Tom Berger responds

TO MEDIA WATCH QUESTIONS

We're considering a segment on Media Watch examining The Australian's coverage of the "widespread errors" in the Bureau of Meteorology's raw climate data, based on your report, of which I see you were the lead author.

Christodoulou.Mario@abc.net.au

Response to your questions of 23 October

1. Qualifications and experience

I am a computer engineer by training. I worked for ICL Computers (later Fujitsu) in Johannesburg and Sydney for 12 years, which was treated as unbroken service. My work there was in computer engineering and systems analysis. I also did 3 years contracting in the field.

Since then I have worked in computer analytics and statistical modelling. I use commercial statistical software (Salford Systems, Netica, JMP) for consulting and analysis. My deception-detection modelling was presented at the Tasmanian Prosecutors Conference in 2017 (verifiable by Tony Jacobs QC).

You may have read from my blog that that I and beat the University of Michigan lie detection benchmark with my model. It is mentioned in the FBI exchange too. I have also done computer analysis of election linguistics for a political leader. The corporate

entertainment work is a separate career path; it is not the basis of the audit.

The audit is a data integrity analysis — deduplication, precision-loss detection, and pattern analysis on a published dataset. https://en.wikipedia.org/wiki/Data_deduplication.

Those are standard computing techniques. They do not require climate science qualifications, and the claim that they do is a *category and labelling error.*

It is documented in any computer science textbook, used in every backup system in the world. It requires comparing data blocks for identity. It does not require a climate science degree, any more than auditing a bank's ledgers requires a degree in economics.

The BOM's own data is published in CSV files. I have attached some for you. The duplication is visible in those files. The question is not whether I am qualified to see it — the question is why the BOM's quality control did not.

The findings are checkable by anyone with the files, and I am happy to walk you through the method.

I would add one factual note. My site analytics show *3 visits to my deception-detection site from the ABC* which started on the *same day* the BOM made a statement.

This blog describes my computer analysis and statistical modelling — in the days before your email.

ABC also downloaded my Audit file at that time. I mention this because the email states that the only publicly available information relates to entertainment.

The technical material is publicly available and was accessed by the ABC.

2. The FBI

I was not employed or contracted by the FBI. In September 2020, Frank J. Marsh, an Instructor at the FBI Academy, contacted me to ask about my methods and to ask whether I had written a book or

training material. He stated he was *"honored and humbled"* by the information I shared and asked to share it with colleagues.

The full exchange is attached. We exchanged ideas and files several times over the years. I share it not to overstate the connection but to show that the methods have been of interest to law enforcement professionals.

3. Response to the BoM statement

Attached is my analysis of the BoM statement, and a one-page summary of key findings. In brief:

The BOM's statement says it has *'detected no further duplicate months in the entire bureau network other than those quoted in the report and then confirmed by us.'*

That claim is contradicted by the BOM's own published data, in four places, all from Sydney (station 066062) and all flagged Quality = Y.

The clearest is June 1943 and June 1944. All thirty days of June 1943 are replicated exactly in June 1944. This falls inside the pre-1946 window the BOM itself cites, so the pre-1946 explanation does not dispose of it — it includes it.

Two complete months of daily observations cannot both be original; one is a copy of the other; and the QC flag cannot be correct for both.

The second is a seam rather than a copy. In December 1935 and December 1936, fourteen days are identical, then two days appear only in 1936, then a value of 20 appears in both columns with the 1935 record carrying no days-of-accumulation flag, then fourteen more days are identical**.**

That is the signature of a block splice — a join point with a two-day insert and a malformed record at the stitch — not an isolated transcription error.

The third is November 1903 and November 1905, where all thirty days are replicated. I accept the BOM may treat this as outside the ACORN-SAT window; I include it for completeness, not as the basis of the complaint.

The fourth is a seven-day replication across the January/February 1910 boundary, in the first year of the Bureau's standardised record.

These are not *'misunderstandings'* of homogenisation. They are countable, checkable, whole-block replications in the public record, **and they falsify the specific sentence in the BOM's statement that says no further duplicate months exist.**

- The BoM characterises the errors as isolated transcription mistakes. A transcription mistake is a single wrong value. Thirty consecutive identical numbers, in the same order, two years apart, is a block copy. It is not what "isolated transcription error" describes.

