BIASES IN THE AUSTRALIAN HIGH QUALITY TEMPERATURE NETWORK

David R.B. Stockwell* and Kenneth Stewart
**Central Queensland University, Rockhampton, Qld 4720, Australia*
d.stockwell@cqu.edu.au

ABSTRACT
Various reports identify global warming over the last century as around 0.7°C, but warming in Australia at around 0.9°C, suggesting Australia may be warming faster than the rest of the world. This study evaluates potential biases in the High Quality Network (HQN) compiled from 100 rural surface temperature series from 1910 due to: (1) homogeneity adjustments used to correct for changes in location and instrumentation, and (2) the discrimination of urban and rural sites. The approach was to compare the HQN with a new network compiled from raw data using the minimal adjustments necessary to produce contiguous series, called the Minimal Adjustment Network (MAN). The average temperature trend of the MAN stations was 31% lower than the HQN, and by a number of measures, the trend of the Australian MAN is consistent with the global trend. This suggests that biases from these sources have exaggerated apparent Australian warming. Additional problems with the HQN include failure of homogenization procedures to properly identify errors, individual sites adjusted more than the magnitude of putative warming last century, and some sites of such poor quality they should not be used, especially under a "High Quality" banner.

1. INTRODUCTION

Many of the claimed adverse effects of global warming are more perception than reality, and require a careful and critical reassessment [1]. Methodological biases [2] have been shown to be behind many adverse effects, such as increasing extinctions [3; 4], unprecedented warming [5], accelerating warming trends [6; 7] and severe Australian droughts [8]. Particularly when scientific claims are the basis for public policy decisions, the highest-quality work must be subjected to third-party checks and the data and computational methods disclosed for independent replication [9].

Of course, biases can act in both directions. In Australian meteorological datasets, the method of calculating mean temperature by fixed hour observations, introduced in 1994, was found to produce overall means between 0.17°C and 0.43°C cooler than those derived from the average of maximum and minimum temperatures [10]. The many poorly sited weather stations, and intermittently collected records may also be a source of bias. It is well known that station quality affects the recorded maximum and minimum temperature trends, with the most poorly sited showing a warmer trend [11]. A through understanding of the origins and magnitude of bias in surface temperature records is needed to properly evaluate the confidence levels of published global warming trends.

In contrast to the problems caused by the more well-known urban heat island (UHI) effect, the homogeneity adjustments used to correct for changes in station location and instrumentation, and splice segments of individual station data into contiguous series, have not yet been exposed to intense scrutiny. While one recent study found adjustments reduce, but do not eliminate, site-specific trend differences in the USA [11], a review of the New Zealand 7-Station Series [12] found adjustments may have exaggerated warming from 0.34 to 0.91°C/Century [13]. Yet a different form of bias may arise inadvertently by excluding urban sites from the Australian annual temperature time series if the urban sites have inherently lower temperature trends, due to their proximity to the ocean.

The design of this study is to quantify the effect of adjustments on temperature trends in a network of 134 Australian temperature series called the High Quality Network (HQN) developed in the doctoral thesis of Torok (1996) [14] and subsequent journal article [15]. A new Australian temperature network was prepared using overlapping sections to join segments into contiguous series, called the Minimally Adjusted Network (MAN). As baseline shifts can be assessed objectively from the overlapping segments, the approach is less subjective. The trends in the MAN were compared with the trends in the HQN.

Here we test two requirements for reliability of national scale average temperature trends: (1) overall biases of the adjustment methodology are neutral, and (2) absence of rural/urban selection bias.

1.1. The High Quality Network

Della-Marta et.al. (2004) noted problems with the HQN, particularly subjectivity of the homogeneity adjustments used to correct for changes in location, instrumentation and other errors in individual station data. If the homogenization adjustments used to correct for changes in location and instrumentation are a significant source of warming bias in the Australian temperature record, they may explain the inconsistency between the change in surface temperature last century of around 0.7°C globally and 0.9°C in Australia.

The HQN data are widely reported as a reliable basis for regional and national climate analysis, including determination of the Australian warming trend. From our point of view, a network of weather stations would only be regarded as reliable if each of the individual weather stations was reliable, or well-validated procedures corrected most known errors. We are not aware of any studies that demonstrate the homogenization methods accurately detect and correctly adjust known errors on simulated data [16; 17; 18; 19], or evidence the adjustments are not biased. Nevertheless, the Australian Bureau of Meteorology (BoM) claims on their website that "inhomogeneity problems have been reduced or even eliminated" in the HQN data

[20].[1] In the absence of empirical evaluation of these factors [21], the claim the BoM have eliminated inhomogenities in the HQN must, therefore, be treated with caution.

The HQN is an Australian network of selected individual stations (or sites), currently composed of 100 rural and 34 urban continuous temperature series dating back to 1910 [14; 15]. Subsequently reviewed by Della-Marta et.al. (2004) [21] (DM), they found on average six homogeneity adjustments per site to control for changes in location, instrumentation or measuring practice at sites. DM noted that "the decision [by Torok] whether or not to correct for a potential inhomogeneity is often a subjective one".[2]

Sites were originally classified as urban by Torok and Nicholls [15] if their population exceeded 10,000. However, Torok et.al. [22], who studied urban heat island (UHI) effects of small Victorian towns and Melbourne, found that "smaller but significant UHI can be detected even in small towns". DM reclassified 15 formerly urban sites as non-urban and included them in the rural network. One site (Innisfail) is classified as urban although its population is less than 10,000. Plausibly, UHI is site specific and not dependent entirely on population size, as Australian regional cities are rarely more densely populated than most country towns. The potential effect of excising urban sites from the average of Australian regional trends needs to be clarified.

1.2. Problems of Homogenization

Statistical studies show the efficacy of well-executed homogenization varies [16; 17; 18; 19], and in practice the accurate execution of a reliable method can also be problematic. A review by the New Zealand Climate Science Coalition (NZCSC) [13], of the New Zealand meteorological authority, NIWA, 7-Station Series [12] claims that because NIWA did not correctly follow statistical techniques in Rhoades and Salinger [23], their ad-hoc method exaggerated warming over the last hundred years.

The Rhoades and Salinger [23] homogeneity adjustments called for comparing each station with neighbouring stations using a monthly, symmetric interval centered on a discontinuity in the temperature record. A local climatology was developed from the averages weighted by correlations with neighbouring stations. A 95% significance level was required before performing adjustments. The NZCSC claim that due to deviation from the exact method of Rhoades and Salinger [23], i.e. by comparing annual records, larger asymmetric neighbourhoods, and no statistical tests, NIWA exaggerated the centennial trend from 0.34°C to 0.91°C.

---

1. The prevailing view of the reliability of the HQN was expressed by Dr David Jones, Head of Climate Monitoring and Prediction, BoM, (pers. comm., 25th April 2011) "On the issue of adjustments you find that these have a near zero impact on the all Australian temperature because these tend to be equally positive and negative across the network (as would be expected given they are adjustments for random station changes)."
2. DM were concerned that subjective decisions (1) whether to adjust and (2) by how much, were irreproducible: "Despite the suitability of the dataset for national and regional-scale analyses, any individual station record within the dataset should still be treated with caution. The subjectivity inherently involved in the homogeneity process means that two different adjustment schemes will not necessarily result in the same adjustments being calculated for individual records. However, if the overall biases of the different approaches are neutral then spatial averages should be highly consistent across a large number of homogenized records" [21].

