# Weather & Climate

JOURNAL OF THE METEOROLOGICAL SOCIETY OF NEW ZEALAND (INC.)

VOLUME 40 • ISSUE 1 • 2020

# WEATHER & CLIMATE

JOURNAL OF THE METEOROLOGICAL SOCIETY OF NEW ZEALAND (INC.)

VOLUME 40 ● ISSUE 1 ● 2020

**ISSN 0111-5499**


Weather and Climate is the official journal of the Meteorological Society of New Zealand (Inc.). Any opinions, statements or recommendations expressed in this journal are those of the respective author(s) and do not necessarily reflect the views of the Meteorological Society of New Zealand (Inc.).

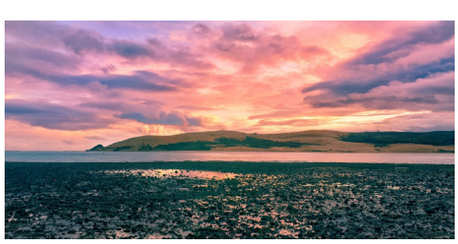
Contributions to Weather and Climate are welcome on any meteorological or climatological subject, but preference will be given to contributions related to New Zealand and/or the Southwest Pacific.

Submissions can be made via Scholastica at: https://weather-and-climate.scholasticahq.com/ or by emailing the editor at: nava.fedaeff@niwa.co.nz


Published: November 2020

EDITOR Nava Fedaeff

COVER IMAGE Hannah Court

*A break in the rain - Opononi, Northland - 20th June 2020*

## THE METEOROLOGICAL SOCIETY OF NEW ZEALAND (INC.)

The Meteorological Society of New Zealand (Inc.) was inaugurated in October 1979. The objects of the Society are to encourage an interest in the atmosphere, the weather and the climate, particularly in the New Zealand and South Pacific regions. Membership is open to all those with an interest in the objects of the Society. Membership comprises both professionals and amateurs including meteorologists, climatologists, geographers, hydrologists, yachting and tramping enthusiasts, glider pilots, people involved in aviation and marine industries, agriculturalists, professional weather forecasters, ecologists, economists, farmers, engineers and weather observers.

**Membership fees:**

| | |
|---|---|
| Ordinary members | $40.00 |
| Student members | $20.00 |
| Institutional members | $120.00 |
| Overseas postage surcharge | $15.00 |

All members receive copies of Weather and Climate and the quarterly Newsletter (which includes details of significant weather events in the previous period).

**Correspondence:**
The Secretary
Meteorological Society of New Zealand
PO Box 6523, Marion Square, Wellington, 6141, New Zealand
Web site: https://www.metsoc.org.nz/

**Officers of the Society for 2020**

| | |
|---|---|
| President | Michael Martins |
| Immediate Past President | Sylvia Nichol |
| Auckland VP | Petra Pearce |
| Hamilton VP | Tim Gunn |
| Wellington VP | *Vacant (President is from Wellington)* |
| Christchurch VP | Jiawei Zhang |
| Dunedin VP | Daniel Kingston |
| Secretary | Katrina Richards |
| Treasurer | Gregor Macara |
| Circulation Manager | Lisa Murray |
| Journal Editor | Nava Fedaeff |
| Newsletter Editor | Bob McDavitt |
| Website Liaison | Tim Gunn |
| General Committee | Silvia Martino, James Renwick, Ethan Dale |

# WEATHER & CLIMATE

JOURNAL OF THE METEOROLOGICAL SOCIETY OF NEW ZEALAND (INC.)

VOLUME 40 ● ISSUE 1 ● 2020

---

# The 1950 Albert Park thermometer screen change: a critical review of previous work

A. M. Fowler[1]

[1]*School of Environment, University of Auckland, Private Bag 92019, Victoria Street West, Auckland 1142, New Zealand.*

Correspondence: a.fowler@auckland.ac.nz

**Key words:** climate change, Albert Park, homogeneity analysis, thermometer screens

## Abstract

In late 1950 the thermometer screen in Auckland's Albert Park meteorological recording site was replaced. Thirty years later, as the issue of global warming began to emerge, the first shots were fired in an ongoing public debate about the magnitude of 20th century warming in the New Zealand region. The long-term Albert Park record has been part of that debate, including (strongly disputed) claims that a large part of ~0.5°C of mid-century warming was simply an artefact associated with the 1950 screen change. This paper summarises and critically reviews the peer-reviewed work related to this issue. The screen artefact argument is rejected as flawed and not sustainable in the light of available information, which indicates that the screen change had only a minor, probably positive, impact on calculated mean surface air temperature. However, it is plausible that an additional undocumented pre-1950 screen change occurred, with potentially significant implications for analysis of early 20th century daily minimum and maximum temperatures and the associated diurnal cycle.

## 1. Introduction

Based on analysis of several land-station records, Salinger and Gunn (1975) presented some of the first evidence of 20th century warming in the New Zealand region. A key finding was that "fluctuations over the whole region are in-phase, though of different amplitudes" (p. 397); this derived from 5- and 20-year running means of surface air temperature for four sites (Auckland, Christchurch, Waimate, Campbell Island). It is noteworthy that three of the four sites experienced a notable temperature increase of about 0.5°C from the mid-1940s to the mid-1950s, the exception being Campbell Island, which had a similar pattern but of reduced amplitude. Salinger and Gunn noted that the sites they used had "minimal site changes" (p. 396), except Auckland where (unspecified) corrections were applied.

Partly in response to Salinger and Gunn (1975), Hessell (1980) explored whether the apparent mid-20th century New Zealand warming might be an artefact caused by inhomogeneities introduced as a result of evolving site characteristics (shelter and urbanisation) and/or abrupt screen changes. He contended that these changes would all have increased temperature and, based on analysis of other sites not impacted by such changes, concluded that there was "…no important change in annual mean temperature since 1930" [up to the late 1970s]. This conclusion has partly fuelled an acrimonious debate related to the so-called 'seven-station series', originally

*Submitted: 13 January 2020*
*Accepted: 29 June 2020*

developed by Salinger (1981) and still updated by the National Institute of Water and Atmopsheric Research (NIWA) (in revised form). For example, de Freitas et al. (2015) cited Hessell (1980) when they questioned the veracity of NIWA's seven-station series – because it includes "unreliable" sites (notably Auckland and Wellington).

Hessell (1980) devoted considerable attention to the Auckland Albert Park record. The park is in the heart of Auckland city and its character changed considerably over the period it operated as a full meteorological site (1909–1989). Increasing urbanisation and sheltering by growing trees are certain to have introduced inhomogeneities into the climate record – including a warming trend, as asserted by Hessell. However, Hessell actually attributed most (+0.4°C) of the apparent warming over the mid-1940s to mid-1950s, not to evolving site changes, but to a screen change in late 1950. In contrast, Salinger (1981) and Mullan et al. (2010) derived much lower estimates of the screen-change impact (+0.1°C and +0.03°C respectively).

This paper reviews previous published work related to the 1950 screen change at Albert Park. Reasons for the cited disagreements are explored to determine if there are conceptual or analytical errors, if the most recent analysis provides the best available estimate, and what residual issues may remain. It is not the intention here to derive a new estimate, and analyses are intentionally limited to reworking and synthesising previously published data/results. A companion paper will present additional work that extends the analysis beyond mean annual surface air temperature that has been the primary focus of previous work.

A significant issue arising in this review is that the details of the 1950 screen change are disputed. The change to a 'Bilham' Screen is agreed, but whether it replaced a large Stevenson Screen or something else is not. Regardless, the previous work used early inter-comparison studies of large and small screens as context for their findings. In view of this, I begin with a review of two of the early screen inter-comparison studies because, as will become apparent, they do indeed provide useful context. The previous Albert Park work is then reviewed in chronological order. As far as possible, any reworking is done to present the results in a consistent format, while preserving the integrity of the original findings. Critical review of the findings follows, including the implications for future work.

## 2. The Bilham screen

The small ('Bilham') Stevenson Screen was designed in the early 1930s in response to the standardised use of 'sheathed' thermometers by the British Meteorological Service in 1931 (Bilham, 1937). Sheathed thermometers allowed within-screen reconfiguration to near-horizontal in a smaller, lighter, simpler, and cheaper screen – replacing the larger standard Stevenson Screen. Due to superior ventilation and lower thermal inertia, it was expected that recorded screen temperatures would more closely follow true air temperature outside of the screen. The details of the new screen and the arrangement of the thermometers are described in detail by Bilham (1937). The New Zealand small double-louvered thermometer screen is substantially the same (Sparks, 1972).

Bilham (1937) compared monthly means of daily minimum and maximum temperatures ($T_{min}$, $T_{max}$) recorded in the new screen with those from an adjacent large Stevenson Screen at Kew Observatory, England, for a 12-month period (October 1932 to September 1933). Gadre and Narayanan (1939) reported results for a similar 24-month experiment (January 1937 to December 1938) at the Meteorological Observatory at Pune (then 'Poona'), India. The Pune experiment was undertaken to test the screen under a markedly different climate regime to Kew: distinct dry and wet seasons and a more extreme diurnal

range during the dry season (Figure 1a, b). Both papers reported results as monthly mean differences (new screen minus old) and associated derived values for the daily mean ($T_{mean}$, ½[$T_{max}$+$T_{min}$]) and the diurnal range ($T_{range}$, $T_{max}$-$T_{min}$). Figure 1 plots the results, converted to °C but retaining the precision of the original °F measurements. These indicate:

**a)** $T_{max}$ *increases*. Most pronounced at Kew (all months, mean 0.17°C). The Pune mean increase is less (0.04°C) and in February and March, the new Bilham screen has a lower $T_{max}$ than the large Stevenson Screen. There is little evidence of seasonal dependence.

**b)** $T_{min}$ *decreases*. Consistent across all months at both sites. Pune differences are more negative (mean -0.26°C) than Kew (-0.08°C). Again, there is little evidence of seasonal dependence.

**c)** $T_{range}$ *increases*. Consistent across all months at both sites and annual means are similar (Kew 0.25°C, Pune 0.29°C). Pune values decline in June and July, near the height of the monsoon season, but the lowest and highest values are both at the height of the dry season, so a simple correspondence to the annual cycle is absent.

**d)** $T_{mean}$ *inconsistent*. Positive for Kew (mean 0.05°C) and negative for Pune (-0.11°C).

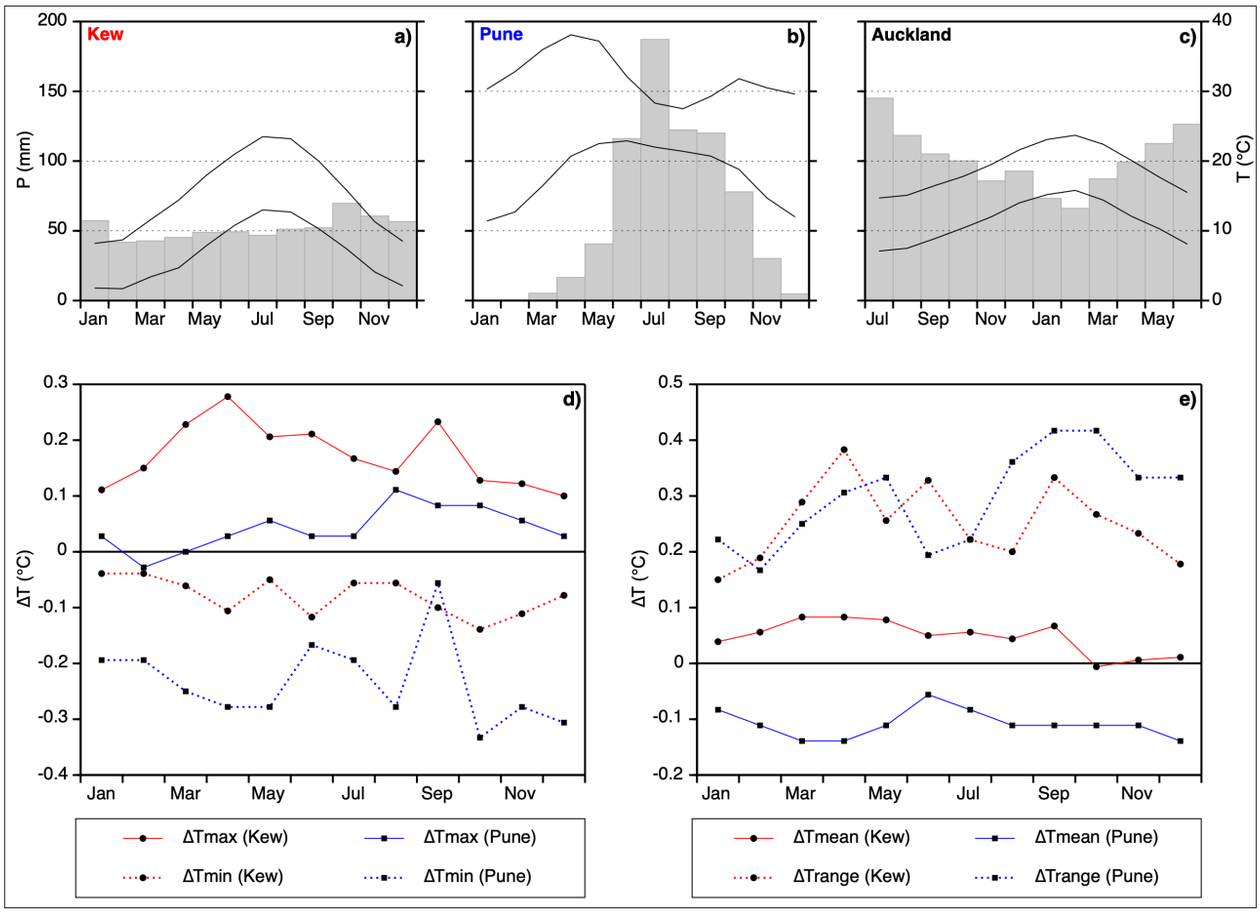



**Figure 1**: Stevenson Screen vs. Bilham Screen inter-comparison results for Kew Gardens, England (Bilham, 1937) and Pune, India (Gadre and Narayanan, 1939). Panels a–c: climographs for Kew (1981–2010, UK Met Office), Pune (1951–80, Indian Meteorological Department), and Auckland (1981–2010, NIWA). Panels d & e: monthly mean differences (Bilham minus Stevenson) in $T_{min}$, $T_{max}$ $T_{mean}$, and $T_{range}$ ($\Delta T_{min}$, $\Delta T_{max}$, $\Delta T_{mean}$, $\Delta T_{range}$).

It is clear from the above that the impacts of changing to a Bilham Screen are dependent on the climate regime. Because the impacts on $T_{min}$ and $T_{max}$ have opposite signs, they at least partially cancel, so $T_{mean}$ is minimally affected. However, the direction of change is variable. This is not the case for $T_{range}$, where the Kew and Pune results are both relatively large and positive.

## 3. Previous work on the 1950 Albert Park screen change

### *3.1 Hessell (1980)*

As part of a wider study challenging emerging ideas about New Zealand warming in the 20th century, Hessell (1980) investigated the impact of the Albert Park screen change in detail. He checked the site files to determine the nature of the change, analysed changes in $T_{min}$, $T_{max}$, $T_{mean}$, and $T_{range}$ for five years either side of the screen-change year, and undertook a paired-site comparison against Riverhead Forest, a more inland site about 20 km northeast of Albert Park.

Table 1a shows the results of Hessell's before and after study (his Table 2). It is worth quoting Hessell here, because these specific results appear to be the basis for his conclusion that a substantial inhomogeneity was introduced as a result of the screen change:

> *"Overall these changes are rather large, and because of the shortness of the periods, they may include a short period synoptic scale secular increase, though the apparent change in the daily temperature range indicates that the screen change is an important contributor to the increase of mean temperature, probably accounting for about 0.4°C of the 0.5°C found"* (Hessell, 1980, p. 4).

Note that the 0.76°C increase in $T_{range}$ is larger than the screen inter-comparison results presented in Figure 1, by a factor of about three. This suggests that the screen that the new Bilham replaced had higher thermal inertia and/or poorer ventilation than a large Stevenson, either of which would be expected to decrease the within-screen diurnal temperature cycle. This is consistent with Hessell's contention that the replaced screen was <u>not</u> a standard Stevenson Screen[1].

Figure 2 reproduces the results of Hessell's (partial) Albert Park – Riverhead Forest paired-site analysis (his Table 5). Three minor modifications are made for plotting clarity and consistency with what follows. First, the sign of the differencing has been reversed, so that it indicates the direction of the screen change impact. Second, the "base difference" for each of $T_{min}$, $T_{max}$, and $T_{mean}$, calculated by Hessell over 1940–60, have been added back in. Third, the original 0.1°C units have been multiplied by 10. These changes do not affect the integrity of Hessell's results. For example, his "-7" result for Riverhead minus Albert Park $T_{min}$ in 1950 indicates a temperature difference of -4.6°C (base difference -3.9°C minus 0.7°C). In Figure 2 it is +4.6°C (sign reversed). Note that annual mean $T_{max}$ is similar for the two sites but $T_{min}$ is about 4°C cooler at Riverhead Forest. Horizontal dotted lines in Figure 2 show mean differences between the two sites for blocks of five years either side of the screen change year in 1950. Differencing these gives the $\Delta T_{min}$, $\Delta T_{max}$, and $\Delta T_{mean}$ statistics in Table 1b. $\Delta T_{range}$ is $\Delta T_{max}$ minus $\Delta T_{min}$. The five-year blocks are consistent with Hessell's before-and-after analysis (Table 1a).

The five-year differencing results (Table 1b) imply that the 1950 screen change resulted in lower minimum and higher maximum screen temperatures (-0.48°C and +0.38°C respectively). These impacts in turn substantially increased the screen diurnal range (+0.86°C) but, because $\Delta T_{min}$ and $\Delta T_{max}$ largely cancel each other out, the impact on $\Delta T_{mean}$ is minor (-0.06°C). The results for $\Delta T_{min}$, $\Delta T_{max}$, and $\Delta T_{mean}$ differ substantially from Hessell's before-and-after study (Table 1a), but $\Delta T_{range}$ is similar.

[1]Hessell (1980, p. 4) included a quote from the site files that states that the replaced screen was "…*not* a standard screen. It was locally made and in a bad state of repair and should be replaced by a standard Stevenson…". A note to Hessell's Table 2 states that the screen was single louvered (Stevenson Screens are double louvered).

**Table 1:** Summary of previous work investigating the impact of the 1950 screen change at Albert Park on recorded and derived screen temperatures. Most values are from the original authors (from tables or extracted from plots). Underlined statistics are calculated here from those values (e.g. most $\Delta T_{range}$ values) or are derived from reanalysis of the original data (row b). Statistics in rows c–i are to one decimal place, following Salinger (1981).

| | Comparison site | Before | | After | | $\Delta T_{min}$ | $\Delta T_{max}$ | $\Delta T_{mean}$ [1] | $\Delta T_{range}$ [2] | Source |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Yrs | Range[3] | Yrs | Range[c] | (C°) | (C°) | (C°) | (C°) | |
| a | Albert Park[4] | 5 | 1945-49 | 5 | 1951-55 | +0.12 | +0.88 | +0.50 | +0.76 | Hessell (1980, Table 2) |
| b | Riverhead Forest[5] | 5 | " | 5 | " | <u>-0.48</u> | <u>+0.38</u> | <u>-0.06</u> | <u>+0.86</u> | Figure 2: reanalysis of Hessell (1980, Table 5) |
| c | Whenuapai[6] | 5 | - | 15 | - | -0.7 | +0.6 | -0.1 | <u>+1.3</u> | Salinger (1981, Table AK.4, p. C9) |
| d | Oratia[6] | 2 | - | 15 | - | -0.2 | +0.3 | 0.0 | <u>+0.5</u> | " |
| e | Ruakura[6] | 11 | - | 15 | - | -0.4 | +0.4 | 0.0 | <u>+0.8</u> | " |
| f | Te Aroha[6] | 15 | - | 5 | - | -0.3 | +0.7 | +0.3 | <u>+1.0</u> | " |
| g | Waihi[6] | 4 | - | 15 | - | -0.6 | +0.6 | 0.0 | <u>+1.2</u> | " |
| h | Tauranga[6] | 10 | - | 10 | - | -0.3 | +0.8 | +0.2 | <u>+1.1</u> | " |
| i | Wellington[6] | 15 | - | 15 | - | -0.5 | +0.8 | +0.1 | <u>+1.3</u> | " |
| j | Waipoua[7] | 10 | 1945-49 | 10 | 1951-60 | - | - | +0.04 | - | Mullan et al. (2010, Figure 6, p. 27) |
| k | Waiuku Forest[7] | 10 | " | 10 | " | - | - | -0.17 | - | " |
| l | Te Aroha[7] | 10 | " | 10 | " | - | - | +0.19 | - | " |
| m | Ruakura[7] | 10 | " | 10 | " | - | - | +0.18 | - | " |
| n | Riverhead Forest[7,8] | 5 | 1945-49 | 10 | " | <u>-0.50</u> | +0.32 | -0.09 | <u>+0.82</u> | Mullan et al. (2010, Figure 6, p. 27 & Figure A3.2, p. 35) |
| o | Five site composite[9] | 6 | 1944-50[10] | 6 | 1950-56[10] | -0.49 | +0.54 | +0.04 | <u>+1.03</u> | Mullan (2012, Figure 5) |

***Notes***

**1.** $\Delta((T_{min} + T_{max}) / 2)$.

**2.** $\Delta T_{max} - \Delta T_{min}$ where calculated here (underlined).

**3.** Range shown if explicitly stated in the original source. Otherwise likely to be calendar years before/after the screen change (1950 excluded), except Row o (see Note 7).

