New Methods for Remodelling Historical Temperatures: Admirable Beginning using AI

A reasonable person might assume that past temperatures – by their very nature of being in the past – cannot be changed. But in climate science, historical temperatures are continually homogenised ostensibly to correct for changes in equipment and its location, and with methods that have a subjective component. A new report by Jaco Vlok, from the University of Tasmania, details an alternative approach using artificial neural networks (ANNs), which are a form of artificial intelligence (AI). The technique can be used for both reconstructing past temperatures and also for infilling missing values.

As Dr Vlok explains on page 79, ANN are particularly applicable to solving the problem of temperature reconstructions in that they have application in function approximation, probability estimation, pattern recognition and prediction.

The output from an ANN will be totally dependent on the quality of the data inputted. In mainstream climate science, and many other areas of science, solutions are found based on mathematical formulae derived from theory. This is not the case when using ANNs, as explained in Appendix 2 with respect to rainfall forecasting.

The ANN technique detailed in this report may have more application for temperature data derived from the United States and Indonesia than from Australia. This is because US and Indonesian measurements from automatic weather stations (AWS) are properly calibrated to international standards. The situation with Australian temperature measurements is somewhat unique, in particular since 2011/2012 there has been no averaging of the one-second spot readings from electronic probes in AWS as detailed in Appendix 1. This means temperature may be reading up to 0.4 degrees hotter for the same weather.

Reasons for Scepticism

A most dramatic example of historical revisionism is the IPCC acknowledging in its first Assessment Report (AR1) that it was about a degree warmer just 1,000 years ago during the Medieval Warm Period. Then removing this warm period from the 2,000 year-long proxy temperature reconstruction in its third Assessment Report (AR3) – remember the acclaimed 'hockey stick'.

This flattening of the proxy temperature record was apparently justified on the basis of remodelling by Michael Mann: remodelling that Mark Steyn described as "fraudulent". That claim resulted in a defamation action being brought by Mann against Steyn, with the dispute so far unresolved despite eight years passing since it was first lodged in the District of Columbia Superior Court, Washington DC.

John Abbot and I published on the application of ANNs for resolving such disputes concerning the reliability of proxy records for historical temperature reconstructions, and the extent to which recent warming may be natural. There was a backlash against our findings, and much lamenting of our use of ANNs.

More recently, in fact just a few months ago, Rowan Dean used the word “fraudulent” on Australia’s Sky TV to describe the Bureau of Meteorology’s remodelling of Australia’s temperature history as recorded at official weather stations by mercury thermometers and more recently electronic probes at automatic weather stations (AWS).

As Graham Lloyd explained about one month earlier in The Weekend Australian newspaper: for the second time in six years the rate of warming has been dramatically increased – by 23 percent between versions 1 and 2 of the official temperature reconstructions for Australia (ACORN-SAT). In the newspaper article Graham Lloyd used my example of Darwin, in northern Australia, to illustrate the nature of this remodelling by the Bureau.

It is one thing to continue to object to current methods and rally against them, but what is ultimately needed is an alternative method for historical temperature reconstructions.

Revisionism Can be Complicated

The revisionism is not denied by the IPCC or the Bureau of Meteorology. Rather it is justified on the basis that it is claimed to be necessary to make changes to past temperature measurements through the process of homogenisation because of equipment changes and site moves.

Gavin Schmidt – the director of the Goddard Institute for Space Studies (GISS), which is part of NASA and responsible for reconstructions using Australian data incorporated into IPCC reports – has explained to me through a public exchange on Twitter that homogenisation involves procedures, specifically algorithms, using temperatures as measured at nearby locations that may be up to 1,000 km away, to correct for discontinuities in the target temperature series.

There is also the issue of missing values and area weightings, with the Bureau calculating temperature anomalies based on daily and monthly gridded data with more than one station contributing towards each value at each grid point with no specific set of weights attached to these. The effective contribution from each weather station thus changes on a daily or monthly basis, depending on which stations did or did not report on any given day or month.

To be clear, both NASA and the Bureau’s method for historical temperature reconstructions is complicated and cannot be replicated. This does not necessarily mean it is fraudulent, or easily disputed, particularly when the theory of human-caused catastrophic global warming (CAGW) has the backing of our most esteemed scientific institutions from NASA in New York to the Royal Society in London. But it does mean that those who are really curious about temperature change over the last 100 or so years and want to be sure reconstructions at the national and global level are reliable, need an independent method for checking.

An Alternative Method

Given all the potential issues with NASA and the Australian Bureau of Meteorology's reconstructions, I have pondered "what if" there were a method for combining individual series in a transparent way such that the resulting historical reconstruction was:

1. Based on actual temperatures as measured, rather than remodelled/homogenised series.

2. Incorporated only temperature measurements from remote and regional locations, and thus *excluded* temperatures measured in cities that have become hotter with more roads, tall buildings and air conditioners as the cities have grown larger. This is known as the Urban Heat Island (UHI) effect.