In their review of the HQN, DM noted the critical influence of the homogenization method on individual trend accuracy. Apparently, in view of this, rather than change the entire HQN dataset, DM retained the pre-1994 homogeneity adjustments of Torok [14], except for about 30 adjustments in which an obvious error had been made, such as the wrong sign or an incorrect year of adjustment, and incorporated additional adjustments. Thus, DM treated the HQN as a legacy, handed down, and not reliable for individual records. This is not an ideal situation, as science disregards irreproducible data [9].

The purpose of the significance tests and weighting of neighbouring series in the homogenization procedure is to ensure that nearby sites provide reliable information on the candidate site. Inadequate significance testing increases the risk of performing adjustments on natural variations. Imposing too large a neighborhood confounds short-term changes with the long-term warming trend, thus biasing the adjustments to exaggerate the trend.[3] The capacity of metadata to provide a warrant for these adjustments is also questionable, as there are many instances of station moves or changes with no apparent discontinuity in the data, and there are many discontinuities without supporting metadata. DM found that 22% of adjustments in the HQN were completely undocumented. It is clear that the introduction of a large number of adjustments, unsupported by statistical significance or metadata, instead of improving the data by correcting errors, introduces a large additional uncertainty into the dataset.

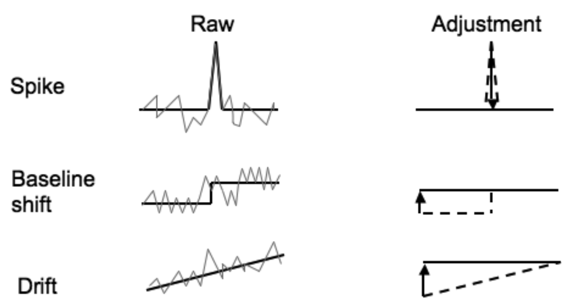

Figure 1. Three types of instrument errors and their adjustments.

In evaluating the efficacy of these homogenization procedures on data, one must first understand the potential types of errors and their effects. Basic instrumentation theory provides a guide. Instruments produce four main types of errors: random noise, spikes, baseline shifts, and drifts. While temperature records can contain more complex seasonally variable errors [24], the primary concern is with the last three as they may be rectified by simple adjustments (Figure 1).

3. As noted by DM: "The objective test has been found to over-estimate the number of discontinuities in a difference series since the nominated distribution of the test statistic, F3,n-4 incorrectly assumes that multiple discontinuities are independent [27]. The critical values of the corrected underlying distribution are typically twice those of the F3,n-4 critical values, indicating that if detected discontinuities in a difference series are applied without metadata support then the resulting series could include adjustments associated with natural climate variability or as we have just discussed artificial trends associated with biased reference series." [21]

The standard homogenization procedure only associates discontinuities in the data (sudden changes) with errors. Discontinuities could be from an error such as a spike (one up, one down) or a baseline shift. However, discontinuities also result from the rectification of an error when the baseline has changed due to a slow drift in temperatures, arising say, from instrument or enclosure deterioration, vegetation growth and creeping urbanization. Simulation studies of homogenization techniques on temperature data appear to have focused solely on the diagnosis of breaks and correction by baseline shifts [16; 17; 18; 19] and drifts are not considered. This is surprising considering drift could be quite prevalent in these data, arising from repeated moves as the surroundings evolved.

Homogenization of baseline shifts, spikes and drifts would tend to affect the long-term trend in different ways. Rectification of drift, by relocating the station to a better, climatologically similar location, would produce a sawtooth pattern, but does not change the long-term trend. Subsequent application of an homogenization procedure to remove discontinuities, by shifting the baseline, would introduce an erroneous trend (Figure 2). Where the drift is due a warming from progressive urbanization, multiple rectification and corrections would amplify the erroneous warming trend.

2. METHODS

In this study, minimal adjustments were made to a series, by only splicing overlapping data of known offset. The purpose was not to fully homogenize the data. The series still contain discontinuities due to spikes, and shifts from station moves and changes. However, the procedure is sufficient for the purposes of this study – to develop a comparative network with reduced potential for bias of the overall trend. While great efforts were made to examine all sites in the HQN for the purposes of comparison, several sites (e.g. Cashmore and Grafton) were of very poor quality.

The raw station data and HQN network data were downloaded from the BoM website. The available data yielded 268 series of annual temperatures from 1910 to 2009 – 134 from the HQN series downloaded from the BoM website, and 134 compiled by minimal adjustment to raw data, composed of 100 rural and 34 urban sites, called the Minimally Adjusted Network (MAN). Nine series with large upward adjustments to their warming trend were identified, and the BoM was asked to explain the apparent bias.[4,5]

4. The communications consisted of an email from Dr Jones on 25th April 2011; a formal complaint to the Minister, Mr Tony Burke on 26 October 2010; a response on 10 February 2011 by Dr Greg Ayers, head of the Bureau of Meteorology, promising information with the reasons for the large adjustments in 9 of the stations, copies of a journal article reviewing potential bias in warming trends to be published "later in the year" and an updated review of operational adjustments after publication in the scientific peer review literature; a response on 1 June 2011 after another letter to BoM with a copy to Mr Greg Hunt, the Shadow Minister, providing the BoM metadata [28] for the nine series. The reasons for the adjustments and review of operational adjustments have not been provided to date.

5. A major development after acceptance of this paper has vindicated our findings. The Australian Bureau of Meteorology has replaced the entire High Quality Network (that we audited in the paper) with a re-analysis called the ACORN-SAT. This has been done without reference to the limitations of the previous HQN. Our paper therefore provides an important reference point and justification for this upgrade. As far as motivations for this are concerned, we suspect the past and continuing efforts of a number of auditors of the Australian surface temperature records meant the HQ data set could no longer be defended and an improved data set was required. We are gratified that our efforts have in some way led the BoM to address major problems with the Australian surface temperature record. Our paper is a necessary component for our next step, which would include an audit of the new ACORN-SAT data set.

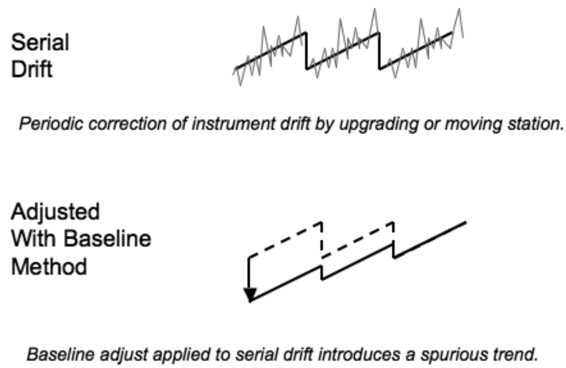

Figure 2. Example of spurious trend introduced by incorrect classification of error.