**4.** Same site (Albert Park) before and after study (see text for details).

**5.** Hessell (1980) contains the relevant data for the paired-site comparison as part of his Table 5 (p. 6) but tabled values are from Figure 2.

**6.** $\Delta T_{range}$ calculated here ($\Delta T_{max} - \Delta T_{min}$).

**7.** $\Delta T_{mean}$ from Mullan et al. (2010, Figure 6, p. 27).

**8.** $\Delta T_{max}$ from Mullan et al. (2010, Figure A3.2, p. 35). $\Delta T_{min}$ & $\Delta T_{range}$ derived from $\Delta T_{mean}$ & $\Delta T_{max}$.

**9**. Built from sites in lines j–n using the Rhoades & Salinger (1993) methodology. $\Delta T_{min}$, $\Delta T_{max}$, $\Delta T_{mean}$ from Figure 5, p. 33.

**10.** 6 x 12 months (Nov–Oct).

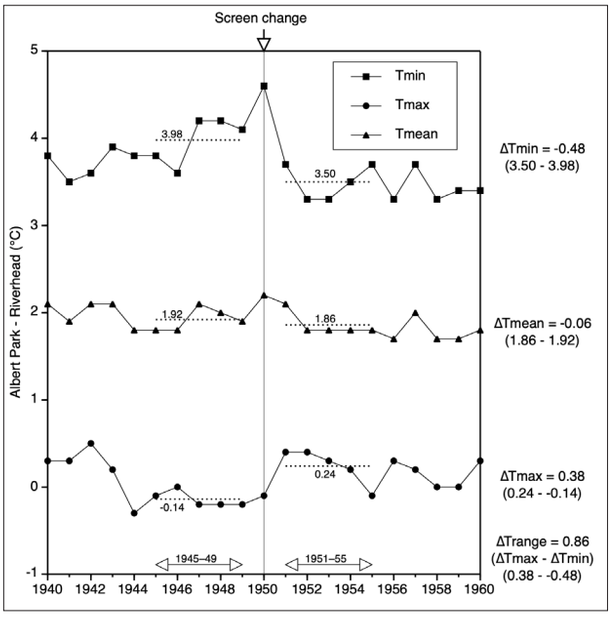



**Figure 2:** Albert Park – Riverhead paired-site analysis of the impact of the 1950 screen change. Solid lines show annual mean differences for $T_{min}$, $T_{max}$, and $T_{mean}$, reworked from data presented in Hessell's (1980) Table 5 (see Section 3.1 for details). Dotted lines show means over 1945–49 and 1951–55. Derived values for $\Delta T_{min}$, $\Delta T_{max}$, and $\Delta T_{mean}$ are the differences across these five-year blocks. $\Delta T_{range}$ is $\Delta T_{max}$ minus $\Delta T_{min}$.

### *3.2 Salinger (1981)*

Salinger (1981) assessed the impact of the Albert Park 1950 screen change using multiple paired-site analyses. Interestingly, he refers to the screen change as being from a large to a small Stevenson Screen, not from a non-standard screen. This would not have affected the analyses undertaken, but may have influenced the interpretation of the $\Delta T_{mean}$ results, because Salinger a priori expected changes "…only as great as 0.1°C between the different types of Stevenson Screen that have been in use in New Zealand…" (Salinger, 1981, p. 67, citing screen inter-comparison results in Sparks (1972)).

Salinger (1981) compared Albert Park with seven other North Island sites (Table 1c–i). Two are local (Whenuapai, Oratia) and all except Wellington are in the northern temperature response region (Salinger and Mullan, 1999). Table 1c–i reproduces Salinger's summary results for $\Delta T_{min}$, $\Delta T_{max}$, and $\Delta T_{mean}$ ($\Delta T_{range}$ is calculated here). For the period before the screen change, annual means are based on different numbers of years. Three sites, including the two local ones, have <6 years and the rest at least 10. Means calculated for the post-change period are for five years (Te Aroha), 10 years (Tauranga) or 15 years (five sites).

In broad terms, the Salinger (1981) results are consistent across all sites: $\Delta T_{min}$ negative; $\Delta T_{max}$ positive; $\Delta T_{mean}$ small (because $\Delta T_{min}$ and $\Delta T_{max}$ largely cancel); and $\Delta T_{range}$ positive. This is similar to Riverhead Forest (Table 1b). Inter-site spreads for $\Delta T_{min}$ and $\Delta T_{max}$ (both ~0.5°C) are similar in magnitude to the respective means (-0.4, +0.6). Only Whenuapai has a negative $\Delta T_{mean}$, three sites are zero, and three are positive (+0.1, +0.2, +0.3). Whenuapai is geographically quite close to Riverhead which also has a negative $\Delta T_{mean}$. Aside from Oratia, all annual change statistics are substantially larger than those reported for Kew and Pune (Figure 1), especially in the case of $\Delta T_{range}$.

### *3.3 Mullan et al. (2010)*

Salinger's PhD research was the genesis of what became known as the New Zealand "seven-station series", derived by combining surface air temperature data from seven sites around the country after homogeneity adjustments (such as the Albert Park screen change adjustment detailed above). Salinger led these developments, first at the New Zealand Meteorological Service (e.g. Salinger et al. 1992) and later at NIWA. Homogenisation methods were revised over time, with a notable development being the adoption of what became known as the Rhoades and Salinger homogenisation method (Rhoades and Salinger, 1993). However, tables of adjustments were not published (Mullan et al., 2018) and the veracity of the series was challenged in the early 2000s by climate change sceptics, including through questions being asked in the New Zealand Parliament in 2009. In response, NIWA made available adjustment tables for each of the

seven sites online and, at the behest of their responsible minister, initiated a major review (Mullan et al., 2010). This included reanalysis of the impact of the Albert Park 1950 screen change.

Mullan et al. (2010) performed paired-site analyses of Albert Park against five North Island sites (Waipoua, Riverhead, Waiuku, Te Aroha, and Ruakura), all drawn from the northern temperature response region. In each case, 10 years before and after 1950 were analysed, except Riverhead which had five prior years to avoid a 1944 screen change at that site. $\Delta T_{mean}$ results (extracted from their Figure 6, p. 27) are reproduced in Table 1j–n. Mullan et al. (2010) also undertook two relevant supplementary analyses. The first repeated their Riverhead and Waiuku analyses, but with before/after overlaps reduced to five years, in order to directly compare with Hessell (1980). The second was a longer paired-site analysis for Riverhead (1935–1965) showing the impacts on $T_{max}$ of the two separate screen changes at Albert Park and Riverhead, in both cases assumed by the authors to be from large to small Stevenson Screens. The latter analysis is the basis for the more complete statistics for Riverhead shown in Table 1n.

The Mullan et al. (2010) $\Delta T_{mean}$ results (Table 1j–n) are inconsistent in terms of the direction of change (-0.17 to +0.19°C). A simple arithmetic mean suggests a small positive impact (+0.03°C), although dropping a single site can change this by as much as ±0.05°C. The Riverhead result is similar to that derived by reworking Hessell's (1980) partial paired-site analysis (Table 1b), but the Te Aroha and Ruakura results are notably different to those presented for the same sites in Salinger (1981). This is presumably due to the different years used for the before/after overlaps. This sensitivity to the number of before/after overlap years is also shown by the Mullan et al (2010) supplementary analysis for Waiuku, where the most negative $\Delta T_{mean}$ for 10-year overlaps (-0.17°C, Table 1k) collapses to -0.01°C for five-year overlaps.

To further explore the sensitivity of $\Delta T_{mean}$ to decisions about the before/after overlap, Figure 3 superimposes the annual paired-site differences presented in Mullan et al. (2010, their Figure 6, p. 27), as anomalies relative to 'base' differences over the common period 1945–1955 (excluding 1950). Positive values indicate annual paired-site differences larger than over the base period (i.e. Albert Park is relatively warm and/or the comparison site is relatively cool). Negative values indicate the reverse. The median (thick grey line) highlights evolving features and the inset graph shows how $\Delta T_{mean}$ estimated from the median line changes as the before/after overlap increases from one year (1951 minus 1949) to 10 years (1951–60 minus 1940–49).

Several interesting features about Figure 3 are worth commenting on here that are relevant to how paired-site analysis results are used to estimate the impact of a thermometer screen change (see discussion):

**a)** Individual anomaly years. The individual paired-site difference lines tend to track quite well. There are subtle year-to-year differences but major deviates are uncommon. Waipoua in 1954 is an exception, possibly indicating a short-term (cool) inhomogeneity at that site.

**b)** Ruakura is volatile. This site is an outlier pre-1950. The difference spread (-0.67 to +0.20°C) is about double that of the other sites. It also has the largest spread post-1950.

**c)** Post-1950 negative trend. Several sites (especially Waiuku and Ruakura) have declining difference trends post-1950, to the extent that the median line declines by 0.4°C. An additional inhomogeneity is suggested here, with Albert Park cooling relative to several other sites and/or those other sites warming.

**d)** Mid-1940s dip. Relative to the base differences, Albert Park is cool compared to all sites over 1944–46. This may be real local cooling, or perhaps an additional Albert Park inhomogeneity.

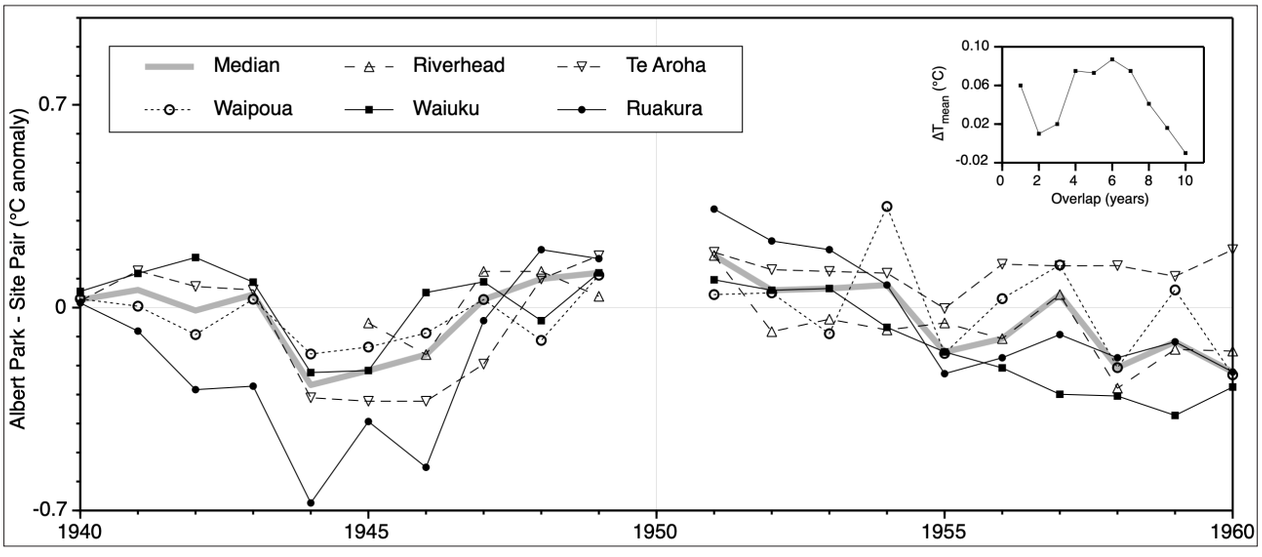



**Figure 3:** Reworking of the Mullan et al. (2010) paired-site analyses for Albert Park (results extracted from their Figure 6). For comparison purposes, annual $T_{mean}$ differences for each comparison site (Albert Park minus that site) are shown as departures from their respective means for the common period 1945–1955, excluding 1950. The thick grey line is the median across available sites. The inset plot shows the impact on $\Delta T_{mean}$, estimated using the median line, as the overlap extending either side of 1950 is increased from one to 10 years.

### *3.4 Mullan (2012)*

From what can be deduced from the original publications, Table1a-n source data were all calculated for calendar years, with 1950 excluded. Salinger (1981) and Mullan et al. (2010) then calculated $\Delta T_{mean}$ as a simple average – in essence undertaking multiple versions of the analysis shown in Figure 2, but with variable before/after overlaps, then averaging the $\Delta T_{mean}$ results. However, since the mid-1990s, through to the Mullan et al. (2010) revision, the New Zealand seven-station series was based on the Rhoades & Salinger (1993) methodology, which differs in two important ways: calculations are on monthly data and variable weights are assigned to sites. The former removes the calendar-year constraint and allows data in the change year to be included in the analysis. The latter, based on inter-site correlations, gives greater weight to (likely local) sites with similar temperature characteristics.

Mullan (2012) reviewed the Rhoades & Salinger (1993) methodology, exploring the sensitivity of site change corrections to site weightings, missing and bad data, site-specific short-term temperature anomalies, and the number of years used in the before/after overlap. Key findings were shown using multiple indicative cases, including the 1950 Albert Park screen change. Mullan's key finding was that "…short-term anomalous periods are relatively common in the New Zealand temperature data sets… [and] …the straightforward solution is to take a long enough comparison period so the effects of short-term anomalies are minimised" (p. 30). He recommended before/after overlaps of at least four years to achieve this, although his assessment of the Hokitika 1943 site shift, which was preceded by several years of bad data associated with a severely degraded Stevenson Screen, indicates that skipping over known bad data is also advisable. Mullan (2012) also found that different approaches to assigning site weightings had little impact, but that choices made about which (high correlating) sites to use were influential. He noted that the latter "… needs to be assessed on a case-by-case basis, especially where there may be a suggestion of a non-climatic trend such as urban heating or exposure degradation" (p. 36).

Although the 1950 Albert Park screen change was only briefly discussed by Mullan (2012), it usefully extended the Mullan et al. (2010) analysis. The same five sites were used, results for $\Delta T_{min}$ and $\Delta T_{max}$ were presented, and error estimates were provided for the first time. Mullan stated that the Bilham screen replaced a Stevenson Screen in November 1950 (citing Fouhy et al., 1992, although that reference does not actually state that the replaced screen was a Stevenson), so his before/after years split on that month are not calendar years. Results for maximum overlaps of six years are shown in Table 1o. Consistent with most other Table 1 results, $\Delta T_{min}$ and $\Delta T_{max}$ are both large and of opposite sign. Because they are of similar magnitude the net effect is a very small increase in $\Delta T_{mean}$. Error estimates for $\Delta T_{min}$, $\Delta T_{max}$, and $\Delta T_{mean}$ are respectively ±0.19°C, ±0.11°C, and ±0.11°C. In addition to the larger error estimates, $\Delta T_{min}$ shows greater sensitivity to the overlap period, increasing from -0.80°C at one year to -0.49°C at six. Because $\Delta T_{max}$ is relatively stable, it is evolving $\Delta T_{min}$ that is responsible for $\Delta T_{mean}$ changing sign as the overlap increases.

## 4.Discussion

### *4.1 Early screen comparisons*

Thermometer screens are necessary, but an undesirable consequence is that the diurnal cycle of recorded screen temperatures is suppressed relative to true surface air temperatures outside the screen. The roles of screen ventilation and thermal inertia in causing this suppression were well understood by the mid-1800s, leading to the development of the Stevenson Screen as a standard. In this context, the introduction of the new small (Bilham) screen was a major development, prompting research to understand how recorded temperatures are affected, such as the two inter-comparison studies reproduced in Figure 1. These indicate that the diurnal range is less suppressed in the Bilham screen, relative to the standard Stevenson, but $T_{mean}$ is little affected because the impacts on $T_{min}$ and $T_{max}$ largely cancel. However, the Kew and Pune results also show that the climate regime is influential in terms of the relative importance of changes in $T_{min}$ or $T_{max}$, and therefore if the (small) net effect on $T_{mean}$ is positive or negative. One implication is that, although a change from standard Stevenson to Bilham screen is likely to have minimal impact on calculated $T_{mean}$, local screen inter-comparisons or paired-site studies are required to deduce the direction of change. In contrast, the $\Delta T_{range}$ results are quite consistent. This suggests that $\Delta T_{range}$ is likely to be particularly useful for homogeneity analyses, such as identifying undocumented screen changes or site shifts which also often affect the diurnal range (Mullan, 2012). Also, the consistency of the Kew and Pune $\Delta T_{range}$ results suggests that it may be useful for checking if a presumed large Stevenson Screen to Bilham Screen conversion yields consistent results. Such screen or site changes would likely manifest as abrupt changes in $\Delta T_{range}$.

### *4.2 The 1950 screen change*

The change to a Bilham screen in late 1950 is not in dispute, although Hessell (1980) has it occurring in September, whereas others say November. With respect to what it replaced, Hessell's quote from the site files is very explicit about it not being a standard Stevenson Screen (see Section 3.1), so why subsequent researchers have stated that it was is something of a mystery. The most likely explanation is that documentation sighted by Hessell was subsequently destroyed or misplaced. Interestingly, Hessell acknowledged that differences between large and small double-louvered screens are "usually negligible", at least in terms of $T_{mean}$, so the replacement of a non-standard screen would appear to be central to his argument concerning a significant impact (see next section). Also, it is noteworthy that the $\Delta T_{range}$ results presented in Table 1 are generally much larger than those for the Kew and Pune inter-comparison experiments

(Figure 1), by a factor of about three. While we have to be careful about transferring the Kew and Pune results to Auckland, this very large discrepancy suggests that the replaced screen was indeed non-standard, perhaps with thermal and ventilation characteristics that suppressed the diurnal temperature range by ~0.5°C more than a standard Stevenson.

### *4.3 Secular change*

Except for $\Delta T_{range}$, the Hessell (1980) before-and-after study (row a) is a clear outlier in Table 1. It has the only positive $\Delta T_{min}$ and the largest values for $\Delta T_{max}$ and, especially, $\Delta T_{mean}$. Mullan et al. (2010) explained the discrepancy with their own findings in terms of Hessell's before/after results being influenced by a strong regional warming trend over the 11-year block straddling 1950. Hessell's analysis of rural sites (his Figure 4) shows this warming and it is clear from the block quote in Section 3.1 that he was aware of the potential implications of secular warming for his before-and-after study. It is also apparent from Hessell's discussion of the results of his partial Riverhead paired-site analysis that he recognised an inconsistency (compare Table 1 rows a and b). He attributed this to "…quasi-parallel trends at both" sites, presumably unrelated to secular warming.

Given the above, it is curious that Hessell (1980) attributed 0.4°C of the 0.5°C increase in $T_{mean}$ to the screen change. In essence, his argument appears to be that Stevenson-type screens tend to overheat (so higher $T_{max}$), which in turn increases $T_{range}$ and $T_{mean}$. The latter part of the block quote in Section 3.1 indicates that he interpreted the conjoint positive and large increases in $T_{range}$ and $T_{max}$ (Table 1a) as evidence that the screen change was the main contributor to the increase in $T_{mean}$, and that secular trend could be largely discounted. However, the statement that the increase in $T_{range}$ (0.76°C) implies that the screen change is mostly responsible for the increase in $T_{mean}$ (0.50°C) is incorrect. It is true that if the screen change were responsible for an apparent increase in $T_{max}$ of several tenths of a degree, then an increase in $T_{range}$ and $T_{mean}$ would result. However, it does not follow that a large increase in $T_{range}$ is evidence that a screen change is responsible for a conjoint increase in $T_{mean}$. We can see this very clearly in Figure 1. In both studies the change to a Bilham screen results in a higher $T_{range}$ but the impact on $T_{mean}$ is minor. Moreover, in the Pune case, the increase in range is associated with about a 0.1°C decrease in $T_{mean}$. $T_{max}$ increases but is offset by a more substantial decrease in $T_{min}$ – roughly the opposite of the Kew results (Figure 1d).

An alternative, and simpler, explanation of the Hessell (1980) before-and-after results (Table 1a) is secular warming of several tenths of a degree, combined with screen-induced amplification of the recorded diurnal cycle. The secular warming would suppress or even reverse the recorded decrease in $T_{min}$ but would amplify the increase in $T_{max}$. Reinterpreting Hessell's results in this way roughly reverses his partitioning of the 0.5°C increase in $T_{mean}$ (0.1°C secular, 0.4°C screen change) to 0.4°C secular and 0.1°C screen change. It follows that the entries for $\Delta T_{min}$, $\Delta T_{max}$, and $\Delta T_{mean}$ in Table 1a are invalid estimates of screen change impacts. $\Delta T_{range}$ remains valid.

