Tennant Creek NT [BoM ID 015135 (airport), 015087 (post office); joined 1 November 1970]
Temperature data from February 1910 to December 2022.

Dr Bill Johnston
www.bomwatch.com.au

Summary

Concise rules-based protocols underpin the scientific method. Using mean annual maximum temperature (Tmax) for Tennant Creek as the case study, this report shows how BomWatch protocols unbiasedly estimate trend and change.

Rainfall alone explained 41.6% of Tmax variation (44.7% if apparent outliers in 1915, 1928, 1930 and 1950 were ignored). Residuals from that relationship were, however not homogeneous. Ignoring outliers, step-changes in rescaled residuals in 1935, 1962, 1990 and 2013 affected trend. While rainfall responses were the same, Tmax ~ rainfall relationships for segments from 1910-1934 & 1962-1989, and 1935-1951 & 1990-2012 were coincident (relationships for respective segments were overlaid). Combining coincident segments and including data for 2013-2022, final-round analysis found step-changes related to site changes caused mean Tmax to increase 1.51°C between 1910 and 2022 independently of rainfall, and that site-changes and rainfall simultaneously explained 73.8% of Tmax variation. *Post* hoc analysis found no residual trend or change was attributable to any other factor including CO_2, coalmining, electricity generation or anything else.

Commencing in 2012, five rounds of homogenisation conducted by the Bureau of Meteorology cooled pre-1935 and pre-1963 Tmax to varying extents and achieved trends approaching 0.15°C/decade. Australian Climate Observations Reference Network – Surface Air Temperature V.2.4 also adjusted data higher from 1999 to 2015 by 0.52°C thereby smoothing the effect of post-2013 site changes. However, while for the Tmax ~ year relationship, significances and R^2_{adj} improved, variation explained by rainfall declined from 46.3% initially, to 16.5% for AcV.2.4. Thus, realignment of the tails of data distributions so they were homogeneous for extremes occurred at the expense of the resulting homogenised series reflecting the weather. Furthermore, mean Tmax for data segments defined by homogenisation changepoints were mostly the same or they overlapped, indicating they were either poorly specified, or data were affected by outliers or other problems. Step-changes embedded in Tmax ~ rainfall residuals indicated homogenised data were not homogeneous.

Conclusions

BomWatch protocols comprise four elements, namely:

- The overall relationship between Tmax and rainfall partitions total variation into that due to rainfall, and the residual non-rainfall part. Linear regression also derives the overall Tmax/rainfall coefficient, and significance (*P*) and goodness of fit (R^2_{adj}) statistics that indicate conformity with the First Law of Thermodynamics.
- Homogeneity analysis of rescaled residuals identifies non-climate impacts on data, which are categorised as step-change or (Sh)ift scenarios
- Segment-by-segment analysis with rainfall detects outliers, lack of fit, and other potential problems, and,
- Categorical multiple linear regression (and interaction analysis) finalises and verifies outcomes.

Segmented trend and graphical analysis confirm and verify that relationships are linear, residuals are normally distributed, independent, with constant variance, and that they are timewise homogeneous.

Based on the First Law Theorem that maximum temperature depends on rainfall, BomWatch protocols provide an unequivocable basis for understanding the effect of non-climate impacts on data, and for objectively assessing the BoM's homogenisation methods.

1. Introduction

Located on the Stuart Highway 1753 km north of Adelaide, 451 km north of Alice Springs and 872 km south of Darwin, Tennant Creek was originally the location of the 1872 repeater station on the overland telegraph between Adelaide and Darwin. A gold-rush town established 16 km south of the repeater station in about 1932 and a post office opened there in 1935. Although now a tourist attraction, the repeater station eventually closed and fell into disrepair.

Bureau of Meteorology (BoM) temperature datasets for Tennant Creek comprise daily observations for the repeater station combined after 1935 with data for the town post office, and also data for the airport commencing from when the new meteorological office opened in 1969. An Aeradio station was established at the aerodrome in 1939 to service the air-corridor between Adelaide and Darwin (callsign VZTC) and while daily temperature data for that site were apparently digitised from 1957, they are currently unavailable.

According to the Australian Climate Observations Reference Network – Surface Air Temperature (ACORN-SAT) Catalogue:

The current site (015135) is at the Meteorological Office at the airport 2 km northwest of the town centre. The instrument enclosure is largely bare red soil with short unwatered grass around. The Meteorological Office buildings are about 40 m to the south (Figure 1).

Figure 1. The weather station at Tennant Creek photographed on 17 January 2012 (BoM photograph from the ACORN-SAT catalogue). The louvered box or Stevenson screen (right of centre) houses thermometers and platinum resistance temperature probes installed in 1990 and used in their place by the automatic weather station (AWS) since 1996, at the standard height of 1.2m above the ground. The screen appears to be an older 230-litre type. Metadata indicates that manual thermometers and the galvanised A-pan evaporimeter on the left of the picture were removed on 1 June 2017.

History

The original site (015087) was based in the town. There was a major move (12 km south) in June 1935 and a smaller move of about 20 m in November 1963. This later site was quite close to a road. New screens were installed to replace damaged ones in May 1945 and November 1958. Observations ceased in 1970.

This airport site has been operating since 1969. An automatic weather station was installed on 8 July 1990 and has been the primary instrument since 1 November 1996. There are no known substantial moves at this location.

Data for a select number of weather stations homogenised by BoM scientists, most recently Blair Trewin are used by CSIRO to monitor climate warming in Australia. Although Tennant Creek was not used in the original 224-station dataset of Torok and Nicholls (1996)[1], it was included in the 133-station High Quality (HQ) dataset updated by Della-Marta et al. (2004)[2]. HQ homogenisation was updated by Blair Trewin and re-named ACORN-SAT v.1 (AcV1) in 2011. Trewin's 112-station AcV1 dataset was updated incrementally until 2017. A major revision of ACORN-SAT was undertaken in 2018 (AcV2.1) with the most recent iteration (to 2023) being AcV2.4. Due its remoteness, data for Tennant Creek is weighted as representing 0.05313 or about 5.3% of Australia's land area.

---

[1] Torok, S.J. and Nicholls, N. 1996. A historical annual temperature dataset for Australia. Australian Meteorological Magazine, 45, 251-260.

[2] Della-Marta, P., Collins, D. and Braganza, K. 2004. Updating Australia's high-quality annual temperature dataset. *Australian Meteorological Magazine*, 53 75-93.

Maximum temperature time-series potentially comprise the signal of interest (Tmax) and parallel non-climate impacts that affect observations. **Non-climate effects are called inhomogeneities, while homogenisation is the process of adjusting for their effect.** Provided methods are physically based and transparent, homogenisation is a necessary prerequisite to deriving unbiased trends. The question is not whether homogenisation is scientifically valid, rather, whether the methods used are scientifically sound, objective and replicable.

ACORN-SAT homogenisation is a complex process that uses metadata (information about the dataset) and *reference series* derived from up to 20 comparator sites, to detect and adjust inhomogeneities in target site data[1]. Tennant Creek *post office* Tmax was used with other sites to homogenise ACORN-SAT data for Alice Springs airport, Boulia airport, Halls Creek airport, Longreach aero, and Tennant Creek airport (AcV1 and AcV2). Data for the *airport* was used to homogenise Boulia airport and Camooweal town (AcV1) and in addition, Halls Creek, Victoria River Downs and Rabbitt Flat (AcV2). As non-ACORN-SAT sites are also used to homogenise Tennant Creek and other aforementioned ACORN-SAT sites, station by station comparisons are not straightforward. Furthermore, stations used to homogenise Tmax are not necessarily used to homogenise Tmin.

Selecting comparator datasets based on Pearsons linear correlation of first-differences with those of the target, potentially undermines objectivity of the process. Also, detection and correction of inhomogeneities using the same reference series is arguably a form of *P-hacking*, which is seeking statistically significant trends in data where none exist (e.g., https://journals.plos.org /plosbiology/article?id=10.1371 /journal.pbio.1002106&). Lack of evaluation of homogenised datasets *post hoc* is also problematic, particularly given the critical role played by ACORN-SAT in underpinning government policy.

As Trewin's methods are obscure and not readily deconstructed, methodical biases are best explored using objective statistical methods and/or by assessing the quality of datasets used in the process. Should data for Tennant Creek be of low quality and not homogeneous, they are of questionable use for determining trend and change in the climate, or for homogenising data for other ACORN-SAT sites. The same applies to comparator-site datasets used to homogenise Tennant Creek. Quality assurance of the data and the process is therefore paramount.

The aim of this report is twofold. Using maximum temperature (Tmax) observed at Tennant Creek as a case study, the overarching aim is to outline step-by-step, protocols developed by BomWatch for analysing trend and change that are consistent with the scientific method. Secondly, to apply the same methods to evaluating homogenised versions of the same dataset.

2. Is Tmax data for Tennant Creek any good?

Naïve analysis using spreadsheet applications such as Excel cannot separate climatic trends and change from those attributable to site and instrument changes. There is therefore a need for a rigorous protocol that can specifically uncover non-climate impacts on data. While metadata is used routinely to identify when site conditions might have changed[2], multiple studies have shown that not all documented site changes are influential, and that not all influential changes are documented. Misleading metadata provides considerable scope to manipulate the homogenisation process to derive pre-determined outcomes. Homogrenisation of Townsville, Queensland is an example[3].