   The Australian Bureau of Meteorology claims to not use temperature series from cities when it estimates Australia-wide climate change because of the UHI effect, but it does. This is because the temperatures as measured in inner city Sydney and Melbourne have been used to remodel the actual temperature measurements from remote and rural locations. For example, in ACORN-SAT version 1, which has been the official temperature reconstruction for Australia up to, and including last year, the temperature series from Cape Otway lighthouse is changed/remodelled/homogenised based on data from Melbourne.

3. Used only on data recorded from mercury thermometers, and *excluded* measurements from the Bureau's new automatic weather stations (AWS), which since 2012 have been recording maxima based on one-second-spot readings from electronic probes. This is an issue that is likely to be unique to Australia.

I was hopeful that such a reconstruction might be achieved through the work of Jaco Vlok at the University of Tasmania. Three years ago, he set out to generate an historical temperature reconstruction for Australia based on the latest artificial neural network technology (ANNs), which is a form of AI.

The new report by Jaco Vlok goes into much detail explaining and demonstrating how an ANN can be used to generate an accurate historical temperature series of data points for a specific location, without actually having any measured values/any data for that location. Specifically, he uses Deniliquin as a case study to show the capacity of an ANN to approximate from other data.

Deniliquin as a Case Study

Deniliquin is a regional centre in New South Wales with a very long continuous temperature series, surrounded by many other locations also with long continuous temperature records. It is thus a good place to start.

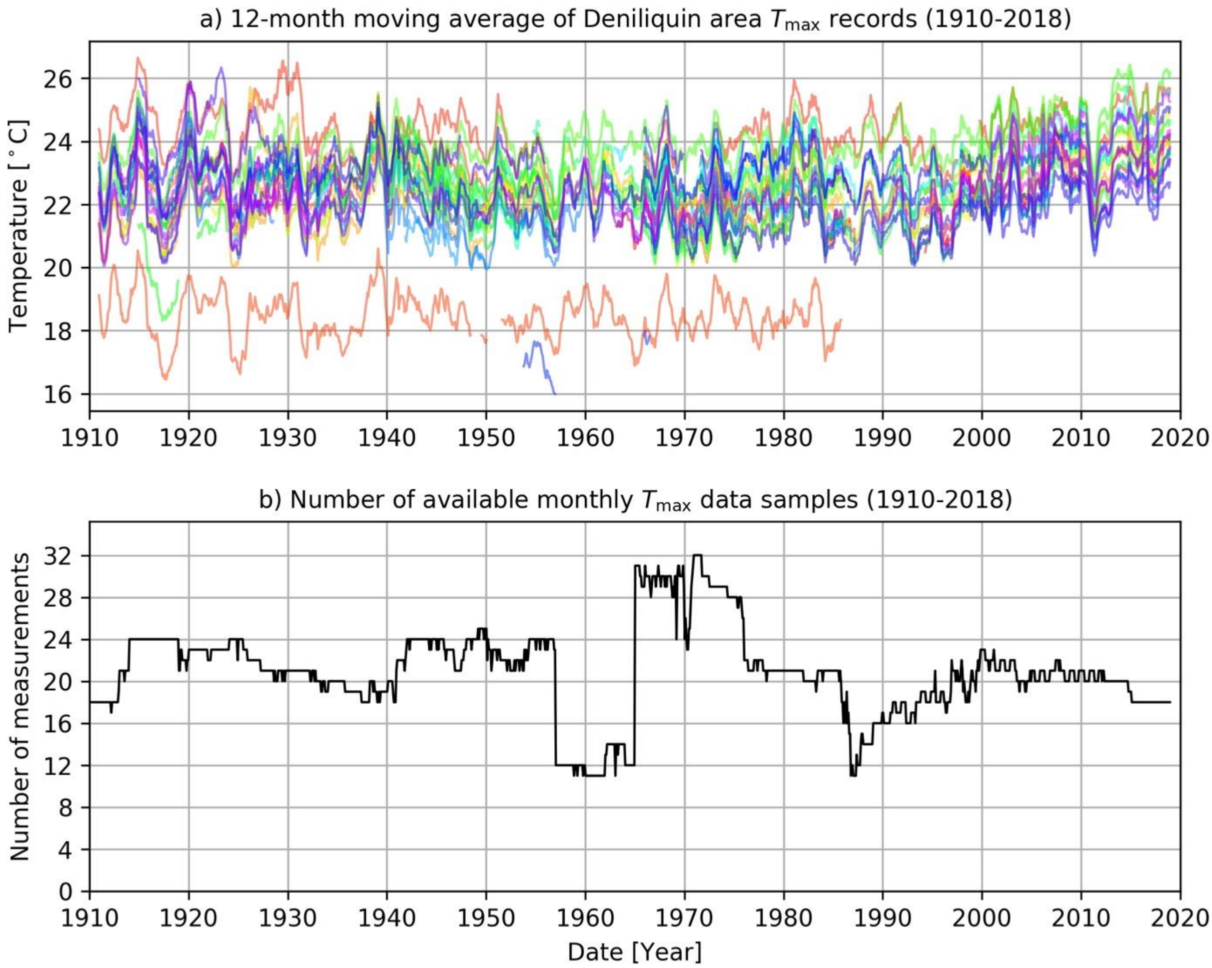


Figure 50. Monthly mean maximum temperature series at all 71 locations used to recreate Deniliquin's temperature profile.