### *4.4 Overlap period*

Mullan (2012) was emphatic in recommending before/after overlaps of at least four years based on convincing evidence. However, how far it is advisable to extend the overlap is debatable. Although extending has the advantage of further reducing the impact of bad data near the change point, it also pushes the analysis into years increasingly removed from that change point, thereby increasing the risk of the results being influenced by additional (possibly undocumented) inhomogeneities. Consider the Albert Park case shown in Figure 3. Extending the overlap period beyond three years, pushes the early-period overlap into a possible additional mid-

1940s Albert Park inhomogeneity and the late-period overlap into a period of marked $\Delta T_{mean}$ decline for two sites (Ruakura, Waiuku). As the overlaps are increased to 10 years the influence of the short-term deviation in the mid 1940s declines, but the continuing declining trend into the late 1950s becomes more influential. If the $\Delta T_{mean}$ trends at Waiuku and Ruakura are a result of local warming at those sites, then overlap extension may well be counter-productive. Such hypothetical speculation is hardly a convincing basis for limiting the overlap period, so a consistent baseline has merit, at least as a first estimate that could be refined after appropriate supplementary analyses are undertaken. In the Albert Park case these might include investigation of:

**a)** A possible undocumented inhomogeneity in the mid-1940s (and earlier). If Hessell (1980) is correct about a non-standard screen being replaced, then an earlier undocumented change from a Stevenson Screen to the non-standard one is a distinct possibility.

**b)** The generic suitability of the Ruakura site. It is an outlier in Figure 3, so independent appraisal against other sites (not limited to those in Figure 3) is appropriate.

**c)** Possible local warming inhomogeneities post-1950 at Waiuku and Ruakura related to increasing shelter and urbanisation.

**d)** A possible local inhomogeneity in 1954 at Waipoua. If 1954 is anomalous at Waipoua compared to other sites, the year could reasonably and objectively be skipped in the Albert Park analysis.

Another point to consider with respect to optimal overlap length, and indeed the specific years to select, is the nature of the site change in question. Consider the case of a screen in a poor state of repair which is replaced by a pristine new one. Using a few years before and after the screen change tells us how the screens in their respective states compare, but it makes no sense to use that relationship to 'correct' all of the data collected by the old screen – because doing so would implicitly treat the old screen as being in a poor state throughout its life. The Hokitika example presented in Mullan (2012) is an excellent example of how badly things can go wrong if that type of correction is applied. Mullan's solution for Hokitika was to skip over two years of bad data. A precautionary approach would suggest doing something similar where the state of repair of the replaced screen is unclear, as is the case for Albert Park[2]. The possibility of a bad data situation towards the end of an old screen's life is also an incentive to extend the overlap period in order to mitigate any undesirable influence. However, this is not the case for the new screen, where extension beyond four years is not relevant in these terms and may be counterproductive, due to the possible homogeneity issues previously noted. The implications are not trivial. For example, using the median line in Figure 3 and before/after overlaps of 1940–48 and 1951–54 would increase the $\Delta T_{mean}$ estimate by about 0.1°C.

### *4.5 Implications for climate change analyses*

Excluding Hessell's (1980) estimates of $\Delta T_{min}$, $\Delta T_{max,}$ and $\Delta T_{mean}$ in Table 1a, because they are significantly impacted by secular temperature increase (see Section 4.3), leaves broadly consistent results. $\Delta T_{min}$ is consistently negative by a few tenths of a degree, $\Delta T_{max}$ is consistently positive by a similar amount, and $\Delta T_{mean}$ is relatively small (because $\Delta T_{min}$ and $\Delta T_{max}$ cancel). The sign of $\Delta T_{mean}$ varies, but the weight of evidence is that the 1950 change to a Bilham thermometer screen resulted in a small positive increase in calculated $T_{mean}$. For the reasons outlined in Section 4.4, Mullan's (2012) best estimate of +0.04°C may be low, but his error estimate of ±0.11°C likely captures the true value.

On the basis of Hessell's (1980) description of the screen

[2] Most of the experiments reported in Table 1 implicitly do this to some degree by excluding 1950 data. Because the screen change was in late 1950, most of a year of potentially 'bad' data is skipped.

replaced in 1950, a plausible hypothesis is that a standard Stevenson Screen installed in Albert Park in 1909 was replaced by an undocumented non-standard screen several years prior to 1950. This cannot be explicitly corrected for, because we don't know when (or even if) it happened, but the uncertainty should be accommodated in reconstructions of $T_{mean}$. Given the fairly minor differences in calculated $T_{mean}$ between different Stevenson Screens (Sparks 1972, Figure 1e), it seems unlikely that the adjustment would be much more than a doubling of Mullan's (2012) error estimate.

The above benign comments pertain only to $T_{mean}$. Suppression of the within-screen diurnal cycle is so much more pronounced in the non-standard screen (Table 1) than a standard Stevenson Screen (Figure 1d) that reconstructions of early 20th century $T_{min}$, $T_{max}$, and $T_{range}$ would likely be substantially in error if the Table 1 results were used to correct data actually collected from a standard Stevenson Screen. Take Mullan's (2012) $\Delta T_{min}$ estimate of -0.49°C (Table 1o). This indicates that the new Bilham screen recorded lower $T_{min}$ than the old screen. To homogenise the record (i.e. bring the early record into line with the new screen) 0.49°C is subtracted from pre-1950 $T_{min}$ observations. However, if that correction is also applied over an early period when temperatures were recorded in a standard Stevenson Screen, the results plotted in Figure 1d suggest we could be over-correcting (subtracting too much) by perhaps 0.3–0.4°C. For $T_{max}$ the over-correction would be similar but in the opposite direction, resulting in reconstructed early 20th century $T_{max}$ a few tenths of a degree too warm. The resulting diurnal range would be about 0.7°C too wide.

## 5. Conclusions

The Albert Park screen change in 1950 was a simple affair. A Bilham thermometer screen replaced an older screen and several decades of daily climate observations were continued. A little over 30 years later, as the issue of global warming emerged, the Albert Park record became a source of controversy, which has continued intermittently to the present. The two main issues relate to what exactly the Bilham screen replaced and whether recorded warming at the time of the change was secular or was mostly an artefact of the screen change itself (due to the changed thermal regime of the screen). On balance, Hessell (1980) was probably correct in asserting that a non-standard screen was replaced. His direct quotation from the site files is convincing in this regard, and the fact that most of the results in Table 1 are inconsistent with Stevenson Screen inter-comparison studies (e.g. Figure 1d,e) provides additional support. However, Hessell was mistaken in attributing the bulk of the apparent warming after 1950 to the screen change itself. Unequivocal secular warming occurred over this period and the paired-site analyses summarised in Table 1, which account for temperature changes common to both sites, consistently point to the screen change having only a minor impact on $T_{mean}$. However, the latter is a somewhat fortuitous consequence of substantial screen-change impacts on $T_{min}$ and $T_{max}$ cancelling out. Screen-change impacts on $T_{range}$ are very large (ca. +1°C) and point to that variable being particularly useful for detecting screen-related homogeneity issues.

If we were to accept that a non-standard screen was replaced in 1950, an unresolved third issue would then be introduced. Because the Stevenson Screen had long been standard in New Zealand (Robertson, 1950), a reasonable hypothesis is that an additional undocumented change occurred some years prior to 1950. The implications for mean temperature are not serious, although error estimates should be expanded. However, direct analysis of Auckland pre-1950 minimum and maximum temperatures and of the diurnal cycle may be seriously compromised, because any such analysis will inevitably rely on the Albert Park record. The hypothesis of a pre-1950 undocumented screen change will be explored in a follow-up paper.

## References


Bilham EG. 1937. A screen for sheathed thermometers. *Quarterly Journal of the Royal Meteorological Society* 63, 309–322.

de Freitas CR, Dedekind MO, Brill BE. 2015. A Reanalysis of long-term surface air temperature trends in New Zealand. *Environmental Modeling & Assessment* 20, 399–410.

Fouhy E, Coutts L, McGann R, Collen B, Salinger MJ. 1992. South Pacific Historical Climate Network. Climate Station Histories: Part 2, New Zealand and Offshore Islands. NZ Meteorological Service, Wellington. ISBN 0-477-01583- 2, 216p.

Gadre KM, Narayanan A. 1939. Comparative observations of temperature in a standard Stevenson and a Bilham screen at the Central Agricultural Meteorological Observatory, Poona. *Quarterly Journal of the Royal Meteorological Society* 65, 450–452.

Hessell JWD. 1980. Apparent trends of mean temperature in New Zealand since 1930. *New Zealand Journal of Science* 23, 1–9.

Mullan AB. 2012. Applying the Rhoades and Salinger method to New Zealand's "Seven-Station" temperature series. *Weather and Climate* 32, 23–37.

Mullan AB, Salinger J, Renwick J, Wratt D. 2018. Comment on "A Reanalysis of long-term surface air temperature trends in New Zealand". *Environmental Modeling & Assessment* 23, 249–262.

Mullan AB, Stuart SJ, Hadfield MG, Smith MJ. 2010. Report on the review of NIWA's 'seven-station" temperature series. NIWA Information Series No. 78, 175 p.

Rhoades DA, Salinger MJ. 1993. Adjustment of temperature and rainfall records for site changes. *International Journal of Climatolology* 13, 899–913.

Robertson NG. 1950. The organization and development of weather observations in New Zealand. In: *Garnier BJ (ed.) New Zealand weather and Climate*. New Zealand Geographical Society, Miscellaneous Series, No 1, 7–25.

Salinger MJ, McGann R, Coutts L, Collen B, Fouhy E. 1992. South Pacific historical climate network: temperature trends in New Zealand and outlying islands, 1920–1990. New Zealand Meteorological Service, Wellington.

Salinger MJ. 1981. *New Zealand climate: the instrumental record*. PhD thesis, Victoria University of Wellington, Wellington, New Zealand.

Salinger MJ, Gunn JM. 1975. Recent climatic warming around New Zealand. *Nature* 256, 396–398.

Salinger MJ, Mullan AB. 1999. New Zealand climate: temperature and precipitation variations and their links with atmospheric circulation 1930–1994. I*nternational Journal of Climatolology* 19, 1049–1071.

Sparks WR. 1972. The effect of thermometer screen design on the observed temperature. World Meteorological Organisation Technical Note No. 315.

# Machine learning-based climate time series anomaly detection using convolutional neural networks

R. Srinivasan[1], L. Wang[1] and J.L. Bulleid[1]

[1]*National Institute of Water and Atmospheric Research (NIWA), Private Bag 14901, Wellington, New Zealand*

Correspondence: Raghav.Srinivasan@niwa.co.nz , +64-4-386-0525

**Key words:** machine learning, deep learning, climate data quality control, image recognition, convolutional neural networks, time series anomaly detection, chart mining

## Abstract

Data from New Zealand's National Climate Network are operationally verified both during data ingest and post data ingestion into the National Climate Database. The quality control process in the database verifies the data in two ways: automated checks, such as where data are automatically checked for 'out-of-expected-range' values, or for consistency with nearby site data (buddy checks), and another approach where a quality control analyst manually scrutinises data for anomalies. In the latter approach, climate timeseries data from sensor networks are plotted for quality control purposes and these plots are manually analysed for anomalies by a quality control analyst on a weekly basis. This manual process is performed in addition to other manual and automated quality checks because it helps to identify additional anomalies or unusual patterns in data that were not caught by the automated quality control processes. As more observational capacity is added to the climate network, manually reviewing quality plots becomes increasingly time-consuming, cumbersome and costly, and has an increasing potential to compromise data quality. In this study, we explore an image-based method to automate the manual anomaly detection process on quality control plots using deep learning. To do this we trained a Convolutional Neural Network (CNN) model with images of time series quality plots. The model learned to identify plots that contained anomalies similar to those a manual reviewer would detect. We have successfully achieved a high anomaly classification score using a CNN with modified VGG-16 architecture. We were also able to successfully identify, and colour-highlight, the classified anomalous regions within the quality plots using Gradient-based Class Activation Mapping. We achieved an overall anomaly classification F1 score of 0.92 and anomaly localisation accuracy of 91%.

## 1. Introduction

### *1.1 Background*

New Zealand's National Climate Database, hosted by the National Institute of Weather and Atmospheric Research (NIWA), stores observational data from climate stations located across the country. These include stations managed by NIWA, MetService, Regional Councils, Fire and Emergency New Zealand and other agencies. The

*Submitted: 21 July 2020*
*Accepted: 3 November 2020*

data serves a multitude of applications, from weather and hazard forecasting, to scientific and economic studies and commercial activities. Around 1.8 billion rows of data are stored for extraction from the database, for both public and private use. There are about 2000 regular users and 50000 registered database users who have extracted data at some time in the past decade. Curation and quality control of these data are key to ensuring the data are fit for purpose for a multitude of uses.

NIWA operates the National Climate Network (NCN) that comprises hundreds of climate monitoring stations. Data telemetered from each station within the NCN are quality checked as part of the process of ingestion into the National Climate Database (CliDB). Quality control (QC) checks are performed on these datasets using both automated and manual verification. As part of the manual QC process, data from the NCN are plotted to produce raw weekly timeseries QC plots (Figure 1) and these plots are used for manual data verification. When verified, the data are securely archived in CliDB. These QC plots are produced by individual climate stations and there are hundreds of QC plots produced every week that require manual review.

### *1.2 Study aims*

In this study, we aim to augment the manual review process by applying a deep learning solution to automatically recognise, and flag potentially erroneous data. We treated this as an image classification task and used a Convolutional Neural Network (CNN) to classify and flag data anomalies to the attention of a QC analyst. This helps minimise the manual effort otherwise needed

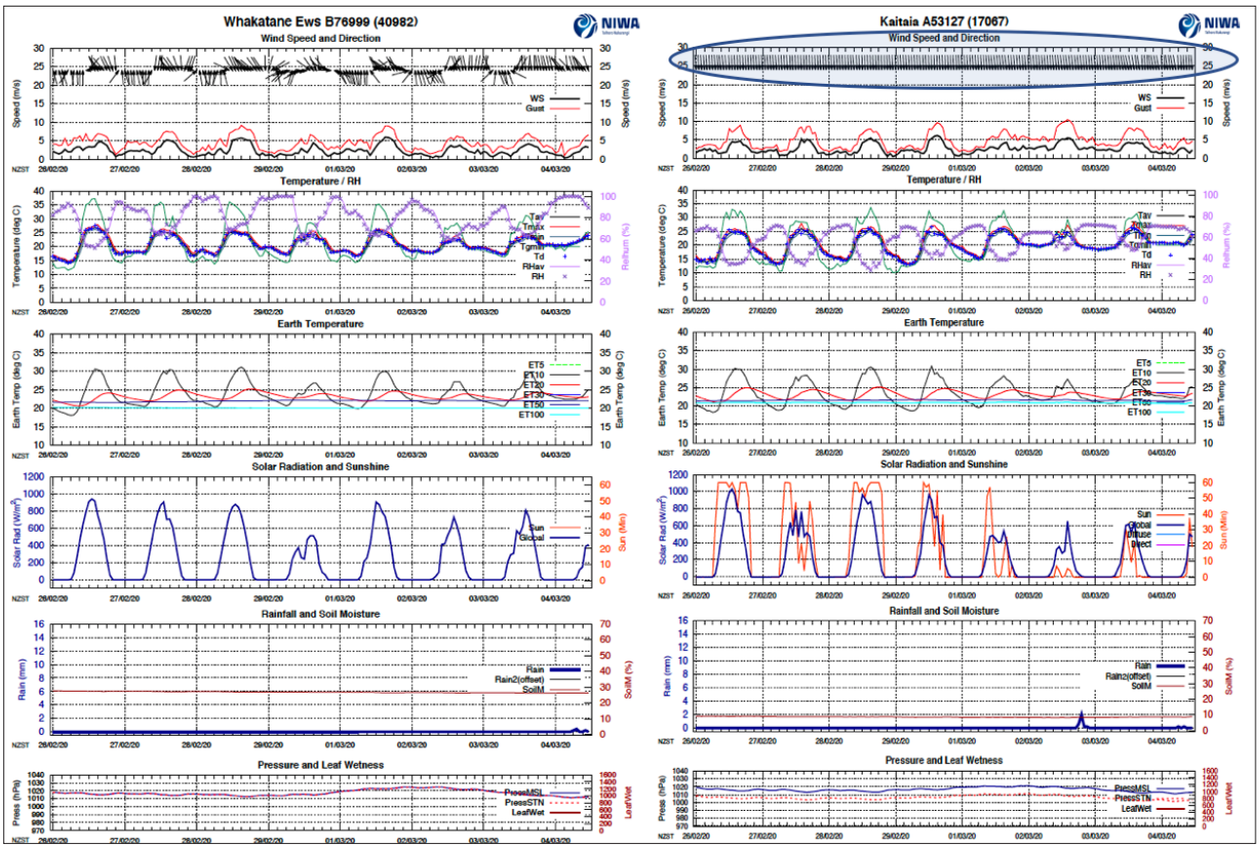



**Figure 1:** Two examples of weekly QC plots produced from individual climate stations before verification and archival. Left: a plot showing no pattern of anomalies. Right: a plot that should be flagged as anomalous because the wind direction sensor (circled) has been stuck in the same position for the whole week.

to review all the plots and reduces the manual task to one of remediating the flagged anomalies.

It is common to use a Recurrent Neural Network (RNN) model in time series analysis problems due to their temporal/sequential nature (Hundman et al., 2018; Park et al., 2018; Karim et al., 2018; Hüsken et al., 2003). In our study, we were trying to reproduce a manual QC review process whereby a person looks at the image of a quality plot and detects anomalous features within it. We treated this as an image classification problem because we intended to replicate the manual review process. Figure 2 illustrates the current QC plots review process and we aimed to retain this overall QC process by replacing the manual component of the review with an automated process. In addition, these images were already operationally produced, and an archive of these plots were readily available to train and test the algorithm. We decided to use a CNN for detecting anomalies as the CNN model was suited to image classification tasks.

There has been some previous work done where CNNs are used on time series images. Zheng et al. (2014); Yang et al. (2015); Cui et al. (2016) successfully used a deep CNN on time series data for image classification. Also, Zhao et al. (2017) similarly successfully used their multi scale convolutional neural network (MCNN) for time series classification tasks. In particular, Wen et al. (2019) developed a transfer learning-based CNN framework where U-net (Ronneberger et al., 2015) inspired CNN was used successfully for anomalous image segmentation in time series data.

In the above-mentioned CNN related works, the input signal was fed into the CNN either through transformations on the sensor signal or as a one-dimensional input of the sensor signal itself. In our study, the input was the 2-dimensional image of the QC plots and the aim was to perform chart mining on these images to detect anomalies. There has been some earlier work done

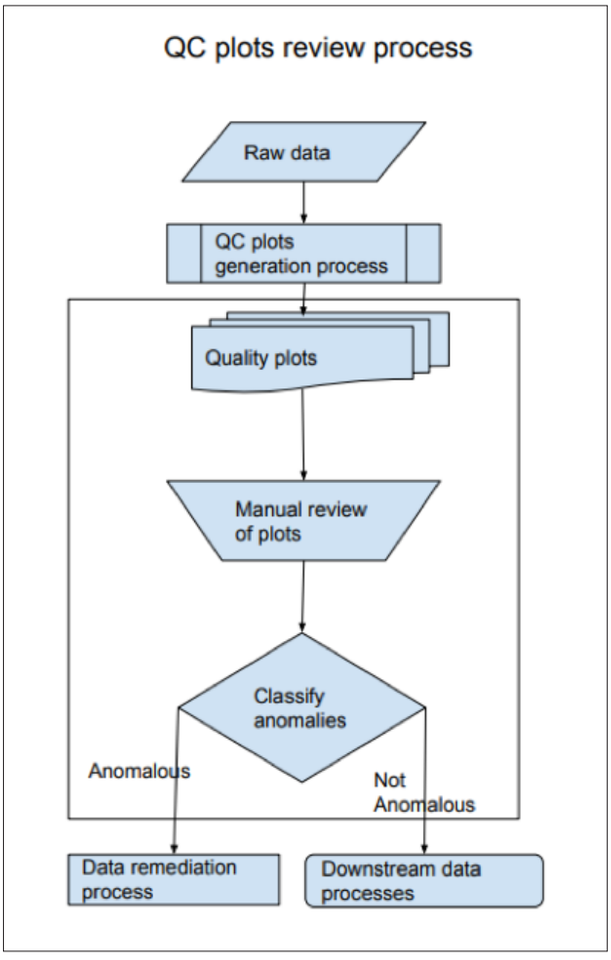



**Figure 2:** A high-level overview of the QC plots review process. The processes inside the box above could be changed to an automated CNN-based image classification task.

on chart mining (Davila et al., 2020). One past study used CNNs for chart type classification and extracted data, and text, from charts using image processing techniques and Optical Character Recognition (OCR) respectively (Balaji et al., 2018). CNNs were used to detect text or numerical elements in images of the charts along with their type (Liu et al., 2019, Cliche et al., 2017). To the best of our knowledge, our work is the first attempt in detecting anomalous patterns from charts using CNNs on sensor data for QC purposes.