Backstopped by the First Law of Thermodynamics, BomWatch protocols are physically based, replicable, objective and cannot be 'fiddled' with an outcome in-mind. In addition, they provide a statistical assessment of data quality, and include checks and balances against which outcomes are assessed. **While objectivity, replicability and *post hoc* verification are key components of the scientific method**,

---

[1] Trewin, Blair (2012). Techniques involved in developing the Australian Climate Observations Reference Network – Surface Air Temperature (ACORN-SAT) dataset. CAWCR Technical Report No. 049. March 2012

[2] Guidelines on Homogenization. World Meteorological Organisation, Geneva. WMO-No 1245 (2021 Edition).

[3] https://www.bomwatch.com.au/wp-content/uploads/2020/02/Townsville-full-paper.pdf

adoption of standardised best-practice procedures, are rarely discussed by either side of the climate debate.

As it is a century-long dataset from at least three locations, and potentially affected by a range of site conditions that might have affected observations, analysis of Tennant Creek Tmax provides useful case study of the problem of uncovering non-climate impacts and other data attributes that may affect trend and change.

2.1 Methods

Daily temperatures and monthly rainfall for the Tennant Creek telegraph/post office and airport were downloaded from BoM's *Climate Data Online* facility. Datasets were joined (abutted) on 1 November 1970. Multiple attributes were summarised each year using code written for the statistical program R (https://www.r-project.org/). Although only Tmax is analysed in this study (Figure 1), Tmin was also summarised. Factcheckers may access the summarised Tmax (.xlsx) dataset separately.

Monthly rainfall datasets were also merged, missing data were infilled using the nearest available site, and with infilled years flagged, annual rainfall was inserted into the analysis dataset. As it was incomplete at the time of writing data for 2023 was ignored.

Analysis was undertaken using the statistical program R and the *Rcmdr* package[1], with rainfall-adjusted segment means calculated using the *emmeans* package[2]. The statistical approach used in the study is an application of covariance analysis and is not controversial.

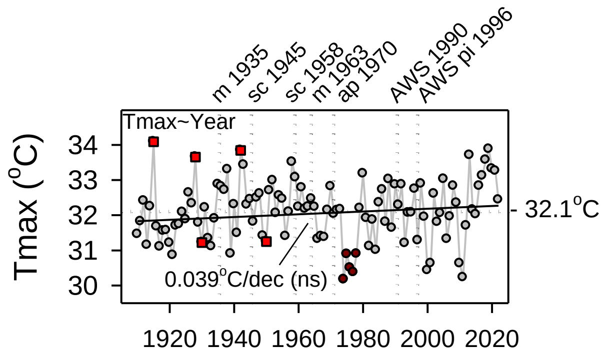


Figure 2. Average annual Tmax for the combined Tennant Creek dataset. Highlighted data (red squares) were likely outliers. Site changes noted in ACORN-SAT metadata include station moves (m), screen changes (sc), the join with airport data (ap), installation of the automatic weather station (AWS), and the AWS becoming the primary instrument (pi) from 1st November 1996. The naïve trend of 0.039°C/decade is not significant (ns) (i.e., *P*>0.05). Relatively wet years from 1974 to 1978 are highlighted.

2.2 The physical basis

The First Law of Thermodynamics that energy is neither created nor destroyed, requires that energy incoming to the environment is balanced by transformations that dissipate energy away. The First Law Theorem is commonly expressed as:

$$R_N = G + H + LE\text{, where:}$$

R_N is net radiation available at the surface [short-wave incoming (R_s) minus long-wave outgoing radiation (R_l)]
G is ground heat flux [heat exchanges with the deeper regolith (soil and rock)]
H is sensible heat exchange between the surface and the near-surface atmosphere
L is latent heat of vaporisation (2.45MJ/kg @ 25°C) and E, the amount of water evaporated (kg).

At annual timesteps:

- Components of R_N (R_S and R_L) depend on Latitude and for a fixed location are effectively constant (numerically-large values are characterised by a relatively small interannual coefficient of variation in the range 2-4%[3]).

---

[1] https://cran.r-project.org/web/packages/Rcmdr/index.html
[2] https://cran.r-project.org/web/packages/emmeans/index.html
[3] Average daily solar exposure (1990-2023) at Tennant Creek = 22.54 MJ/m^2/day, SD = 0.622, thus CV = 2.76.

- As cyclical surface heating (when $R_S > R_L$) is balanced by opposite-phase cyclical cooling (when $R_L > R_S$), ground heat flux, G, is zero.
- Under dryland (unwatered) conditions evaporation cannot exceed the rainfall. Hence, as 1mm of rain = 1kg/m^2, flux units are MJ/m^2/yr.

With G=0 and R_N held constant, a dynamic balance is expected between latent heat accounted for by evaporation of rainfall (LE), and sensible heat (H) advected to the local atmosphere. Advection is proxied during the heat of the day by Tmax measured by thermometers (or electronic temperature probes) held 1.2m above the ground in Stevenson screens. Evaluated using ordinary least squares regression of the form Tmax ~ rainfall, the First Law Theorem predicts that dry years are warm and the drier it is the warmer it gets. The Tmax/rainfall coefficient (^{o}C/100mm) is therefore expected to be negative and significant ($P_{slope=0}$ <0.05) and for a well-maintained site, variation explained (R^2_{adj}) is expected to exceed a benchmark of 0.50 (meaning that annual rainfall explains >50% of variation in mean Tmax).

Lack of significance and/or a positive coefficient indicates something is wrong, with the inference that as Tmax and its causal covariable are poorly related or uncorrelated, Tmax data are unlikely to be useful for tracking trend and changes in the climate. Irrespective, should R^2_{adj} not exceed 0.50, a variable may be missing or the relationship may be affected by outliers or other problems requiring further investigation.

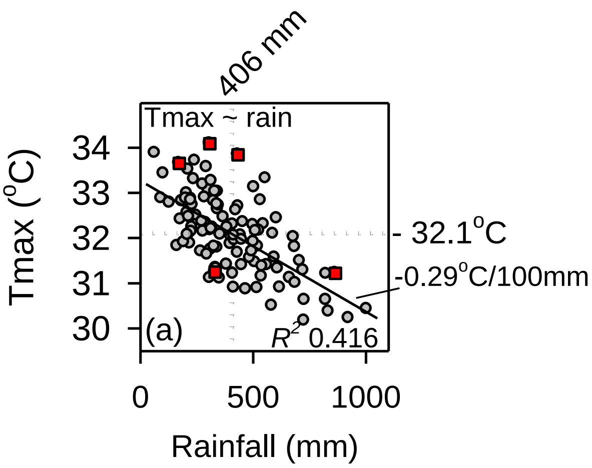
The advantages of analysing Tmax in the rainfall-domain are that:

- Tmax is expected to depend on rainfall, thus the methodology is physically-based and robust;
- Problems of autocorrelation are avoided;
- The Tmax/rainfall coefficient partitions variation due to rainfall from residual variation that is not explained, including random effects and underlying changes that may be systematic;
- Timewise inhomogeneities in Tmax ~ rainfall residuals are evaluated using independent statistical methods
- Final-round, covariance analysis and *post hoc* tests validate the outcome.

2.3 Results

2.3.1 Dependence of Tmax on rainfall

Consistent with the First Law Theorem, the overall relationship between Tmax and rainfall is highly significant (P<0.001). Rainfall reduces Tmax 0.29^{o}C/100mm and explains 41.6% of Tmax variation (R^2_{adj} = 0.416), 44.7% if apparent outliers in 1915, 1928, 1930, 1942 and 1950 are ignored (see Table 1, Case (i)). *Rainfall-domain* residuals were normally distributed, random, with equally dispersed variance, thereby validating key assumptions.



However, as R^2_{adj} is less than the benchmark of 0.50 there may be problems, including that a missing variable may affect the Tmax ~ rainfall response.

Figure 3. Naïve linear regression partitions the deterministic rainfall effect from variation attributable to other factors. *Rcmdr* influence plots Fox, 2022[1]) indicate data for 1915, 1928, 1930, 1942 and 1950 (red squares) were outliers. While the slope coefficient (-0.29^{o}C/100mm) was unaffected, ignoring outliers increased R^2_{adj} to 0.447.

[1] https://socialsciences.mcmaster.ca/jfox/Misc/Rcmdr/

Linear regression partitions variation due to rainfall from the non-rainfall, residual, or error portion of the Tmax response, which is shown in Figure 4.

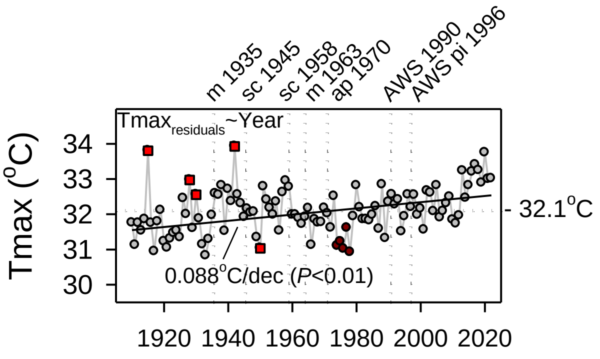


Figure 4. The non-rainfall, residual portion of the Tmax signal, re-scaled to be comparable by adding the Tmax grand mean of 32.1°C. Compared to Figure 2, variation is considerably less. Also, the steeper linear trend of 0.088°C/decade is significant. As they affect trend apparent outlier data for 1915, 1928, 1930, 1942 and 1950 (red squares) were ignored.