2.1. HQN Metadata

The BoM provided updated metadata (not available on the BoM website) on the adjustments to the nine series. The information provided on the timing of the adjustments, their size, and the method of identifying them were largely identical to Torok [14] and adjustments post 1996 but not after 2004, presumably from DM. The metadata did not contain sufficient information to reproduce or justify the adjustments as they omit: (1) justification for the size of the adjustments; (2) the threshold criteria for objective tests; (3) lists of neighbouring series used in the objective and median tests; and (4) the reference climatology.

The metadata provided by the BoM lists the tests that potentially justify adjustments. These adjustments are of two types: 'type 0' that applies to all previous years, and 'type 1' that applies only to one year. The justifications for adjustments are:

Overlap data (l): The most reliable adjustment method. Ideally, where a station has moved there would be at least two years, and even one year of parallel monthly observations if the data were good [21].

Detect (d): Statistical differences between candidate sites and two neighbours indicate spurious trends or discontinuity. This test is dependent on the existence of an overlap of a reasonable length, and local sites with similar climate.

Objective test (o): This method depends on the development of a reference climatology, requiring the selection of series at large distances from the candidate site and long temporal overlaps, and statistical detection of change points in the difference series (between candidate and reference climatology). Despite its name, this method can be unreliable, as Torok [15] concedes that the reference series' overall trend and variability can be biased by small differences in the reference station selection process.

Median test (f): The reference climatology is based on the median inter-annual temperature of all non-urban series within 6 degrees separation, or 8 degrees in remote areas [21], no more than 300m difference in altitude, and at least 10 years of overlap. In addition to the problems of contamination with trends, the median reference series can have biases introduced when converting the median interannual differences back into an absolute reference series by the accumulation of rounding errors in the interannual differences [25].

Range (r): Visual analysis of changes in the diurnal temperature range, the difference between minimum and maximum temperatures.

After detection with one or more of these methods, the baseline is shifted or one-year spike reduced (shown as types 0 and 1 in the metadata). As noted above, none of these methods provide definitive proof of an instrument error. Moreover, the removal of discontinuities does not detect or adjust for potential drift.

2.2. Analytical Approach

The MAN was developed as a minimal adjustment version of the HQN. The only adjustments applied in the MAN were those needed at some sites to join short records into continuous series, by combining records from nearby closed sites with those of the currently open sites (Table 1). Usually only two data records were combined, but up to four were used. The procedure was to visually examine annual data from each station where the records overlapped. If the data showed a difference between two records, one (usually the earlier) was adjusted to match the other. The adjustment was calculated from the mean difference in the data during the overlap (from two to 10 years).

This was done on the maximum and minimum series for each location. Mean series were calculated from the average of annual minima and maxima series. The available data yielded 268 series of annual temperatures from 1910 to 2009 – 134 from the HQN series downloaded from the BoM website, and 134 compiled from raw data, the Minimal Adjustment Network (MAN). These were further divided to two groups of 100 rural and 34 urban sites producing four groups: HQUrban, MANUrban, HQRural, MANRural.

Table 1. Adjustments to those sites that required splicing (using the overlap method) in the Minimal Adjustment Network. Other sites did not require splicing.

Station	Year	Baseline	Adjustments	Type
Halls Ck	1951	0.70	1	Overlap
Broome	1939	-0.27	1	Overlap
Geraldton	1941	-0.40	1	Overlap
Carnarvon	1948	-0.10	1	Overlap
Kalgoorlie	1952	-0.50	1	Overlap
Sthn Cross	2007	-0.50	1	Overlap
Wandering	1998	-0.51	1	Overlap
Albany	1964	-0.70	1	Overlap
Atherton	1992	0.30	1	Overlap
Ayr	1951	-0.06	1	Overlap
Bowen	1986	-0.90	1	Overlap
Rockhampton	1939	-1.00	1	Overlap
Gladstone	1956	0.82	1	Overlap
Charleville	1952	-0.39	1	Overlap
Roma	1993	-0.40	1	Overlap
Mildura	1945	-0.85	1	Overlap
Ararat Prison	1968	-0.30	1	Overlap
Cashmore			3	Overlap
Sale	1945	-0.30	1	Overlap
Strathalbyn	1996	-0.40	1	Overlap
C Borda	2002	-0.10	1	Overlap
Mudgee	1991	-0.90	1	Overlap
Cowra	1965	-0.54	1	Overlap
Richmond RAAF	1993	-0.24	2	Overlap
	1952	-0.30		
Wagga Wagga	1941	-0.75	1	Overlap
Low Head	1997	-0.26	1	Overlap
Launceston	1939	-1.20	1	Overlap
Darwin	1940	-0.40	1	Overlap
Kent Town	1997	-0.28	1	Overlap
Mackay			2	Overlap
Brisbane	1949	0.07	1	Overlap
Dalby	1991	-1.43	1	Overlap
Moree			3	Overlap
Inverell	1964	-0.26	2	Overlap
	1995	-1.16		
Grafton	1939	-1.75	1	Overlap
Bathurst	1942	-0.44	2	Overlap
	1964	0.44		
Orange	1944	-0.80	3	Overlap
	1945	-1.47	1	Overlap
Geelong			4	Overlap
Warnambool	1971	-0.49	4	Overlap
	1972			
	1984	-0.15		
	1998	0.15		

A temperature network Y can be described mathematically as the time evolution of a set of $k = 1, ..., K$ variables (i.e. temperature recording stations) over the sample period $t = 1, ..., T$. These values can be written as an array where the temperature at a station k in a specific year is an entry y_{kt} in a matrix.

$$Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1T} \\ \cdots & \cdots & \cdots & \cdots \\ y_{K1} & y_{K2} & \cdots & y_{KT} \end{bmatrix}$$

It must be remembered that different calculation methods can produce different average trends, particularly in the presence of missing values. Table 2 lists the results from simple alternatives, described as follows using pseudo-code.

Table 2. Comparison of trends in degrees C/Century from 1910 to 2009 of the Rural, Urban, HQN and MAN groups using different averaging methods.

	AvYrs	AvYrsZero	MedYrsZero	AvSeries	AvYrDiff
HQNUrban	0.80	0.79	0.76	0.77	0.77
MANUrban	0.40	0.40	0.43	0.39	0.43
HQNRural	0.87	0.91	0.91	0.92	0.90
MANRural	0.59	0.62	0.60	0.70	1.27

The mean (or average) temperature for Australia in any year, is a series written as $Mean_k Y$, calculated over each column in the network Y, indicated by the subscript k. The number of observations in the mean varies according to the number of missing values. The obvious way to calculate the average trend of Y is to calculate the linear trend on the series of annual means, written $Trend(Mean_k(Y))$ (method AvYrs). The problem with calculating the trend in this way is that a warming bias is introduced by the increase in the number of stations established in hotter, northern parts of Australia during the later part of the period (during and post WWII). One way to reduce this bias is to normalize each station before taking the mean, i.e. subtract the mean temperature for a station from each one of its values, then take the average and the trend (method AvYrsZero). Alternatively, taking the median instead of the mean temperature in each year of the normalized series may be more robust (method MedYrsZero). An alternative to normalization is to calculate the change in temperature from year to year for each station, calculate the mean across years, and then cumulatively sum the change in each year (method AvYrDiff).