- Separately, the Lost Decimals Report documents 51,249 station-days of precision loss across 23 ACORN-SAT stations between 1995 and 2011. This is a modern-era data failure, not a historical quirk, and the BoM statement does not address it.

- The audit does not allege deliberate manipulation by individuals. It documents systematic data-handling failures that are statistically significant and structurally patterned.

4. The "innocent errors" question

The question offers two options — small-scale innocent errors, or deliberate manipulation — and neither describes what the audit found. A single duplicate could in principle be innocent. The pattern of fifty block events across seventeen stations, with 42% post-1990, and a Monte Carlo result of $p < 0.0001$, is not consistent with random innocent error. Whether the cause was a software fault, a manual copy, or a digitisation process is a question for the BoM. What it is not is a random transcription slip.

I am happy to provide the full analysis, the data, or a walkthrough of the method. I am confident the findings stand up to scrutiny.

I have attached the relevant csv's that directly contradict BOM's statement.

Extra questions:

The BoM says the roughly 200 errors it identified are likely transcription mistakes from punch-card records dating to the 1960s, and are statistically irrelevant. How do you respond?

You are merging 2 statements into one causal claim.
The BOM says the duplicates are *"all from handwritten records of pre-1946 data that have been manually digitised."*

Separately, the BOM says records *"were digitised using methods including punch-card systems in the 1960s, which may have contributed to isolated transcription errors."*

You have merged these into one causal claim: the errors were caused by 1960s punch-card transcription. The BOM does *not* say that. It says the errors are from handwritten pre-1946 records, and that punch-cards may have contributed. The BOM is explicitly hedging (*"may have contributed," "is confirming if it was the cause"*).

So you have taken a hedged *maybe* and turned it into a definite cause. That is a real distortion.

The BOM says *"so small as to be statistically irrelevant."* This is the BOM's own phrase.

What your *"statistically irrelevant"* paraphrase omits:
The BOM's statement also says *"Further investigation has detected no further duplicate months in the entire bureau network other than those quoted in the report and then confirmed by us."*

You have dropped the completeness claim entirely. That is the claim my data actually falsifies. By omitting it, you have removed the one part of the BOM's statement that is checkably wrong. The BOM's claim is falsified and can be checked with the attached csv's.

A transcription mistake means: the person reading the paper record wrote down the wrong number, or the keypunch operator punched the wrong card, or a digit was transposed. It implies an isolated, random, human error — one value wrong, in one place, at one time.

That is *not* what the data shows. Latest Sydney data shows:

1. June 1943 / June 1944. Whole-month replication, inside the BOM's own pre-1946 window, both flagged Quality = Y. This is the one that falsifies the BOM's "no further duplicate months" claim without any caveat.

2. December 1935 / December 1936 (strongest evidence of a seam). 14 identical days, a two-day insert, a malformed shared value, 14 more identical days.

This is the one that shows the mechanism — a block splice — rather than just a copy.

3. November 1903 / November 1905 (strong but pre-1910). Whole-month replication, but outside the ACORN-SAT window, so the BOM can set it aside.

4. January 1910 / February 1910. A seven-day replication across the January/February 1910 boundary, in the first year of the Bureau's standardised record.ay month (Jun 1943 = Jun 1944), and a 7-day block copied from January into February 1910 are not transcription errors.

A transcription error does not reproduce thirty consecutive numbers in the same order two years apart.

That is a copy operation — a block of data duplicated as a block. **The distinction matters enormously.**

"Given the seriousness of the claims you are making, I'm seeking to better understand the expertise and experience on which they are based. Especially, in circumstances where experts in climate science and the BoM itself, have dismissed your allegations."

You say experts in climate science and the BoM itself have dismissed my allegations. I'd ask you to note two things.

First, *the BoM is the subject of the audit, not an independent assessor of it — its dismissal is the expected response of any institution to an audit of its own records, and it does not tell us whether the audit is correct.*

Second, the climate scientists who have commented have, as far as I can tell, commented on media reports about the audit rather than on the audit itself. If any of them have examined the block-count data or the precision-loss windows and found an error in the

method, I would genuinely like to see it — that is exactly the kind of scrutiny the audit should receive.

But 'dismissed' is not the same as 'assessed and found wanting,' and the difference matters.

