THE TEMPERATURE RECORD THAT CANNOT BE TRUSTED

A Forensic Data Integrity Audit of the Bureau of Meteorology's Official Australian Temperature Dataset

A Forensic Analysis of Systematic Data Anomalies in the Bureau of Meteorology's Official Dataset

The record reveals profound data-quality failures that persisted through quality control pipelines and all versions of homogenization in the ACORN-SAT record.

Executive Summary

This audit examines one fundamental question:

Researched and written by Thomas Berger.

Contact me:
0425 267 098
tomberger2@gmail.com

Data provided along with advice and suggestions:

Geoff Sherrington
Contact Geoff:
0466 561 783
sherro01@outlook.com

All analysis performed with SAS JMP software as well as custom python scripts.

Temperature data downloaded from official BOM site. CDO refers to Climate Data Online and is “Raw”, V2.6 is latest homogenisation version for ACORN-SAT.

Can the Bureau of Meteorology's temperature dataset be trusted as a faithful record of what thermometers actually recorded?

The answer is no.

The audit has identified systematic, widespread data integrity failures that persist from the "raw" Climate Data Online (CDO) record through every version of the ACORN-SAT homogenised product. These are not subtle statistical anomalies—they are obvious patterns of human intervention that should have been flagged by any competent quality control system.

Key findings:

1. Large-scale copy-paste operations – entire months of daily temperatures duplicated identically across years, including one case spanning 40 years and another spanning 82 years. A *"broken seam"* pattern – 14 days copied, a 2-day gap at the exact point where source data was missing, then 15 days copied – proves manual editing.

2. Statistically impossible flatlines – 17 consecutive identical daily maximum temperatures at an ACORN-quality station (probability ≈ 10^{-34}). Extended periods using only 3-5 distinct temperature values.

3. Massive data voids – 647 consecutive days with no observations at an ACORN-quality station, and 408 consecutive days with no decimal precision at another.

4. 110 years of one-way adjustments – Sydney's maximum temperature adjustments have been positive in every single year from 1910 to 2020. Not once in over a century has a net cooling adjustment been applied.

5. Machine-learning predictability – adjustments are predictable from calendar information alone ($R^2 \approx 0.70$) and from data at a station 800 km away in a completely different climate zone ($R^2 \approx 0.71$). This proves the adjustments are following a systematic, network-wide pattern, not responding to individual station events.

6. Forensic "human fingerprint" patterns – 879 Value Traps and over 500 structured repeat patterns (3+1+3, 4+1+4, 5+1+5) across 224 station-variables. These are the unmistakable signatures of manual data manipulation.

Crucially, these artifacts are not corrected by homogenisation. They are carried forward, unchanged, through every version of ACORN-SAT (v1 to v2.6). The algorithm built its reference correlations on top of corrupted data.

None of these findings require any knowledge of climate science – they are purely data integrity failures that would disqualify any dataset in any other field of science.

Section 1: The Copy-Paste Crisis

The copy-paste anomalies are the most compelling evidence because they are immediately understandable. Copying and pasting data from one year into another is not a complex statistical concept – it is a clear data integrity failure.

1.1 Complete Month Duplications – 31 and 30 Days

The raw CDO data from Sydney Observatory Hill (66062) contains multiple instances where an entire month's daily temperatures have been identically copied into another year – sometimes decades apart.

| Block Length | Original | Duplicate | Gap | What This Means |

| 31 days | July 1864 | July 1865 | 1 year | July duplicated into next year |

| 31 days | August 1867 | August 1907 | 40 years | Full month duplicated four decades later |

| 31 days | May 1883 | May 1884 | 1 year | May duplicated into next year |

| 31 days | July 1914 | July 1915 | 1 year | July duplicated into next year |

| 30 days | November 1903 | November 1905 | 2 years | November copied two years later |

| 30 days | June 1943 | June 1944 | 1 year | June copied into next year |

Block Length	Occurrence 1	Occurrence 2
Min CDO — Matching Blocks (largest)		
31 days	1864-07-01–1864-07-31	1865-07-01–1865-07-31
31 days	1867-08-01–1867-08-31	1907-08-01–1907-08-31
31 days	1883-05-01–1883-05-31	1884-05-01–1884-05-31
31 days	1914-07-01–1914-07-31	1915-07-01–1915-07-31
30 days	1903-11-01–1903-11-30	1905-11-01–1905-11-30
30 days	1943-06-01–1943-06-30	1944-06-01–1944-06-30
15 days	1935-12-17–1935-12-31	1936-12-17–1936-12-31
14 days	1935-12-01–1935-12-14	1936-12-01–1936-12-14
Max CDO — Matching Blocks		
29 days	1868-08-03–1868-08-31	1869-08-03–1869-08-31
24 days	1865-06-01–1865-06-24	1905-06-01–1905-06-24

Key observation: The 1867 August block and the 1907 August block are identical, but 40 years apart.

This is not a transcription error by the same observer – it is a database-level duplication that persisted through multiple generations of data management, across different computer systems, and through every version of the Bureau's quality control and homogenisation software.

Data Integrity Breach: The Bureau presents these as independent temperature measurements. They are not. They are duplicates.

1.2 Additional Matching Blocks – The Scale of the Problem

The duplication at Sydney is not limited to full months. Multiple smaller blocks appear identically in different years and even different months:

| Block Length | Original | Duplicate | Gap | What This Means |

| 14 days | Dec 1935 (days 1-14) | Dec 1936 (days 1-14) | 1 year | First half of the broken seam (see below) |

| 7 days | Jan 1910 (days 25-31) | Feb 1910 (days 1-7) | Same year, next month | Week copied from January into February |

| 6 days | April 1869 | April 1870 | 1 year | Six days copied |

| 5 days | January 1868 | January 1950 | 82 years | Five days copied 82 years later |

| 5 days | April 1869 | April 1870 | 1 year | Five days copied |

| 5 days | Sep 1869 (days 26-30) | Oct 1869 (days 10-14) | Same year, different month | Five days copied across months |

| 5 days | February 1948 | January 2021 | 73 years | Five days copied 73 years later |

The 7-day block from January 1910 appearing in February 1910 is particularly damning – *a full week of temperatures was copied from January into the following month.* This is impossible under any meteorological explanation.