$AvYrs = Trend(Mean_k(Y))$
$AvYrsZero = Trend(Mean_k(Norm_t(Y)))$

$MedYrsZero = Trend(Median_k(Norm_t(Y)))$
$AvYrDiff = Trend(Sum(Meank(Diff_t(Y))))$
$AvSeries = Mean(Trend_k(Y))$

Each of the above calculation methods produces a single, average temperature trend for Australia. A final method (method AvSeries) is to calculate an average trend for each station individually, and then average the individual station trends. The variation in results between the methods was larger than the standard error of any given method (0.1 vs 0.05 °C). One method, AvYrDiff, produced a uniquely higher trend for the MANRural group (Table 2). This method should be treated with caution. It could be argued that the AvSeries method of calculating the trend of a network – the average of the trends of each of the stations – is more parsimonious as it requires the fewest processing steps, and provides a standard error for comparing the difference between means.

Analysis of variance (ANOVA) is a basic statistical method in which the observed variance in trend is partitioned into components – here the adjustments are homogenization method (HQN vs. MAN) and urbanity (Urban vs. Rural). ANOVA provides a statistical test of whether or not the means of groups are all equal, similar to a t-test without the need for multiple tests and decreased chance of committing a type I error.

Tables 3 and 4 list the trends of each of the mean HQN series, the trends of each of the MAN mean series, and their differences. The mean series are calculated from the average of the station minimum and maximum series. From this table of differences, four rural series with the largest difference between HQN and MAN were selected for detailed examination (see Figure 3). Table 5 lists the characteristics of the four rural series: the trends of their MAN and HQN minimums and maximums, the number and type of adjustments, and the cumulative adjustments based on the BoM provided metadata.

Table 3. Trends of the urban series, ordered by difference between HQN and MAN networks.

	HQN	MAN	diff
Newcastle	-0.22	0.56	-0.78
Gunnedah.Pool	-0.16	0.56	-0.72
Sydney	0.76	1.33	-0.57
Hamilton.Aero	0.90	1.45	-0.54
Bairnsdale.Aero	0.44	0.86	-0.42
Warnumbool.Aero	0.28	0.61	-0.33
Pt.Macquarie	0.82	1.03	-0.21
Hobart	0.70	0.83	-0.13
Dalby.Aero	0.81	0.87	-0.05
Grafton.Pool	1.31	1.31	0.00
Mackay.MO	0.85	0.78	0.07
Innisfail	1.06	0.91	0.14
Perth.Aero	0.97	0.77	0.20
Moree.Aero	-0.04	-0.31	0.28
Bowral	0.61	0.32	0.28
Inverell	0.72	0.43	0.29
Broken.Hill	0.98	0.65	0.33
Geelong.Aero	0.57	0.24	0.33
Tamworth.Aero	0.05	-0.31	0.35
Brisbane.Aero	0.43	-0.10	0.53
Orange.Aero	0.68	0.14	0.54
Bathurst.Ag.Stn	0.63	0.08	0.55
Melbourne	0.70	0.08	0.62
Lismore	0.80	0.15	0.65
Maryborough.Vic	0.91	0.23	0.68
Pt.Lincoln	1.02	0.33	0.69
Kent.Town	1.67	0.81	0.86
Ballarat.Aero	1.14	0.15	0.99
Darwin	1.36	0.25	1.11
Maryborough.Qld	1.46	0.34	1.12
Dubbo.Aero	0.93	-0.37	1.30
Benalla	0.72	-0.72	1.44
Echuca.Aero	1.20	-0.30	1.50
Wangaratta.Aero	1.05	-0.66	1.71

Table 4. Trends of the rural series, ordered by difference between HQN and MAN networks.

	HQN	MAN	diff	Sites	HQN	MAN	diff
Barcaldine	0.85	1.54	-0.69	Wilsons.Prom	0.98	0.72	0.26
Derby	0.13	0.65	-0.52	Gladstone	0.99	0.72	0.27
Pt.Perpendicular	0.15	0.63	-0.48	Meekatharra	1.28	0.99	0.29
Tennant.Ck	0.57	1.04	-0.46	Cape.Borda	1.28	0.99	0.29
Ceduna	0.75	1.17	-0.42	Cairns.Aero	0.94	0.64	0.30
Hughenden.Aero	0.85	1.24	-0.39	Cape.Naturaliste	1.48	1.16	0.33
Atherton	0.28	0.64	-0.35	Cape.Moreton	1.19	0.83	0.36
Roebourne	1.07	1.42	-0.35	Walgett.Aero	0.91	0.53	0.38
Stathalbyn	1.01	1.33	-0.32	Carnarvon	1.13	0.75	0.38
Jarrah	0.67	0.98	-0.31	Broome	0.61	0.21	0.40
Cardwell	1.35	1.66	-0.31	Cape.Leeuwin	0.96	0.56	0.40
Mudgee.Aero	0.17	0.45	-0.28	Wandering	1.20	0.78	0.42
Robe	0.39	0.67	-0.28	Bridgetown	1.27	0.85	0.42
Cape.Northumberland	0.93	1.20	-0.27	Cotway	0.98	0.54	0.44
Richmond.RAAF.NSW	0.74	1.00	-0.26	Halls.Ck	1.34	0.89	0.44
Low.Hd	0.96	1.21	-0.25	Charters.Twrs	1.26	0.79	0.46
Tibooburra.PO	0.87	1.11	-0.24	Swan.Hill	0.84	0.37	0.47
Cape.Bruny	1.28	1.49	-0.21	Ayr	0.75	0.27	0.48
Normanton.Aero	0.59	0.80	-0.21	Jerrys.Plains.PO	0.55	0.06	0.49
Gabo.Is	0.53	0.74	-0.21	Katanning	0.82	0.32	0.50
Georgetown.PO	0.83	1.03	-0.20	Southern.Cross	0.82	0.32	0.50
Glen.Innes.Aero	0.55	0.70	-0.15	Bourke.Aero	0.77	0.25	0.51
Pt.Hedland	0.82	0.96	-0.14	Rottnest.Is	1.20	0.68	0.52
Gayndah	1.23	1.36	-0.13	East.Sale.Aero	0.79	0.27	0.53
Rockhampton	1.27	1.38	-0.10	Miles	1.08	0.53	0.54
Meredin	1.09	1.18	-0.08	Newman	1.03	0.49	0.55
Mt.Gambier	1.45	1.54	-0.08	Boulia.Aero	1.63	1.08	0.55
Tarcoola	1.66	1.73	-0.06	Yamba.Pilot.Stn	1.23	0.68	0.56
Amberley	0.09	0.15	-0.06	Longreach	1.31	0.75	0.56
Hay	0.82	0.88	-0.06	Kerang	1.14	0.58	0.56
Woomera	2.15	2.20	-0.04	Goondiwindi	0.90	0.31	0.59
Geraldton	0.66	0.70	-0.04	Ararat.Prison	0.62	0.02	0.60
Bowen.Aero	0.85	0.88	-0.03	Cashmore	0.68	0.03	0.65
Mildura	0.80	0.82	-0.03	York	0.80	0.14	0.65
Moruya.Heads	0.49	0.50	-0.01	Kalgoorlie	1.00	0.33	0.67
Camooweal	1.12	1.12	-0.00	Strahan.Aero	0.47	-0.23	0.71
Cunnamulla	0.93	0.93	0.00	Yongala	1.50	0.78	0.72
Cobar	0.59	0.59	0.00	Sandy.Cape	1.19	0.46	0.73
King.Is	0.65	0.63	0.02	Marble.Bar	0.94	0.20	0.74
Laverton	0.79	0.77	0.02	Launceston.Aero	0.85	0.05	0.80
Charleville	0.59	0.54	0.04	Alice.Springs	1.70	0.89	0.81
Giles	1.15	1.11	0.04	Roma	1.50	0.68	0.82
Richmond.Qld	1.02	0.97	0.05	Rutherglen	0.82	0.00	0.82
Palmerville	0.72	0.65	0.07	Cowra.Aero	1.31	0.47	0.85
Esperence	0.64	0.56	0.08	Kellerberrin	1.00	0.10	0.90
Rayville	0.76	0.64	0.13	Cooktown	0.95	0.05	0.90
Wilcannia	0.63	0.45	0.19	Nhill	0.75	-0.19	0.93
Bollon	0.90	0.68	0.22	Wagga.Wagga	1.32	0.38	0.94
Albany	0.72	0.50	0.22	Omeo	1.03	0.02	1.01
Tenterfield	0.41	0.15	0.26	Deniliquin.PO	0.65	-0.57	1.21