Dr Vlok undertakes a reconstruction for the location of Deniliquin using temperature records from a large number of weather stations in the Deniliquin area – but not Deniliquin as such (see page 96). Specifically, he shows how an ANN can skilfully recreate an historical temperature profile for Deniliquin *without* actually using any temperature measurements from Deniliquin (see page 100).

Jaco Vlok's new series for Deniliquin can be compared with the actual measurements for Deniliquin, and a skill score generated to see how accurate the ANN is in its modelling of historic temperatures.

In his new report, Jaco Vlok goes into some detail showing how this reconstruction for Deniliquin can be further optimised. But this is already an exceptionally skilful reconstruction made possible by the many surrounding rural and regional locations with long temperature series. What is needed, if this ANN technique is to have practical application for historical temperature reconstructions more generally, is for the ANN to be able to recreate series where there are significant amounts of missing data and a limited number of surrounding stations – the more usual situation.

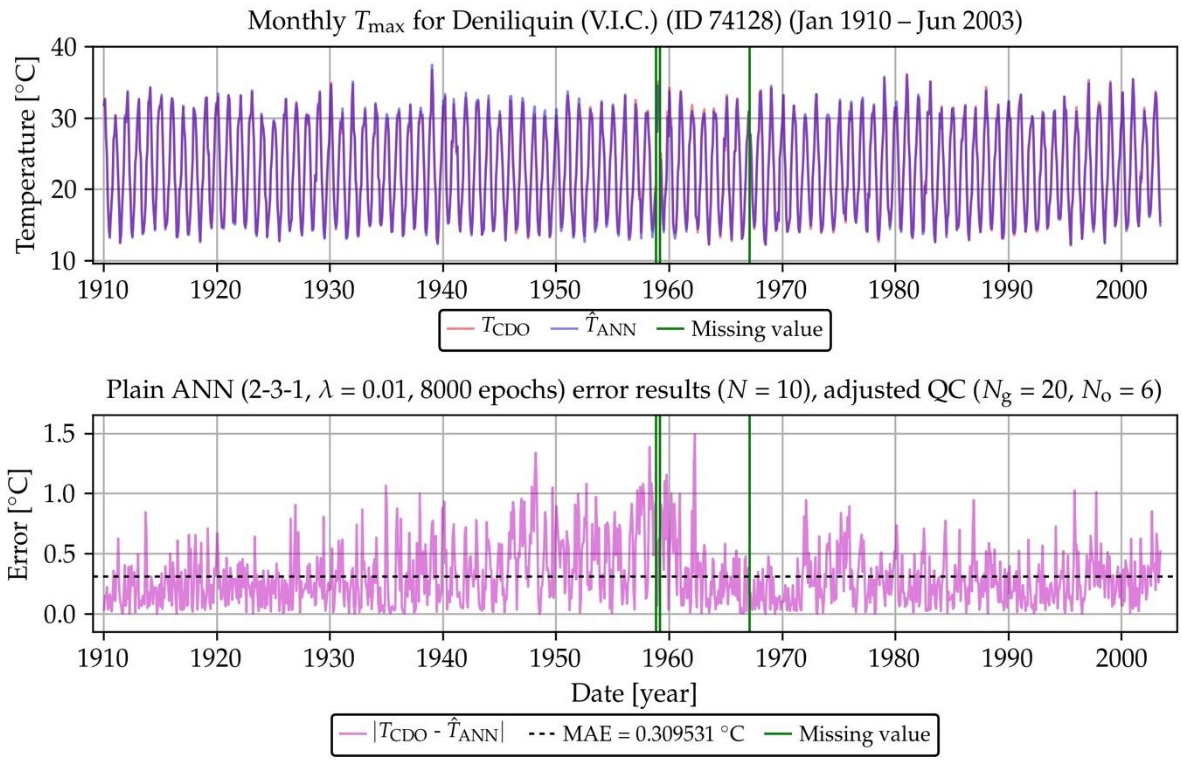


Figure 58. Estimation results for Deniliquin using ANN.

Computational Propaganda

It is the case that ANNs can generate content by mimicking and recreating patterns in such a way that even the expert may be confused about what might have been – as opposed to what the computer can simulate. Indeed, at one level this 'infilling' using AI solves many problems, as I explain in more detail in the next section. At another level it shows the potential for deception when AI is eventually applied to historical temperature reconstructions – and this will eventually happen, whether we like it or not.

Sean Gourley gave a talk about this issue dubbed 'computational propaganda' at a recent AI conference in New York. He explained how algorithms can be used to generate photographs of all types of apparently real people, except they don't exist: blurring the lines between what is real and what is fake. This is the same ANN technology used by Uber in driver-less cars, and by Facebook to determine which advertisements to deliver onto your unique Facebook page, and by Google in language translation applications. It's relevance to climate science more generally is detailed in Appendix 2.

Improving Temperature Series

Dr Vlok's 'invented' series for Deniliquin might be praised as the ultimate in 'infilling' – should you have ever been as frustrated by missing values when working with historical temperature series, as I have.

Over the last year, I worked with the Indonesian Bureau of Meteorology (BMKG), evaluating options for forecasting monthly and seasonal rainfall using historical data and ANNs. A limitation to more skilful forecasts is often the quality of the temperature and rainfall data with significant gaps/missing data – consistently in the 1950s.