Site-change effects (inhomogeneities) are evidenced by abrupt step-changes in the re-scaled mean, which are detected as described below.

2.3.2 Unbiased detection of inhomogeneities

Sequential t-test analysis of regime shifts (STARS) was developed by Sergi Rodionov in 2004 to evaluate regime changes, or abrupt re-organisations in Bearing Sea ecosystems[1]. Adjusted for autocorrelation, STARS is also useful for detecting inhomogeneities in the mean of temperature timeseries. Detected changes ((Sh)ifts or step-changes) can be cross-referenced to known events and verified *post hoc* using objective statistical methods. Rodionov subsequently included additional features and preserving the acronym, re-named the updated software *Sequential Three-part Analysis of Regime Shifts*. The program is freely available as a standalone Excel workbook (*.xlsx) together with supporting information, a range of references and peer-reviewed papers at: https://sites.google.com/view/regime-shift-test/home.

STARS is used iteratively to detect sequential shifts in the mean of time-ordered data, with iterations examined as scenarios (Figure 5). Scenario-segments are evaluated and compared independently using linear regression, then verified using categorical multiple linear regression.

Iterative step-change analysis involved holding the *target* cutoff length (*l*) constant while varying the *P*-level, and/or holding the target *P*-level constant while varying *l*. Alternative Sh(ift) scenarios were also developed depending on whether outliers were included or not.

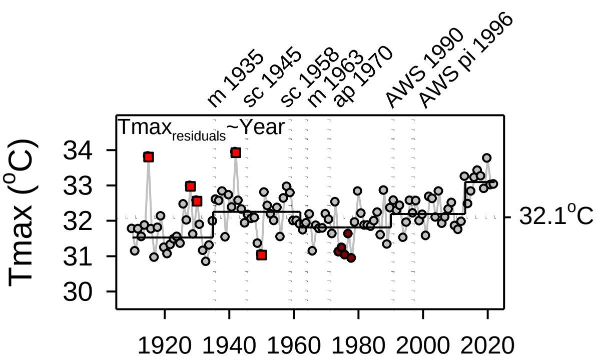


Figure 5. One of several shift-analysis iterations of rescaled Tmax ~ rainfall residuals identified using STARS (horizontal line). As they affected the magnitude of step-changes, outlier data for 1915, 1928, 1930, 1942 and 1950 (red squares) were omitted.

For a dataset of (N>30) annual means, a *target* cutoff-length of *l* = 10 and *P* = 0.1 is a useful starting point for evaluating decade-scale changes. Other settings include Huber's weight parameter (for handling outliers) = 5, IP4 for red noise estimation, no pre-whitening, and a cutoff of 5-years before the end of the series (for details see relevant documentation). Deciding which scenarios to accept or dismiss is informed by their timing compared to changes documented in metadata, *P*-levels associated with individual changepoints, and graphical (and statistical) analysis of changepoint residuals, which are expected to be random, independent and homogeneous.

STARS-analysis found the overall Tmax mean of 32.1°C consisted of five 'regimes' whose sequential means were statistically different at the *P* <0.01 level of significance. Importantly, changepoints identified by STARS cannot be prescribed in advance.

[1] https://www.beringclimate.noaa.gov/regimes/

2.3.3 Visualisation and verification of scenarios

It is evident from Figure 5 that the original site at the repeater station was cooler than after it moved to the post office. However, site-summary or ACORN-SAT metadata presents no clue as to why Tmax abruptly increased 0.73°C (P<0.001) and then declined 0.44°C (P=0.001) around the time the Stevenson screen reportedly moved in 1963. A likely explanation is that to alleviate harsh conditions surrounds at the repeater station were cooled artificially by watering. Also, that a structure or building affected exposure of the Stevenson screen at the post office prior to when the problem was rectified during a site audit in 1963.

Relationships between Tmax and rainfall for segments partitioned by step-changes in 1935, 1962, 1990 and 2013 (in Figure 5) are shown in Figure 6. In each case Tmax depends on rainfall (Table 1, Case (iv)). However, the Tmax response from 1910-1934 overlaid that for 1962-1989. Similarly, the response from 1935-1961 was coincident with that from 1990-2012 (i.e., slope coefficients and segment means were the same – Table 1, Case (ii)). Figure 6 also shows that responses successively stepped-up relative to the intersection of average rainfall (406mm, $\overline{x}$) and mean Tmax (32.1°C, $\overline{y}$). (Should rainfall alone explain Tmax, the least-squares line would overlie $\overline{x}, \overline{y}$). Suspected outliers are also identified as exceeding likely data ranges.

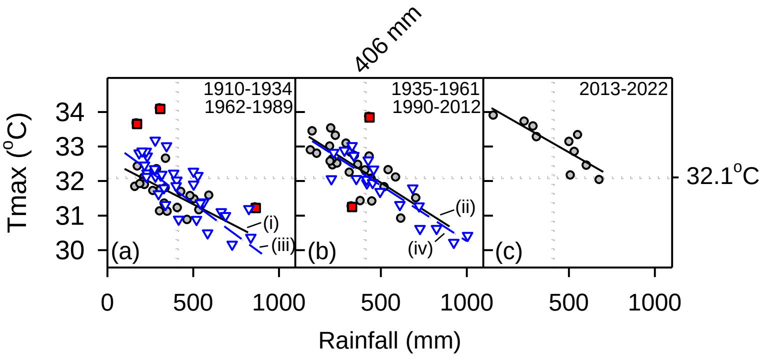
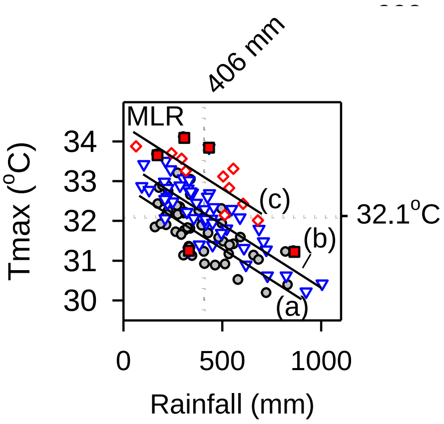


Figure 6. Analysis of the segments identified by STARS in Figure 5. In (a) regression line (i) (1910-1934 (grey circles) is coincident with line (iii) 1962-1989 (inverted blue triangles). Likewise, in (b) for (ii) 1935-1961 and (iv) 1990-2012 (refer Table 1, Case(ii)). (Outliers indicated by red-squares were ignored.)

Factorised as categories and with outliers ignored, first-round multiple linear regression analysis found that mean Tmax and regression coefficients were the same for the first and third, and the second and fourth segments (Figure 6 and Table 1, Case (ii)). Data for 1910-1934 and 1962-1989, and 1935-1961 and 1990-2012 were therefore combined into two response categories and with data from 2013-2022 were reanalysed (Figure 7, Table 1, Case (iii)).



Responses to rainfall were the same across categories (interaction was not significant, thus lines were parallel). Rainfall adjusted category means were different, thus, with rainfall held at 406mm, lines were significantly offset (Table 1, Case (iii)). Rainfall reduced Tmax 0.31°C/100mm and together with step-changes, simultaneously explained 73.8% of variation in Tmax.

Figure 7. Final-round multiple linear regression of the response of Tmax to rainfall grouped by categories as explained in the text.

A second scenario (not shown) of shifts in 1936 (when the site moved), 1962 (the year preceding the suspected site audit), 1980 (for which there is no supporting evidence), and 2013 (when the 60-litre screen/wind-profiler was installed) was also a candidate. Ignoring outliers, R^2_{adj} was slightly less (0.719 vs. 0.738), the Akaike information criterion indicated a slightly less parsimonious fit (129.66 vs. 121.75), and RMSE was 0.421 vs. 0.406. Further investigations including overlay plots, and considering the lack of evidence supporting a site-change in 1980 and that HQ and ACORN-SAT made no adjustments at that time, the scenario was dismissed. Moreover, although Tmax in 1979 was possibly under-range it was not flagged as an outlier.

Table 1. Tennant Creek statistical summary (following an independent statistical evaluation data for 1915, 1928, 1930, 1942 and 1950 were ignored).