Table 5. Number of adjustments of each type (baseline or spike), trends of the MAN and HQN series, and the stated adjustment in the BoM metadata in degrees C for the maximum and minimum temperature series at four High Quality stations

	No	Type 0	Type 1	MAN	HQN	Cum Adj
Omeo Max	83025	8	1	-0.53	0.61	1.10
Omeo Min	83025	6	4	0.38	1.41	0.80
Nhill Max	78031	2	3	0.55	0.84	0.10
Nhill Min	78031	4	4	-1.00	0.73	2.00
Deniliquin Max	74128	4	15	0.11	0.39	0.40
Deniliquin Min	74128	4	6	-0.64	1.64	1.60
Wagga Wagga AMO Max	72150	3	0	0.11	0.39	0.30
Wagga Wagga AMO Min	72150	2	2	-0.71	0.88	1.00

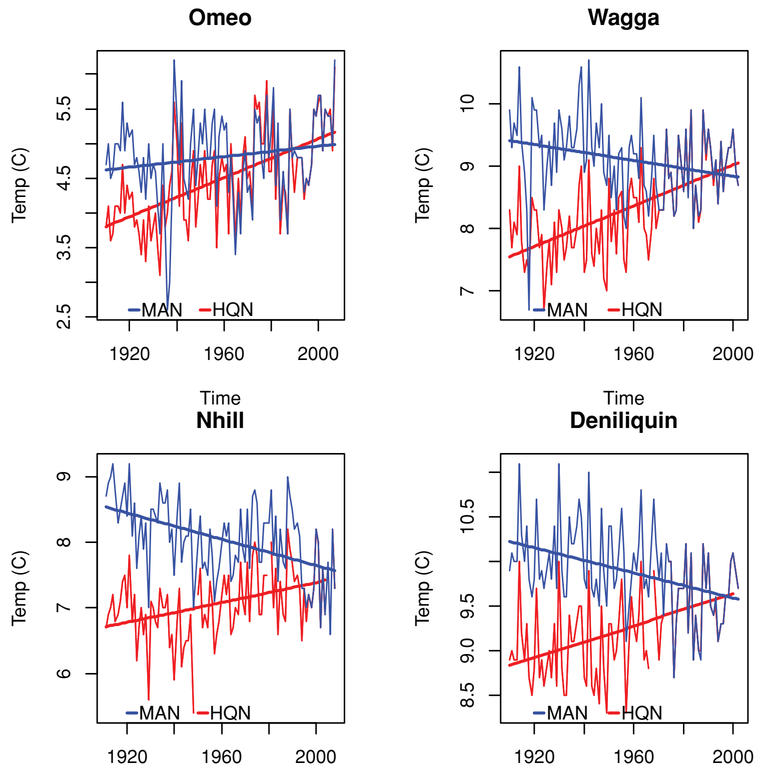

Figure 3. Examples of large warming adjustment to trends at four locations: MAN (blue) and adjusted HQN (red) minimum temperature series.

Figure 4 illustrates the MAN series for these four rural series (blue) in relation to their neighboring stations (grey), and the adjustments in the HQN series (in red). The regional climatology is defined as the 1 degree Celsius range either side of the average trend of the neighboring sites, adjusted to a common mean, representing the two-sigma range, and illustrated as a grey band on Figure 4.

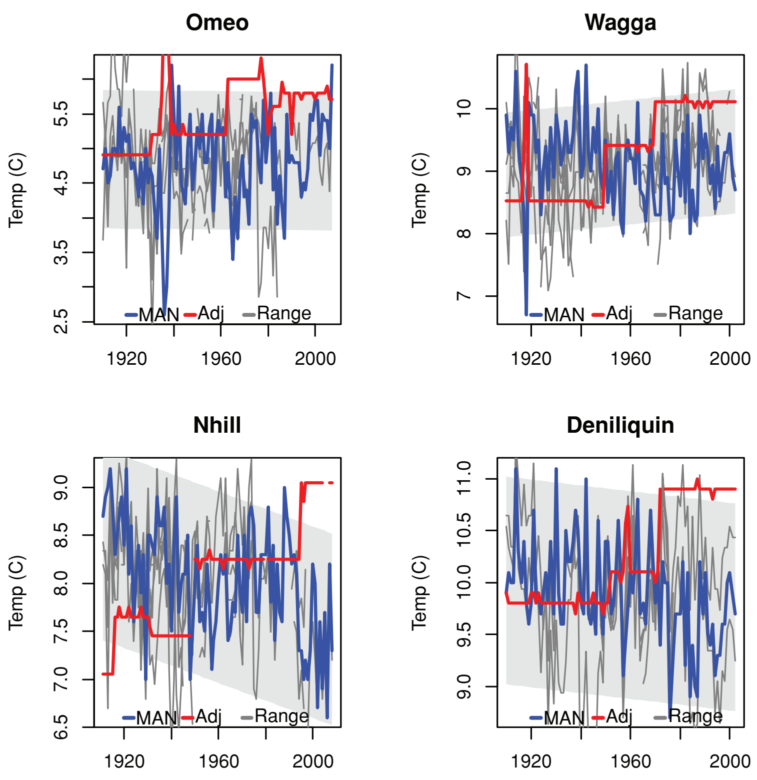

Figure 4. The MAN series (blue) relative to the range of variation in nearby stations (grey), and the adjustments in the HQN series (in red).

3. RESULTS

Table 1 lists 27 non-urban and 12 urban sites that required at least one adjustment (the others required no adjustments). Of the 55 adjustments required to construct the MAN series, 31 were for a single adjustment for the site, with the remainder having multiple adjustments, an average of 0.4 adjustments of -0.5 °C per series per 100 years.