We aimed to use the VGG-16 based CNN architecture for the purposes of this study. VGG-16 (Simonyon et al., 2014) network architecture was developed by the Visual

Geometry Group (VGG), Department of Computer Science, University of Oxford. Their architecture consists of 16 weighted layers (dubbed VGG-16). This approach replaced the large filter sizes of previous CNN architectures, such as AlexNet (Krizhevsky et al., 2012), with small 3 pixel x 3 pixel filters. VGG-16 uses an input image size of 224 pixels x 224 pixels x 3 channels (RGB). The network consists of five convolutional blocks with max pooling at the end of each convolutional block. The first two blocks have two convolutional layers each. The next three blocks have three convolutional layers each (Figure 3). The VGG-16 network was trained and tested on the ImageNet (Deng et al., 2009) dataset, comprising more than 14 million images and 1000 different classes. VGG-16 has a good reputation having secured first place in the ILSVRC - ImageNet Large Scale Visual Recognition Competition (Russakovsky et al., 2014) for low localisation error, and won second place for low classification error.

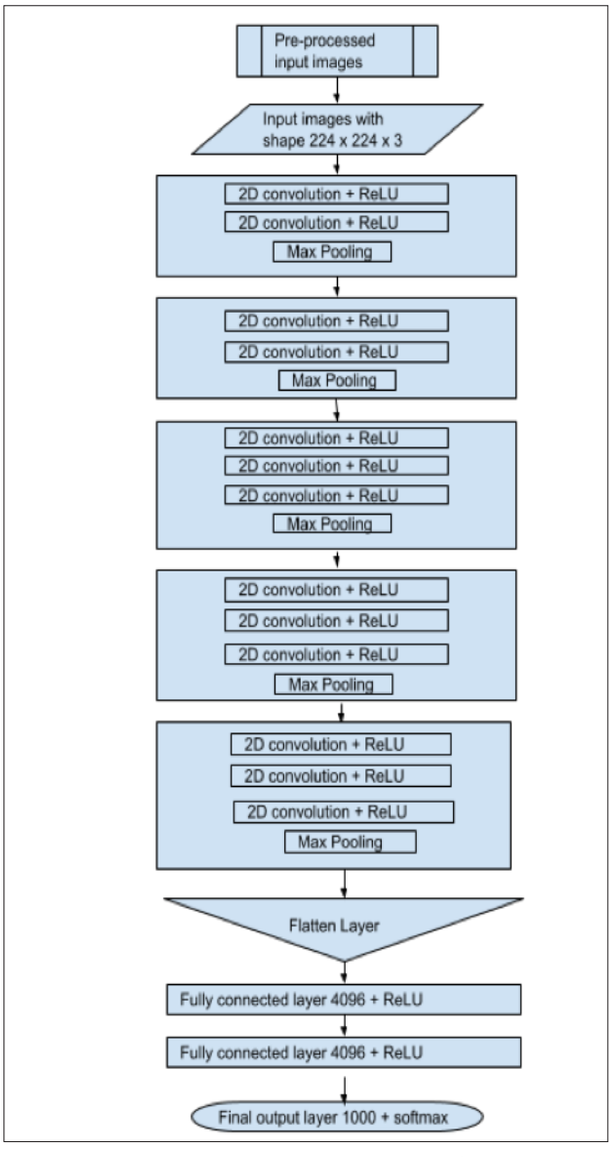



**Figure 3:** VGG16 architecture diagram that shows the five convolutional blocks along with its two fully-connected and softmax output layers.

## 2. Train/test data selection

This Section explains the QC plots, how the training and testing datasets for the model were chosen, and describes the pre-processing steps involved before feeding the training images into the model.

### *2.1 QC plots*

Most of the data archived into CliDB are ingested in near-real time. Some of the QC checks are performed during data ingestion and some checks are performed as a batch process, post-ingestion. As explained in Section 1.1, an aspect of NIWA's CliDB QC process involves generation of timeseries plots for each station, on both a daily and weekly schedule (developed by NIWA Climate Database Technician Errol Lewthwaite). These plots either contain hourly data, or data at 10-minute intervals if available. These QC plots are generated from the pre-ingest data and this helps to check if the routine data ingest QC processes have correctly identified the quality issues. Currently, these QC plots are manually reviewed, post data ingestion, to ensure the quality of the archived data. This manual review of plots is one of the many aspects of manual quality control. The QC plots capture many basic climate variables in a single plot, for each station. Figure 1 shows the following variables: wind direction, wind speed, wind gust, maximum temperature, minimum temperature, mean temperature, grass minimum temperature, relative humidity (RH), earth temperature profile (5 cm, 10 cm, 20cm, 30 cm, 50 m, 100 cm), sunshine, solar radiation,

rainfall, soil moisture, leaf wetness and mean barometric pressure at sea-level. Representing all of these variables on every plot enables the analyst to explore any inter-dependency and thereby add 'context' to clarify observations that facilitate QC decisions. For example, a sudden upward spike in soil moisture could be related to a rainfall event; an increase or decrease in air temperature could be related to a change in wind direction. Also, there are subtle relationships between solar radiation and sunshine and between wind and RH. Since these inter-dependencies play an important role in QC decision making, we have decided to use the entire plot to train our model, hence encompassing all the variables and their format. Not all stations measure all of the above climate variables. Hence, while the format is the same for all stations, some variables may be absent.

### *2.2 Data selection*

The QC plot generation process has been used for several years and weekly QC plots spanning the last six years are archived in the CliDB database system. In our study we have used these archived plots to identify various scenarios and flag the corresponding images. We manually reviewed these plots and chose around 1000 (both valid and anomalous) to train the model. We included plots over different seasons and from different monitoring stations.

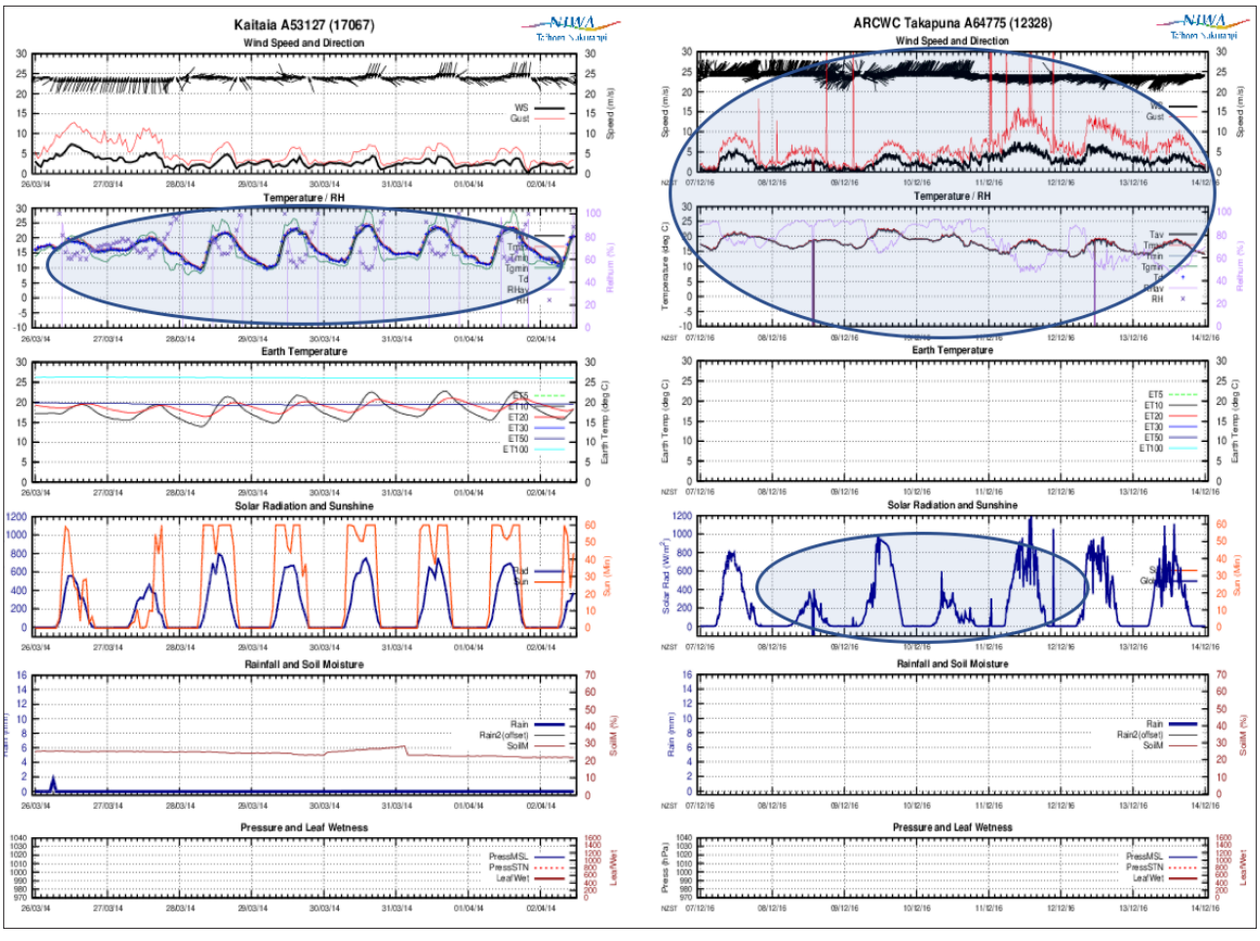



**Figure 4:** Left: the relative humidity (RH) time series shows sudden drops to zero; Right: there was an issue with sudden peaks in wind gust to beyond the displayed axis scale, and sudden drops in RH, temperature and radiation values (which may be deliberate default settings so that missing values can be recognised).

We chose plots representing various anomaly features for different variables. These anomalies are characterised by their patterns and broadly defined as:

a. Sudden peaks/drops;

b. Stuck values;

c. Uncorrelated behaviour of related variables;

d. Unusual pattern/fluctuation;

e. Incomplete timeseries.

**a. Sudden peaks/drops**

A large sudden drop can occur when there is a missing value in a timeseries, and the instrument channel defaults to reporting a pre-set extreme low value. A large sudden peak could be a value that is out of range or within a valid range but resulting from some other issue. These sudden peaks or drops could be due to a range of issues such as instrument calibration, communication, or power supply faults. The QC plots are generated from this pre-ingested data to highlight these issues (Figure 4). The data ingest process checks the data for minimum and maximum range before ingesting into CliDB. If there is an out of range value observed, the observation will be assigned 'suspect' on ingest into the climate database. These range thresholds are set for each site and parameter observed.

Also, a sudden peak may sometimes be valid if, for example, it is the result of an extreme weather event. In some cases of extreme values, we would classify those data as valid if the correlated variables were also displaying similar patterns.

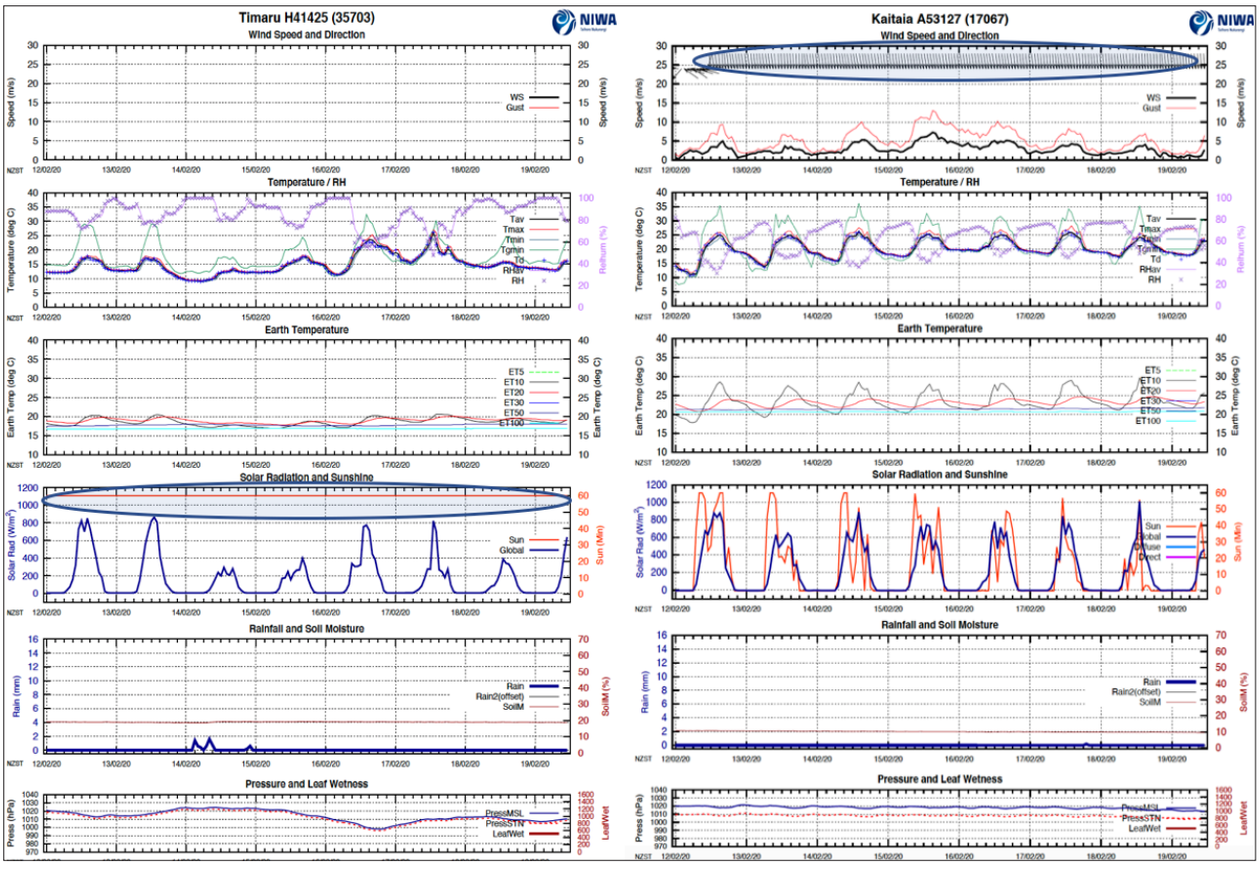



**Figure 5:** Left: sunshine is stuck, as indicated by the straight line in the timeseries; Right: wind direction appears to be stuck in one direction.

**b. Stuck values**

Stuck values may be indicated by a straight line or an unchanging/repetitive pattern. It could be stuck for a short duration within the timeseries, or for the whole plot period. Except for wind direction, remaining variables' stuck values could be identified by a horizontal line. Wind direction stuck values were identified by wind barb vectors pointing in the same direction (Figure 5). In other variables the lines could appear horizontal, vertical or slanted. Periods of zero rainfall and radiation lie within valid ranges, so are not interpreted as stuck values.

**c. Uncorrelated behaviour of related variables**

This anomaly relates to dissimilar behaviour of a certain variable with respect to its correlated variable(s). For example (Figure 6), an increase in soil moisture, without a corresponding rainfall event, should be flagged for further investigation. Also, there was an example where the 10 cm earth temperature was consistently higher than its corresponding 5 cm, 20 cm, 30 cm, 50 cm, 100 cm temperatures. We expect the 10 cm earth temperature to be closer to the 5 cm and 20 cm values.

**d. Unusual fluctuations/blips**

This issue relates to unusual patterns or fluctuations in the timeseries. These need not be extreme but could result from measurement errors, potentially caused by faulty instruments, even when values fall within the defined acceptable range. For example, Figure 7 shows a site whose 20 cm earth temperature is perturbed by small blips that occur too frequently and too briefly to be a natural environmental event.

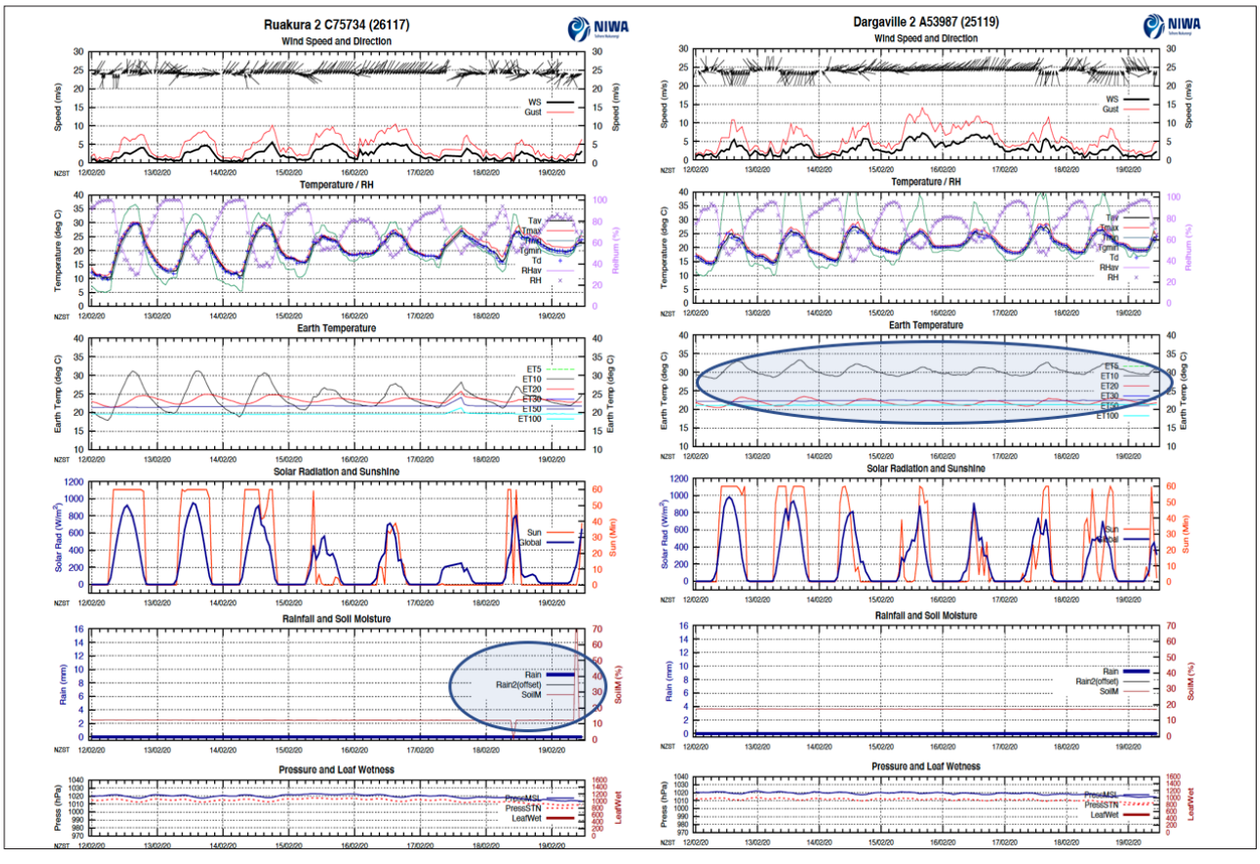



**Figure 6:** Left: a sudden increase in soil moisture value without a corresponding rainfall event; Right: the 10 cm earth temperature time series differed abnormally from the corresponding 5 cm and 20 cm values.

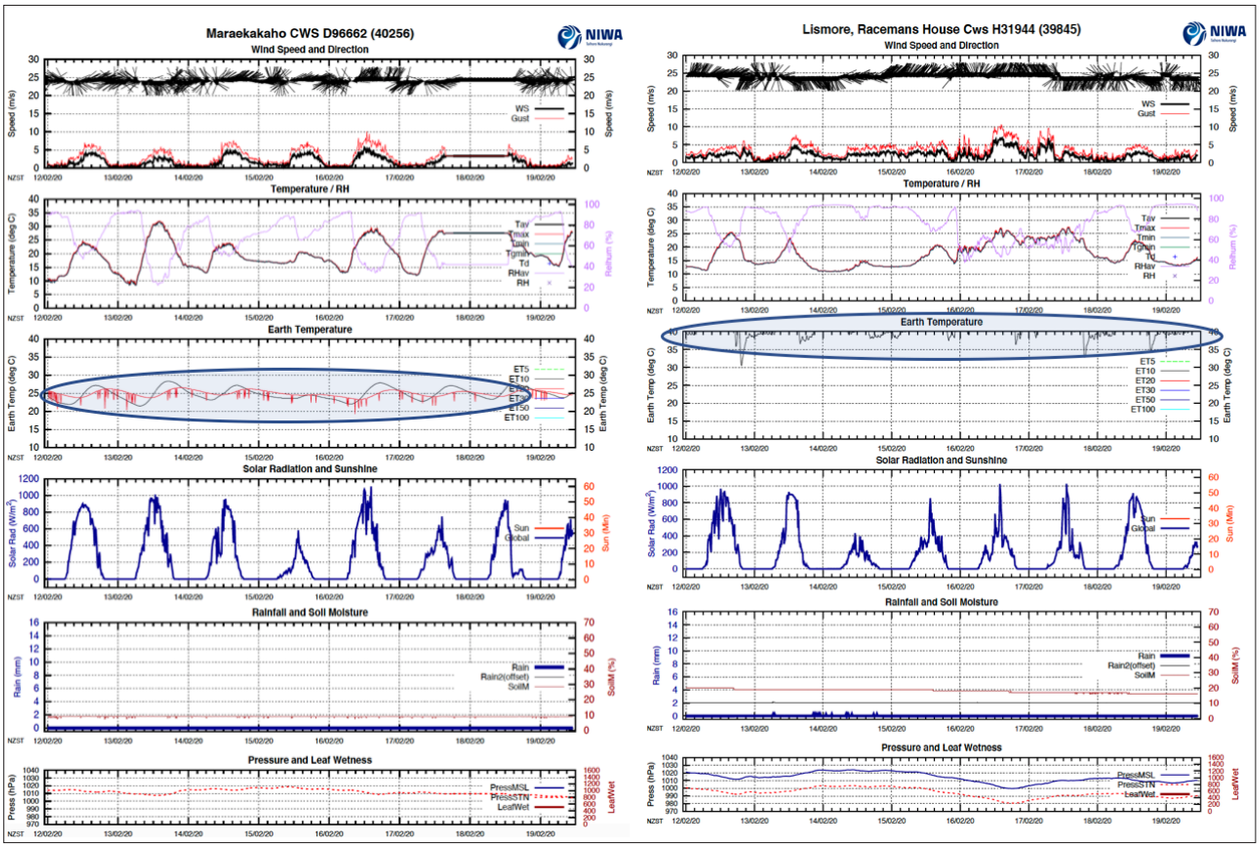



**Figure 7:** Left: the 20 cm earth temperature (red line) had small and frequent blips indicating a faulty instrument. Right: earth temperature is abnormally high and has an unusual pattern.