Model	Coef. (°C/100mm)	*P*	R^2_{adj}	Segment	RainAdj°C (SEM)	AIC	Notes
(i) Tmax ~ rain	-0.289	<0.001	0.447			200.73	
(ii) Tmax ~ Sh_{res} + rain	-0.318	<0.001	0.739	1910-1934	31.5$^{(a)}$ (0.09)	123.51	Means having the same superscript combine into groups, and reanalysed as Case (iii).
				1935-1961	32.3$^{(b)}$ (0.09)		
				1962-1989	31.7$^{(a)}$ (0.08)		
				1990-2012	32.2$^{(b)}$ (0.07)		
				2013-2022	33.1$^{(c)}$ (0.13)		
(iii) Tmax ~ Sh_{res1} + rain	-0.312	<0.001	0.738	1910-1934&	<<- combined	121.7	Interaction
				1962-1989$^{(1)}$	31.6$^{(a)}$ (0.07)		Tmax ~ Sh_{res}1 * rain
				1935-1951&	<<- combined		*P* = 0.83 (ns)
				1990-2012$^{(2)}$	32.2$^{(b)}$ (0.06)		
				2013-2022$^{(3)}$	33.1$^{(c)}$ (0.13)		
				$Delta_{(1 vs 2)}$	0.61 (0.09)		
				$Delta_{(1 vs 3)}$	1.51 (0.15)		
				$Delta_{(2 vs 3)}$	0.91 (0.14)		
(iv) Tmax ~ rain							
1910-1934	-0.256	0.001	0.370				
1935-1961	-0.317	<0.001	0.558				
1962-1989	-0.366	<0.001	0.696				
1990-2012	-0.322	<0.001	0.758				
2013-2022	-0.284	<0.001	0.668				
(v) Tmax ~ Year	(°C/decade)			Segments	Coefficient		
Overall (1910-2022)	0.101	0.214	ns	1910-1934			*P* = 0.26 (ns)
				1935-1961			*P* = 0.65 (ns)
				1962-1989	-0.78°C/decade		*P* = 0.03 (<0.05)
				1990-2012			*P* = 0.19 (ns)
				2013-2022			*P* = 0.62 (ns)
(vi) $Tmax_{residuals}$ ~ Year	0.088	<0.001	0.223				Refer Figure 4

[**Notes**: Case numbers (i) to (v) are provided for reference; RainAdj refers to rainfall adjusted means (with standard errors) calculated by the *emmeans* package with turkey *P*-level adjustments for multiple comparisons; AIC refers to the Akaike Information Criterion; segment refers to data-segments defined by step-changes identified using sequential t-tests (STARS); ns indicates non-significance (*P* >0.05).]

2.3.4 *Post hoc* evaluation and quality assurance

Multiple linear regression residuals were examined graphically for linearity, normality, homogeneity of variance across categories (homoscedasticity) and leverage (the potential effect of outliers on regression coefficients) and found to be consistent with underlying assumptions. Influence plots confirmed that data for 1915, 1928, 1930, 1942 and 1950 were outliers. In standard deviation units, they exceeded 2-times the expected range of residuals and according to Cooks Distance, may have influenced trend. While ignoring data could be regarded as a form of data-hacking, provided they are identified objectively, omitting faulty data avoids potential bias.

In addition to multiple linear regression and outlier tests, segmented data were also analysed for trend. Except for 1962-1989, where trend was spuriously negative, within-segment trends were not significant (Table 1 Case (v)). So, while the cumulative effect of step-changes caused Tmax to increase 1.51°C since 1910 (Table 1, Case (iii)), removing their effect together with that of rainfall left no residual trend that could be attributed to CO_2, coal-mining, electricity generation or anything else. [The cumulative effect of residual step-changes calculated by STARS was 1.57°C.]

Scatter plots of *observed data* (on the y or vertical axis) verses values fitted (or predicted) by statistical models (on the x axis)[1] are one of the richest forms of data visualisation (https://stats.stackexchange.com/ questions/104622/what-does-an-actual-vs-fitted-graph-tell-us). For example, should rainfall fully explain Tmax, fitted/predicted values would align with those observed along a 1:1 diagonal (Figure 8(a)). Lack of fit is evidenced by 'spread' from the 1:1 line, bias by apparent or clustered skew from the 1:1 line, while outliers are peripheral to the data-cloud.

Rainfall alone poorly explained the dataset with data for 2013-2022 clustered high relative to the 1:1 line (Figure 8(a), Table 1, Case (i)). Inclusion of the Sh(ift) variable (Table 1, Case (iii)) re-aligned fitted/predicted values, which as indicated by the reference arrow and the square of Pearson's coefficient of linear correlation (r^2) considerably reduced 'spread', while outliers, which were analysed separately, were dispersed outside the data cloud (Figure 8(b)).

Multiple linear regression residuals were independent, normally distributed with equal variance across categories, and not unduly affected by outliers. As residuals were also homogeneous with no detectable trend (Table 1, Case (v)), accounting for step-changes and rainfall simultaneously left no residual trend or change attributable to another factor.

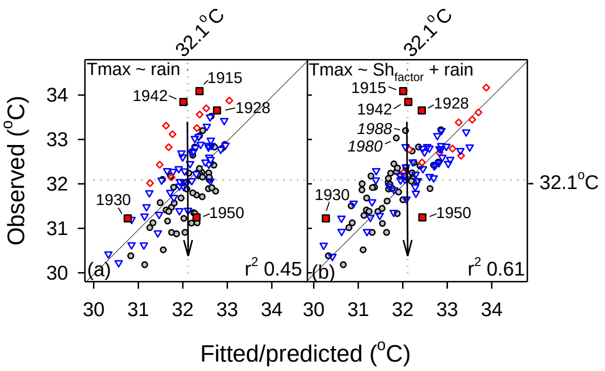


Figure 8. Observed Tmax vs. values fitted/predicted by: in (a), rainfall alone, and in (b) (Sh)ift$_{factor}$ + rainfall), where Sh$_{factor}$ corrects for step-changes. For a near-perfect fit paired datapoints would align closely along the 1:1 line. Grey circles represent data for 1910-1934 & 1962-1989, blue triangles, 1935-1951 & 1990-2012, and red diamonds, data for 2013-2022. Outliers are indicated by red squares. Average Tmax provides a reference, while arrows indicate spread at Tmax = 32.1°C. Data for 1980 and 1988 were not identified as outliers. Pearsons r^2 measures linear association between x and y.

While the analysis phase is complete, it remains to attribute step-changes in the data to physical changes at the sites. As previously discussed, it was likely that the site at the telegraph station was watered (Figure 9) whereas that at the post office was not. Also, the Stevenson screen at the post office may not have been moved away from a building or construction immediately, but later, following a site audit in 1963.

While the AWS was installed in 1990, due to poor-quality metadata and lack of collaborative evidence such as photographs etc., the step-change at that time cannot be strictly attributed. It is possible for instance, that to accommodate additional instruments thermometers were rearranged within the screen, or some other activity that affected observations but was not reported.

Figure 9. The Tennant Creek Telegraph station in 1872, showing a wind vane and Stevenson screen on the far right of the photograph (National Archives of Australia Item ID: 11774998). As the site was close to buildings, it was probably watered by bucket or watering-can during the seasonal 'dry'.

[1] https://sci-hub.se/https://doi.org/10.1016/j.ecolmodel.2008.05.006

A building plan for the airport shows the original Aeradio met office was beside the operations building facing the apron and the hydrogen generator shed (used to fill weather balloons) and met-store adjacent to a fenced area about 100m to the southwest (Figure 10). The old buildings and the foundations of the H_2 and met-store are visible in the March 2003 Google Earth Pro satellite image. The overlay of the image and the map shows that the current weather station is in the same place as it was when the Aeradio office opened in 1939. Being possibly the only Australian weather station that remained in the same place since it was established 84-years ago and that daily Aeradio station data were digitised post-1957 but are not available, it is strange that Aeradio data has not been used in ACORN-SAT.

The ACORN-SAT adjustment file indicates the former 230-litre Stevenson screen was probably replaced by a 60-litre wooden or plastic screen on 27 September 2012 about the time Google Earth Pro satellite images show that the site was sprayed with herbicide. Then after September 2012, a wind-profiler array occupying a gravelled area of about 0.38ha was installed just 45m west of the screen. Although the site has remained in the same place, spraying out the grass, installing the smaller screen, and the effect of the wind-profiler array on the local surface energy balance would be sufficient to explain the abrupt rainfall-adjusted Tmax up-step of 0.91°C in 2013 (Table 1, Case iii).

Figure 10. Google Earth Pro satellite image showing the 1969 met office and surrounds in September 2012 (upper left). Upper right, the image for March 2013 overlaid by a re-scaled 1950s aerodrome plan (100% opacity) showing the location of the Aeradio office (mo), operations centre (op), terminal (T), H_2 shed for preparing weather balloons (H_2) and met-store (ms) beside the instrument enclosure (x). Lower left is the same image with the overlay map at 60% opacity and on the right, at 30%.

2.3.5 Effect of site changes on maximum temperature extremes

Defined as values that occur in the tails of data distributions, warm and cool extremes occur throughout the year, typically daily values that exceed the 95^{th} and are less than the 5^{th} day-of-year (1-366) percentiles (Hi and Lo extremes respectively) calculated across all observations. Ten percent of daily observations are expected to occur as short-tail/long-tail extremes.

Frequencies of high and low extremes and their ratio (Hi_N/Lo_N) are expected to vary randomly over time and exhibit no persistent deviations from their long-term mean. Expressed as logarithms so they are symmetrical and amenable to parametric tests, homogeneity of the $\log_{10}(Hi_N/Lo_N)$ ratio is evaluated using STARS in Figure 11.

