative. It was found from this that there were numerous cases when the overlap period between older manual stations and newer automatic stations was not representative of the longer-term record (most often because the manual site’s exposure deteriorated in its final years). In such cases, the overlap period was not used (or only used for a limited period) in v2, with other stations instead being used as references. The methods used in v2 were also more effective in filtering out potential reference stations which had inhomogeneities of their own. In v1, these factors had tended to mask some of the impact of the 1990s/early 2000s moves to out-of-town locations (suggesting, in turn, that the close correspondence between the post-1960 v1 and AWAP data was the result of two roughly offsetting biases). Such methodological improvements also resulted in the introduction of new pre-1950 adjustments at Eucla, previously masked by a large inhomogeneity at the nearest neighbour (Cook); because of Eucla’s remoteness, this had a large effective ‘footprint’ in national means.

- New data have been added to the data set; v1 ended in 2011 (being appended thereafter with real-time, unadjusted data), with no new adjustments after 2009. The post-2009 period saw a renewed wave of site moves out of town, particularly in New South Wales and Western Australia, with several of these moves leading to substantial negative breakpoints in minimum temperature.

### 4.2 Spatial assessment of trends

Maps of trends for the 1910–2016 period for ACORN-SATv2 and v1 are shown in Figure 11. These show warming through virtually all of Australia. The strongest warming trends in mean temperature (over 0.15°C per decade) are in central

---

**FIGURE 9** 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)

**FIGURE 10** Percentage of sites within the built-up area of towns (of any size), by decade
Australia, whilst the weakest trends (mostly between 0.05 and 0.10°C per decade) are in north-western Australia—which has seen a large increase in rainfall since the 1960s (Bureau of Meteorology, 2018)—in parts of the south-east (especially near the coast), and in northern Tasmania. In general, trends in v2 show more spatial coherence than those for v1, especially for minimum temperature. This suggests that the homogenization process has been effective in reducing local, often artificial, noise. The only negative trends for any element are for minimum temperatures in a small part of north-western Australia, and for maximum temperatures around Moree. The latter may be associated with the development since the 1960s of irrigated agriculture in the region, a phenomenon previously also documented at Mildura (Trewin, 2012).

The Australian land temperature data in ACORN-SATv2 also now show a somewhat stronger warming trend than Australian region SST data, consistent with global patterns which generally show stronger warming trends over land areas than over

**FIGURE 11** Linear trends (in °C/decade) for annual maximum temperature (top row), minimum temperature (middle row) and mean temperature (bottom row), from the ACORN-SAT version 1 (left column) and version 2 (right column) data sets, across the period 1910–2016.
the oceans (Hartmann et al., 2013). There is a weak tendency (Figure 11) for temperature trends to be weaker near the coast, particularly in the south, than in inland areas. The Australian temperature changes in ACORN-SATv2 are very similar to the global average for land areas in data sets used in the IPCC Special Report on Climate Change and Land (IPCC, 2019).

5 | CONCLUSIONS AND FUTURE DIRECTIONS

The ACORN-SATv2 temperature data set shows consistent and spatially coherent warming across almost all of Australia over the last century, particularly in the last 50 years. The stronger warming trends, relative to unhomogenized data, are likely to be primarily attributable to the tendency over time for sites to move from in-town to out-of-town locations. Whilst these findings are only specifically attributable to Australia, it is likely that broadly similar network changes have occurred in many countries and need to be properly considered in land-based data sets, although at a global scale, homogenized data show weaker long-term warming trends than unhomogenized data because of cool biases in most sea surface temperature (SST) data prior to the Second World War (Kennedy et al., 2011) which offset biases which may exist on land.

The updated ACORN-SAT data will support ongoing analysis of temperature extremes in Australia, one example being the relatively weak trends in extreme low minimum temperatures in parts of southern Australia since the 1980s, associated with a strengthening of the subtropical ridge (Pepler et al., 2018). The new data set will be supported by a comprehensive uncertainty assessment which is currently in preparation.

As with any data set, ongoing use of the ACORN-SAT data set will require regular updates, as the network continues to evolve. Based on the last decade, typically one to three sites per year will close and be replaced by other sites nearby. It is expected that there will also be further technological changes, such as the ultimate replacement (possibly with associated site changes) of the remaining manual sites by automated instruments. There are also approximately 15 sites in the network which are believed still to have undigitized daily data. It is intended to carry out regular updates annually, with more substantial version updates every few years or when warranted by large-scale changes affecting significant parts of the network.

6 | LOCATION OF DATA AND ASSOCIATED MATERIAL

The data set is archived in the National Computational Infrastructure (NCI) repository. The citation for the data set is: Australian Bureau of Meteorology, 2018. Australian Climate Observations Reference Network – Surface Air Temperature (ACORN-SAT) data set, version 2. Australian Bureau of Meteorology, https://doi.org/10.25941/5d28a5d352de7.