**e. Incomplete timeseries**

If an instrument stops measuring during the plot period, the QC plots would display gaps, or sometimes a plot might rescale the time axis to fit only the available observed values. Gaps between measurements in a particular time series are typically indicated by prescribed error default values. These are set during the data ingest process, to aid easy recognition of missing data in cases where remaining sensors are reporting normally. We classified both categories of missing data as incomplete timeseries (Figure 8).

### *2.3 Pre-processing*

In order to meet input requirements for the VGG-16 based model, we converted the archived pdf quality plots to PNG. During conversion, we applied lossless compression in our plots, to retain any uniquely representative features that could affect classification accuracy. The resultant size of the PNG images was 792 x 612 x 3 pixels. We used this image size as the input to our VGG-16 based model. The PNG files produced were 8-bit images. This included 8-bit RGB channels and an 8-bit alpha channel. The images we used for training were rendered within the sRGB colour profile. We used ImageMagick v6.9.9-26 Q16 x86_64 in conjunction with Ghostscript v9.07 to generate the PNG files. In addition, we normalised the pixel values by dividing by 255. The images were presented with plot axis labels and the station name for training with the assumption that the algorithm would either learn with the labels or learn to ignore them. The Y axis scales in the plot images are not always constant, as the scale changes with season and station location (e.g. NZ North Island, NZ South Island and Antarctica).

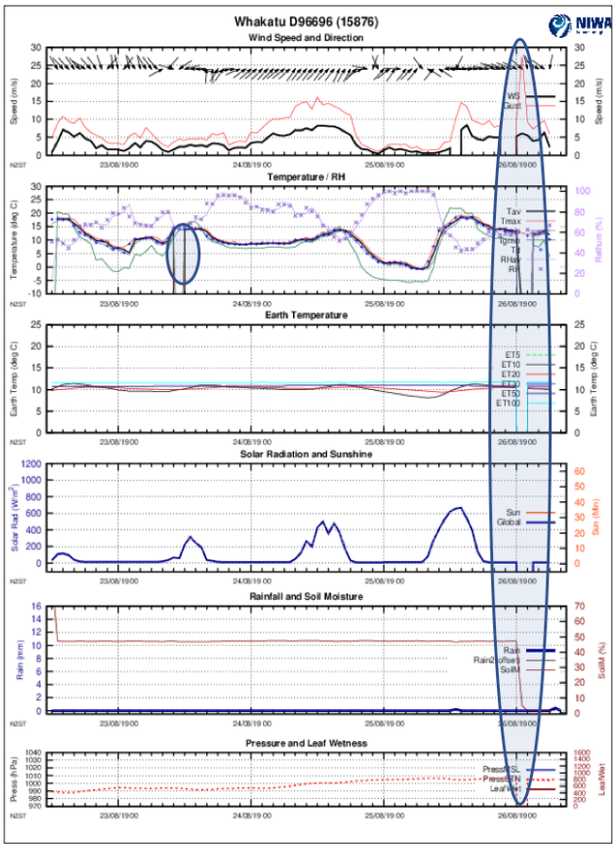



**Figure 8:** This station reported only four days of data for the whole week, and there were also missing values within the reported time series during the four days.

## 3. Network architecture, transfer learning and results

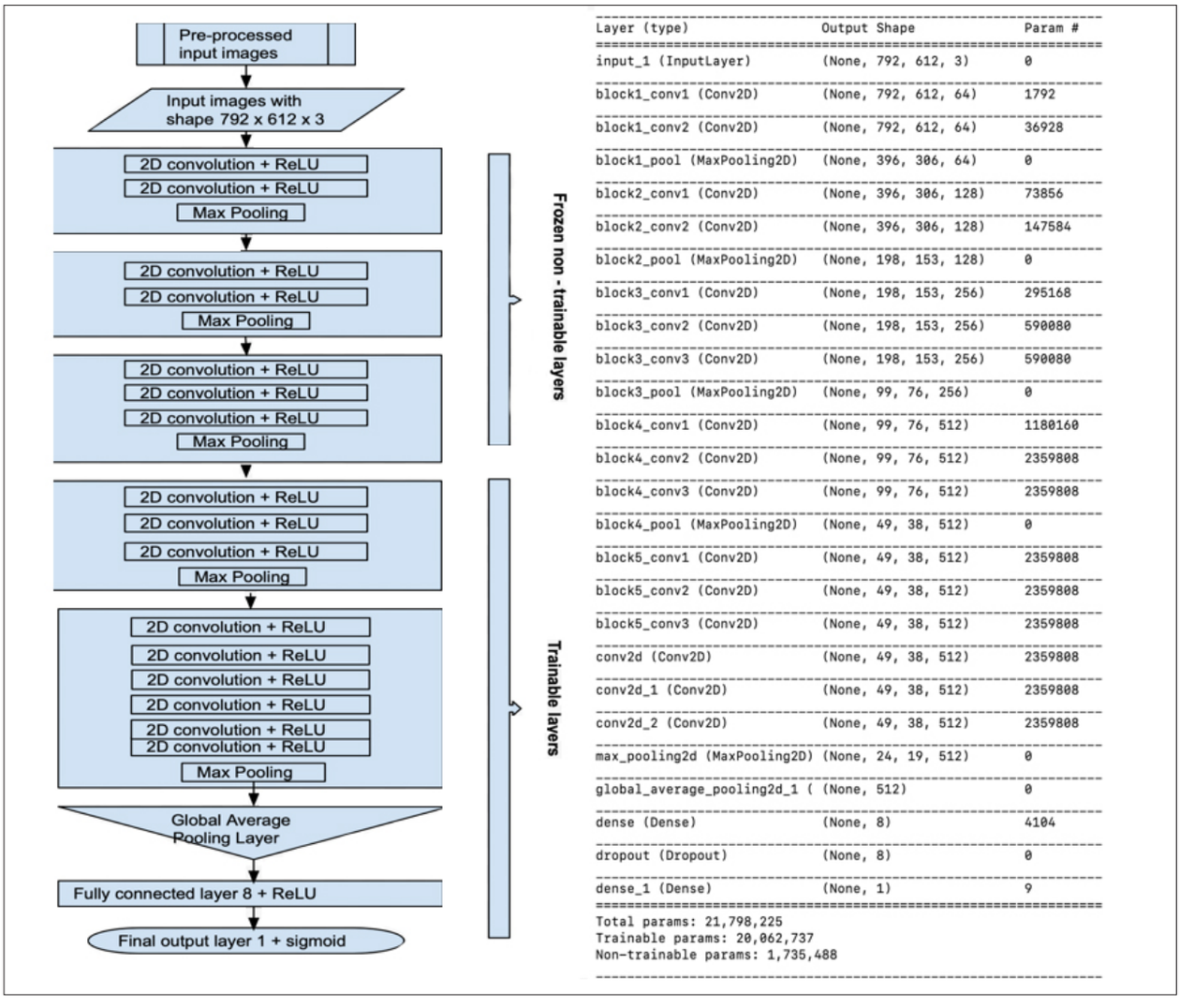
In this section we describe, in detail, the network architecture that we have used for this study, along with the results of tests done on our validation dataset. Also, we discuss our anomaly localisation process, using Gradient-based Class Activation Mapping (Selvaraju et al., 2016).

### *3.1 Network architecture*

We use the 2-dimensional QC chart images as an input into CNN for anomaly classification. As described in Section 1.2, VGG-16 has a good reputation for low classification and localisation errors. Since we are interested in both anomaly classification and localisation, we chose VGG-16 (Figure 3) network architecture for our study. This architecture is known for its simplicity and accuracy. We have used transfer learning, which is a method of fully, or partially, applying weights (knowledge) from an existing model that has been trained on a large set of images for a different problem, and customising it to solve a related problem (West et al., 2007).

We have modified the last convolutional block of VGG-16 to include three more convolutional layers with 512 filters because it improved our training and testing loss values. The filter size for these layers was 3 x 3 pixels. We have replaced the flatten layer with a global average pooling (GAP) layer (Lin et al., 2013) as it helped to minimise overfitting. Lin et al. (2013) in their study used a GAP layer instead of the flatten layer and fully connected layer, and they were able to minimise overfitting. We chose GAP as we thought that it would perform well in extracting the feature maps due to the large amount of white space in our images. On top of the GAP layer, we have added a fully connected layer of eight neurons with ReLU activation (Agarap et al., 2018) followed by a dropout and the final output layer with a sigmoid activation function. We have used Stochastic Gradient Descent (SGD) (Bottou et al., 2018) as the optimizer and binary cross-entropy loss function for training this model. Figure 9 represents the network architecture that was used in this study. Keras coding library was used for the purposes of this study (Chollet et al., 2015).

As discussed in Section 2.3, our input image size was set at 792 x 612 x 3 pixels. In our study, we have frozen the weights from VGG-16 that was trained on the ImageNet dataset for the first nine layers in the architecture (Figure 9) and retrained the remaining layers. We tried different frozen and trainable layer combinations and this combination yielded the best accuracy during our training process.




| Layer (type) | Output Shape | Param # |
|---|---|---|
| input_1 (InputLayer) | (None, 792, 612, 3) | 0 |
| block1_conv1 (Conv2D) | (None, 792, 612, 64) | 1792 |
| block1_conv2 (Conv2D) | (None, 792, 612, 64) | 36928 |
| block1_pool (MaxPooling2D) | (None, 396, 306, 64) | 0 |
| block2_conv1 (Conv2D) | (None, 396, 306, 128) | 73856 |
| block2_conv2 (Conv2D) | (None, 396, 306, 128) | 147584 |
| block2_pool (MaxPooling2D) | (None, 198, 153, 128) | 0 |
| block3_conv1 (Conv2D) | (None, 198, 153, 256) | 295168 |
| block3_conv2 (Conv2D) | (None, 198, 153, 256) | 590080 |
| block3_conv3 (Conv2D) | (None, 198, 153, 256) | 590080 |
| block3_pool (MaxPooling2D) | (None, 99, 76, 256) | 0 |
| block4_conv1 (Conv2D) | (None, 99, 76, 512) | 1180160 |
| block4_conv2 (Conv2D) | (None, 99, 76, 512) | 2359808 |
| block4_conv3 (Conv2D) | (None, 99, 76, 512) | 2359808 |
| block4_pool (MaxPooling2D) | (None, 49, 38, 512) | 0 |
| block5_conv1 (Conv2D) | (None, 49, 38, 512) | 2359808 |
| block5_conv2 (Conv2D) | (None, 49, 38, 512) | 2359808 |
| block5_conv3 (Conv2D) | (None, 49, 38, 512) | 2359808 |
| conv2d (Conv2D) | (None, 49, 38, 512) | 2359808 |
| conv2d_1 (Conv2D) | (None, 49, 38, 512) | 2359808 |
| conv2d_2 (Conv2D) | (None, 49, 38, 512) | 2359808 |
| max_pooling2d (MaxPooling2D) | (None, 24, 19, 512) | 0 |
| global_average_pooling2d_1 ( | (None, 512) | 0 |
| dense (Dense) | (None, 8) | 4104 |
| dropout (Dropout) | (None, 8) | 0 |
| dense_1 (Dense) | (None, 1) | 9 |

Total params: 21,798,225
Trainable params: 20,062,737
Non-trainable params: 1,735,488

**Figure 9:** Left: The VGG-16 network-based architecture with modifications to its last block and the fully connected layers. The first three blocks were frozen and existing VGG-16 ImageNet weights were used. The remaining layers were trained. Right: Model summary of different layers with corresponding shapes.

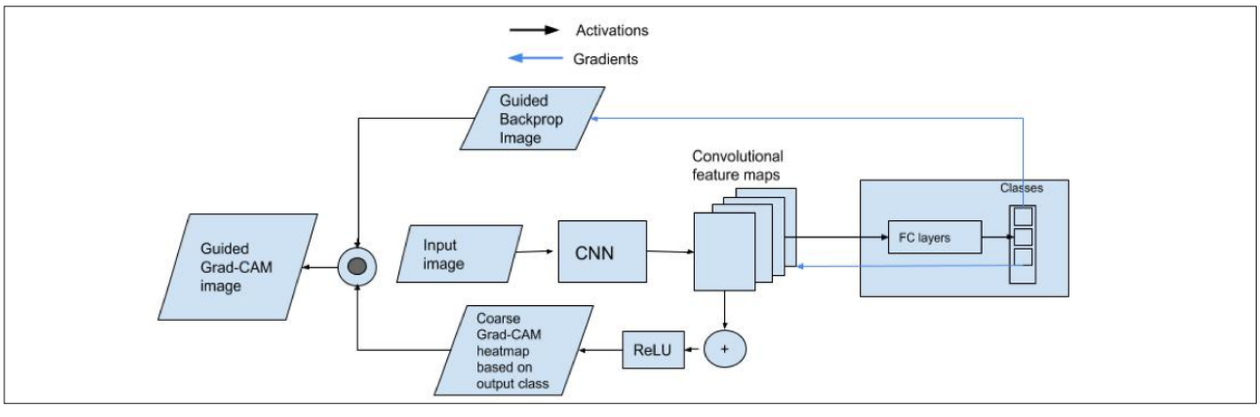



**Figure 10:** An overview of the Grad-CAM process (Selvaraju et al., 2016). As can be seen, the gradients that are backpropagated based on the output class are combined with the convolutional feature maps to generate the coarse Grad-CAM localisation. After which, this heatmap is pointwise multiplied with the guided backpropagation image to get the final Guided Grad-CAM visualisation.

### *3.2 Anomaly localisation using Gradient based Class Activation Mapping*

In this section, we discuss our anomaly localisation process. In addition to classifying anomalous images, we aimed to identify the anomalous regions within the plots. We considered the class activation map (CAM) based technique for anomaly localisation.

Zhou et al. (2015) demonstrated in their study using their CAM technique that high accuracy classification and object localisation could be achieved using a GAP layer without training on any bounding box annotation. In that study, they performed GAP on the last convolutional layer that outputs the spatial average of the feature map and uses these as features for the final output fully-connected layer. Based on this, the important regions of the image can be identified by projecting back the weights of the class in the output layer to the convolutional feature maps of the last convolutional layer. This highlights the regions of the images based on the final classification derived by the network. But this technique could be applied to a convolutional network with no fully connected layers and where the output of the GAP layer was directly fed into the output layer.

In the following year, Selvaraju et al. (2016) proposed a Gradient-based Class Activation Mapping (Grad-CAM) method to visualise classes in an image. This approach is a generalisation of CAM and could be applied to a variety of CNN networks which also includes fully connected layers. Here, the gradient of the final output class score, with respect to the feature map activations of the last convolutional layer, was computed and the global average pooled to get the important weights for regions of interest. We can then produce a visual heatmap overlaid on the actual images indicating the region of interest to the model (Figure 10).

We have used this Grad-CAM based approach to localise and highlight anomalies in our network because we had a fully-connected layer in our network. The heatmap produced in this approach indicates an anomalous region, if the classification output is flagged as anomalous. An example of a Grad-CAM output is shown in Figure 11, where the issue with soil moisture is identified by the coloured heat map lines overlapping the plot lines.

### *3.3 Results and discussion*

We evaluated the results of this model using a blind validation set not used as part of the training or testing datasets. The validation set consisted of 477 weekly plots, generated over a period of three weeks from 159 different monitoring stations. We manually identified all the potential anomalies in the validation set and compared them against the results from the model. We used >=0.5

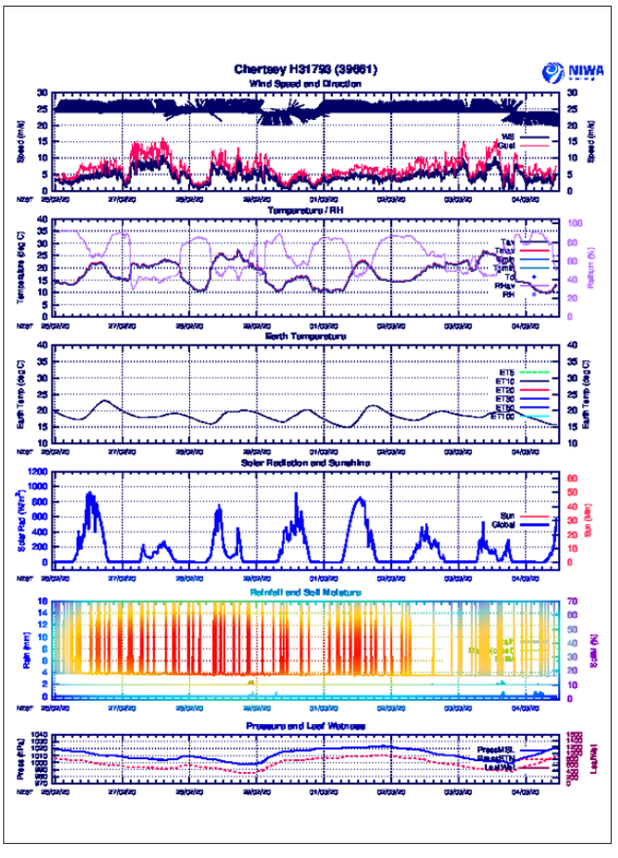



**Figure 11:** An example of a Grad-CAM output - a heat-mapped image. The anomalous region is highlighted in colour.

as the threshold probability score to classify an image as anomalous.

This threshold value was derived by plotting a Precision-Recall curve on the validation dataset for a range of thresholds (Figure 12). The plot showed that threshold values between 0.39 and 0.6 had high recall and good precision values. We chose the mid-point value of 0.5 as our threshold as that point had a higher recall value that did not significantly compromise the precision. As can be seen in the plot, the threshold could be lowered to around 0.4 to increase the recall at the expense of the precision. For operationalisation, we used a threshold of 0.5.

The Receiver Operating Characteristic (ROC) curve was plotted on the validation set and we calculated the area under the curve (Figure 13). The area was 0.98. This indicated that the model performance on the validation set is highly accurate.

In addition, we have used four scalar metrics, including F1 score and Matthews Correlation Coefficient (MCC) for scoring and evaluating the overall results (Table 1).

A perfect classifier would return an MCC score of +1; a classifier that always misclassifies would return a score of -1. In the table, all weeks have a high recall score, indicating that false negative rates were very low. Our overall precision was 0.88 indicating good positive predictive value. Precision was relatively low when compared to recall. Since we were building an anomaly detection system to identify issues with the data, our tolerance for false negatives was lower than for false positives. However, we are unable to keep the false positive tolerance too low because that would again involve considerable manual effort to go through the plots to reject false positives especially when we increase the scope of this process to cover additional stations and different time frequency data such as 10 minute and daily. Since the goal of the automated process was to save time and improve the efficiency of a QC review process, we have currently set a 20-25% tolerance on false positives and a 5-10% tolerance on false negatives. As seen in Figure 12, the threshold of 0.5 yields a precision of 0.88 and recall of 0.95. In future, we plan to test against new increasingly diverse validation datasets. During a test, if the precision of 0.75-0.8 increases the recall value significantly, we would choose an appropriate corresponding threshold. For operational runs, we aim to retain a recall score of at least 90% and a precision of at least 75%. As shown in Table 1, high values of F1 and MCC scores indicated a strong correlation between predicted and true classes.

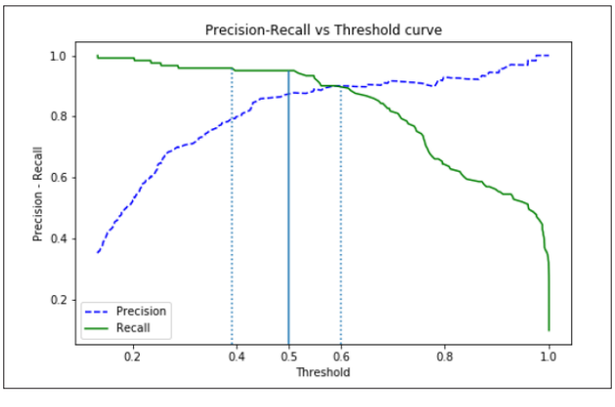



**Figure 12:** The precision and recall values plotted with respect to a range of thresholds. Dotted blue lines indicate the optimal range that could be used for thresholding that could improve either precision or recall accordingly. Threshold of 0.5 was chosen based on its optimal position with respect to good precision and a high recall value.