The development of this technique for skilful infilling could thus significantly improve the forecasts for droughts and floods by extending the length of the arrays available for inputting.

My rainfall forecasting work for Australian locations with John Abbot using ANNs has demonstrated that the longer the temperature and rainfall series available for input, the more skilful the forecast. So, it follows this new technique for infilling could be applied to enable and/or improve monthly rainfall forecasting for locations where the availability of temperature and rainfall series is currently a limitation.

Contrasting Methods for an Australia-wide Infilling

Jaco Vlok joint authored a chapter with me in the best-selling book 'Climate Change: The Facts 2017'. In that chapter, in which we use the state of Victoria as a case study, we show how simply combining all the data from every series from 1910 will not necessarily give an accurate representation of climate change because the number of sites in the hot Mallee region has decreased, while the number of stations in cooler alpine areas has increased. So, even if there has been no overall change in the climate, linear regression through the mean of all the stations shows an overall cooling trend.

A solution to avoid the bias from the changing distribution of the stations would be to use all the alpine and Mallee series from the beginning to the end but first infill for the missing values. This is what Dr Vlok does in the new report, in order to create an Australia-wide infilling.

The ANN technique for infilling using nearest-neighbours has been undertaken without adequate quality assurance of the data, and without excluding UHI affected locations or correcting for equipment changes. So, Dr Vlok's reconstruction has all the limitations of the official ACORN-SAT reconstruction undertaken by the Bureau, and perhaps for this reason is so similar.

An alternative approach to using the nearest neighbour method for infilling across every available series as Dr Vlok has done, would be to create an ANN-reconstruction for Australia based only on rural and regional locations and excluding measurements from AWS. This would require the individual temperature series to be first segmented and categorised by the equipment used and also by population.

Because so many of the measurements since 2012 from Australia are from electronic probes recording in a way that is not consistent with calibration (see Appendix 1), this approach may prove impractical – for Australia. It may be better to demonstrate the technique using data from rural and regional United States because there are many long and continuous raw temperature series that have integrity – in so much as the one-second readings from electronic probes have been averaged over a five-minute period.