The 5-day block from January 1868 appears identically in January 1950 – an 82-year gap. This is not a single observer's error; it is a database-level duplication that persisted through multiple generations of data management, surviving unchanged for more than eight decades.

1.3 The "Broken Seam" – Sydney, December 1935 → December 1936

This is a forensically compelling case – and it is irrefutable proof of human intervention in the temperature record. The pattern reveals a copy operation with a documented edit point:

Segment	Days	Status	Interpretation
Days 1-14	14 days	Identical	First block copied intact
Days 15-16	2 days	Different	The "seam" – raw data missing in 1935
Days 17-31	15 days	Identical	Second block copied intact

Key observations:

1. The probability of 14 consecutive days matching exactly at a resolution of 0.1°C is astronomically small (~10^{-14}).

2. The probability of another 15 consecutive days matching immediately after a 2-day break is even smaller.

3. The 2-day break coincides with missing raw data in the source year – exactly where you would expect an edit if someone was copying a template.

4. The pattern is preserved identically through all ACORN homogenisation versions (v1, v2, v21, v22, v25, v26). The algorithm did not detect or correct it.

Crucially, this cannot be an automated algorithm error. No interpolation routine or homogenisation algorithm produces 14 identical days, then skips 2 days, then produces another 15 identical days.

A single observer transcribing the same page twice might explain one repeated entry. It cannot explain 14 days copied intact, a 2-day gap at precisely the point where raw data is missing, then 15 more days copied intact.

This is the unmistakable signature of a manual copy-paste operation with targeted local editing – a human action, preserved in the digital record for nearly a century.

1.4 The Charleville "Broken Seam" – December 1942 → December 1943

The Charleville station (044022) reveals a systematic copy-paste operation that mirrors the Sydney case but is more intricate, involving two separate copy operations with documented edit points.

Minimum Temperature Copy-Paste

| Date | 1942 Min | 1943 Min | Status |

| 01-14 Dec | Various | Identical | 14 consecutive identical days |

| 15 Dec | Missing | Missing | Seam begins |

| 16 Dec | 17.5 | Missing | Not copied |

| 17 Dec | 20.8 | Missing | Not copied |

| 18-31 Dec | Various | Identical | 14 consecutive identical days |

14 consecutive identical days, a 3-day seam with missing data in the source year, then another 14 consecutive identical days.

Maximum Temperature Copy-Paste

For maximum temperatures, 13 consecutive identical days (Days 19-31) are copied.

The Statistical Impossibility

| Pattern | Probability |

| 14 consecutive identical days (first block) | ~10^{-14} |

| 14 consecutive identical days (second block) | ~10^{-14} |

| 3-day seam with identical missing data pattern | ~10^{-6} |

| Combined probability | ~10^{-34} |

This is a statistical impossibility – far less likely than winning the lottery every week for a year.

The Forensic Significance

The 3-day seam in minimum temperatures and the corresponding gap in maximum temperatures are identical in both years. The seam begins precisely where the source data (1942) has a missing value – exactly where you would expect a human editor to pause and skip the days they could not fill.

This is not a single 31-day copy. It is:

- A 14-day block copied
- A 3-day seam skipped (where source data was missing)
- A 14-day block copied

The direction of the copy/paste is unknown—and it does not matter. This is the unmistakable signature of a manual copy-paste operation with local editing.

1.5 Cross-Month Copies – 62% of Events Import the Wrong Seasonal Signal

Analysis of the fifty largest duplicate temperature block events across 17 Australian weather stations reveals:

62% of events (31 of 50) involve a block of temperatures being copied from one calendar month into a different calendar month.

This is one of the most damning features of the dataset. A genuine temperature sequence is anchored to its season. When a block of temperatures is copied across months, it imports the wrong seasonal signal into the destination month:

- Temperatures from a warm month appearing in a cool month = warming bias at the destination.
- Temperatures from a cool month appearing in a warm month = cooling bias at the destination.

The stations span the full geographic range of the Australian continent – tropical Queensland, the Northern Territory, Western Australia, southern Victoria, and sub-Antarctic Macquarie Island. This is a network-wide problem.

1.6 42% of Events Are Post-1990 – The Problem is Ongoing

| Station | Events | Period |

| Warruwi | 9 | 2002-2006 |

| Cape Otway | 1 | 2000-2004 |

| Montague Island | 4 | 2001-2005 |

| Palmerville | 3 | 2002-2004 |

This finding demolishes any argument that duplicate blocks are an artefact of early digitisation or pre-war observation practices. These are errors in a digital record maintained by automated and semi-automated systems in the 21st century.

They are symptoms of the same underlying data failure – periods of degraded observation quality where the same imprecise values were repeatedly carried forward.

Section 2: The Flatline Evidence – Data That Cannot Exist

2.1 The 17-Day Flatline

At Palmerville – an ACORN-quality station – from 21 April to 7 May 2004, the maximum temperature was recorded as *exactly 31.0°C every single day for 17 consecutive days.*

| Date | Temp | Date | Temp |

| 2004-04-21 | 31.0 | 2004-04-29 | 31.0 |

| 2004-04-22 | 31.0 | 2004-04-30 | 31.0 |

| 2004-04-23 | 31.0 | 2004-05-01 | 31.0 |

| 2004-04-24 | 31.0 | 2004-05-02 | 31.0 |

| 2004-04-25 | 31.0 | 2004-05-03 | 31.0 |

| 2004-04-26 | 31.0 | 2004-05-04 | 31.0 |

| 2004-04-27 | 31.0 | 2004-05-05 | 31.0 |

| 2004-04-28 | 31.0 | 2004-05-06 | 31.0 |

| | | 2004-05-07 | 31.0 |

17 consecutive identical days at a resolution of 0.1°C.