Of the nine HQN series for which we obtained metadata from the BoM, the metadata indicated 6 adjustments of -0.2C, producing an average downward adjustment of past temperatures (warming adjustment) of -1.2 °C. As the minimal adjustments performed in the MAN increased the warming trend, there may be some basis for claiming that some downward adjustment of past temperatures is necessary to create a contiguous record.

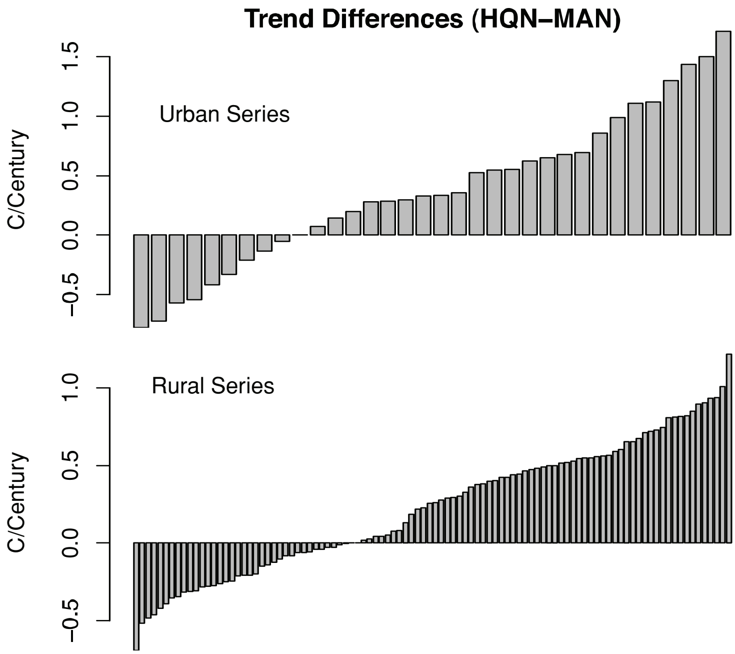

Figure 5. Adjustments to both urban (upper) and rural (lower) series are predominantly warming.

The majority of HQN trends exceeded the corresponding MAN trends in both the urban and rural series (Figure 5). The largest increases in trend were 1.7 and 1.2 °C/Century to urban and rural sites respectively, while the largest decreases in trend were -0.8 and -0.7 °C/Century respectively (Tables 3 and 4).

The effects of the Rural/Urban and HQN/MAN factors on trends of the stations were highly significant, as shown by the ANOVA analysis (Table 6). The HQN/MAN factor was the most significant ($P<0.00005$), while the Rural/Urban factor was also significant ($P<0.0002$). The interaction term was not significant.

The average of the HQNRural trends was 0.92 ± 0.01°C/Century (based on the AvSeries method, with uncertainty quoted as the 95% CL of the mean). The average of the MANRural trends was 0.70±0.01°C/Century. By comparison, the global trend over the same period estimated by the Climate Research Unit at the Hadley Center (using the file hadcrut3vgl.txt referred to as HadCRU) is 0.75 ± 0.05°C/Century (Figure 6) over the same period. The 31% difference in average trend between the HQNRural and the MANRural is highly significant ($P \approx 0$) by a difference of means test. The average trend of the HQNRural group is also significantly greater than the global temperature trend (P=0.0003), but the trends of the MANRural and HadCRU do not differ significantly (P=0.16).

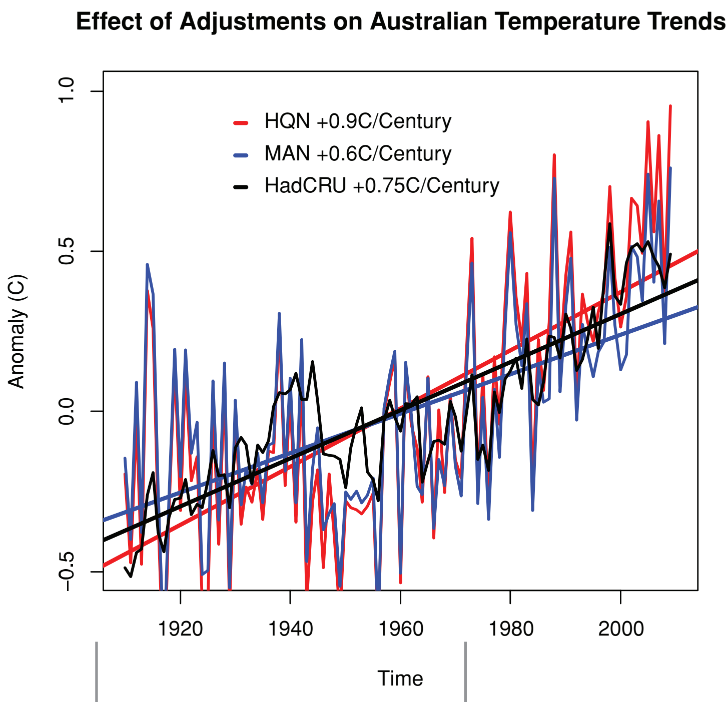

Figure 6. Comparison of average temperatures and their trends since 1910: the rural Australian High Quality Network (HQNRural), the rural Minimal Adjustment Network (MANRural) and the global temperature series from the Hadley Center Climate Research Unit (hadcrut3vgl).

On Figure 6 is shown the average trend of the HQNRural, MANRural series (based on the AvYrsZero method) and their trends, relative to the global average temperature series HadCRU. Here, the trend of the MANRural series (≈ 0.6°C/Century) is even lower than the HadCRU global trend, and similar to the AvYrs and MedYrsZero (refer to Table 2 for results of all methods).

The average trend of the rural group is greater than the average trend of the urban group in both HQN and MAN series. In Table 2, the highest MAN urban trend is 0.43, and the lowest MAN rural trend is 0.59. MAN rural trends are greater whichever method is used, and the effects of adjustments on HQN urban series are also much greater (approaching 100 %).

The rural series with the largest discrepancies between MAN and HQN data were Omeo, Deniliquin Post Office, Nhill, Wagga Wagga AMO, and Kellerberrin. The discrepant urban series were Wangaratta Aero, Echuca Aerodrome, Benalla Shadford St, and Dubbo Airport AWS. Table 6 lists the number and magnitude of adjustments to the MAN and HQN trends at the four most adjusted rural series. Both the minima and maxima of HQN at Omeo exceed the MAN, while the minima of Deniliquin, Nhill, Wagga Wagga, and Kellerberrin are the main cause of the deviation (Figure 3).

Figure 4 shows the temperature series of the four series above, relative to the nearby stations in grey, the two-sigma range of their common trend. The MAN series for these sites (in blue) and the adjustments applied to the HQN series (in red) are also shown.

Ninety-nine adjustments were made to the nine disparate series, with an average change to the baseline of -0.17 per adjustment with a standard error of 0.08. Therefore there is a significant tendency for downward adjustment of past years ($P<0.05$). Detailed descriptions of the adjustments to the four illustrated series follow.