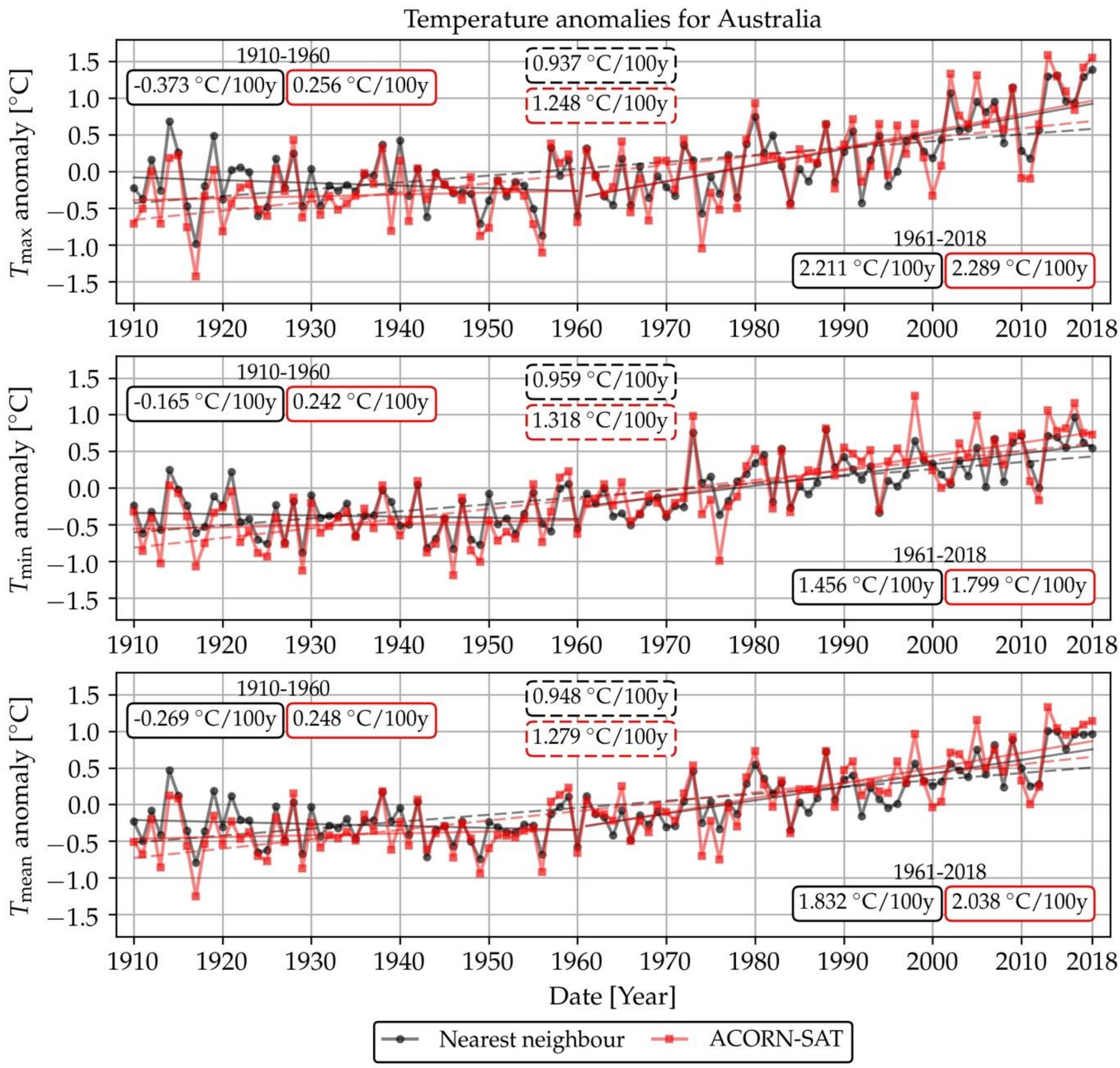


Figure 42. Average anomalies with trends indicated.

Quality Assurance Issues

Towards the end of the new report Dr Vlok acknowledges both issues with the raw temperature data in terms of UHI and also the calibration issues with measurements from the AWSs. Dr Vlok has used the nearest neighbour technique for infilling for his Australia-wide reconstruction and this will potentially firmly embed an artificial warming trend from the AWS and also UHI.

Dr Vlok suggests under 'Future Work' that a way-around this could involve "excluding AWS measurements". He goes on to acknowledge that this will involve removing a significant amount of recently-recorded data from the analysis. Hindsight is a great thing.

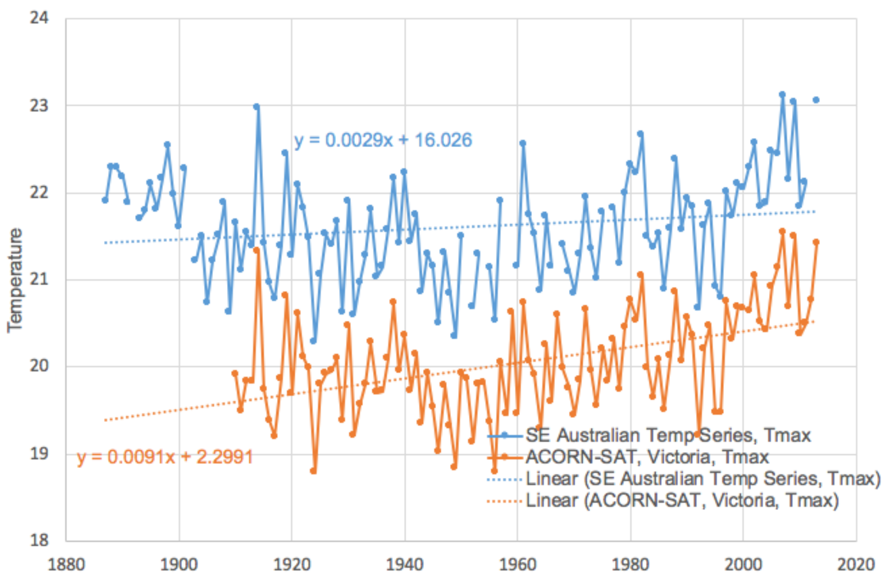
Anyone beginning such a study should note his very last dot point (page 109) where he writes: Further details are obtainable in the basic climatological station metadata, which should be investigated individually for each weather station to uncover potentially important information regarding site moves. He provides a link to the metadata for Deniliquin.

This is not only important information for finding otherwise 'undocumented' site moves, but also important for knowing when there were major changes in the actual type of equipment used to record temperatures; including the specific date the transition was made to an AWS, and any subsequent changes in the type of electronic probe – all affecting the potential to record warmer days for the same weather.

Dr Vlok does some limited comparisons with a quality assurance technique that I have been promoting since at least 2016, when I published a very detailed report on Rutherglen. Specifically, I favour the use of control charts. He does not adequately compare this method with his preferred nearest neighbour method or show how an ANN reconstruction might be different if control charts rather than the nearest neighbour technique is used for an Australia-wide reconstruction.

Those familiar with the Bureau's infamous reconstruction of Rutherglen's minimum temperatures for ACORN-SAT versions 1 would know that Blair Trewin justifies turning a cooling trend into warming through QA using a similar nearest neighbour technique … except Rutherglen's nearest neighbours all show cooling, like Rutherglen. All pretence is apparently discarded with ACORN-SAT version 2, and the Bureau just cools the past – even more.

In a book chapter that I co-authored with John Abbot that was published by Elsevier in 2016 control charts were used for quality assurance, and then combined with a straightforward area weighting to calculate annual average temperatures for south eastern Australia from 1887.



The blue line is my weighted-area average for south eastern Australia from 1887 to 2013, based on work undertaken back in 2014, and subsequently published in a book edited by Don Easterbrook.

At the time, I was unaware of the issues with the AWS data for Australia (see Appendix 1), and so did not correct for, or exclude measurements from electronic probes. This issue may have affected the last one or two years of temperature recordings in this study.