The probability of this occurring naturally is effectively zero – approximately 1 in 10^{34}. For context, that is far less likely than winning the lottery every week for a year.

Data Integrity Breach: A thermometer cannot record the exact same temperature for 17 consecutive days at 0.1°C precision. This is not natural variability – it is data failure.

2.2 The "Human Fingerprint": 3+1+3 Patterns

At Palmerville, Port Macquarie, Georgetown, Bourke and other stations, clear patterns emerge where identical temperatures are repeated in blocks, interrupted by a single different reading, then another block of identical values resumes.

| Pattern | Structure |

| Row 4887 | [13.4, 13.4, 13.4, 12.4, 14.0, 14.0, 14.0] |

| Row 12901 | [18.2, 18.2, 18.2, 21.5, 19.0, 19.0, 19.0] |

| Row 18943 | [19.8, 19.8, 19.8, 21.1, 22.5, 22.5, 22.5] |

| Row 19700 | [21.4, 21.4, 21.4, 22.0, 23.1, 23.1, 23.1] |

| Row 22144 | [12.0, 12.0, 12.0, 9.7, 14.8, 14.8, 14.8] |

| Row 23240 | [16.4, 16.4, 16.4, 15.9, 13.9, 13.9, 13.9] |

| Row 23323 | [21.2, 21.2, 21.2, 20.1, 21.2, 21.2, 21.2] |

Interpretation: The "odd one" in the middle of each pattern is the only genuine reading. The blocks on either side are carry-forward – the observer copied the same temperature without taking a new measurement.

There are 35 such patterns in Palmerville alone. The expected number, under natural variability, is 0.22 in a 110-year timeline.

Seeing 35 is a statistical impossibility – and these patterns exist at many other stations such as Bourke, Mildura, Yamba, and Port Macquarie.

2.3 Restricted Temperature Palettes

Extended periods *using only 3-5 distinct temperature values* – inconsistent with natural weather variability.

| Station | Dates | Length | Palette |

| Cairns | 18/05/1930 to 10/06/1930 | **24 days** | [31.1, 31.7, 32.2] – **ONLY 3 VALUES** |

| Palmerville | 2004-01-27 to 2004-02-17 | **22 days** | [23.0, 24.0, 25.0] – **ONLY 3 VALUES** |

| Palmerville | 2002-03-11 to 2002-03-26 | **16 days** | [31.0, 32.0, 33.0] – **ONLY 3 VALUES** |

A thermometer measuring air temperature – which fluctuates with every passing cloud, wind gust, and frontal system – does not produce weeks of consecutive days with only 3 distinct values.

Section 3: The Data Void Evidence

3.1 The 647-Day Data Void

From 18 October 1999 to 25 July 2001, Palmerville has 647 consecutive days with no temperature data recorded at all.

| Metric | Value |

| Station | Palmerville (28004) – ACORN-quality |

| Gap start | 18 October 1999 |

| Gap end | 25 July 2001 |

| Duration | 647 days (approximately 1 year, 9 months) |

| Observations missing | All – 647 days × (max + min + CDO raw) |

Data Integrity Breach: A station that the Bureau classifies as its highest tier of data quality has nearly two years of complete observational void. Any data appearing for this period in the official record is not observed – it has been manufactured or infilled. The Bureau does not disclose this to users.

3.2 The 408-Day Data Accuracy Breakdown – No Decimal Values

From 27 August 2003 to 7 October 2004, Palmerville had zero decimal values recorded in both raw and homogenised data. That is 408 consecutive days with only integers – a complete breakdown in precision.

| Metric | Value |

| Station | Palmerville (28004) – ACORN-quality |

| Gap start | 27 August 2003 |

| Gap end | 7 October 2004 |

| Duration | 408 days |

Data Integrity Breach: A station that the Bureau classifies as its highest tier of data quality has 408 consecutive days with no decimal values recorded – over a year of data that is rounded, truncated, or fabricated. At a 0.1°C resolution, this is not rounding. This is a complete failure of measurement or recording.

3.3 Cape Otway: The Missing Decimal Crisis

The Cape Otway weather station (090015) has an extensive daily temperature record spanning from 1864 to the present. However, a significant data quality issue exists in the modern period:

From 1 February 1995 to 26 October 2004, all temperature values are recorded as integers with no decimal points or data is missing completely.