While the frequency of upper-range extremes was high in 1942, 1943, 1958 and 1990 (Figure 11(a)), effects did not persist. The up-step detected by STARS in 2013 aligned with herbicide use, installation of the 60-litre screen and the wind-profiler array, which could not be regarded as due to the climate.

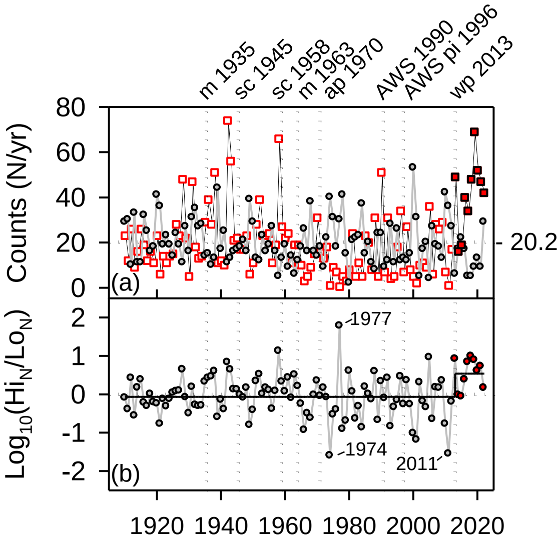


Figure 11. Frequency (N/yr) of daily values <5th and >95th day-of-year Tmax percentiles ((a), grey circles and red-squares respectively), and their log_{10}-transformed (Hi_N/Lo_N) ratio (b), analysed using STARS (horizontal line). The up-step corresponds with replacing the Stevenson screen and installation of the wind-profiler array (wp) between September 2012 and March 2013. Before 2013 the average number of Lo-extremes was 20.5/yr vs. 18.6 Hi-extremes, averages after were 13.2 vs. 41.6. (Effects due to rainfall have not been removed.)

2.4 Interim discussion

Using the century-long dataset for Tennant Creek as a case study, the preceding section outlined the steps involved in evaluating trend and change in Tmax. While some parts of the protocol: linear regression of Tmax on rainfall, calculating residuals and undertaking STARS analysis can be accomplished using EXCEL, spreadsheet programs are not suited to undertaking categorical multiple linear regression, interaction analysis and calculating rainfall-adjusted means, which is essential for verifying that segmented means are different and slopes are parallel. The added complication in this case was that two pairs of segments were coincident. Consequently, with those combined into single categories, only three out of the original five segments remained to be compared in final-round analysis.

In addition to partitioning between variation due to rainfall and variation that is not explained, linear regression provides an objective assessment of how well data conform to the First Law Theorem (*P* and R^2_{adj}) and a provisional estimate of the Tmax/rainfall coefficient. While Tmax is not a unit of heat *per se*, the coefficient is analogous to Bowen ratio, which partitions between heat advected to the near-surface atmosphere and heat expended as latent heat, usually in the context of estimating evaporation from soils, vegetation and open-water surfaces.

As rainfall occurs in sequences of wet/dry years it is imperative at the outset that the deterministic rainfall portion of the Tmax signal is removed. **Variation not due to rainfall** is thereby partitioned into Tmax ~ rainfall residuals, which is the zero-centred *error* or unexplained part of the Tmax signal. The second BomWatch protocol step evaluates if Tmax residuals (rescaled to be comparable by adding grand-mean Tmax) are homogeneous in time. The STARS t-test evaluates if the mean of incoming data is significantly different to the cumulative mean of those before, flagging the significance of a change if one occurs. However, without validation, the procedure could be regarded as a form of data-fishing (i.e., seeking changepoints in order to support an *a priori* hypothesis).

Validation involves, firstly, evaluating goodness of fit with rainfall, of segments defined by STARS, checking for outliers, and visually comparing the same relative to average rainfall and mean Tmax. Segments that bear no relationships with rainfall, or whose relationships are markedly skewed or off-scale may require further investigation: problems may be due to excessive missing daily data, miss-alignment of data-pairs, or a changepoint may be miss-identified, for example.

With segments specified as factors, categorical multiple linear regression confirms that:

- Segmented data are offset (rainfall-adjusted segmented means are statistically different, thus segmented responses to rainfall are not coincident)
- Segmented responses to rainfall are the same (interaction is not significant, thus individual Tmax ~ rainfall relationships are parallel, or more strictly that coefficients are the same)

- MLR residuals are independent, normally distributed, with equal variance across categories and outcome are not unduly affected by outliers.

Post hoc analysis also includes analysis of segmented timewise trends, which confirms that data consist of non-trending segments separated by step-changes, and cross-referencing shifts in rescaled residuals to site changes documented in metadata and other sources including maps, plans, aerial photographs and satellite images where available. Multiple Linear Regression residuals are also examined for inhomogeneities and possible trend that may due to additional factors such as CO_2, electricity generation, coal mining or anything else.

3. Homogenisation of Tennant Creek Tmax

Homogenisation of temperature data aims to correct for the effect of discontinuities caused by site-changes and other data problems, so homogenised data reflected the climate alone (Della-Marta et al, 2004, *op. cit.*). Della-Marta et al (and previously Torok and Nicholls (1996)) homogenised a set of high-quality (HQ) datasets that were used to calculate trends in annual mean temperature anomalies. Such trends were widely reported in various publications, such as annual climate summaries, CSIRO and Climate Commission reports as well as in scientific literature. The HQ series was also the first of several used to support models developed by the BoM, CSIRO and universities that underpinned *The Science*.

Handed down by from the lofty heights of academia by *professors* including Will Steffan, Tim Flannery, David Karoly and Lesley Hughes, and by the WWF-associated Climate Council, *The Science* elicits fear-based, irrational responses in vulnerable people to specific events such as the ‘unprecedented’ Millennium drought, the ‘record’ 2011 Brisbane flood, ‘once-in-a-generation’ bushfires, and natural climate changes related to the El Niño Southern Oscillation (ENSO), including warm days and protracted runs of dry or wet weather.

Flannery’s fiction *We are the weather makers* (Candlewick Press, ISBN 9780763636562 (ISBN10: 0763636568)) was funded for distribution to school children in 2005/06 by the Purves Environment Fund (https://uploads.prod01.sydney.platformos.com/instances/403/assets/Annual-Reports/Annual%20Review%202006.pdf?updated=1619513888). Rich-man Robert Purves was President of WWF (Australia) from 1999 to 2006. Purves’ environment fund also financed The Wentworth Group’s assault on farmers in the Murray-Darling Basin, the Australian Youth for Climate Coalition, Farmers for Climate Action etc. With links to GetUp! and spearheaded by Purves, WWF has long-been the political wing of environmental alarmism and other causes in Australia and at Conferences of the Parties, and IPCC gatherings overseas.

While previous homogenisation iterations applied annual or monthly adjustments, ACORN-SAT[1] aimed to align data distributions so daily data are “*more homogeneous for extremes as well as for means*”[2]. Changes included that breakpoints were detected using multiple tests, and that effects of site changes on extremes were adjusted by matching percentiles of data distributions with those of reference series using complex transfer functions[3].

In contrast to STARS, which detects sequential shifts in dataset means using the same settings with outcomes verifiable, the greater number of statistical tests used by Trewin cannot increase the confidence that change occurred (i.e., *post hoc* verification is more important than the number of tests deployed). HQ and ACORN-SAT lacked verification that homogenised data reflected the climate.

HQ homogenisation of Tennant Creek Tmax cooled the past by up to 1°C, particularly the period from 1935 to 1945, then to 1969/70 when the site allegedly moved to the airport (Figure 12).

---

[1] Blair Trewin (2012). Techniques involved in developing the Australian Climate Observations Reference Network – Surface Air Temperature (ACORN-SAT) dataset. CAWCR Technical Report No. 049, March 2012.

[2] See: BRR-032.pdf (bom.gov.au)

[3] Trewin, B., Braganza, K., Fawcett, R., Grainger, S., Jovanovic, B., Jones, D., Martin, D., Smalley, R. and Webb, V. (2020). An updated long-term homogenized daily temperature data set for Australia. *Geoscience Data Journal*, 7, pp.149-169.

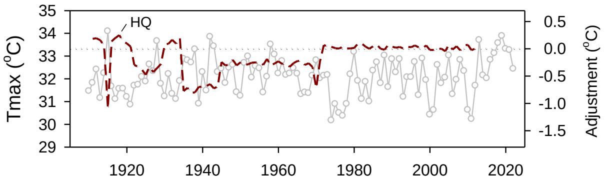


Figure 12. HQ homogenisation of Tmax data from 1911 to 2012 cooled the past (dashed line, right axis) relative to post-1965 data. The reason for the adjustment in 1970 of -0.71°C is unclear.

AcV1 homogenisation cooled data before 1965 more aggressively than HQ, particularly the period from 1936 to 1964 (Figure 13).

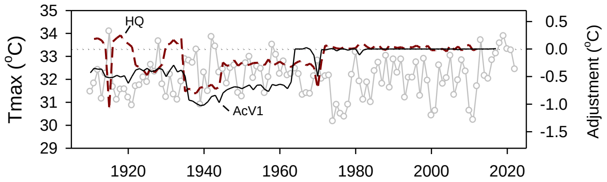


Figure 13. AcV1 homogenisation (1911 to 2017, continuous line right-axis) overlaid on HQ adjustments shown previously.