Even without considering the AWS issue, our quality assurance reduced the number of locations with suitable data to just five. We then applied a straightforward weighting method to generate a temperature trend for the entire south east of Australia. This weighted mean of the five highest-quality maximum temperature time series shows statistically significant cooling of −1.5°C per century from 1887 to 1950, followed by relatively rapid warming of 1.9°C per century to 2013. It needs updating.

In Conclusion

When Jaco Vlok first started at the University of Tasmania there were plans to work on an Australia-wide temperature reconstruction using temperature series that the Bureau considers to be the highest quality – because it is homogenised: that is the ACORN-SAT series drawn from 112 stations. Full marks to Dr Vlok for not taking the easy path and using these already remodelled temperature series from the ACORN-SAT database.

Instead over the last three years Dr Vlok has worked with raw data. His mistake perhaps has been to not be more sceptical of this data and realize that individual raw series vary in quality: that within each series with the same ID number there are embedded measurements recorded in very different ways (mercury thermometer versus electronic probe) and that over time the environment at the one location may have significantly changed (e.g. Melbourne has grown into a large city, Rutherglen has become irrigated).

His new report, humbly entitled 'Temperature Reconstruction Methods', does shine a light in an important new direction for understanding climate variability and change, demonstrating the potential value of AI for historical temperature reconstructions.

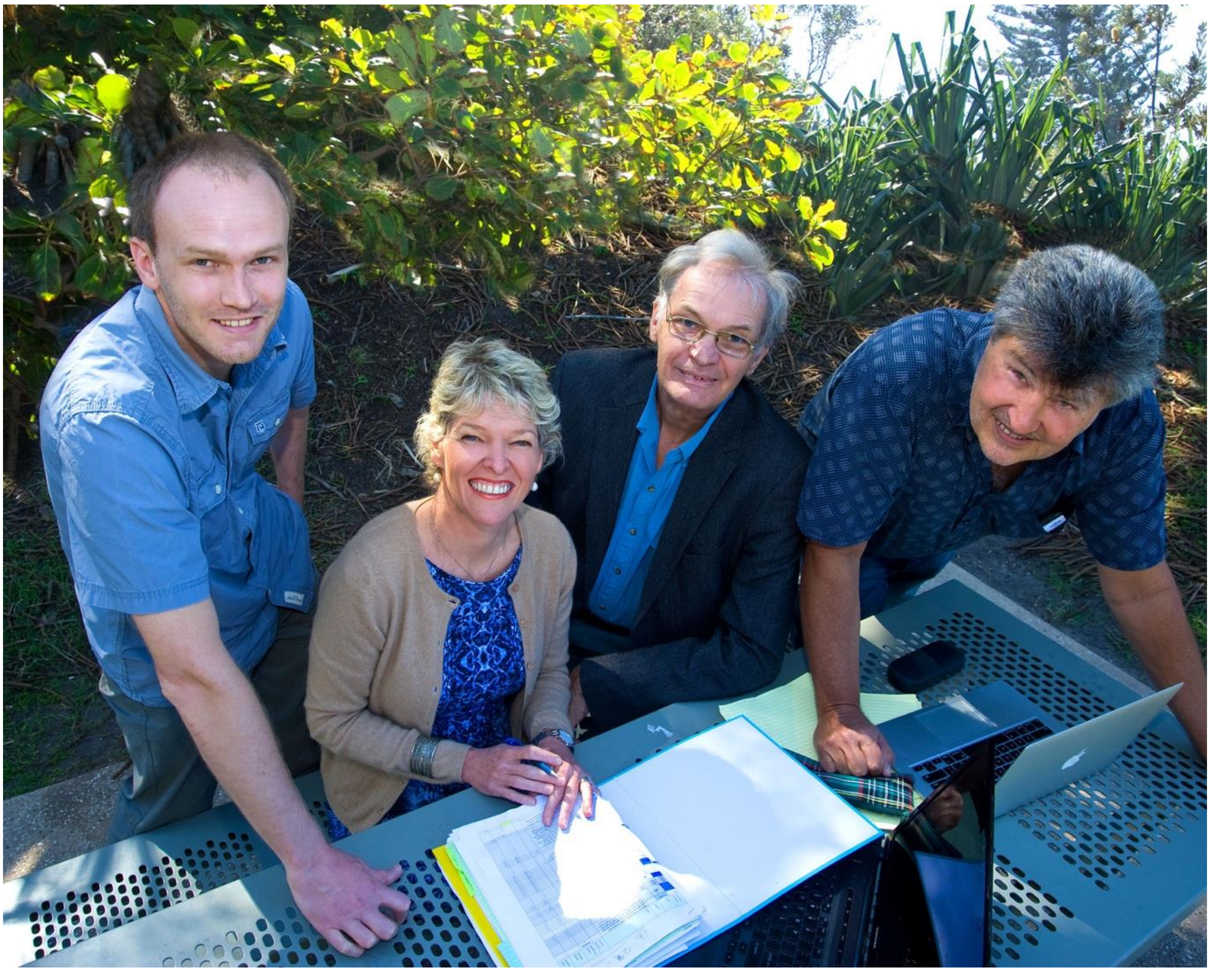
Dr Vlok has identified and documented major limitations in the Bureau's recording and archiving of historical temperatures. This is a particular problem for ANNs that are totally data driven. Any next attempt at an Australian-wide reconstruction using AI needs to be cognisant of these issues – particularly the change to electronic probes in AWS and also the UHI effect.