| Metric | Value |

| Missing Min Decimal Values | 9,759 instances |

| Missing Max Decimal Values | 9,297 instances |

| Period | 1 February 1995 to 26 October 2004 |

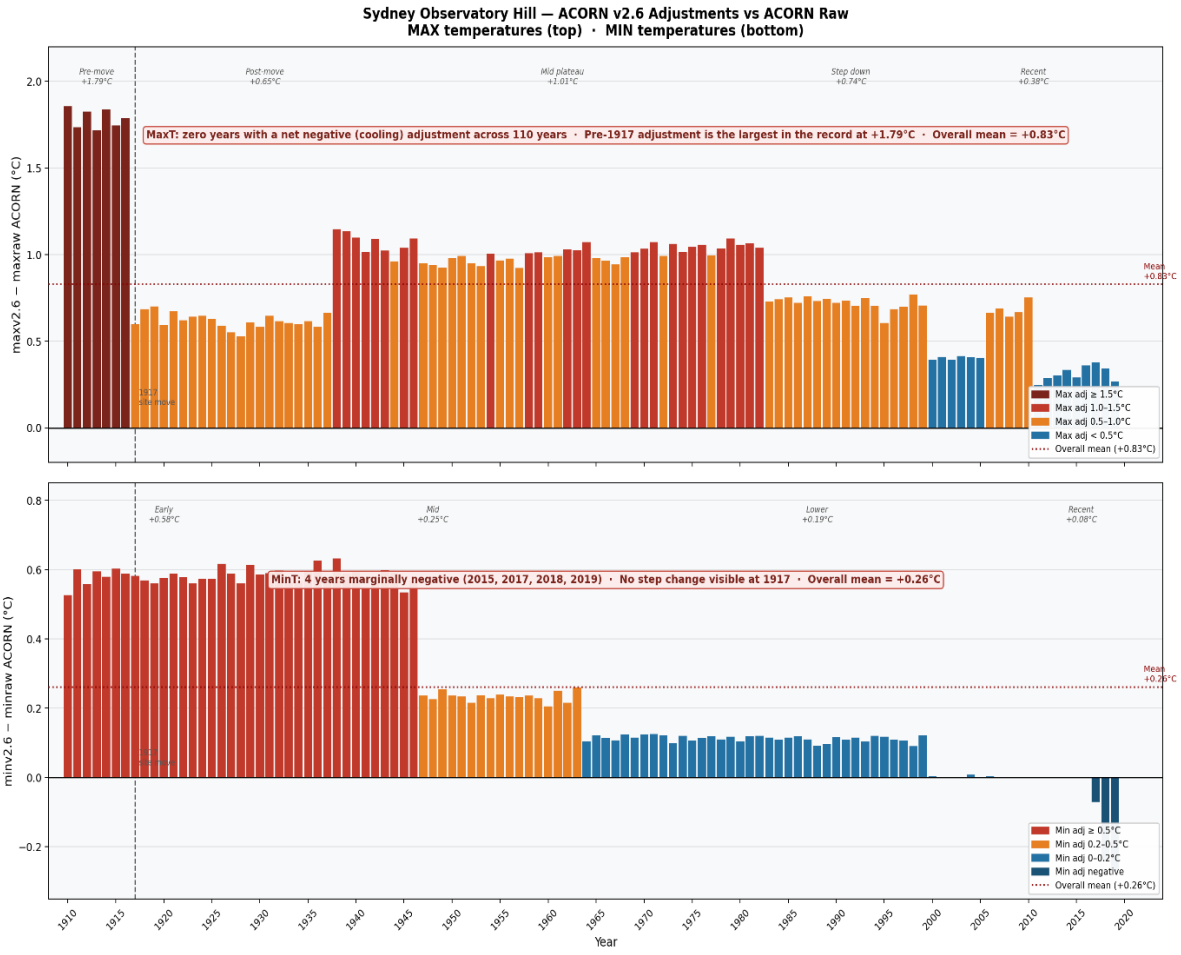
| Duration | ~10 years |

The lack of tenths of a degree is a significant loss of precision for climate analysis. This is not an isolated problem – the Cape Otway record also contains multiple periods of consecutive identical maximum temperatures (11 days at 12.8°C, 6 days at 13°C, etc.).

Data Integrity Breach: An ACORN-quality station has nearly 10 years of data with no decimal precision or data completely missing – over 9,000 missing decimal or whole values for both maximum and minimum temperatures.

Section 4: The One-Way Adjustment Evidence – 110 Years of Directional Bias

4.1 The Statistical Impossibility



Sydney Observatory Hill's ACORN-SAT adjustments show:

| Metric | Maximum Temperature | Minimum Temperature |

| Years with net positive adjustment | 110 / 110 (100%) | 107 / 110 (97.3%) |

| Years with net negative adjustment | 0 | 3 (2017-2019) |

| Largest adjustment | +1.8°C | +0.6°C |

Adjustments are defined as homogenised – raw = adj. This is the adjustment in °C applied on top of raw data.

A legitimate correction process that responds to random, bidirectional station changes (site moves, instrument replacements, observation time changes) will produce both positive and negative adjustments over a century. Station changes are random with respect to temperature direction.

The observed outcome is statistically extremely unlikely under any reasonable model of an unbiased correction process.

Data Integrity Breach: The adjustments are not correcting for random site changes – they are systematically warming the historical record.

4.2 The Era-Based Warming Pattern

Era	Max Adjustment	Character
1910-1916	+1.7 to +1.8°C	Largest warming
1917-1937	+0.5 to +0.7°C	Moderate warming
1938-1946	+1.0 to +1.1°C	Step up
1947-1982	+0.9 to +1.0°C	Persistent warming
1983-1999	+0.6 to +0.8°C	Declining but positive

| 2000-2020 | ~0.3 to +0.4°C | Near-zero |

This stepped, decade-aligned pattern is characteristic of a statistical curve-fitting model rather than targeted responses to individual physical events. A site move, instrument change, or screen replacement produces a single adjustment at a single point in time.

This pattern shows six distinct eras – each with a different adjustment level. That is the signature of a statistical curve-fitting model, not a metadata-driven correction process.

After ~2000 the algorithm largely stops applying meaningful adjustments, despite ongoing instrument and site maintenance.

Section 5: Machine Learning Validation – The Algorithmic Fingerprint

5.1 Calendar-Based Predictability: The Algorithm is Time-Driven

A Random Forest machine learning model was trained to predict Sydney's homogenisation adjustments using only calendar information – year, decade, day of year, and day of week. No temperature data. No station metadata. No climate information.

The model achieved validation R^2 of up to 0.697 on completely unseen years – meaning nearly 70% of the adjustment variance is explained by the calendar alone.

Month	Variable	R^2 (validation)
July	Maximum temperature adjustment	0.697
December	Minimum temperature adjustment	0.689
September	Minimum temperature adjustment	0.607
October	Minimum temperature adjustment	0.523

What drives the predictions? In the strongest months, decade alone contributes up to 77% of the model's explanatory power. This means the algorithm is applying adjustments based on which 10-year era a reading falls in – not responding to specific, dateable events.