Table 6. Anova analysis from 1910 to 2009 of the UHI (Urban/Rural) and Adjustments (MAN/HQ)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
UHI	1	2.69	2.69	14.21	0.0002
Adjust	1	4.64	4.64	24.54	0.0000
UHI:Adjust	1	0.29	0.29	1.54	0.2163
Residuals	264	49.93	0.19		

3.1. Omeo

Omeo is a small settlement high in the NE ranges of the state of Victoria. It is very difficult to determine which series should be used for comparison, as many nearby stations are over the ranges, on the coast, much higher up mountains, or far to the north (over more mountainous country) and thus disqualified by similar climate criteria [15]. The comparison series of Orbost (near the coast) and Beechworth (inland Victoria) have records which overlap in the second half of the Century, and only Hotham Heights and Kosciusko Hotel are mountainous sites, with records in the early part of the Century. Extrapolation of data from distant sites is risky, even if the correlation coefficients between sites is high, which may be more due to a common trend and not a similar local climatology.

The elevated data for 1936-38 is obvious and may be justified by the move from “newspaper to PO” in 1935, but there is no corresponding metadata justifying the adjustment at the end of the spike around 1938. Other than this period, Omeo’s anomalies track its neighbours closely: slightly high relative to Orbost and Beechworth in the 1910s, and low from the mid 60s to mid 70s. Similarly, the Kosciusko record is close to Omeo in the late 1920s, while Hotham Heights is completely different.

The adjustment of -0.8 °C in 1962 was justified by the BoM with a move noted in the metadata as “Move and new screen as old one had an iron roof since 1910”. However, nearby stations also fell sharply, and so the adjustment may be exaggerated. Like Omeo, the overall trend of the neighbouring series is flat throughout. However, cumulative adjustment of Omeo by -1.6 °C produces a significant warming trend.

The adjustment documentation indicates a station move in 1930, but there is no move documented in the metadata, contradicting the rule that adjustments must be verified by metadata [21]. Omeo is an example of a station with no overlapping records from stations with a similar climate, and neither the metadata nor neighbouring stations justify the large individual and cumulative adjustments.

3.2. Wagga Wagga

Wagga Wagga is a city in the state of New South Wales, on the Murrumbidgee River. The Wagga Wagga AMO HQN station is a compilation of the records for Kooringal and AMO with the good overlap in the 1940s. In our compilation of the raw series, overlap in the minima record indicates reducing Kooringal by 0.1C, the average difference for the first 5 years of overlap. However, the HQN record adjusts Kooringal down by 1.0C. This may be due to inclusion of longer time frame of neighbouring sites pre-1940 that have temperatures falling in the lower half of the band. Overlap is however a superior detection and adjustment method and should take precedence.

There is an obvious spike in 1917 and 1918. Comparison with overlapping nearby stations, Adelong PO, Deniliquin, and Cootamundra, and the Resource Centre shows Wagga is high in the first half of the record and low in the second half but within the variation in the neighbours. The trend of neighbouring sites is slightly warming, however, the cumulative adjustments of the HQN series produce a strongly warming trend.

These adjustments of -1.0 °C in 1949 and -0.7 °C in 1969 appear to be unnecessarily large and are not justified by metadata, annotated only with an asterisk indicating “documentation unclear”.

3.3. Nhill

Nhill is a town in the Wimmera, in western Victoria, halfway between Adelaide and Melbourne. Large adjustments were made to the HQN minimum record for recorded moves in 1931, 1949, and 1995, but an adjustment was made with no supporting metadata in 1915.

There are no overlaps that would have justified adjustments for these moves. In fact, there are gaps. In December 1994, the observer died, the station closed, and there were no observations from 17th December until 17th January after moving 5 km. Normally BOM does not give a monthly mean if there are more than 7 days missing, and does not give an annual mean for any year with a missing month. Therefore, 1994 and 1995 should not have annual means, and there is a gap of two months with no overlaps at all.

Due to this gap, the site does not qualify for the HQN network [21]. The closest sites with overlap – Rainbow PO, and Jeparit, Stawell, Horsham, and Ararat Prison – show Nhill was low from 1942-45, high from 1947-54. The period 1992-94 needs adjusting, but the neighbouring records around 1989-92 are close. When the adjustments are applied, the HQN record is strongly warming, reversing the cooling of the neighbouring stations.

3.4. Deniliquin

Deniliquin is a town in the Riverina region of New South Wales close to the border with Victoria. The minimum was adjusted in 1971 by -0.8. However, the neighbours of Deniliquin – the urban site Echuca (included due to long record), Kerang and Falkiner Memorial – are comparable, with only Kerang being too low from the 1920s to the 1960s. The neighbouring anomalies show some discrepancy in 1971, large ones in 1972, and to a lesser extent 1973. However, the years before this, except for 1960-61, do not appear to have any marked discrepancies to justify the 1971 adjustment. Moreover, there does not appear to be anything unusual about the diurnal range around 1971.

Like Deniliquin, the neighbouring stations are cooling. The Deniliquin adjustments produce strong warming that does not appear to be justified.

4. DISCUSSION

A new compilation of Australian temperature series (MAN) based on minimal adjustments to achieve contiguous series was compared with the heavily adjusted Australian HQN series. The average of trends was 0.92 ± 0.01°C/Century and 0.70 ± 0.01°C/Century for the rural groups, HQNRural and the MANRural respectively, a significant 31% reduction in estimated warming over the period 1910-2009. ANOVA analysis shows the adjustment factor HQN/MAN is the main source of trend variability. Moreover, as the MANRural trend does not differ significantly from the global mean trend for HadCRU of 0.75±0.05°C/Century, it may be concluded that unjustified homogeneity adjustments are the cause of the inconsistency between global and Australian temperature trends to within statistical error.

It is not clear why the homogenization adjustments should produce a warming bias. It is suggested the bias may be due to unrecognized, falsely rectified drift from UHI. It is not clear, a priori, that the strategy of shifting the baseline is always the correct remedy for a discontinuity. In the case of multiple station moves with serial drift, adjusting the baseline would introduce a strong warming bias. As it is generally assumed that errors are recognizable by discontinuities in the data, correction of drift, which cannot be discriminated by discontinuities alone, could produce the bias.

Rural vs urban differences are also a significant source of bias. Table 2 shows the urban trend is less than the rural trend in all comparisons, and adjustments made by the BoM have been much greater (in AvSeries, by 0.38, or 97%). This may be due to higher occurrences of baseline shifts in the urban data set, or due to comparing candidate series with reference series of mostly rural sites. Therefore, exclusion of the lower trend urban sites from the overall Australian average may also exaggerate warming.

There may be no valid reason for excluding urban sites. It may be that UHI contamination occurred early in the growth of towns in the urban series, thus leading to a shallower warming. However, the neighborhood of the urban sites also shows less warming, and it is not possible to identify when UHI contamination occurred in large or small towns, except by overlap comparison, as the metadata is so poor. Possible UHI contamination is just as likely in small towns as in towns over 10,000 population. As the effects of UHI appear to be subtle, unclear, difficult to diagnose and possibly more related to site-specific problems, there should be no reason to exclude urban sites.