In addition to high classification quality, we also achieved high anomaly localisation accuracy using Grad-CAM. In

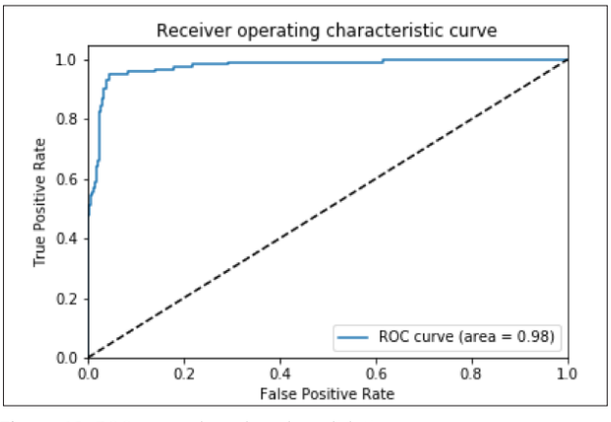



**Figure 13:** ROC curve plotted on the validation set

our validation set, the model identified 116 anomalous plots and all anomalous regions were successfully identified in 106 of these plots. So, we achieved 91% accuracy for anomaly localisation. In the remaining 10 images, anomalous regions were partially identified in nine images and, in one image, an anomalous region was incorrectly identified. In addition, anomalous regions were successfully identified for all the error types defined in Section 2. Examples of anomaly localisation can be seen in Figure 14.

Also, with the grad-CAM outputs of the validation set of all 477 images, we could see that the algorithm learnt with the plot labels present and that these were not detrimental to the performance of the algorithm.

As mentioned in section 2.3, Y axis scales were not constant in our dataset and the algorithm was able to classify and localise anomalies across all stations with varying scales.

These results suggest that the VGG-16 inspired anomaly classification model could be developed further and applied, on an operational basis, to minimise manual processing. We aim to further improve the model by detecting and adding more diverse scenarios to the training set and retrain the model. We could do this by regularly testing the output of the model, identifying scenarios where the model did not outperform the manual analysis, and then adding them to the training dataset.

During the manual review process, the QC issues that are identified or missed depends on the amount of the time spent reviewing each of these plots, along with the expertise of the manual reviewer. Currently there is no tracking of the QC issues missed during manual review and so the manual review process could not be scored directly. However, the labels of the validation set used in this study, to compare the ML process against, was manually labelled by an expert reviewer. So, Table 1 is an indirect, but fair comparison of a manual review process with this ML algorithm, with one caveat that a reviewer might not be in a position to spend sufficient time identifying and labelling these anomalies every week, as was done during this study. This manual labelling of the validation dataset by an expert reviewer was similarly done in Hundman et al. (2018), where the study used expert labelled dataset to test their LSTM based anomaly detection process.

**Table 1:** Statistics derived from the validation set, separated into individual weeks, and combined overall scores

| | Precision | Recall | F1-score | MCC |
|---|---|---|---|---|
| Week 1 | 0.94 | 1.0 | 0.97 | 0.96 |
| Week 2 | 0.87 | 0.91 | 0.89 | 0.85 |
| Week 3 | 0.88 | 0.96 | 0.91 | 0.88 |
| Overall | 0.88 | 0.95 | 0.92 | 0.89 |

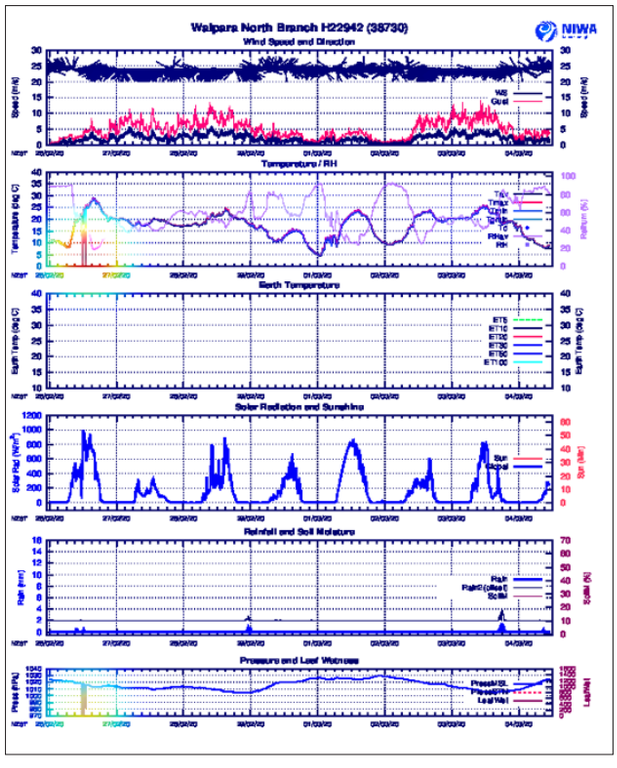



**Figure 14:** Examples of the different anomaly types defined in section 2.2. This image: example of sudden peak anomaly type correctly classified and highlighted by the Grad-CAM heatmap process.

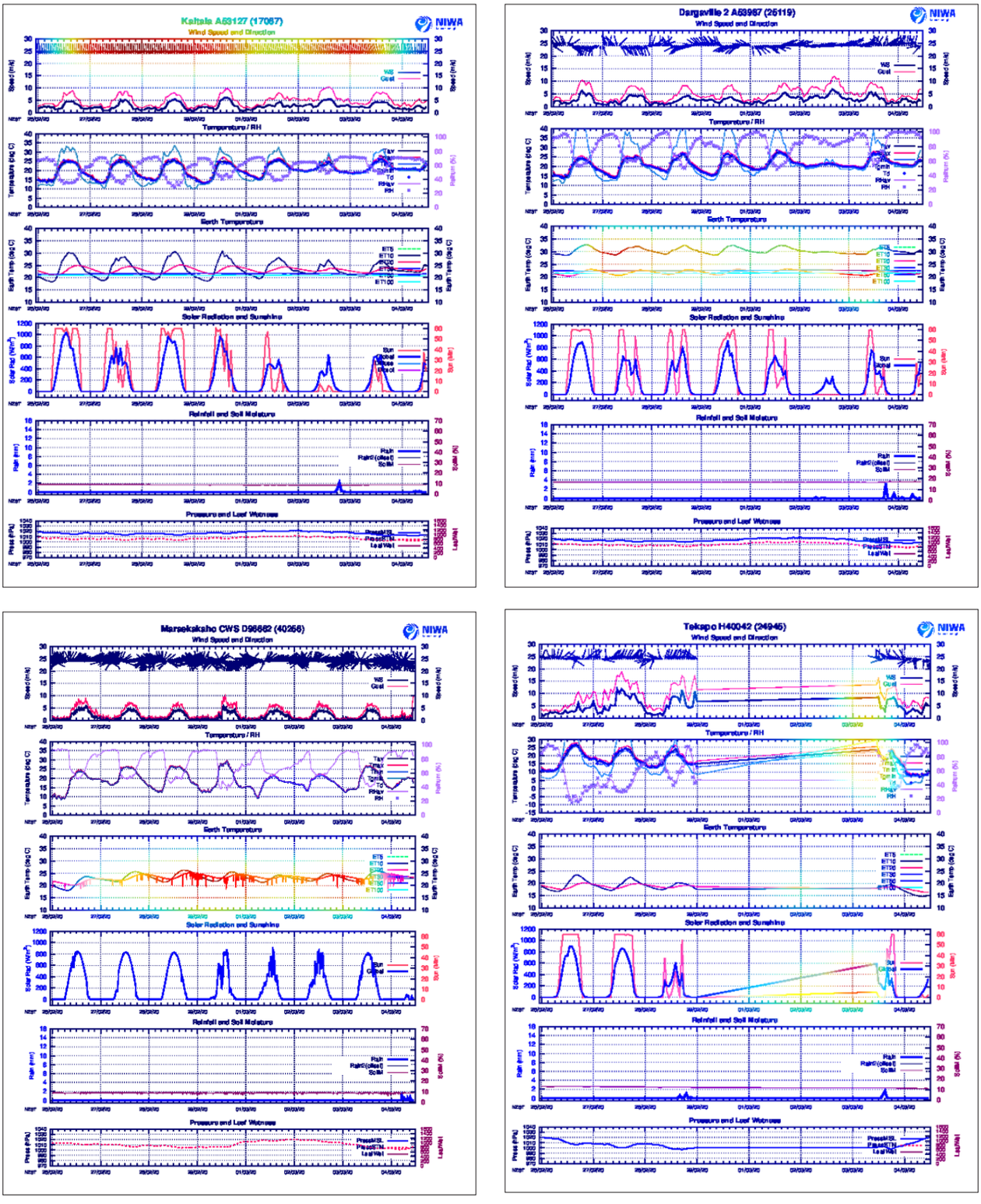



**Figure 14 (continued):** Examples of the different anomaly types defined in Section 2.2. From top left, examples of stuck instrument, uncorrelated behaviour, unusual blips and, bottom, incomplete timeseries. All these error types were correctly classified and highlighted by the Grad-CAM heatmap process.

## 4. Summary and Conclusion

In this study, we used a CNN image-based model to classify anomalies in the images of quality plots. To the best of our knowledge, this is the first attempt to use CNN to detect anomalous patterns in charts for QC purposes. From the archive of quality plots, we created a class-balanced dataset of around 1000 images for training our CNN model. This study investigated with a model using VGG-16 based architecture and were able to successfully train it using transfer learning. We were able to achieve high classification accuracy on our validation dataset, with an overall recall score of 0.95 and F1 score of 0.92. In addition, we were able to use a Grad-CAM based approach to successfully identify anomalous regions within images with an accuracy of 91%.

The above results indicate that this approach could be used on an operational basis to detect and identify anomalous plots and the specific anomalous regions within it. This method will reduce the significant effort of manually reviewing all the plots and thereby save significant time for the QC analyst. In addition, this will help to identify anomalies that might otherwise be overlooked or missed. This will improve the efficiency of the overall data quality process, as this enables the QC analyst to focus on data remediation instead of the identification of anomalies. Identification of data anomalies using this process will contribute to the improvement of the overall quality of the national climate data collection. This will in turn increase the reliability of climate data products and reports that are produced from the data extracted from the climate database. The procedure will also act as an early warning process to identify instrument issues and thereby enable more efficient planning of site visits.

This algorithm is currently operating within NIWA's CLiDB system as an 'assistant' for a weekly review of QC plots mainly on hourly and 10 min data for selected stations. This algorithm could be scaled up to detect anomalies at a greater number of sites. Also, this could be trained on different frequencies of timeseries like 1 min, 10 min, daily, sub-hourly. The algorithm could be scaled up to predict anomalies on different frequencies of a timeseries. The possibility of using this algorithm on a near real-time basis needs to be explored as this involves image generation and prediction. This algorithm falls short in generating an expected value for a variable in case of a missing observation or anomaly. As a next step, variable-specific individual timeseries algorithms can be explored that could complement this process by generating an expected value in the case of an anomalous observation. We also plan to compare the results of this study against standard variable specific timeseries-based algorithms to evaluate the relative performance of this CNN-based approach.

The work completed under this study so far has enabled us to operationalise this process in the climate database QC system. The introduction of this process has already considerably improved the efficiency of the anomaly detection process in the climate database system. As a next step, we intend to identify more scenarios of anomalies, such as those that this algorithm did not capture, and add them periodically to the training dataset for further model training. This feedback loop of identifying gaps and retraining will improve the overall accuracy of the model and the anomaly detection process. This retraining process will ensure the model can continue learning to identify increasingly diverse scenarios of anomalies. In addition, we are planning to expand this into a multi-class output model which would enable us to classify different error types for different variables. This would enable us to automatically produce anomaly reports based on different error types and parameters.

## Acknowledgements

The authors would like to thank Andrew Harper, Alan Porteous and Errol Lewthwaite for their support of this work. We would also like to thank Dr. Kameron Christopher for his valuable inputs throughout this study.

## References


Agarap, A.F. (2018). Deep Learning using Rectified Linear Units (ReLU). *ArXiv*, abs/1803.08375.

Balaji, A., Ramanathan, T., & Sonathi, V. (2018). Chart-Text: A Fully Automated Chart Image Descriptor. *ArXiv*, abs/1812.10636.

Bottou, L., Curtis, F.E., & Nocedal, J. (2016). Optimization Methods for Large-Scale Machine Learning. *SIAM Review*, 60, 223-311.

Chollet, F., & others. (2015). Keras. GitHub. Retrieved from https://github.com/fchollet/keras

Cliche, M., Rosenberg, D.S., Madeka, D., & Yee, C. (2017). Scatteract: Automated Extraction of Data from Scatter Plots. *ArXiv,* abs/1704.06687.

Cui, Z., Chen, W., & Chen, Y. (2016). Multi-Scale Convolutional Neural Networks for Time Series Classification. ArXiv, abs/1603.06995.

Davila, K., Setlur, S., Doermann, D., Bhargava, U. K., & Govindaraju, V. (2020). Chart Mining: A Survey of Methods for Automated Chart Analysis. I*EEE transactions on pattern analysis and machine intelligence*, 10.1109/TPAMI.2020.2992028. Advance online publication.

Deng, J., Dong, W., Socher, R., Li, L., Li, K., & Li, F. (2009). ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248-255.

Hundman, K., Constantinou, V., Laporte, C., Colwell, I., & Söderström, T. (2018). Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding. P*roceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.*

Hüsken, M., & Stagge, P. (2003). Recurrent neural networks for time series classification. *Neurocomputing*, 50, 223-235.

Inada, M., & T. Terano (2005). QC chart mining: extracting systematic error patterns from quality control charts. 2005 IEEE International Conference on Systems, Man and Cybernetics, Waikoloa, HI, 2005, pp. 3781-3787 Vol. 4,

Karim, F., Majumdar, S., Darabi, H., & Chen, S. (2018). LSTM Fully Convolutional Networks for Time Series Classification. *IEEE Access*, 6, 1662-1669.

Krizhevsky, A., Sutskever, I., & Hinton, G.E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *NIPS*.

Lin, M., Chen, Q., & Yan, S. (2014). Network In Network. *CoRR*, abs/1312.4400.

Liu, X., Klabjan, D., & Bless, P.N. (2019). Data Extraction from Charts via Single Deep Neural Network. *ArXiv*, abs/1906.11906.

Park, J. (2018). RNN based Time-series Anomaly Detector Model Implemented in Pytorch. Published in website URL: https://github.com/chickenbestlover/RNN-Time-series-Anomaly-Detection.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *ArXiv*, abs/1505.04597.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M.S., Berg, A.C., & Li, F. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115, 211-252.

Selvaraju, R.R., Das, A., Vedantam, R., Cogswell, M., Parikh, D., & Batra, D. (2016). Grad-CAM: Why did you say that? Visual Explanations from Deep Networks via Gradient-based *Localization*.

Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. *CoRR*, abs/1409.1556.

Wen, T., & Keyes, R.W. (2019). Time Series Anomaly Detection Using Convolutional Neural Networks and Transfer Learning. *ArXiv*, abs/1905.13628.

West. J., Venture. D., and Warnick. S., (2007). Spring research presentation: A theoretical foundation for inductive transfer. Brigham Young University, College of Physical and Mathematical Sciences.

Yang, J., Nguyen, M.N., San, P.P., Li, X., & Krishnaswamy, S. (2015). Deep Convolutional Neural Networks on Multichannel Time Series for Human Activity Recognition. *IJCAI*.

Zhao, B., Lu, H., Chen, S., Liu, J., & Wu, D. (2017). Convolutional neural networks for time series classification. *Journal of Systems Engineering and Electronics*, 28, 162-169.

Zheng, Y., Liu, Q., Chen, E., Ge, Y., & Zhao, J.L. (2014). Time Series Classification Using Multi-Channels Deep Convolutional Neural Networks. *WAIM*.

Zhou, B., Khosla, A., Lapedriza, À., Oliva, A., & Torralba, A. (2016). Learning Deep Features for Discriminative Localization. *2016 IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), 2921-2929.

# Surface temperature trends and variability in New Zealand and surrounding oceans: 1871-2019

M.J. Salinger[1], H.J. Diamond[2] and J.A. Renwick[1]

[1] *School of Geography, Environmental and Earth Sciences, Victoria University of Wellington, P. O. Box 600, Wellington, New Zealand*
[2] *NOAA/Air Resources Laboratory, College Park, Maryland 20740, USA*

Correspondence: jimbosalinger09@gmail.com ORCID: 0000-0002-5782-1411

**Key words:** New Zealand, surface temperature trends, temperature variability, sea surface temperature, anthropogenic global warming, CMIP5, Southern Annular Mode, Interdecadal Pacific Oscillation, El Niño Southern Oscillation, volcanic eruptions.

## Abstract

We compare homogenised series of maximum, minimum, and mean air temperature averaged over New Zealand (NZ) for the period 1871-2019, with surrounding ocean surface data. Temperatures over the New Zealand Exclusive Economic Zone exhibit an increase (linear trend) of 0.66°C from 1871-2019. As well as the anthropogenic warming signal (identified by CMIP5 simulations), interannual to decadal variability is also examined. Significant volcanic eruptions have caused temporary cooling and the positive trend in the Southern Annular Mode is linked to warming over NZ. The influences of the Interdecadal Pacific Oscillation (IPO) and the El Niño/Southern Oscillation (ENSO) are also evident in the temperature series. All warm years (>+0.45°C above the 1981-2010 normal) occur from 1998 onwards, and all of the nine cold years (<-0.84°C below the 1981-2010 normal) occur prior to 1933. We conclude that the climate teleconnections that cause interannual to decadal variability (ENSO and IPO) are key factors in these results, beyond the anthropogenic warming signal.

## 1. Introduction

The New Zealand (NZ) region, including the entirety of its Exclusive Economic Zone (EEZ), an area of 4 million km², represents a significant area in the southwest Pacific, and is of similar size to the Indian subcontinent. NZ lies in the South Pacific Ocean, largely in the temperate zone, but extending into the sub-Antarctic zone in the south.

This study examines variations and trends in air and sea surface temperature (SST) for New Zealand's Exclusive Economic Zone (EEZ) area using a network of 22 surface air temperature series (NZ22T) and SST from the surrounding oceans in the New Zealand Exclusive Economic Zone, from 1870 to 2019. Land and SSTs are combined to form a combined temperature series (NZEEZT). New Zealand land surface temperature data, which have been adjusted for inhomogeneities, have been demonstrated to be consistent with SSTs from 1870 (Folland and Salinger, 1995). In this paper we document trends and variability in NZ22T combined with Extended Reconstructed Sea Surface Temperature version 5 (ERSST, Huang et al., 2017) for the New Zealand region to NZEEZT on annual to multidecadal time scales. We investigate relationships with anthropogenic global warming (AGW) associated with human-induced increases in atmospheric

*Submitted: 28 November 2019*
*Accepted: 13 November 2020*

greenhouse gas concentrations by comparing with temperature series from CMIP5 simulations, with major volcanic eruptions, and with the Southern Annular Mode (SAM), El Niño/Southern Oscillation (ENSO) and the Interdecadal Pacific Oscillation (IPO).

The extent to which historic temperature data are adequate to describe regional trends and variations over the land and the oceans is an open question. For the land surface, issues arise because artificial changes can be introduced, such as variations in the exposure of thermometers (Parker, 1994), alterations in observing methods and times (Karl et al., 1986) and changes in the environment surrounding climate sites, such as those due to urbanization (Jones et al., 1990; Karl et al., 1993). Sea surface temperature data have also suffered from artificial changes caused mainly by differences in the methods of observation. In the mid- to late nineteenth century wooden buckets were often used to collect seawater and were largely replaced by uninsulated canvas buckets by the early 20th century. A further change to the predominant use of engine intake thermometers occurred in the early part of the Second World War, followed by the development of insulated sea-water buckets (Bottomley et al., 1990; Folland and Parker, 1995). The parametric uncertainty, including measurement and sampling errors, and variability of the (ERSSTv5) dataset is depicted in Figure 1 of Huang et al. (2018). The uncertainties are generally larger in the earlier period (1854–1900) than the latter periods of 1900–50 and 1950–2010. For the area of study in this paper, the uncertainties in SST were less than 0.4°C from 1854-1900, around 0.2°C from 1900-1950, and less than 0.2°C from 1950 to the present.

Previous work on New Zealand land temperature trend has focused on the "seven-station series" (7SS); Mullan et al. (2012) noted that the linear trend in the 7SS was +0.91°C/century in the period 1909-2009. The 95% confidence interval on the calculated linear trend was ±0.29 °C/century. Three global products that apply to land temperature (GISS, NOAA, HadCRU4), over the period 1909-2015 show linear trends of +1.12°C, +1.16°C and +1.03°C/century respectively. A significant contribution to the warming can be attributed to greenhouse gas increases (Dean and Stott, 2009).