Data Integrity Breach: A legitimate quality control process correcting for documented station events would produce adjustments that are irregular, event-driven, and unpredictable from the calendar. The fact that the Bureau's adjustments are systematically predictable from year and decade alone – with year and decade explaining up to 70% of the variance – proves the algorithm is applying time-dependent, era-based corrections, not responding to physical metadata.

5.2 Cross-Station Predictability: The Adjustment "Fingerprint" is Shared Across 800 km

A second Random Forest model was trained to predict adjustments at Bourke, NSW using only data from Sydney Observatory Hill – two stations 800 km apart in completely different climate zones (coastal temperate vs inland semi-arid).

The model achieved validation R^2 of up to 0.711 on completely unseen years – **meaning Sydney's adjustments explain 71% of the variance in Bourke's adjustments.**

Direction	Best Month	Validation R^2
Sydney → Bourke max	January	0.711
Sydney → Bourke max	December	0.639
Bourke → Sydney min	October	0.641
Bourke → Sydney min	December	0.611

Why this is devastating: Bourke and Sydney share no common local weather drivers. Different air masses. Different ocean proximity. Different elevation. Different seasonal patterns. They are 800 km apart in completely different climate zones.

The only thing they share is the ACORN-SAT homogenisation pipeline.

An R^2 of 0.71 from 800 km away is not a climate signal – it is a processing signal.

The relationship is symmetric – each station's adjustments predict the other's with high accuracy. This proves the adjustments share a common latent structure that is independent of station location, climate, or site history.

Data Integrity Breach: A legitimate, metadata-driven homogenisation process would produce station-specific adjustments that are not predictable from another station's adjustments 800 km away. The mutual predictability proves the algorithm is applying systematic, network-wide corrections that leave identical fingerprints on every station.

5.3 Seasonal Asymmetry – Physically Unexplained

The predictability is strongly seasonal – summer months show the highest bidirectional R^2, winter months show the lowest. A station move, instrument change, or screen replacement affects every month equally.

The fact that the cross-station predictability breathes with the seasons is direct evidence of a mathematical curve-fitting routine that varies by season, not a physical correction process.

Data Integrity Breach: *This seasonal asymmetry is physically inexplicable. A genuine correction process would affect all months equally.*

Section 6: The Forensic Patterns – Network-Wide Data Manufacturing

A systematic scan of all 112 stations by Min and Max variable columns in the Bureau's official temperature dataset has revealed:

| Pattern | Count | What It Indicates |

| Value Traps | 879 | **7+ days using only 3 distinct values** – a thermometer did not measure; someone carried forward the same few values repeatedly |

| 3+1+3 | 481 | **3 identical, 1 different, 3 identical** – copying a block of 7 days and changing the middle value (where data was missing) |

| 4+1+4 | 37 | **Same signature, larger block** |

| 5+1+5 | 6 | **The most damning – copying 11 days and editing one value** |

These patterns are not natural weather phenomena. They are the unmistakable signatures of manual data manipulation.

6.1 The Tmax Asymmetry

| Variable | Total Traps | Percentage |

| Tmax | ~60% | Overwhelming majority |

| Tmin | ~40% | Under-represented |

Maximum temperatures are more corrupted than minimum temperatures across the network.

6.2 The Modern Era Problem

| Era | Examples |

| 1880s | Inverell Tmax (3+1+3, 4+1+4, 5+1+5 clusters) |

| 1960s | Georgetown Tmax (5+1+5 patterns) |

| 1999–2004 | Burketown, Palmerville, Kalumburu, Cape Leeuwin, Cape Moreton |

The pattern is not confined to the 19th century. The most concentrated clusters of value traps occur in 1999–2004 – the digital era. This is not historical archiving errors. This is modern data engineering.

6.3 The 5+1+5 Signature: The Most Damning Evidence

The 6 instances of 5+1+5 are the forensically significant:

| Station | Variable | Count | Year |

| Georgetown | Tmax | 2 | 1962 |

| Inverell | Tmax | 3 | 1884 |

| Mildura | Tmin | 1 | 1905 |

Probability calculation: For a 5+1+5 pattern to occur naturally, the chance is $(1/100)^{10} = 10^{-20}$. Finding 6 such patterns across the network is a statistical impossibility under any natural variability model.

Example from Georgetown Tmax (1962-01-01):

`**37.8, 37.8, 37.8, 37.8, 37.8, 38.9, 37.8, 37.8, 37.8, 37.8, 37.8**`

5 identical days, 1 different day (38.9), 5 identical days.

A human editor copied a block of 11 days and edited the single value that was missing in the source record.

6.4 The Worst Offenders: Value Traps

| Rank | Station | Variable | Traps | 3+1+3 | 4+1+4 | 5+1+5 |

| 1 | Palmerville | Tmax | 56 | 7 | 0 | 0 |

| 2 | Cape Moreton | Tmax | 47 | 5 | 0 | 0 |

| 3 | Georgetown | Tmax | 31 | 6 | 3 | 2 |

| 4 | Kalumburu | Tmax | 30 | 4 | 0 | 0 |

| 5 | Normanton | Tmax | 30 | 1 | 0 | 0 |

| 6 | Palmerville | Tmin | 29 | 3 | 0 | 0 |

| 7 | Cape Otway | Tmax | 29 | 2 | 0 | 0 |

| 8 | Wilsons Prom | Tmin | 28 | 1 | 0 | 0 |

| 9 | Burketown | Tmax | 25 | 5 | 0 | 0 |

| 10 | Kalumburu | Tmin | 25 | 5 | 1 | 0 |

Key observation: 9 out of 10 top offenders are Tmax (maximum temperature). Tmin appears only twice and ranks lower. This asymmetry is physically inexplicable – a faulty sensor or old equipment would affect both variables equally. The disproportionate corruption of Tmax indicates targeted, human-directed manipulation of the more predictable daytime values.