While minimal adjustments were made by AcV1 after 1964, AcV2.1 adjusted data by an average of 1.25°C before 1962, and -0.41°C up to 1998 (Figure 14).

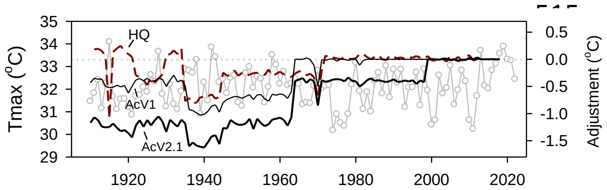


Figure 14. AcV2.1 (1910-2018) cooled data by an average of 1.25°C before 1962.

While reversing previous AcV1.1 adjustments to 1964, AcV2.4 applied a positive adjustment averaging 0.52°C from 1999 to 2015, thereby blending the effect of spraying-out the grass and installing the 60-litre screen with the effect of wind-profiler array before March 2013 (Figure 15).

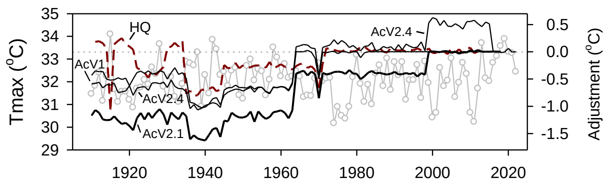


Figure 15. Adjustments made by AcV2.4 (1910 to 2022).

Except for AcV2.4, timing of adjustments generally coincides; however, their magnitudes (Figures 12 to 15) are vastly different. Best described as a trend-hack having the ultimate aim of maintaining trends that are both statistically significant and in-line with predictions made by CSIRO's climate models of around 1.5°C/Century (Table 2), continuing to cool the past as data accumulates is not viable going forward. Thus, to preserve the warming trend established by AcV1, AcV2.4 reduced the previous AcV2.1 adjustment before 1965, while data from 1999 to 2015 were adjusted higher by 0.52°C.

Simultaneous adjustment of both ends of the dataset merged site-change effects noted previously with data going forward, thereby maintaining a seemingly consistent trend (Figure 16).

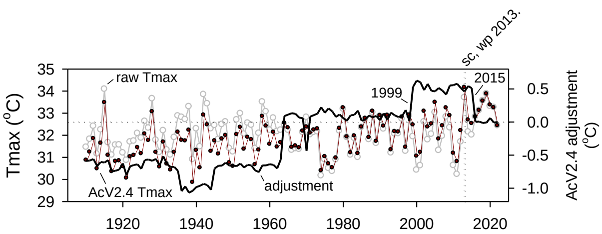


Figure 16. AcV2.4 adjustments (horizontal line, right axis) overlaid as solid circles on raw data, relative to the timewise installation of the 60-litre screen and the wind-profiler array (sc and wp). While smoothing and erasing the effect of 2012/13 site changes, negative adjustments prior to 1963 and positive adjustments from 1999 to 2015 maintained trend.

Table 2 shows that trend was not significant (i.e., not different to zero-trend) until AcV1 superseded HQ. Trend was also little changed between AcV2.1 and AcV2.4. However, while significances and R^2_{adj} for the Tmax ~ year relationship improved, variation explained by rainfall declined from 46.3% for HQ, to 16.5% for AcV2.4. Thus, ACORN-SAT adjustments that ostensibly realigned tails of data distributions so they reflected the climate, occurred at the expense of the resulting homogenised series reflecting the weather.

Table 2. Least squares trend analysis and the response of Tmax to rainfall, for the various rounds of homogenisation of Tennant Creek Tmax that were analysed. (Not all homogenised datasets are shown in Figures 11 to 14.) Tennant Creek Tmax (all data, Av°C_{all}) is included as the control.

Model **(Tmax ~ Year)**	Years	Coef. **(°C/decade)**	*P*	R^2_{adj}
Av°C_{all} ~ Year	1910-2022	0.024	0.339 (ns)	(ns)
HQAv°C ~ Year	1911-2012	Zero	(ns)	(ns)
AcV1Av°C ~ Year	1910-2017	0.068	0.006	0.060
AcV2.1Av°C ~ Year	1910-2018	0.153	<0.001	0.262
AcV2.3Av°C ~ Year	1910-2021	0.151	<0.001	0.284
AcV2.4Av°C ~ Year	1910-2022	0.150	<0.001	0.288
(Tmax ~ Rain)		**°C/100mm**		
Av°C_{all} ~ Rain	1910-2022	-0.290	<0.001	0.416
HQAv°C ~ Rain	1911-2012	-0.264	<0.001	0.463
AcV1Av°C ~ Rain	1910-2017	-0.246	<0001	0.322
AcV2.1Av°C ~ Rain	1910-2018	-0.209	<0.001	0.171
AcV2.3Av°C ~ Rain	1910-2021	-0.202	<0.001	0.171
AcV2.4Av°C ~ Rain	1910-2022	-0.197	<0.001	0.165

Applying the same methodologies outlined in Section 2.3.3, ACORN-SAT changepoints listed in the data-table for AcV1 (to 2009), and in the most recent ACORN-SAT Catalogue for AcV2.1 (to 2018) and AcV2.4 (to 2022) were analysed as scenarios using categorical multiple linear regression as outlined previously in Section 2.3. A statistical summary of those investigations is given in Table 3.

While for each scenario, R^2_{adj} was higher than the base Tmax ~ rainfall case (Table 3, Case(i)), there were no clear or consistent differences between rainfall-adjusted segment means. Thus, while responses to rainfall (the coefficients) were similar and parallel (interaction was not significant), in most instances, segmented regressions were coincident, inferring that over their respective time periods, the step-changes that adjustments aimed to correct were miss-specified. (Note that outlier data were not excluded.) Furthermore, applying changepoints that were not relevant inflates both R^2_{adj} and $Shift_{factor}$ sums-of-squares giving the impression that $Shift_{factors}$ were highly significant when they were irrelevant.

Table 3. Statistical evaluation of ACORN-SAT homogenisation scenarios. In all cases segmented regressions were parallel (Sh_{factor} by rainfall interaction not significant). Superscripts indicate differences between rainfall adjusted means at the *P* = 0.05 significance level. (Note that outlier data were not excluded.)

Model	Coef. (°C/100mm)	P_{reg}	R^2_{adj}	Segment	Cause	RainAdj-°C (SEM)	Notes
(i) Tmax ~ rain	-0.290	<0.001	0.416				All data
(ii) Tmax ~ Sh_{AcV1} + rain (1910 to 2009)	-0.295	<0.001	0.506	1910-1934		$31.8^{(a)}$ (0.11)	No clear differences between segment means at *P*<0.05 level of significance
				1935-1944	move	$32.6^{(b)}$ (0.18)	
				1945-1962	move	$32.1^{(a,b)}$ (0.13)	
				1963-2009	move	$32.0^{(a)}$ (0.08)	
(iii) Tmax ~ $Sh_{AcV2.1}$ + rain (1910 to 2017)	-0.315	<0.001	0.555	1910-1934		$31.7^{(a)}$ (0.11)	No clear differences between segment means at *P*<0.05 level of significance
				1935-1944	move	$32.5^{(b,c)}$ (0.17)	
				1945-1962	move	$32.1^{(a,b)}$ (0.13)	
				1963-1998	move	$31.9^{(a)}$ (0.09)	
				1999-2011	statistical	$32.2^{(a,b,c)}$ (0.16)	
				2012-2017	screen	$32.9^{(c)}$ (0.23)	
(iv) Tmax ~ $Sh_{AcV2.4}$ + rain (1910 to 2022)	-0.310	<0.001	0.600	1910-1934		$31.7^{(a)}$ (0.11)	No clear differences between segment means at *P*<0.05 level of significance
				1935-1944	move	$32.5^{(b,c)}$ (0.17)	
				1945-1962	move	$32.1^{(a,b,d)}$ (0.13)	
				1963-1998	move	$31.9^{(a.d)}$ (0.09)	
				1999-2011	statistical	$32.2^{(a,b,d)}$ (0.16)	
				2012-2015	screen	$32.6^{(b,c,d)}$ (0.27)	
				2016-2022	vegetation	$32.2^{(c)}$ (0.20)	

3.1 *Post hoc* evaluation of homogenised datasets

BomWatch protocols used naïve linear regression to remove the effect of rainfall on homogenised Tmax so shifts in the residual non-rainfall portion of the signal could be assessed objectively using STARS. Step-change scenarios were then verified using categorical multiple linear regression as described in Section 2.3.3.

Using identical methods, highly significant step-changes were found in HQ-Tmax residuals in 1920, 1926 and 1979. Also, in AcV1 residuals in 1918, 1927 and 1980, in AcV2.1 in 1951, 1979 and 2002 and AcV2.4 in 1958, 1980 and 2003. Although residual changepoints reflected the combination of adjustments shown in Figures 12 to 15, possibly mis-allocated changepoints and other embedded problems, STARS analysis showed homogenised data were not homogeneous.