The traditional approach of comparing a candidate site with a regional climatology, while superficially appealing, is fraught with problems. Firstly, in the case of the HQN, too many adjustments have been applied, many of which may be to natural climate variations. Promiscuous adjustments encourage the introduction of bias. Secondly, the size of the neighborhood in a regional climatology is arbitrary, and in the HQN case extends to sites 890 km away. While it is claimed that such large neighborhoods are justified by correlation over that distance, the correlation may be partially due to the continental warming trend. In that case, regions that depart from the overall warming trend are biased upwards. Thirdly, adjustments are only justified by purely statistical considerations, as in most cases the metadata are inadequate to provide much of a warrant. Finally, differences between averaging methods shown in Table 2 are an additional source of uncertainty, and so any regional climatology must be used with caution.

It is apparent that a good overlap of monthly data for 2 years or more provides a reliable estimate of the offset between stations, and there is no real substitute for overlapping data to make a reliable splice. The methods using regional climatology, and the use of distant sites up to 70km away to infill missing data by interpolation are too risky to use. The development of a regional climatology is not essential for detecting discontinuities, as broken linear regression models may be a robust alternative approach [26].

Are the trends derived from the Australian HQN of surface temperatures reliable? There are several reasons why not. The claim that the HQN dataset is suitable for national analysis stands or falls on the magnitude of bias. Robust evidence of significant bias from homogeneity adjustments and exclusion of urban sites contradicts the claims of reliability of trends or regional averages.

Individual sites in the HQN are even less reliable. Close examination of several sites shows that adjustments to the spikes in the raw data appear obvious and needed, but there appears to be little justification for most of the large baseline shifts. The series examined were largely within the range of regional variations and shared similar flat or declining trends. Most of the series in the HQN showed similar sorts of problems with large individual and cumulative adjustments and barely adequate supporting metadata. The few sites providing definitive compilations are not distinguished from the poor sites.

5. ACKNOWLEDGEMENTS

Reviews of drafts by Anthony Cox and Geoff Sherrington, the comments of peer reviewers, and the past and continuing efforts of a number of auditors of the Australian surface temperature records are gratefully acknowledged. Data and code for analysis in R are available from the author (go to http://landshape.org/enm/datacode/).

REFERENCES

1. Khandekar M. Has the IPCC exaggerated adverse impact of global warming on human societies? *Energy & Environment,* 19[5]:713–719, 2008.
2. Stockwell D.R.B. and Peterson A.T. Effects of sample size on accuracy of species distribution models. *Ecological Modelling*, 148[1], 2002.
3. Thomas C.D., Cameron A., Green R.E., Bakkenes M., Beaumont L.J., Collingham Y.C., Erasmus B.F., De Siqueira M.F., Grainger A., Hannah L., Hughes L., Huntley B., Van Jaarsveld A.S., Midgley G.F., Miles L., Ortega-Huerta M.A., Peterson A.T., Phillips O.L., and Williams S.E. Extinction risk from climate change. *Nature*, 427[6970], 2004.
4. Stockwell D.R. Guest Editorial: Biased Towards Extinction. *CO2 Science*, 7[19], 2004.
5. McIntyre S. and McKitrick R. Hockey sticks, principal components, and spurious significance. *Geophysical Research Lett*ers, 32[3], 2005.
6. Rahmstorf S., Cazenave A., Church J.A., Hansen J.E., Keeling R.F., Parker D.E., and Somerville R.C.J. Recent climate observations compared to projections. *Science* (New York, N.Y.), 316[5825]:709, May 2007.
7. Stockwell D.R.B. Recent climate observations: disagreement with projections. *Energy & Environment*, 20[4]:2, 2009.
8. Stockwell D. Critique of Drought Models in the Australian Drought Exceptional Circumstances Report (DECR). *Energy & Environment*, 21[5]:425–436, 2010.
9. McCullough B. and McKitrick R. *Check the Numbers: The Case for Due Diligence in Policy* Formation. 2009. Fraser Institute. ISBN:1918-8323, Retrieved from http://www.fraserinstitute.org/researchandpublications/publications/6511.aspx
10. Trewin B. Effect of changes in algorithms used for the calculation of Australian mean temperature. *Australian Meteorological Magazine*, 53:1–11, 2004.
11. Fall S., Watts A., Nielsen-Gammon J., Jones E., Niyogi D., Christy J.R., and Pielke R.A. Analysis of the impacts of station exposure on the U.S. Historical Climatology Network temperatures and temperature trends. *Journal of Geophysical Research*, 116[D14]:D14120, Jul. 2011.
12. NIWA. *Report on the Review of NIWA's 'Seven Station' Temperature Series*. Tech. rep., National Institute of Water and Atmospheric Research, 2010.
13. NZCSC. *Statistical Audit of the NIWA 7-Station Review*. Tech. rep., The New Zealand Climate Science Coalition, 2011.
14. Torok S. *The development of a high quality historical temperature database for Australia*. Ph.D. thesis, University of Melbourne, 1996.
15. Torok S. and Nicholls N. A historical high quality temperature database for Australia. *Australian Meteorological Magazine*, 45:251–260, 1996.
16. Easterling D.R. and Peterson T.C. A new method for detecting undocumented discontinuities in climatological time series. *International Journal of Climatology*, 15[4]:369–377, Apr. 1995.

17. Ducr-Robitaille J.F., Vincent L.A., and Boulet G. Comparison of techniques for detection of discontinuities in temperature series. *International Journal of Climatology*, 23[9]:1087–1101, Jul. 2003.

18. DeGaetano A.T. Attributes of Several Methods for Detecting Discontinuities in Mean Temperature Series. *Journal of Climate*, 19[5]:838–853, Mar. 2006.

19. Menne M.J. and Williams C.N. Homogenization of Temperature Series via Pairwise Comparisons. *Journal of Climate*, 22[7]:1700–1717, Apr. 2009.

20. Australian Bureau of Meteorology. Australia's high-quality climate change datasets, 2011 (http://www.bom.gov.au/climate/change/datasets/datasets.shtml).

21. Della-Marta P., Collins D., and Braganza K. Updating Australia's high quality annual temperature dataset. Australian Meteorological Magazine. Australian Meteorological Magazine, 53:75–93, 2004.

22. Torok S., Morris C., Skinner C., and Plummer N. Urban heat island features of southeast Australian Towns. *Australian Meteorological Magazine*, 50:1–13, 2001.

23. Rhoades D.A. and Salinger M.J. Adjustment of temperature and rainfall records for site changes. *International Journal of Climatology*, 13[8]:899–913, Dec. 1993.

24. Founda D., Kambezidis H.D., Petrakis M., Zanis P., and Zerefos C. A correction of the recent air-temperature record at the historical meteorological station of the National Observatory of Athens (NOA) due to instrument change. *Theoretical and Applied Climatology*, 97[3-4]:385–389, Dec. 2008.

25. Trewin B.C. *Extreme Temperature Events in Australia*. Ph.D. thesis, University of Melbourne, 2001.

26. Zeileis A., Kleiber C., Kramer W., and Hornik K. Testing and dating of structural changes in practice, *Computational Statistics & Data Analysis*. 44[1-2]:109-123, Oct. 2003.

27. Lund R. and Reeves J. Detection of Undocumented Changepoints: A Revision of the Two-Phase Regression Model. *Journal of Climate*, 15[17]:2547–2554, Sep. 2002.

28. Stewart K. Data and Metadata for High Quality Stations http://landshape.org/enm/datacode/), 2010.