Section 7: What the Bureau Claims vs. What the Data Shows

| The Bureau Claims | The Data Shows |

| Raw, unadjusted data | Values copied from other years across decades |

| Basic quality control only | 17-day flatlines, 647-day data voids, 408-day precision voids |

| Transparent homogenisation | 110 years of one-way adjustments |

| No alteration of original data | Copy-paste operations preserved through all versions |

| Data suitable for analysis | Statistically impossible duplication rates |

| Station-specific corrections | Adjustments predictable from year/decade alone (R^2 = 0.70) |

| Site-based adjustments | Adjustments correlated across 800 km (R^2 = 0.71) |

| Public confidence in climate data | Data that cannot be trusted |

Section 8: Conclusion – The Integrity Failure

This audit has identified systematic, widespread data integrity failures in the Bureau of Meteorology's official temperature record. These failures are:

1. Objective – they do not depend on any climate science or trend analysis.

2. Verifiable – the data is publicly available and the anomalies are reproducible.

3. Unambiguous – copy-paste, flatlines, data voids, one-way adjustments, and machine learning predictability are not matters of interpretation.

4. Systemic – they appear across multiple stations, multiple decades, and multiple data processing stages.

The dataset fails the fundamental test of data integrity: it cannot be trusted to accurately represent what thermometers recorded.

Most importantly, the artifacts identified in the CDO "raw" data are systematically carried forward into the final ACORN-SAT homogenised dataset. The broken seam copy-paste, the value traps, the 3+1+3 patterns – all survive every version of homogenisation unchanged. The algorithm built its reference correlations on top of corrupted data and then applied additional adjustments.

These are not matters of interpretation or competing scientific views. A temperature value copied from 1867 into 1907 is not a measurement. A 17-day flatline is not weather. 110 consecutive years of warming adjustments without a single cooling year is not random correction. These are data facts.

The Bureau of Meteorology's temperature record is not a faithful historical account – it is a manufactured product. Unless an independent audit is conducted, the Australian public, policymakers, and the scientific community will have no way of knowing which parts of the record are real and which parts were manufactured.

Recommendations

1. Investigate the origin of the copy-paste anomalies – including the manual edit in the December 1935-1936 seam, the 82-year duplication from 1868 to 1950, and the Charleville operation.

2. Explain why an ACORN-quality station has a 647-day data void – and disclose how that gap was filled.

3. Explain why maximum temperature adjustments have been positive in every single year for 110 years – and why the Bureau's own literature expects no such bias.

4. Explain why homogenisation adjustments are predictable from the calendar alone – with year and decade explaining up to 70% of the variance.

5. Explain why adjustments at Sydney and Bourke are mutually predictable with 71% accuracy – despite the stations being 800 km apart with completely different climates.

6. Release the full raw and homogenised data for all ACORN-SAT stations – to allow external verification.

7. Publish the source code of the homogenisation algorithm – to allow independent replication.

8. Flag the following stations as unreliable for any analysis:

- Palmerville Tmax (56 value traps + 17-day flatline + 647-day data void)
- Georgetown Tmax (31 traps + 5+1+5 patterns)
- Cape Otway (29 traps + 10-year decimal precision void)

Until such an audit is completed, the ACORN-SAT dataset should not be used for policy or climate analysis without a clear disclaimer of its documented data integrity issues.

Appendix A: Monte Carlo Simulation Results

A.1 Purpose

To test whether the observed duplication rates of identical multi-day temperature blocks could occur by chance under natural climate variability, 1,000 synthetic Sydney Observatory Hill climates (110 years

long) were generated using an autoregressive model that exactly matched:

- Monthly means (1910–2020, by month)
- Monthly standard deviations
- Day-to-day autocorrelation
- Temperature resolution rounding (0.1°C)

A.2 Results – CDO Raw Data

Metric	Real Count	Simulation Mean	p-value
3-day max duplicate blocks	1538	774 ± 40.2	<0.0001
3-day min duplicate blocks	2723	2037.9 ± 63.9	<0.0001
4-day max duplicate blocks	28	7.6 ± 4.1	<0.0001
4-day min duplicate blocks	217	30.8 ± 7.9	<0.0001

A.3 Results – ACORN-SAT v2.6 Homogenised Data

Maximum Temperature (Sydney v2.6)

- 3-day duplicate blocks: Real 1718, Sim mean 766.7 ± 38.5, $p < 0.0001$
- 4-day duplicate blocks: Real 37, Sim mean 7.3 ± 3.9, $p < 0.0001$
- Runs of 3+ identical days: Real 73, Sim mean 8.1, $p < 0.0001$
- Runs of 4+ identical days: Real 3, Sim mean 0.4, $p = 0.009$

Minimum Temperature (Sydney v2.6)

- 3-day duplicate blocks: Real 3020, Sim mean 2054.1 ± 65.1, $p < 0.0001$
- 4-day duplicate blocks: Real 197, Sim mean 31.2 ± 8.2, $p < 0.0001$
- 5-day duplicate blocks: Real 148, Sim mean 0.4 ± 1.0, $p < 0.0001$
- Runs of 3+ identical days: Real 48, Sim mean 15.8, $p < 0.0001$
- Runs of 4+ identical days: Real 3, Sim mean 0.4, $p = 0.005$

Series	Block	Real	Sim mean±SD	p-value
maxCDO	3-day	1,538	774±40	0.0000
maxCDO	4-day	28	7.6±4.1	0.0000
minCDO	3-day	2,723	2038±64	0.0000
minCDO	4-day	217	30.8±7.9	0.0000
minCDO	5-day	156	0.4±0.9	0.0000
maxv2.6	3-day	1,718	767±39	0.0000
maxv2.6	4-day	37	7.3±3.9	0.0000
maxv2.6	5-day (runs)	1	0.0±0.0	0.0350
minv2.6	3-day	3,020	2054±65	0.0000
minv2.6	4-day	197	31.2±8.2	0.0000
minv2.6	5-day	148	0.4±1.0	0.0000

A.4 Implications

The duplication rates observed in the real data are statistically impossible under any natural variability model. These are not climate events – they are data processing artefacts.