The ANN technique that Dr Vlok has detailed has application beyond temperature reconstructions with their political dimension and could make the use of ANN for rainfall forecasting more generally applicable including for regions that currently lack long and continuous rainfall and temperature series.

Dr Jennifer Marohasy
17 May 2019. -26.408474, 153.073310

****

If you have an interest in sponsoring future work where ANN technology is used to address the most pressing problems in climate science, consider getting in touch with me at j.marohasy@climatelab.com.au



The Blue Team: Jaco Vlok, Jennifer Marohasy, John Abbot and JC Olivier (left to right) discussing temperature reconstructions in Noosa.

____

Appendix 1.

THE BUREAU ABANDONDED ONE MINUTE AVERAGING IN 2011/2012

Historically maximum air temperature was measured by mercury thermometers. But over recent decades there has been a transition to electronic probes in automatic weather stations.

There is a lot of natural variability in air temperature (particularly on hot sunny days at inland locations in Australia), which was smoothed to some extent by the inertia of mercury thermometers. In order to ensure some equivalence between measurements from mercury thermometers and electronic probes it is standard practice for the one-second readings from electronic probes to be averaged over a one-minute period – or in the case of the US

National Weather Service the averaging of the one-second readings is over a 5 minute period.

The Australian Bureau began the change-over to electronic probes as the primary instrument for the measurement of air temperatures in November 1996. The original IT system for averaging the one-second readings from the electronic probes was put in place by Almos Pty Ltd, who had done similar work for the Indian, Kuwaiti, Swiss and other meteorological offices. The software in the Almos setup (running on the computer within the on-site shelter) computed the one-minute average (together with other statistics). This data was then sent to what was known as a MetConsole (the computer server software), which then displayed the data, and further processed the data into 'Synop', 'Metar', 'Climat' formats. This system was compliant with the World Meteorological Organisation (WMO) and the International Civil Aviation Organisation (ICAO) standards. The maximum daily temperature for each location was recorded as the highest one-minute average for that day.

This was the situation until at least 2011– I have this on good advice from a previous Bureau employee. It is likely to have been the situation through until perhaps February 2013 when Sue Barrell from the Bureau wrote to a colleague of mine, Peter Cornish, explaining that the one-second readings from the automatic weather station at Sydney Botanical Gardens were numerically-averaged. At some point over the last seven years, however, this system has been disbanded. All, or most, of the automatic weather stations now stream data from the electronic probes directly to the Bureau's own software. This could be an acceptable situation, except that the Bureau no-longer averages the one-second readings over a one-minute period.

Indeed, it could be concluded that the current system is likely to generate new record hot days for the same weather –because of the increased sensitivity of the measuring equipment and the absence of any averaging/smoothing.

To be clear, the highest one-second spot reading is now recorded as the maximum temperature for that day at the 563 automatic weather stations across Australia that are measuring surface air temperatures. This is not generally understood. Most meteorologists and university professors in Australia appear to be working from the wrong assumption that the old system is still in place. Given this data is also used by thousands of other scientists and technologists, not just in Australia but across the world, this needs investigation.

My assessment is based on scrutiny of actual measurements from the probe at Mildura, in north western Victoria, and also January 2019 data from Canberra airport – that I am yet to publish.

The data from Mildura was made available to me following a directive from the then Minister for the Environment, Hon Josh Frydenberg MP, to Andrew Johnson, CEO and Director of Meteorology at the Bureau. This has enabled me to confirm that the automatic weather station at Mildura is logging: The last one-second reading in each one-minute period; The highest one-second reading for the previous 60 seconds, and the lowest one-second reading for the previous 60 seconds.

I have corresponded with the Bureau's CEO, Andrew Johnson, about the situation. He has assured me that because the electronic probe is housed in a metal sheath which provides thermal mass, each measurement is actually the integration of the previous 40 to 80 seconds. I have requested the manufacturer's specifications, specifically for the probe at Mildura (Rosemount ST2401 S/N –654). Dr Johnson has not provided this information, insisting that this is not available because the probes are purpose-designed: "The Bureau purpose-designed the temperature sensors to closely mirror the behavior of mercury in glass thermometers, including the time constant. The manufacturer then manufactured the sensors to the Bureau's design."

The policy of the Bureau of Meteorology is that when there is a change of equipment, parallel measurements must be taken for a period of at least three years, preferably five. At most locations were there has been a change there are no parallel measurements – where there are measurements these are accessed with difficulty. For example, when an AWS with an electronic probe was installed at Rutherglen on 29 January 1998, the mercury thermometer was removed on the same day. If the mercury thermometer and electronic probe had been left in the same screen at Rutherglen, and parallel measurements taken for a period of time, it would now be possible to calculate the equivalence of the measurements from the different measuring devices.

There are no publicly available specifications for the custom-built electronic probes currently used by the Bureau to measure air temperature across Australia. A report published in 2012 shows some pictures of the first probes developed for the Bureau.