The contention that data have been adjusted for non-climate impacts, infers homogenised data truly reflect weather and climate. If that be the case, observed Tmax would be highly correlated with, and closely aligned along a 1:1 line with homogenised values fitted/predicted by rainfall (vis-à-vis Figure 8). Biases would be indicated by a lack of fit, namely either that a line of best-fit between data-pairs would be offset but parallel to the 1:1 line, or skewed high or low depending on the distribution of observed vs. predicted values.

Figure 17 shows homogenisation has not resulted in fitted/predicted values aligning more closely with those observed. While HQ-Tmax is reasonably well-aligned (Figure 17(a)), excessive cooling of past data skewed the least-squares line of best-fit away from the 1:1 line, especially in the case of AcV2.1 (Figure 17(c)). Data do not therefore represent the combination of site changes and rainfall expected if they were adjusted in proportion to the step-changes they aim to correct. Homogenised data for 1915 and 1946 also appear to be outliers.

It is apparent from multiple BomWatch studies that scientists relying on data to calibrate their models undertake no evaluation of the data they use. In the same vein, warming trajectories presented in CSIRO's Sate of the Climate reports lack verification at the individual weather station level. The circularity of using data that are regularly adjusted to agree with models, to calibrate models is a perversion that may have reached its limits. Tmax at Tennant Creek, which is representative of 5.3% of Australia's continental land area is not increasing or likely to increase in the future due to CO_2, electricity generation, coal mining or anything else. Fabricating data to agree with *The Science* is abhorrent, unscientific, a breach of personal ethics and underserving of public trust.

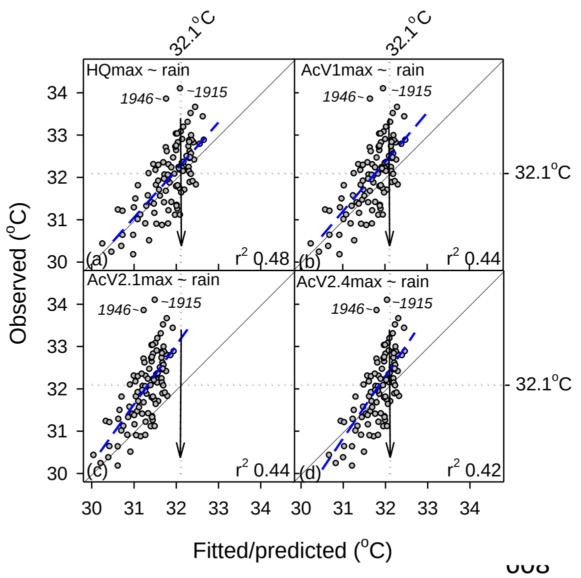


Figure 17. Scatterplots of Tmax observed verses HQ, AcV1, AcV2.1 and AcV2.4 fitted/predicted by rainfall. Correlation is indicated by the dashed line of best fit relative to the 1:1 line, and the square of Pearsons linear correlation coefficient (r). Range indicated by the arrow is the same as that in Figure 8.

4. Discussion

Climate scientists generally ignore that the datasets used to verify *The Science* were not collected in the first-place to monitor trend and change, but rather to monitor day to day weather, assist in making short-term forecasts and to describe and classify regional climates. The 'bolt-on' climate-change experiment commenced in about 1989 in response to calls by WMO to find data that supported IPCC narratives going forward. While Neville Nicholls and his protégées commenced homogenising data in the early 1990s, it was not until 1998 that a consensus approach developed[1]. Even at that late stage, methods used by Australian scientists were arbitrary and subjective, and lacked a statistically disciplined approach (see line 76 in : https://www.bomwatch.com.au/wp-content/uploads/2022/12/Marble-Bar-back-story-with-line-Nos.pdf).

Metadata is a crucial issue; however, BoM scientists make no attempt to separate the rainfall portion of the Tmax signal from effects due to site and instrument changes. Thus, changes that were documented but made no difference, and effects on data that were not documented, could be manipulated to support of pre-determined narratives. Introducing a '*statistical*' Tmax change in 1990 at Tennant Creek that could not be verified, adjusting for the screen-change in 2012, but ignoring the impact of the wind-profiler array before March 2013 are cases in-point. Trend in homogenised Tmax was created and maintained by cooling data to varying extents before 1963, and for AcV2.4, adjusting the warming trajectory post 1990 (Figures 12 to 15).

4.1 BomWatch protocols

Protocols, which are essentially a description of the research question (hypothesis) and the means by which it would be addressed, lie at the heart of the scientific method. Concisely stated, inflexible procedures and methods bring structure to chaos, and serve to adjudicate conflicting approaches to analysing the same or similar datasets. *A protocol should, therefore, convey a reasoned and appropriate research design using rigorous and feasible methods*[2].

The First Law of Thermodynamics, that energy is conserved, is a fundamental tenant of atmospheric physics that is central to arguments about the greenhouse effect and warming of the planet. It is therefore appropriate to apply the same Theorem to evaluating Tmax datasets for individual weather stations. Should Tmax be unrelated, or positively or poorly correlated with local rainfall, it is the data that are suspect, not the alternative, which is that the First Law Theorem does not work or has broken down. Further, as effects are additive and the Tmax ~ rainfall relationship only describes the deterministic rainfall part of the Tmax signal, the remainder or residual portion *must* embed unaccounted-for random as well as possible systematic non-rainfall effects that have occurred in parallel with observations.

Ordinary least squares regression straightforwardly partitions variation in Tmax into the portion related to rainfall, and the residual, non-rainfall part. While *rainfall-domain* residuals may be independent, normally distributed, with equally dispersed variance (and be unaffected by outliers), residuals may nevertheless embed timewise inhomogeneities indicative of parallel site-change effects.

Shifts or step-changes in the mean of rescaled residuals detected by STARS are verified using categorical multiple linear regression. Segmented Tmax ~ rainfall relationships are expected to be offset and parallel (i.e., that rainfall-adjusted segment means are different (segmented regressions are not coincident) and responses to rainfall (slope coefficients) are the same). The possibility of false-positives (applying changepoints to the data that inflate R^2_{adj}, but which are not relevant) is thereby minimised. Whether segments are combined and which scenarios are dismissed or accepted is determined transparently and objectively, not subjectively which is a major flaw in homogenisation methods used by the BoM. Verification that Tmax consists of non-trending segments bisected by step-changes related to site changes, and accounting for step-changes and rainfall simultaneously, provides assurance that factors affecting Tmax are fully explained.

---

[1] Peterson TC, Easterling DR, Karl TR, Groisman P, Nicholls N, Plummer N, Torok S, Auer I, Boehm R, Gullett D, Vincent L, Heino R, Tuomenvirta H, Mestre O, Szentimrey T, Salinger J, Forland EJ, Hanssen-Bauer I, Alexandersson H, Jones P, Parker D. 1998. Homogeneity adjustments of in situ atmospheric climate data: a review. *International Journal of Climatology* 18: 1493–1517.

[2] Chapter 8, in: P. Agger et al., *A Practical Guide to Biomedical Research* (Springer International Publishing AG 2017) DOI 10.1007/978-3-319-63582-8_8

4.2 Homogenisation of Tennant Creek Tmax

The same BomWatch protocols expose major weaknesses in methods used to homogenise the same Tmax data. Trewin's claim that ACORN-SAT aimed to align data distributions so daily data are "*more homogeneous for extremes as well as for means*" implies that homogenised Tmax should more closely reflect weather and climate than the data from which it was derived.

However, this was not the case. Table 2 shows that maintaining trend by arbitrarily cooling the past, then, in addition, for AcV2.4 by warming post-1999 data, caused relationships between $Tmax_{homogenised}$ and rainfall to become progressively less precise. Also, adjustments and some changepoints prescribed by ACORN-SAT did not align with the effect of site changes on Tmax data (Table 3). While their fit with raw data was biased and offset (Figure 17), homogenised data were also not homogeneous. Thus, from those three perspectives, the homogenisation process is clearly faulty.

Trewin's homogenisation methods, which are similar to those advocated by the WMO lack statistical control, replicability, and *post hoc* and other inbuilt checks and balances. There is no indication to what extent Tmax data are correlated with reference series, or *post hoc* tests that verify homogenised data are homogeneous, or that they reflect local weather. Reliance on *reference series* consisting of data that are not homogeneous, to correct faults in target site data is biased, unscientific and should be abandoned.

4.3 Policy driven science

Since the Liberal-National coalition government created the Australian Greenhouse Office in 1998, Australians have been subject to an intense, bipartisan, government-sponsored, climate-focussed misinformation campaign led primarily by atmospheric physicists and modellers within CSIRO. Data homogenisation provided the perfect foil for Neville Nicholls and protégés at the BoM, to create the datasets that underpinned predictions of looming catastrophe, while off to the side, elitest professors churned-out papers and commentary. With the help of *The Conversation*, the ABC *et al*. and activist groups such as WWF, incessant marketing of catastrophism institutionalised the narrative, politically and more broadly within the community.

Climate modelling commenced in Australia in 1969 with a joint CSIRO/Bureau of Meteorology venture focussing on numerical weather prediction[1]. Leading from that, global climate modelling commenced in 1981 with development of the *CSIRO-2* atmospheric general circulation model. A new modelling venture, the Australian Community Coupled Earth System Simulator (ACCESS) commenced in 2004. According to its website, ACCESS aims to provide more detailed and less uncertain projections of future climate change at regional scales, and improved understanding of some of the 'earth system' components including carbon-cycle feedbacks, dynamic vegetation, atmospheric chemistry and oceanic oxygen content. However, much of that *understanding* is based on scant data, modelled data, including satellite data, or no data at all.