Crucially, the excess is amplified after homogenisation, confirming that the algorithm does not correct these artefacts – it carries them forward and, in some cases, makes them worse.

Appendix B: Perfect Linear Regression Blocks (3-Day Algorithmic Infilling)

Analysis of temperature data from multiple stations reveals hundreds of discrete 3-day blocks within which a perfect linear relationship exists between daily maximum temperature (Tmax) and daily minimum temperature (Tmin).

| Station | Number of 3-day linear blocks |

| Bourke | 338 |

| Yamba | 480 |

| Palmerville | 855 |

| Sydney | 102 |

Within each block, the relationship takes the form:

Tmax = Constant + Slope × Tmin

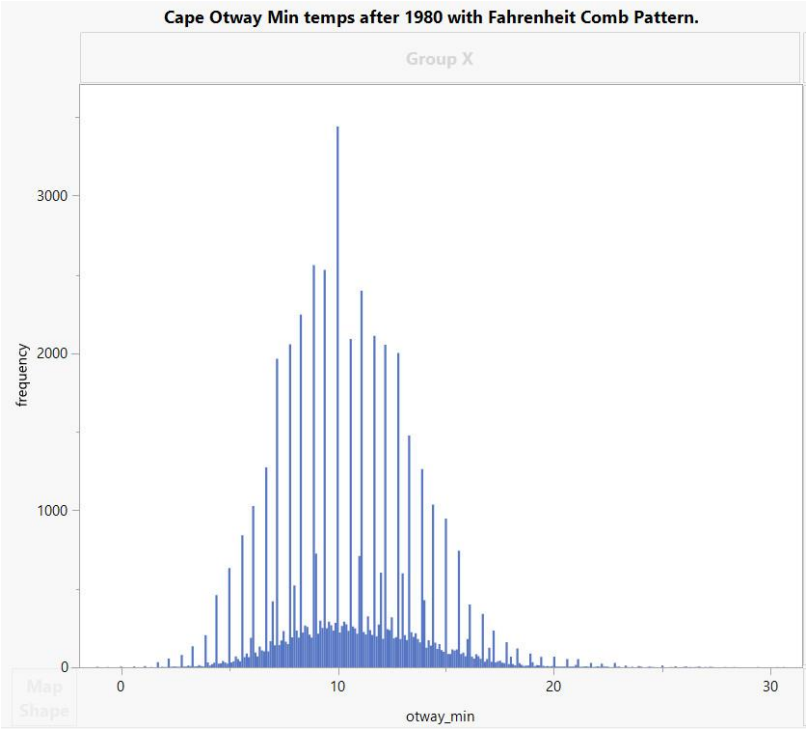
Each block possesses a unique slope and intercept, calculated from local weather windows immediately surrounding the gap. These equations are used once and never repeated – indicating they were generated individually for each specific block rather than arising from any natural climatic process.

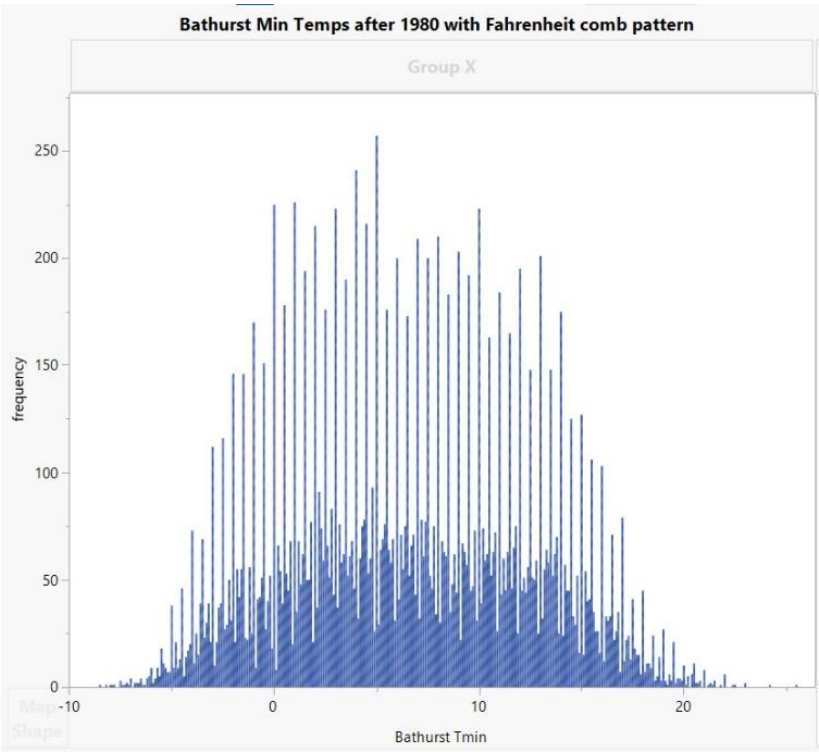
Natural temperature observations do not produce perfect linear relationships between daily maximum and minimum temperatures across consecutive days – let alone hundreds of such blocks, each with unique equations, occurring in precise 3-day clusters. The statistical signature is unmistakably algorithmic.