Several studies have described temperature variability in the NZ region associated with various climate teleconnections. The Southern Annular Mode (SAM) operates on shorter than annual time scales and describes north-south meandering of the eddy-driven jet over the Southern Oceans and the associated storm track (Kidston et al., 2009). A positive SAM phase is associated on average with temperatures at least 0.5°C above normal throughout western parts of the North and South Islands, resulting from weaker than normal westerly winds. In the negative phase, the SAM shows the opposite, with cooler temperatures in the west of both islands. Kidston et al. (2009) note that these temperature anomalies are much stronger in summer (December-February) than in winter (June-August). The SAM has trended increasingly positive during the 20th century contributing to a warming trend over NZ (Arblaster et al., 2011).

For interannual relationships, Gordon (1986) found that land surface air temperatures in NZ were positively correlated with the Southern Oscillation Index (SOI, Troup, 1965). In the El Niño phase, NZ experiences more frequent and stronger than normal southwesterly winds. This generally results in lower temperatures for New Zealand. The La Niña phase is essentially the opposite of El Niño. New Zealand experiences more northeasterly flows, higher temperatures and air pressure tends to be higher than normal over the South Island. The most notable El Niño years with cooler than normal temperatures occurred in 1905, 1912, 1919, 1941 and 1965. Much warmer than normal years associated with La Niña episodes occurred in 1917, 1971, 1999 and 2018.

Interdecadal climate variability in the Pacific is driven

by the IPO (Parker et al., 2007; Power et al., 1999). The partition between recent IPO phases occurred at 1945, 1977 and 1998 (Salinger et al., 2001, Henley et al., 2015). The IPO phases induce decadal variations in NZ climate variability, especially temperature change. Changes in mean annual surface air temperatures from the IPO positive to IPO negative phases (1946-1976 compared with 1930-1945, and 1998-2019 compared with 1977-1997) show accelerated warming over much of the region. During the two later 20th century IPO positive phases regional warming slowed (Salinger and Mullan, 1999).

Finally, explosive volcanic eruptions (Kelly et al., 1996) cause climate impacts for up to 30 months. Examination of six eruption events (Krakatau (1883); Tarawera (1886), Pele, Soufrire and Santa Maria (all 1902); Agung (1963); El Chichon (1982) and Pinatubo (1991)) found good agreement between the spatial patterns of temperature anomalies associated with these events. The composite response shows cooling of just under 0.2°C on average globally; the response is strongest and statistically significant over a 2-year period starting early in the year after the eruption. The total area in which statistically significant departures are found is considerable. Volcanic eruptions in tropical latitudes can be global in their effect, whilst those in temperate latitudes only affect those regions in the eruption hemisphere (Robock, 2003). Robock and Free (1995) conclude that for both hemispheres there is no evidence of an impact of volcanic eruptions on El Niño/Southern Oscillation (ENSO) events. Salinger (1998) documented the impacts of major volcanic eruptions which inject significant amounts of dust and sulphate aerosols into the atmosphere on atmospheric circulation and temperatures for the six major volcanic events above on NZ. The effects commenced rapidly, in the first few months after the volcanic eruption and lasted 24 months on average, with surface temperatures depressed in the region by 0.3°C to 0.4°C from one to 21 months after the eruption. Atmospheric circulation anomaly patterns show more patterns of south westerlies and troughs.

## 2. New Zealand's temperature data

***Land Surface Temperature***

Daily maximum and minimum temperature measurements in New Zealand were irregular until 1859, when the Colonial Secretary of the time, Mr Stafford, supplied standard instruments to observers at eight locations. Responsibility for meteorological stations was transferred to Sir James Hector, Director of the Colonial Museum and Geology Department, in 1867, and he introduced unusually uniform and rigorous methods of observation (Hector, 1869). These included the use of Stevenson screens throughout the network, probably the earliest attempt to do this in any country. The temperatures measured used high-quality sheathed thermometers read at 0930 New Zealand Standard Time (NZST) which were regularly checked for calibration. Since then, instruments and observation methods have remained the same except for a change in the definition of NZST from the mean noon fixed at 172°E to use mean noon at 180°, and an observation time change to 0900 NZST. Standard reporting forms and meteorological notebooks for entering the recordings at the time of observation were also introduced. Indeed, every effort was made to make the observations at best international standards, which has lasted to the present.

The concept of New Zealand's seven-station series (7SS) was developed by Salinger (1980, 1981) based on long term stations. Salinger et al. (1992) homogenised temperature for 24 reference climate stations, which included the 7SS as a subset. The homogenisation of climate data is a process of calibrating old meteorological records to remove spurious factors which have nothing to do with the actual temperature change.

Daily maximum and minimum temperatures, using mercury, glass and alcohol in glass self-registering thermometers, were homogenised at these NZ climate

stations including 22 stations, termed NZ22T (Figure 1). The homogenisation was necessary because some of these locations had observations at more than one site. These are distributed evenly over the land area of New Zealand, 11 in the North Island, and 11 in the South Island, and is representative of New Zealand regional climate. The two stations not used of the 24 were remote offshore islands far away from the main islands of NZ. Records were screened for inhomogeneities by examining station histories and comparisons were made between neighbouring stations to identify unrecorded site changes or other environmental changes near the climate station sites. Procedures used to homogenise the data (Rhoades and Salinger, 1993) included cumulative sum plots and neighbouring station comparisons. For a few early records where neighbouring stations do not exist, other techniques were used to evaluate the significance of, and make adjustments for, suspected inhomogeneities (Rhoades and Salinger, 1993). This included the estimation of the size and standard error of a change point (discontinuity) in the temperature time series by (i) using manual data before and after the change points, (ii) using monthly data for specified symmetric intervals before and after the discontinuity, and (iii) estimating the most likely change points using the size of the change from an appropriate average.

Mullan et al. (2010, 2012) re-analysed the 7SS over the period 1909-2010 with the key result is that the NZ-wide warming trend from the 7SS series trend is +0.91 °C/century, which is identical to the previous estimate of +0.91 °C/century from Salinger et al. (1992). From 1870-80 10 stations were available, this reduced to six from 1881-1895, seven from 1895-1909, increasing to 14 from 1910–1930, 18 from 1931-1940 then 22 from 1940 onwards. The correlation between 7SS and NZ22T on all time scales starting across all decadal intervals (viz 1871-2019, 1881-2019 etc.) from 1871-2019 is 0.99.

The first conjoint analysis of trends and variability of both land and marine surface temperatures in the NZ region was by Folland and Salinger (1995) with later publications by Folland et al. (1997, 2003). Folland and Salinger (1995) concluded that the observed trend and shorter-term variations in the 7SS are in excellent agreement with those of nearby SST. Folland et al (2003) note *non detrended* standard deviations and correlations with annual land data from eight New Zealand stations, and marine series, are 0.35°C, 0.83 for SST and 0.32°C, 0.85 respectively with night marine air temperature series over the period 1870-1998. Folland et al. (2003) concluded that consistency between the land and marine temperature series both seasonally and interannually is very good. Mullan et al. (2010) found that the spatial pattern in the warming is consistent with changes in SST around NZ. More warming occurred in the north of the country, and less warming in the south.

***Sea Surface Temperature***

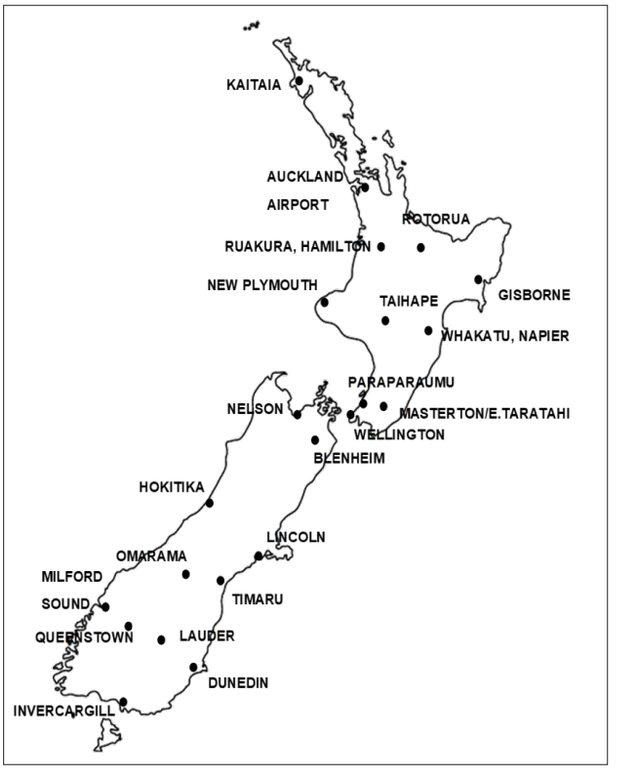



**Figure 1**: Location of 22 stations to measure land surface temperature (NZ22T) after Salinger (1992). The 7SS is a subset of NZ22T, which includes Mangere (Auckland), Kelburn (Wellington), Masterton, Nelson, Hokitika, Lincoln and Dunedin.

The SST data set used for the NZ region was taken from the ERSSTv5 of Huang et al. (2017, 2018). The monthly global ERSSTv5 has been revised to incorporate a new release of ICOADS release 3.0 (R3.0), a decade of near-surface data from Argo floats, and a new estimate of centennial sea ice from HadISST2. The resulting ERSST estimates have more realistic spatiotemporal variations, better representation of high-latitude SSTs, and ship SST biases are now calculated relative to more accurate buoy measurements, while the global long-term trend remains about the same. These data extend back to 1854. SST data from ERSSTv5 from 2°x2° grid points (83) from within New Zealand's Exclusive Economic Zone (EEZ) (Figure 2) between 32-54°S, 162°E-176°W were averaged to form a monthly ERSST series around New Zealand. The averaging done in the analysis here was an area-weighted average by using the cosine of latitude, given the significant convergence of meridians across the latitude range of 32-54°S.

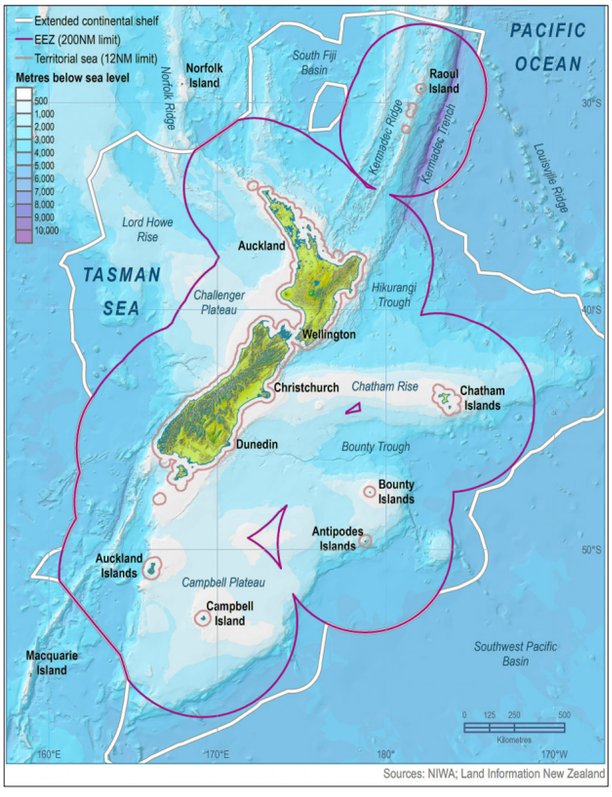



**Figure 2:** New Zealand's exclusive economic zone. Source: https://www.mfe.govt.nz/publications/marine-environmental-reporting/our-marine-environment-2016-introduction-our-marine

The accuracy of SST analyses is mostly dependent on how the biases of observations from different instruments are corrected (Kent et al., 2017). Huang et al. (2017) suggested that the globally-averaged SST in the ERSSTv5, is systematically about 0.1°C lower than the previous version, ERSST.v4 (Figure 1, from Huang et al. (2018)-solid black and red lines). The lower SST in ERSST.v5 results from the biases of ship observations adjusted to more accurate or homogeneous buoy observations. The quality-controlled SST data were bin-averaged to 2°x2° grid boxes at a monthly timescale from 1920 to 2016, and globally integrated numbers and area coverages of observations were calculated. The area coverage is a ratio of the total gridbox area containing observations over the total ocean area. Calculations show that numbers of observations are solely from reversing thermometers (RT) attached to Nansen–Niskin hydrographic bottles; conductivity–temperature–depth (CTD RT/CTD) over the timeframe from 1920–40, dominantly from mechanical bathythermographs (MBT) and RT/CTD over 1940–70, and mostly from expendable bathythermographs (XBT) and RT/CTD over 1970–2010 (as noted in Figure 2a of Huang et al., 2017). Bottomley et al. (1990) found that a monthly 5°x5° SST value is considered usable if it contains at least one quality-controlled data value; and with the ERSSTv5 dataset meeting this threshold at a resolution of 2.5°x2.5° resolution, we feel confident in the validity of the data prior to 1920. Uncertainties in the ERSSTv5 dataset are generally less than 0.4°C from 1854-1900, around 0.2°C from 1900-50, and less than 0.2°C from 1950 to the present (Huang et al., 2016, 2019).

### *Coupled Model Inter-comparison Project Phase 5 (CMIP5) simulations*

CMIP5 experiments with coupled atmosphere-ocean global climate models, most of which were reported on in the IPCC Fifth Assessment Report, Working

Group I, used the average of all suitable CMIP5 model simulations for 1900-2015 which had AGW and natural variability runs (NAT) (B. Mullan pers. comm., for sea surface temperatures for the New Zealand region (35-50°S, 160°E-180°), 13 CMIP5 models were suitable). The methodology is that of Flato et al. (2013). The CMIP5 (AGW) trends and 95% confidence interval are shown in Figure 3. These simulations show the magnitude of the AGW warming signal compared with NAT over time, and do not relate to any time period.

***Circulation indices***

Indices of variability used were the extended Gong and Wang (1999) SAM index (https://www.esrl.noaa.gov/psd/data/20thC_Rean/timeseries/monthly/SAM/sam.20crv2c.long.data). These were extended for 2012-2019 by regressions with the Marshall (2003) SAM index (https://legacy.bas.ac.uk/met/gjma/sam.html), with a correlation coefficient r = 0.87. The IPO from Henley et al. (2015) (https://www.esrl.noaa.gov/psd/data/timeseries/IPOTPI/tpi.timeseries.hadisst11.data) and the SOI (Troup (1965) (http://www.bom.gov.au/climate/current/soihtm1.shtml) were used, all of which extended back to the 1870s. The SAM, SOI and IPO are standardised indices (unit standard deviation).

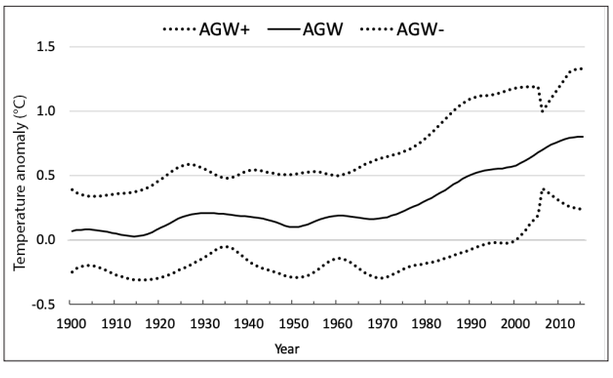



**Figure 3:** Warming in the New Zealand region (°C), (35-50°S, 160°E-180) from AGW 1900-2015. Time series are CMIP5 (AGW) models (solid) SSTs from 13 CMIP5 (AGW) models (solid) and 95% confidence interval (dotted) (AGW+, AGW-).

Atmospheric circulation patterns over New Zealand were characterized (as in Salinger et al., 2019) by correlating annual values of indices from the Kidson (2000) three regimes (Trough, Zonal and Blocking) for the period 1948-2019 and two of the Trenberth (1976) circulation indices Z1 (west-east zonal circulation) and M1 (south-north meridional circulation) from 1896-2019 (Table 1) with annual temperature anomalies. The values of the latter have been standardised for the 1896–2019 period.

***Reanalyses***

Two reanalysis data sets, ERA-Interim (Dee et al., 2011) and the Twentieth Century reanalysis (20CR, Compo et al., 2011) were used. ERA-Interim fields were obtained for the period 1979-2019 and 20CR for the period 1871-2014. Monthly mean anomaly fields were calculated as differences from monthly averages over the 30-year period 1981-2010 (used for both reanalyses).

## 3. Methods

***Land and sea surface temperatures***

The homogenised land data sets of Salinger et al. (1992) were updated, checked for homogeneity, and combined with the 7SS of Mullan et al. (2010). The pre-1909 records were rechecked with the homogenisations of Salinger (1981), Salinger et al. (1992) and Mullan (2012), by examination of maximum and minimum temperatures with the Rhoades and Salinger (1993) methodology, and careful examination of the limited metadata summary from Fouhy et al. (1992). The re-homogenised data were used to form a continuous monthly series of NZ22T from 1870–2019 with anomalies from the 1981-2010 climatological normal period. Table 2 shows the pre-1909 series adjustments compared with Salinger et al. (1992) and Mullan (2012). The twenty-two station values were averaged. The land and sea surface temperature series were combined to produce a weighted anomaly series from the

**Table 1:** Definitions of large-scale climate teleconnections, regional circulation indices (Trenberth, 1976) and Kidson (2000) Circulation Regimes affecting New Zealand

| | Index | Definition | Wind anomaly (+/- index) |
|---|---|---|---|
| Climate Teleconnections | SAM | The zonal mean sea level pressure difference between 40° and 65°S | Weaker/stronger westerlies over the South Island. |
| | SOI | Normalised sea level pressure difference between Tahiti and Darwin | Northeast/southwest (over NZ). |
| | IPO | Pacific decadal SST anomalies - a long-term oscillation of sea surface temperatures in the Pacific Ocean that can last from 20 to 30 years. | West-southwest/east-northeast. |
| Trenberth Indices | Z1 | Pressure difference between Auckland and Christchurch | Measures strength of westerlies over NZ: positive stronger, negative weaker westerlies. |
| | M1 | Pressure difference between Hobart and Chatham Islands | Positive southerly, negative northerly airflow over NZ. |
| Kidson Regimes | Trough | Frequent troughs crossing the country. | |
| | Zonal | Highs to the north with strong zonal flow to the south of NZ | |
| | Blocking | Blocking patterns with highs more prominent in the south. | |

1981-2010 period according to the following surface areas (ERSSTv5 4,083,744 km$^2$, NZ22T 268,680 km$^2$ or 15:1), where NZEEZT is 15/16*ERSST + 1/16*NZ22T. Vose et al. (2012) addresses this in creating a global combined ocean/land dataset. This dataset is well-recognized and states the following: "The reconstruction process dictates that Land Surface Temperatures (LST) and SSTs should be processed separately and then merged together into a single reconstruction".

There are several reasons for performing separate reconstructions, the first being major differences in spatial coverage between the land and ocean surface (the latter having systematic gaps early in the record). Another motive for separate reconstructions is that LST and SST observations are fundamentally different; a land grid box represents at least one month of daily observations from at least one station, whereas in an ocean grid box observation frequency is quite different. Although the combined series are heavily weighted by the ERSSTv5, it is valid to combine these for the NZ region so as to obtain values representative of all the area NZ has jurisdiction over for its economy, particularly for fisheries. The LST is relevant for the agricultural economy. The work documented by Vose et al. (2012) coupled with the work of Smith and Reynolds (2005) and Smith et al. (2008) document how to merge a land and ocean database as to what has been done in this paper.

**Table 2:** Adjustments of the pre-1910 time series for the 7SS compared with Salinger et al. (1992) and Mullan (2012) in °C

| Time period | Salinger et al. (1992) | Mullan (2012) |
|---|---|---|
| 1870-1880 | -0.08 | -0.10 |
| 1880-1900 | +0.03 | -0.09 |
| 1900-1910 | +0.03 | +0.07 |

***Trends, Variability and Extremes***

Detrended annual values were also generated for the NZEEZT area by subtracting out linear trends (Mullan et al., 2016) from NZ22T, ERSST and NZEEZT. Temporal relationships with atmospheric teleconnections of variability, using actual and linearly detrended data, were examined as follows: (1) major volcanic eruptions; (2) the

SAM for both trends and annual periods; (3) the SOI for interannual periods, and (4) the IPO for decadal periods. Bivariate correlations were used to identify relationships between both forcing of temperature by AGW, and teleconnections with SAM, SOI and the IPO. Multiple linear regression was used to explore multivariate relationships between NZEEZT and these four factors.

For the analysis of the ERSSTv5 data in the NZ EEZ region we performed a fairly straightforward (averaged over the New Zealand box) for the following 5 timeframes of 1871-1900; 1901-30; 1931-60; 1961-90; and 1991-2019; and examined the distribution of those anomalies (not shown). The distributions do not change significantly, and the variances across all five time periods are all comparable. A consultation with the ERSSTv5 developer (personal communication, Huang, September 2020) concurred with our finding that the SST trend is relatively strong in the NZ region, with the difference between the 1901-1930 and 1990-2019 30-year time periods being at +0.65°C, and, when 20-year averages are used, the difference between the 1901-1920 and 2001-2019 time periods is +0.82°C. This is also confirmed in a study of global SST trends in a prior version of ERSSTv5 (e.g., ERSSTv4), and that analysis indicated that the SST trend from 1901-2014 was relatively strong as well (Huang et al., 2016).