The data may also be obtained through the Bureau of Meteorology, at http://www.bom.gov.au/climate/data/acorn-sat/. The following items are also available through this site:

- A station catalogue, containing details of adjustments and summarized metadata.
- The transfer functions used for adjustments.
- Breakpoints found by the various methods used in section 3.1 and their consolidation into a single set of breakpoints for testing.
- Raw data (quality-controlled but not adjusted) from the stations which contributed to the ACORN-SAT dataset.

ACKNOWLEDGEMENTS

The assistance of Skie Tobin, Ben Hague and three anonymous reviewers in reviewing earlier drafts of this manuscript is gratefully acknowledged. Linden Ashcroft contributed to the assessment of parallel manual and automatic observations.

OPEN PRACTICES

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at https://doi.org/10.25941/5d28a5d352de7 Learn more about the Open Practices badges from the Center for OpenScience: https://osf.io/tvyxz/wiki

ORCID
Blair Trewin https://orcid.org/0000-0001-8186-7885
David Martin https://orcid.org/0000-0003-1542-6162

REFERENCES

Aguilar, E., Auer, I., Brunet, M., Peterson, T. and Wieringa, J. (2003). Guidelines on climate metadata and homogenization. Geneva, Switzerland: WMO/TD No. 1186, World Meteorological Organization.

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. International Journal of Climatology, 28, 389–400. 1-.1002/joc/1524

Bureau of Meteorology (2018) State of the Climate. Melbourne, Vic.: Bureau of Meteorology.

Commonwealth of Australia (2015) Australian Climate Observations Reference Network – Surface Air Temperature (ACORN-SAT), Report of the Technical Advisory Forum, Canberra: Commonwealth of Australia. Retrieved from http://www.bom.gov.au/climate/data/acorn-sat/documents/2015_TAF_report.pdf

Commonwealth of Australia (2017) Australian Climate Observations Reference Network – Surface Air Temperature (ACORN-SAT), Report of the Technical Advisory Forum, Canberra: Commonwealth
of Australia. Retrieved from http://www.bom.gov.au/climate/data/acorn-sat/documents/2017_TAF_report.pdf

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.

Hartmann, D.L., Tank, A.M.K., Rusticucci, M., Alexander, L.V., Brönnimann, S., Charabi, Y.A.R. et al. (2013) Observations: atmosphere and surface. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, U.K. and New York, U.S.A.

IPCC (2019) Climate change and land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems. In press.

Jones, P. (2016) The reliability of global and hemispheric surface temperature records. Advances in Atmospheric Sciences, 29, 269–282. https://doi.org/10.1007/s00376-015-5194-4

Jones, D.A. and Trewin, B.C. (2002) On the adequacy of digitised historical Australian daily temperature data for climate monitoring. Australian Meteorological Magazine, 51, 237–250.

Jones, D.A., Wang, W. and Fawcett, R. (2009) High-quality spatial climate data-sets for Australia. Australian Meteorological and Oceanographic Journal, 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. Journal of Applied Meteorology and Climatology, 25, 145–160. https://doi.org/10.1175/1520-0545(1986)025<0145:AMETTT>2.0.CO;2

Kennedy, J.J., Rayner, N.A., Smith, R.O., Parker, D.E. and Saunby, M. (2011) Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. Biases and homogenization. Journal of Geophysical Research, 116, D014104. https://doi.org/10.1029/2010JD015220

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

Klein Tank, A.M.G., Wijngaard, J.B., Können, G.P., Böh m, R., Demarée, G., Gocheva, A. et al. (2002) Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. International Journal of Climatology, 22, 1441–1453. https://doi.org/10.1002/joc.773

Lenssen, N., Schmidt, G., Hansen, J., Menne, M., Persin, A., Ruedy, R. et al. (2019) Improvements in the GISTEMP uncertainty model. Journal of Geophysical Research: Atmospheres, 124, 6307–6326. https://doi.org/10.1029/2018JD029522

Melillo, J.M., Richmond, T.C. and Yohe, G.W. (Eds.) (2014) Climate Change Impacts in the United States: The Third National Climate Assessment (p. 841). U.S. Global Change Research Program. https://doi.org/10.7930/J0Z31WJ2

Menne, M.J. and Williams, C.N. (2009) Homogenization of temperature series via pairwise comparisons. Journal of Climate, 22, 1700–1717. https://doi.org/10.1175/2008JCLI2263.1

Menne, M.J., Williams, C.N. and Gleason, B.E. (2018) The Global historical climatology network monthly temperature dataset, version 4. Journal of Climate, 31, 9835–9854. https://doi.org/10.1175/JCLI-D-18-0094.1

Mestre, O., Domonkos, P., Picard, F., Auer, I., Robin, S., Lebarbier, E. et al. (2013) HOMER: HOMogenisation software in R – methods and applications. Időjárás, 117, 47–67.