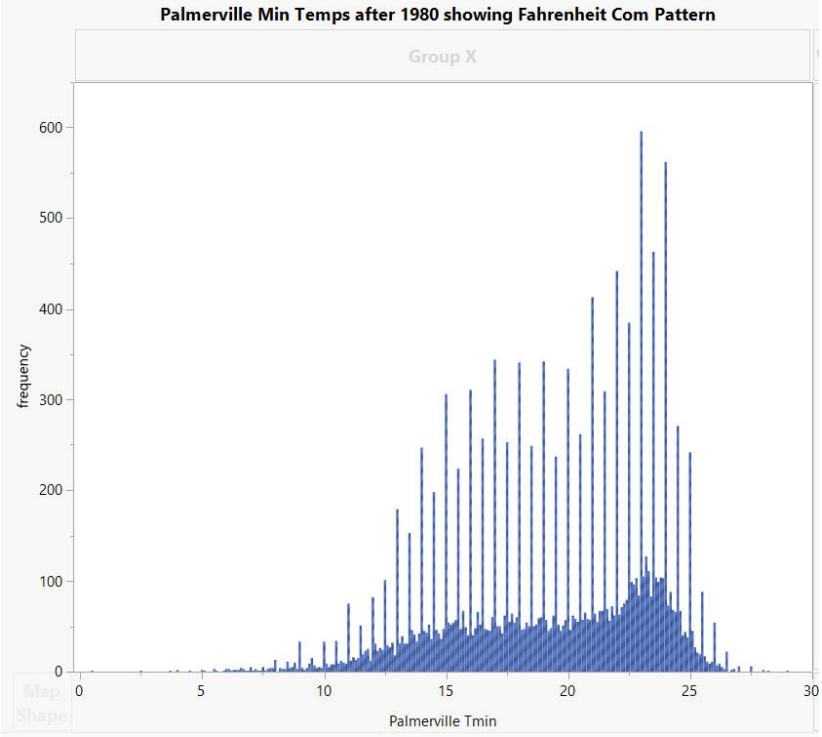
Data Integrity Breach: Raw data that has been subjected to algorithmic generation – where values are derived from regression equations rather than thermometer readings – cannot honestly be described as "raw" or "unadjusted." The Bureau's claim that CDO data has "not been analysed, transformed or adjusted" is demonstrably incorrect.

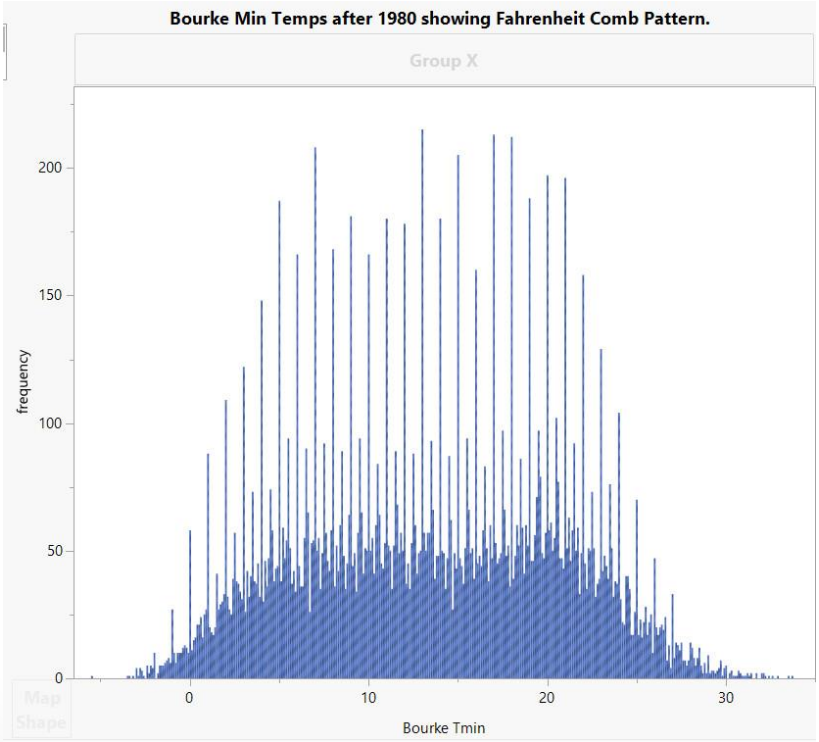
Appendix C: Persistent Fahrenheit Conversion "Ghost" Signatures

Structural analysis of the raw record reveals evidence of historical conversion from Fahrenheit to Celsius that has left persistent artefacts:


Cape Otway Min temps after 1980 with Fahrenheit Comb Pattern.
Group X
3000
2000
1000
0
frequency
0
10
20
30
otway_min
Map Shape


Bathurst Min Temps after 1980 with Fahrenheit comb pattern
Group X
250
200
150
100
50
0
frequency
-10
0
10
20
Bathurst Tmin
Map Shape


Palmerville Min Temps after 1980 showing Fahrenheit Com Pattern
Group X
frequency
600
500
400
300
200
100
0
0
5
10
15
20
25
30
Palmerville Tmin
Map Shape


Bourke Min Temps after 1980 showing Fahrenheit Comb Pattern.
Group X
frequency
200
150
100
50
0
0
10
20
30
Bourke Tmin
Map Shape

| Artefact | Description |

| Scale compression | 62 years of data confined to 18% of the available Celsius scale |

| Comb pattern | Immediately visible in any frequency plot – values cluster around whole-degree Fahrenheit conversions |

| Truncation evidence | Values show truncation rather than rounding in at least part of the record |

| Distributional distortion | Statistical properties systematically distorted by the conversion process |

This is not a minor technical footnote. The Fahrenheit ghost is baked into the foundation upon which all subsequent versions of homogenisation software (v1 through v2.6) have been built. Every adjustment, every correction, every trend calculation inherits this structural distortion.

You cannot homogenise your way out of a problem that exists in the grid itself.

This report is based on a forensic analysis of the Bureau of Meteorology's Climate Data Online (CDO) "raw" temperature data and ACORN-SAT homogenised data. All patterns are reproducible using the publicly available dataset.