There are no published studies that provide any indication of the equivalence of measurements from the electronic probes with mercury thermometers.

More recently, in correspondence to David Coad from Sydney after he queried the lack of correspondence between a claimed heat wave in Canberra and temperature data that he purchased from the Bureau – the Bureau explained on 22 March 2019 (Ticket J2EL264594) that:

"The data you were provided with were the 1 minute air temperature observations. These are effectively instantaneous measurements and do not capture the temperature between each observation. It is most likely that the maximum temperature occurred between the observed times. I have attached another dataset to this email, which includes one minute air temperature, along with the one minute maximum temperature and the one minute temperature. The maximum temperature field should capture the absolute maximum for each day."

I will upload the different datasets to the Climate Lab Pty Ltd website in due course.

Appendix 2.

ANN ARE TOTALLY DATA DRIVEN

A reasonable test of the value of any scientific theory is its utility – its ability to solve some particular problem. For example, the invention of the incandescent light bulb in 1870 followed the development of a practical theory of electricity, with lighting becoming one of the first publicly available applications of electrical power.

There has been an extraordinary investment in scientific research into climate change over the last three decades, yet it is unclear whether there has been any significant improvement in the skill of weather forecasting, and specifically the capacity of climate scientists to forecast droughts and floods.

Mainstream climate scientists, and meteorological agencies, generally rely on simulation modelling for their forecasts, and have been dismissive of the potential application of AI. A simulation model attempts in every instance to mimic actual physical processes from a first principles understanding of atmospheric physics and chemistry.

Some forecasters claim it has become harder to forecast weather and climate because of anthropogenic global warming (AGW), while others claim that weather is inherently chaotic and that it will never be possible to forecast more than a few days in advance – others claim weather forecasting has improved over recent decades but provide no empirical evidence.

Since June 2013, the Bureau has used output from the simulation model, POAMA. POAMA is a global coupled ocean-atmosphere ensemble seasonal prediction system developed jointly by the Australian Bureau of Meteorology and the Commonwealth Scientific and Industrial Research Organisation (CSIRO).

Forecasts from POAMA are provided in a two-category-format of above or below median rainfall. For example, in May 2016 the Bureau issued a three-month rainfall forecast for a region known as the Murray Darling Basin indicating that there was a 70-80% chance of above median rainfall for the period June to August. Farmers were pleased at the prospect of good rains, only to experience record flooding with crops washed-away on many farms. There has been no quantification of the level of skill of POAMA, including relative to the earlier statistical method used by the Bureau.

According to Microsoft founder Bill Gates, any community can achieve incredible progress if it is prepared to set a clear goal and find a measure that will drive progress towards that goal. What would be an appropriate measure of a skillful weather forecast? What are appropriate goals for weather and climate forecasting?

Should all forecasts be delivered in a deterministic form, for example, as the amount of rain forecast by the model for a specific interval – with some error values?

Arguably the world's most widely used atmospheric model is the Weather Research and Forecasting (WRF) model. Like POAMA, WRF is a simulation-type model used for weather prediction and research. The model has facilitated the development of a community of 39,000 users in 160 countries with annual get-togethers. A 2017 review paper concludes that future research concerning the development of this model will focus on further improvements in the representation of physical processes and particularly their integration into an all-encompassing Earth system model.

But will this result in improved weather and climate forecasts? Indeed, how can the community that uses and promotes the WRF model know that any of the model's simulations are an accurate representation of weather and climate processes?

Conventional simulation weather models, for example WRF and POAMA, rely on rules that attempt to simulate the actual physical weather system, they are not data driven. In the early days of AI, the expectation was that a computer would learn how to apply rules. For example, in the early 1960s computers using AI were expected to provide language translation services after learning rules of grammar and using lists of individual words already translated.

Then in the late 1980s, with the renewed interest in ANNs, a team at IBM threw out the grammar rule books and attempted to generate translations between English and French by providing computers with only examples, specifically the Hansard from the Canadian parliament, which is available in both English and French. The ANNs were given no information about the meaning of individual words, or the rules of grammar. Instead the ANNs had to rely on finding patterns through statistical modelling.

This is now a common method for language translation. Google Translate, which supports over 100 languages and serves over 500 million people daily, is based on a statistical model. It was developed by providing ANNs with the United Nations and European Parliament transcripts. Rather than translating language directly, Google Translate first translates text to English and then to the target language. Sometimes it gets translations wrong, and it is apparently better at translating European languages to and from English, which is not surprising given it has learnt from these examples.

When you think about it, this is how we all learnt our mother tongue – by example.

Yet some high-profile linguists, such as Noam Chomsky, claim that this approach using statistical modelling for language translation is essentially misguided because in the process we are not caring about the rules of language – or linguistic concepts. In particular, according to Chomsky we are not learning about how language works.

A very similar criticism is levelled at ANNs by mainstream climate scientists. For example, Francis Zwiers and Hans Von Storch have claimed that ANNs cannot improve our ability to 'synthesise knowledge' and therefore have limited value in weather and climate forecasting.

It could be the case, however, that farmers and others dependent on rain for their livelihoods care less about how climate scientist's synthesise knowledge and more about the accuracy of the weather forecasts.

Ends.