Morice, C.P., Kennedy, J.J., Rayner, N.A. and Jones, P.D. (2012) Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: the HadCRUT4 dataset. Journal of Geophysical Research, 117, D08101. https://doi.org/10.1029/2011JD017187

Nicholls, N., Tapp, R., Burrows, K. and Richards, D. (1996) Historical thermometer exposures in Australia. International Journal of Climatology, 16, 705–710. https://doi.org/10.1002/(SICI)1097-0088(199606)16:6<705:AID-JOC329>3.0.CO;2-S

Nordli, P.O. (1997) Adjustments of Norwegian monthly means of daily minimum temperature. Oslo, Norway: KLIMA Report 6/97, Norwegian Meteorological Institute.

Pepler, A.S., Ashcroft, L. and Trewin, B. (2018) The relationship between the subtropical ridge and Australian temperatures. J. Southern Hemisphere Earth Systems Science, 68, 201–214. https://doi.org/10.22499/3.6801.011

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. International Journal of Climatology, 27, 267–276. https://doi.org/10.1002/joc.1381

Peterson, T.C., Easterling, D.R., Karl, T.R., Groisman, P., Nicholls, N., Plummer, N. et al. (1998) Homogeneity adjustments of in situ atmospheric climate data: a review. International Journal of Climatology, 18, 1493–1517. https://doi.org/10.1002/(SICI)1097-0088(199811)18:13<1493:AID-JOC329>3.0.CO;2-T

Quayle, R.G., Easterling, D.R., Karl, T.R. and Hughes, P.Y. (1991) Effects of recent thermometer changes in the cooperative station network. Bulletin of the American Meteorological Society, 72, 1718–1723. https://doi.org/10.1175/1520-0477(1991)072<1718:EORTCO>2.0.CO;2

Rohde, T., Muller, R.A., Jacobsen, R., Muller, E., Perlmutter, S., Rosenfeld, A. et al. (2013) A new estimate of the average Earth surface land temperature spanning 1753 to 2011. Geoinformatics & Geostatistics: An Overview, 1, 1. https://doi.org/10.4172/2327-4581.1000101

Squintu, A.A., van der Schrier, G., Brugnara, Y. and Klein Tank, A. (2018) Homogenization of daily temperature series in the European Climate Assessment and Dataset. International Journal of Climatology, 39, 1243–1261. https://doi.org/10.1002/joc.5874

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. (2001) Extreme temperature events in Australia. Ph.D thesis, School of Earth Sciences, University of Melbourne.

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. International Journal of Climatology, 33, 1510–1529. https://doi.org/10.1002/joc.3550
Trewin, B.C. (2018a) The Australian Climate Observations Reference Network – Surface Air Temperature (ACORN-SAT) version 2. Bureau of Meteorology Research Report 32, Bureau of Meteorology, Melbourne.
Trewin, B.C. (2018b) 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.
Trewin, B.C. and Trevitt, A.C.F. (1996) The development of composite temperature records. International Journal of Climatology, 16, 1227–1242. https://doi.org/10.1002/(SICI)1097-0088(199611)16:11<1227:AID-JOC82>3.0.CO;2-P
Venema, V, Mestre, O., Aguilar, E., Auer, I., Guijarro, J.A., Domonkos, P. et al. (2012) Benchmarking homogenization algorithms for monthly data. Climate of the Past, 8, 89–115. https://doi.org/10.5194/cp-8-89-2012
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. Journal of Applied Meteorology and Climatology, 48, 2160–2168. https://doi.org/10.1175/2009JAMC2191.1
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. International Journal of Climatology, 38, 692–707. https://doi.org/10.1002/joc.5203
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. Journal of Applied Meteorology and Climatology, 49, 2416–2436. https://doi.org/10.1175/2010JAMC2376.1
Warne, J. (1998) A preliminary investigation of temperature screen design and their impacts on temperature measurements. Instrument Test Report 649, Bureau of Meteorology, Melbourne.
World Meteorological Organization (WMO) (2017) Challenges in the transition from conventional to automatic meteorological observing networks for long-term climate records. Geneva, Switzerland: WMO No. 1202, World Meteorological Organization.
World Meteorological Organization (WMO) (2018) Guide to climatological practices. Geneva, Switzerland: WMO No. 100, World Meteorological Organization.

SUPPORTING INFORMATION
Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Trewin B, Braganza K, Fawcett R, et al. An updated long-term homogenized daily temperature data set for Australia. Geosci. Data J., 2020;7:149–169. https://doi.org/10.1002/gdj3.95