The audit is a data integrity analysis. It identifies duplicated blocks, precision loss, and infill patterns in a published dataset. Those findings are checkable by anyone with the files — *and they do not require a climate science degree to check, any more than auditing a set of accounts requires an accounting degree.*

The question is not whether the BoM and its defenders are qualified to dismiss the audit. It is whether the data supports the audit's findings, and it does.

You (Mario) mentioned on the phone that a climate expert said a 17 day flatline at Palmerville was *"possible".*

I would be interested to know which expert said that, and what they were shown, because 'possible' and 'probable' are different questions and the audit is about the second one. A 17-day exact flatline at 0.1°C resolution is not impossible in the way that a perpetual motion machine is impossible — it is statistically

extraordinary. The Monte Carlo work in the audit addresses exactly this: the probability of a 17-day exact flatline at that resolution is on the order of 10^{-34}, and the frequency of flatlines across the network is far above what natural variation produces.

If an expert has examined the surrounding data and the rest of the network and concluded that the observed frequency is consistent with natural variation, I would genuinely like to see that analysis — that is the kind of scrutiny the audit should receive. *But **'it is possible'** is not that analysis.* It answers a different question.

There is also a second point. *The audit's finding is not that one flatline is impossible.* It is that flatlines of 10, 13, and 17 days appear across the network at a frequency that is not consistent with natural variation, *and that ACORN itself creates flatlines* — 10 days at Charters Towers and 13 days at Palmerville — from raw data that was not flat. A single 'possible' instance does not address that pattern.

Your report alleges "large-scale copy-paste operations", "10 years of one-way adjustments" and "human fingerprint" patterns.

"On 'large-scale copy-paste operations': the audit identified over 1,000 duplicated blocks, with block sizes down to five days. That is

large scale, and it is the tip of the iceberg — the audit did not analyse the whole network.

The BOM's statement that it has found *'no further duplicate months'* is therefore not a finding of absence; it is a statement that the search was not done at the block level.

On *'10 years of one-way adjustments'*: that figure is wrong, and it is wrong in a way that understates the finding. It is 110 years.

The audit sets out the annual adjustment series; the ABC has the document and can check it. A 110-year sequence of same-direction adjustments is not what physical site changes produce — site moves and screen replacements produce random, bi-directional adjustments.

A 110-year one-way series is the signature of curve-fitting, and that is why the Random Forest result matters: 70% of the adjustment variance is predicted by calendar decade alone, with no weather data input.

On 'human fingerprint' patterns: the audit documents 3+1+3, 4+1+1 and 5+1+5 structures at a scale that natural variation does not produce, over 800 value traps, and hundreds of three-day linear-regression blocks between maximum and minimum temperatures.

These are structural signatures — they are the shape that editing, infilling and carry-forward operations leave in a dataset. Whether any individual case is natural or not is a question the BOM can answer for each one; the point is that the frequency and regularity of the structures across the network is not consistent with natural variation. That is what *'human fingerprint'* was intended to convey, and the data supports it."

You told me: "I didn't get [my report] peer reviewed ... because I'm only interested in getting an audit."

The full sentence should be: *"I didn't get [my report] peer reviewed ... because I'm only interested in getting an audit, it's not an Academic Report."* Again, detail matters.

"On approaching the BOM for comment:

I did not publish the findings. I sent the report to my federal Member of Parliament, Andrew Wilkie MP. Mr Wilkie forwarded the report to the Bureau, together with questions for the Bureau to respond to. The Bureau's reply to Mr Wilkie, signed by Dr Stuart Minchin, is dated 29 September 2026 and is on the public record. The Bureau

thanked me twice for my findings. So the Bureau was approached through the appropriate parliamentary channel, it responded in writing, and that response is precisely what my analysis of the BOM statement addresses.

The characterisation that the BOM was not approached is incorrect.

The parliamentary channel is the appropriate one for a constituent raising a matter with a government agency, and it ensures the response is formal and on the record.

You asked me twice about my FBI correspondance, it is attached. You asked about my Computer Employment history, it follows at the end of the pdf.

For the avoidance of doubt: I intend to publish this exchange — your questions and my answers — in full on my blog, after your deadline has passed and after the segment has aired. I am telling you this now so there are no surprises, and so the record of what I was asked and what I said is available to anyone who wishes to check it. I have no objection to you quoting from my answers; I would simply ask that any quotation be in context, and publishing the full exchange is the simplest way to ensure that.