Under the auspices of a joint research facility - the Centre for Australian Weather and Climate Research (CAWCR), ACCESS is also used for weather forecasting, seasonal forecasting and climate variability and climate change simulations. Colourised deliberately to stir irrational fear and loathing about the climate, daily ACCESS *projections* have found their way into BoM '*products*' and are increasingly dominating its marketing efforts. Groups including *Weatherwatch*[2] and *Brisbane Storm Chasers*[3] compete in the same space.

To make projections seem plausible, in about 1990 a program of digitising temperature, rainfall, sea-level and river flow records commenced, with many previously state-run investigations coming under the umbrella of the BoM, CSIRO, and the National Tide Facility. Supported by a grant from the National

---

[1] Smith, Ian (2007). Global climate modelling within CSIRO: 1981 to 2006. *Aust. Met. Mag*. 56 153-166

[2] https://www.weatherwatch.net.au/climate

[3] https://www.bsch.com.au/

Greenhouse Advisory Committee, homogenisation of Australian temperature data, initially as annual time-series, was one such project[1].

At about the same time, changes were made in the way data were acquired. Automatic weather stations (AWS) and accurate, rapid-sampling electronic temperature probes and tipping-bucket raingauges progressively displaced manual observations, with AWS becoming primary instruments after 1st November 1996. As a consequence, most former BoM-run weather stations are now unmanned.

The other major network-wide change was the staged replacement of previously standard 230-litre Stevenson screens with 60-litre wooden screens specified in 1973[2], which accelerated after 1990. More recently, some of those screens have been replaced by plastic (PVC) screens. Although not noted in metadata, ACORN-SAT adjustments suggest a 60-litre screen was installed at Tennant Creek on 27 September 2012.

The AWS has been operating at Tennant Creek since 8 July 1990, and while data should be available, no parallel AWS and thermometer data are available for the 7-year overlap before 1 November 1996 when the AWS became the primary instrument. Instrument comparisons are therefore not possible. Furthermore, while adjusting for the screen in 2012, ACORN-SAT makes no adjustments for installation of the wind-profiler array before March 2013. Data reported after 2013 are higher and more extreme due to changes at the site, not changes in the climate.

While no trend or change was evident in Tmax data for Tennant Creek (Section 2.3), guided by Figures 13 to 16, BomWatch protocols were also used to evaluate AcV1, Acv2.1 and AcV2.4 changepoint scenarios (Table 3). While segmented relationships with rainfall were parallel (coefficients were the same), there were no clear differences between rainfall-adjusted segment means. This indicates that changepoints were poorly specified, data were affected by outliers (which were included in the analysis) or data adjusted using transfer functions and reference series were faulty.

To account for site moves, AcV1 (to December 2009) made Tmax adjustments in 1935, 1945 and 1963. AcV2.1 (to December 2018) made additional adjustments in 1999 (statistical), and 2012 (screen). However, no evidence could be found in this study of a down-shift in 1999 that would warrant the positive AcV2.4 adjustment of 0.52°C shown in Figure 15, and the compensating upward adjustment relative to AcV2.1, for data prior to 1965. STARS would have detected a downshift of that magnitude if it occurred. Furthermore, as change at that time could not be confirmed by multiple linear regression (Table 3, Case (iv)), the '*statistical*' adjustment appears to have been made arbitrarily.

As additional data accumulates, it is not feasible to maintain trend by continuing to cool the past. Thus, AcV1 had reached the limit to which pre-1963 data could be plausibly cooled. Consequently, AcV2.3 adjusted post-1999 data warmer, so merging with the 2013 step-change, with pre-1963 data adjusted less-severely to compensate. Smoothing over the screen-size change and the effect of the wind-profiler array maintained the AcV2.1 Tmax trend of 0.15°C/decade to 2022. However, this occurred to the detriment of relationships between homogenised Tmax and rainfall (highlighted in Table 2).

5. Synopsis and conclusions

This report outlines the basis for, and steps involved in undertaking unbiased analysis of trend and change in Australian maximum temperatures datasets using data for Tennant Creek as a case study.

BomWatch protocols embody key components of the scientific method, namely that based on energy balance principles, they are transparent, objective, replicable and cannot be influenced by *a priori* narratives. Protocols also include quality assurance and *post hoc* tests that evaluate assumptions and verify outcomes.

---

[1] Torok SJ (1996). "The development of a high-quality historical temperature data base for Australia". PhD Thesis, School of Earth Sciences, Faculty of Science, The University of Melbourne.

[2] Australian Climate Observations Reference Network – Surface Air Temperature (ACORN-SAT) Observation practices. (ACORN-SAT_Observation_practices_WEB.pdf (bom.gov.au))

While data in general may be described in terms of averages, ranges, percentiles and other attributes, it is highly unlikely that the location or the conditions under which observations have been made have remained consistent in the long-term. Also, as a basis for determining changepoints, site-summary and ACORN-SAT metadata are notoriously unreliable. By removing the rainfall portion of the Tmax signal BomWatch protocols specifically detect the timewise likelihood (*P*-level) and magnitude (°C) of site-related changes on temperature trends. The statistical approach is a straightforward application of covariance analysis and is not controversial.

While some analyses can be undertaken using spreadsheet applications such as Excel, categorical multiple linear regression is best undertaken using a statistical framework such as R, and packages such as *Rcmdr* and *emmeans*, or using Minitab etc. The statistical application PAST from the University of Oslo (https://www.nhm.uio.no/english/research/resources/past/) is useful for exploratory data analysis; however, current versions do not include categorical multiple linear regression, or interaction analysis.

While Figures 2 to 7 show the steps involved, in order to be defendable against factcheckers, principal outcomes must also be supported by statistical summaries (Tables 1 to 3) and *post hoc* evaluations of outcomes (e.g., Figure 8 vs. Figure 16). In addition to inferior fits, segmented regressions based on arbitrary changepoints, or changepoints that do not significantly impact on data are unlikely to satisfy the condition that segmented responses be offset and parallel. Segment by rainfall interactions may also be significant, and residuals may show outliers and other features that are inconsistent with assumptions.

Applied to the problem of evaluating ACORN-SAT, BomWatch protocols determined that although segmented responses to rainfall were parallel, segment means were mostly the same, or they overlapped (Table 3). Changepoints were therefore either poorly specified, data were affected by outliers or homogenised data were faulty. Homogenised data were also not homogeneous.

At all stages in the process (naïve analysis of Tmax ~ rainfall, STARS analysis of rescaled residuals, segment-by-segment analysis, categorical multiple linear regression (and interaction analysis)) properties of residuals were examined graphically or statistically. Residuals and fitted/predicted values were also exported for further evaluation as necessary (e.g., Figure 17).

While BomWatch protocols have not been wrapped into a seamless R package, they provide a robust, unbiased approach to assessing trend and change in individual weather station datasets, and also for assessing homogenised HQ and ACORN-SAT datasets.

Conclusions

BomWatch protocols comprise four elements, namely:

- The overall relationship between Tmax and rainfall partitions total variation into that due to rainfall, and the residual non-rainfall part. Linear regression also derives the overall Tmax/rainfall coefficient, and significance (*P*) and goodness of fit (R^2_{adj}) statistics indicating conformity with the First Law of Thermodynamics.
- Homogeneity analysis of rescaled residuals identifies non-climate impacts on data, which are categorised as step-change or (Sh)ift scenarios
- Segment-by-segment analysis with rainfall detects outliers, lack of fit, and other potential problems, and,
- Categorical multiple linear regression (and interaction analysis) finalises and verifies outcomes.

Segmented trend and graphical analysis confirm that relationships are linear, residuals are normally distributed, independent, with constant variance, and that they are timewise homogeneous.

Based on the First Law Theorem that maximum temperature depends on rainfall, BomWatch protocols provide an unequivocable basis for understanding the effect of non-climate impacts on data, and for objectively assessing the BoM's homogenisation methods.

At each step, protocols are supported by residuals analysis, and segmented trend and graphical analysis to confirm relationships are linear, residuals are normally distributed, independent, with constant variance and that they are timewise homogeneous.

Bill Johnston

06 February 2024

Preferred citation:

Johnston, Bill 2021. Is homogenisation of Australian temperature data any good? Part 7. Tennant Creek, Western Australia. http://www.bomwatch.com.au/ 20 pp.

Disclaimer

Unethical scientific practices including the homogenisation of data to support political narratives undermines trust in climate science. While the persons mentioned or critiqued may be upstanding citizens, which is not in question, the problem lies with their approach to data, use of poor data or their portrayal of data in their cited and referenceable publications as representing facts that are unsubstantiated, statistically questionable or not true. The debate is therefore a scientific one, not a personal one.

Those who contributed ideas and discussion over many years included Geoffrey Sherrington, Ken Stewart (https://kenskingdom.wordpress.com/) and Tom Berger (http://www.elastictruth.com/). David Mason-Jones is gratefully acknowledged for providing invaluable editorial assistance.