***Analogues***

To determine circulation features associated with extremely warm and cold years, sea-level pressure and 500hPa height anomalies for the two reanalyses were extracted for the warmest and coldest years in the actual and NZEEZT-AGW series. Anomaly values chosen for the 5$^{th}$ and 95$^{th}$ percentile values respectively (>+0.45°C and <-0.84°C) for the NZEEZT series which yielded 7 warm (1971, 1999, 2001, 2013, 2016, 2017, 2018) and nine (1902, 1903, 1906, 1912, 1919, 1926, 1930, 1931, 1932) cold cases respectively. For the NZEEZT-AGW series, the 5$^{th}$ and 95$^{th}$ percentile values were also chosen ranged from >+0.15°C and <-1.05°C, yielding four warm (1916, 1917, 1962, 1971) and six cold (1902, 1906, 1930, 1931, 1992, 1993) cases respectively. For this sample the four warm years were different, and the cold years the same.

## 4. Results

***Homogenisations***

Comparisons of NZ22T with the 7SS from Mullan (2012) and Salinger et al. (1992) show the very small adjustments made during the pre-1909 period (Table 2). These are -0.10°C and -0.08°C lower than Mullan et al. (2012) and Salinger et al. (1992) respectively for the 1870-1880 period, and -0.09°C lower with Mullan et al. (2012) and +0.03°C higher than Salinger et al. (1992) for the 1881-1900 period. For the period 1900–1910 the differences are +0.07°C and +0.03°C compared with Mullan et al. (2012) and Salinger et al. (1992) respectively. From 1909 the series is similar to Mullan et al. (2012) with the use of data from more stations to a maximum of the 22 stations.

***Variability***

From 1870-1895 temperatures were 0.4°C below the 1981-2010 normal, then decreased to 0.8°C below normal in the early 1900s (Figure 4). It was cooler by a similar amount again in the early 1930s; both these periods being the coolest in the temperature record. During the 1910s, 1920s and 1940s mean temperatures were 0.4°C below normal. Temperatures subsequently were near normal in the 1950s, 1970s and 1980s, with a brief cooler excursion in the 1960s. Temperatures subsequently decreased sharply to around 0.5°C below normal in early 1990s, then increased rapidly to 0.1°C above normal, then averaged 0.4°C above normal for the 2010-2019 period. On an annual basis the impacts of major volcanic eruptions clearly show a decrease in most cases by about 0.5°C for 12 to 18 months from the relevant decadal averages. The six major volcanic eruptions that affected New Zealand climate depressed temperatures briefly between 0.3 to 0.5°C.

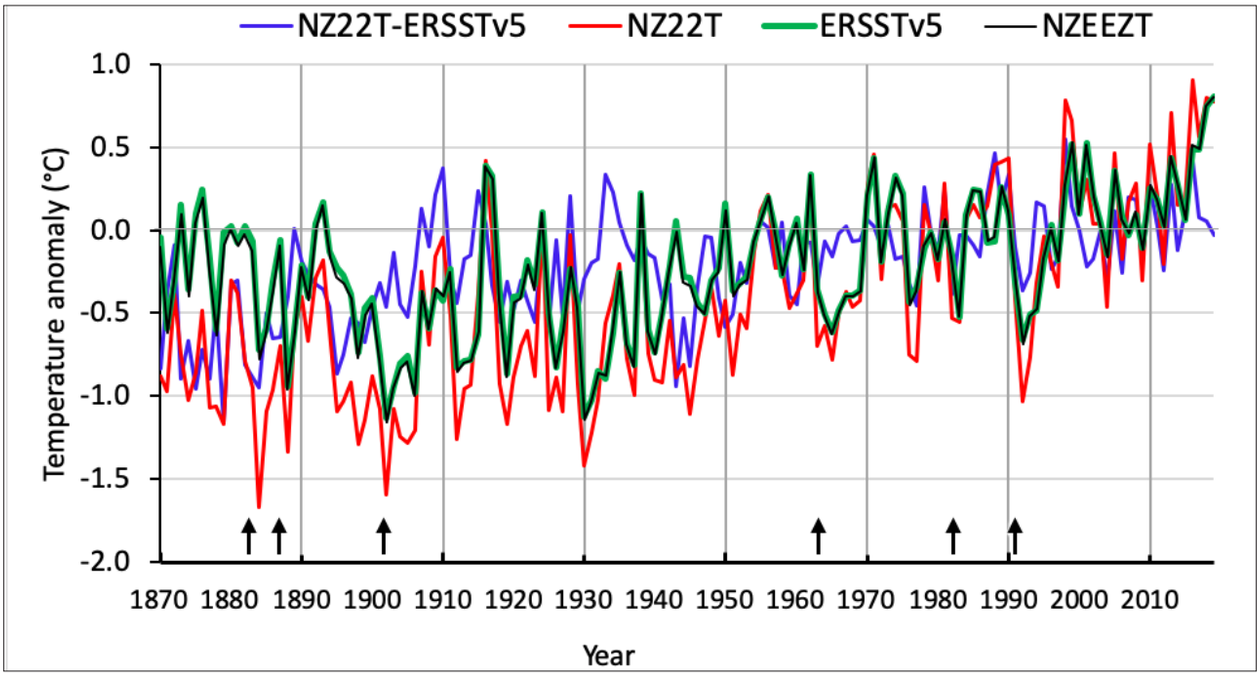



**Figure 4**: New Zealand temperature 1870-2019 (°C). Values are expressed as anomalies in mean annual temperatures from the 1981-2010 climatological period, for NZ22T(red), ERSSTv5 (green), NZEEZT (black) and the difference between NZ22T and ERSSTv5 (blue). ERSSTv5 is very similar to NZEEZT. The arrows indicate dates of major volcanic eruptions that affected New Zealand climate ( August 1883, June 1886, October 1902, April 1963, April 1982 and June 1991).

Table 3 shows serial correlations between atmospheric indices, regional circulation indices and regimes with NZT series, both actual and detrended. SAM and AGW are primary amongst them for the centennial trend. Relationships with regional atmospheric circulation and indices were very significant with both the Z1 and M1 indices and the blocking regime, as well as with related trough regimes (Figure 5 top, Table 3b). Warm (cold) years are associated with more northerly (southerly) airflow over New Zealand with more (less) blocking and fewer (additional) troughing. For the detrended cases associations were highly significant with detrended Z1, M1, blocking and to a lesser extent with zonal regimes. In these cases, warm (cold) years are associated with more northeasterly (southwesterly) airflow and more (less) blocking.

With AGW the most significant climate teleconnection is the SAM (Table 3, Figure 5 bottom) on a centennial scale. On interannual to interdecadal timescales the SOI and the IPO are the most important (Table 3, Figure 5 middle). In the detrended series, the SAM correlation values lower whereas the IPO and SOI correlation values strengthen demonstrating the importance of these on interannual to decadal timescales. For both periods the correlation with the detrended AGW is not significant.

***Anthropogenic Regional Warming***

The New Zealand regional temperature signal, NZEEZT, shows a highly significant linear warming trend of +0.66°C over the 150-year period ($p<0.01$) between 1870 –2019. The trend in CMIP5 for AGW (Figure 3) shows an increase of 0.67°C representing the magnitude of the warming signal for the 1900-2015 period, compared with increases of 1.16°C, 0.81°C and 0.87°C for NZ22T, ERSST and NZEEZT respectively.

**Table 3:** Serial correlations between atmospheric indices, regional circulation indices and regimes with New Zealand temperature anomalies using annual means for calendar years for actual and detrended series. Bolded italicized values are significant at the 1% confidence level and bold at the 5% level of significance. (a) 1900-2015, and (b) 1948-2015. Z1 and M1 indices commence in 1896, and Kidson regimes in 1948. AGW comes from the CMIP5 runs. All series, both indices and temperature, have been detrended linearly.

a) 1900-2015

| | | SAM | SOI | IPO | Z1 | M1 | AGW | NZEEZT-AGW |
|---|---|---|---|---|---|---|---|---|
| Actual | ERSST | ***0.49*** | ***0.38*** | ***-0.43*** | ***-0.43*** | ***-0.38*** | ***0.47*** | ***0.85*** |
| | NZ22T | ***0.55*** | ***0.34*** | ***-0.32*** | ***-0.36*** | ***-0.49*** | ***0.54*** | ***0.71*** |
| | NZEEZT | ***0.51*** | ***0.38*** | ***-0.42*** | ***-0.43*** | ***-0.41*** | ***0.50*** | ***0.84*** |
| | Smoothed NZEEZT | ***0.58*** | 0.04 | ***-0.43*** | ***-0.26*** | ***-0.58*** | ***0.70***[1] | ***0.47*** |
| Detrended | ERSST | ***0.23*** | ***0.47*** | ***-0.45*** | ***-0.33*** | ***0.47*** | -0.14 | |
| | NZ22T | ***0.28*** | ***0.47*** | ***-0.32*** | ***-0.29*** | ***-0.62*** | -0.05 | |
| | NZEEZT | ***0.23*** | ***0.48*** | ***-0.44*** | ***-0.33*** | ***-0.48*** | -0.14 | |

[1] Correlation between smoothed time series of NZEEZT and AGW from CMIP5 models shown in Figure 3.

b) 1948-2015

| | SAM | SOI | IPO | Z1 | M1 | Trough | Zonal | Block | AGW |
|---|---|---|---|---|---|---|---|---|---|
| Actual NZEEZT | ***0.54*** | ***0.50*** | **-0.23** | ***-0.43*** | ***-0.46*** | ***-0.38*** | -0.06 | ***0.53*** | ***0.44*** |
| Detrended NZEEZT | ***0.37*** | ***0.60*** | ***-0.41*** | ***-0.39*** | ***-0.54*** | **-0.28** | **-0.27** | ***0.61*** | 0.05 |

***Variability***

Multivariate analysis demonstrated the relative importance of the climate teleconnections of SAM, ENSO and IPO and forcing via AGW. Table 4 shows the order of importance in affecting NZEEZT is AGW, followed by the IPO, SAM and ENSO as expressed by the SOI. The β values, which compares the strength of the effect of each individual independent variable to the dependent variable (Tiemann and Mahbobi, 2015) are all highly significant and the variance explained by the multiple linear regression is 52.1%. The multicollinearity measures show that the independent variables are not correlated to a high extent in the multivariate analysis (Table 4), even though the individual correlation between AGW and SAM is 0.67. The regression equation is:

$$NZEEZT = -0.482 + 0.67*AGW - 0.46*IPO + 0.149*SAM + 0.143*SOI \quad (1)$$

with a standard error of 0.28. Actual and predicted time series are shown in Figure 6, which shows generally good agreement, although the periods in the 1930s, 1980s and since 2006 are less consistent. The correlation between AGW and filtered NZEEZT (for decadal variability) is 0.70 (Table 3).

**Table 4:** Multivariate relationships between atmospheric teleconnections (SAM, IPO and SOI) and the AGW forced in NZ temperatures (NZEEZT) 1900-2015. The standardized beta (β ranges from 0 to 1 or 0 to -1), and the closer the value is to 1 or -1, the stronger the relationship. b is the slope of the regression relationship. using annual means for calendar years. Multicollinearity is measured by variance inflation factors (VIF) and tolerance (TOL). A TOL of less than 0.20 or 0.10 and/or a VIF of above 5 above indicates a multicollinearity problem.

| | β | b | Std. Error | t | Prob. >t | VIF | TOL |
|---|---|---|---|---|---|---|---|
| SAM | 0.24 | 0.10 | 0.03 | 2.92 | 0.004 | 1.67 | 0.60 |
| IPO | -0.29 | -0.46 | 0.11 | -4.03 | 0.000 | 1.20 | 0.83 |
| AGW | 0.38 | 0.67 | 0.15 | 4.48 | 0.000 | 1.68 | 0.60 |
| SOI | 0.27 | 0.14 | 0.04 | 3.77 | 0.000 | 1.23 | 0.81 |

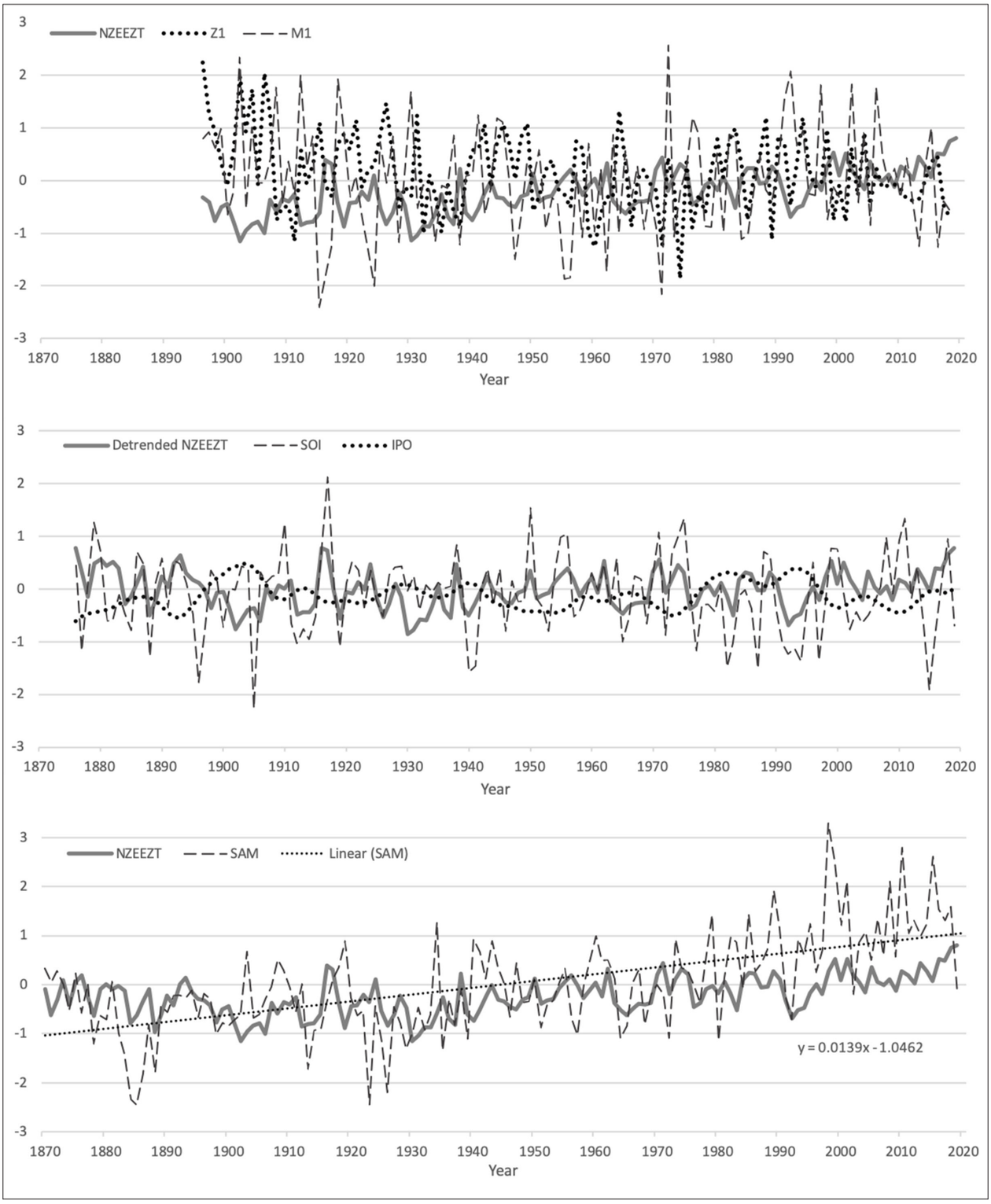



**Figure 5**: NZEEZT (°C) and circulation indices. Top: NZEEZT (solid), standardised circulation indices Z1 (dotted) and M1 (dashed), 1896-2019 Middle: Detrended NZEEZT (solid), circulation indices SOI (dashed) and IPO (dotted), 1876-2019 and Bottom: NZEEZT (solid) and the SAM (dashed), with a linear fit (dotted) 1870-2019.

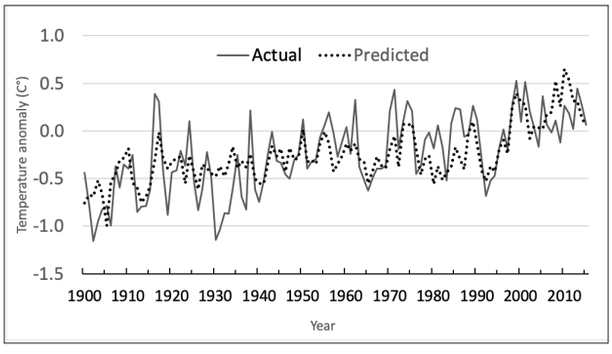



**Figure 6:** Actual (bold) predicted (dotted) NZEEZT (°C) 1900-2015. The predicted values are from multivariate analysis using the multiple regression equation. NZEEZT = -0.48+0.67*AGW -0.46*IPO +0.10*SAM +0.14*SOI, where AGW is anthropogenic global warming, IPO is the Inter-decadal Pacific Oscillation, SAM is the Southern Annular Mode and SOI the Southern Oscillation Index. The standard error is 0.28°C. $R^2$ =0.52.

### *Analogues*

As was documented in Salinger et al. (2019), a subset of past analogue annual periods was chosen from both the NCEP and ERA-Interim reanalysis 500 hPa anomaly fields. The analogues were chosen based on the highest anomaly correlation and lowest RMS difference (RMSD) across the New Zealand/Tasman Sea region. In all cases with AGW, the extreme warm years occurred in the period 1998–2019 (Figure 7a, Table 5). A subset of past analogue (similar) annual periods was chosen from both the ERA-Interim (Dee et al., 2011) and 20th Century reanalysis (20CR, Compo et al., 2011). Analogues that exhibited anomaly correlations of at least 0.65 and RMSD of 16 gpm or less were selected. A common 500hPa feature for warm year analogues were positive 20 to 40 geopotential metre (gpm) anomalies to the east of the country. In all cases Z1 and M1 were negative, demonstrating north easterly flow anomalies over NZ. Kidson (2000) weather regimes did not show specific trends over time. In all cases SAM was significantly positive with ENSO in the La Niña state on three occasions. M1 showed northerly airflow anomalies. The cold year analogues were less consistent,

**Table 5:** Temperature departures (°C), index values and circulation regimes for the warmest and coldest groups of years for actual and detrended NZEEZT. The mean ($\bar{x}$) and standard deviation ($\sigma$) are given for each time period for each index. Bolded values are significant at the 5% level of confidence. All values in the table have been standardised ($\bar{x}$ = 0, and $\sigma$ = 1). (a) 1900-2019 and (b) 1948-2019. Z1 and M1 indices commence in 1896.

a) 1900-2019

| | NZEEZT | SAM | SOI | IPO | Z1 | M1 |
|---|---|---|---|---|---|---|
| Actual ($\bar{x}\pm\sigma$) | -0.25±0.4 | -0.77±0.9 | -0.11±0.8 | -0.07±0.2 | 0.0±1.0 | -0.0±0.0.8 |
| Warmest 7 | **1.98** | **1.30** | 0.61 | **-0.70** | -0.92 | -0.83 |
| Coldest 9 | -1.90 | -0.58 | -0.15 | 0.85 | 0.96 | 0.64 |
| NZEEZT-AGW | -0.53±0.36 | | | | | |
| Warmest 4 | **0.29** | -0.90 | **1.08** | -0.28 | -0.61 | **-1.45** |
| Coldest 6 | **-1.25** | -1.22 | -0.31 | 0.26 | **1.17** | **0.90** |

b) 1948-2019

| | NZEEZT | SAM | SOI | IPO | Z1 | M1 | Trough | Zonal | Block |
|---|---|---|---|---|---|---|---|---|---|
| Actual ($\bar{x}\pm\sigma$) | -0.04±0.35 | -0.41±0.9 | -0.11±0.8 | -0.13±0.3 | 0.0±1.0 | 0.0±1.06 | 0.0±1.0 | 0.0±1.0 | 0.0±1.0 |
| Warm 10 | **1.51** | 0.60 | 0.59 | -0.33 | -0.96 | -1.46 | -0.55 | -1.08 | 0.99 |
| Cold 8 | **-1.46** | -0.58 | -0.76 | 0.76 | 0.23 | 0.09 | 0.52 | -0.65 | -0.63 |
| Detrended | -0.43±0.31 | | | | | | | | |
| Warm 6 | **0.14** | -0.89 | **0.90** | -0.36 | **-1.17** | **-1.11** | 0.17 | **-1.38** | **1.35** |
| Cold 8 | **-0.94** | -0.25 | **-1.06** | 0.13 | 0.77 | **1.12** | 0.35 | -0.41 | -0.89 |

but all occurred between 1902-1932, in the early 20th century. In all cases SAM was negative, and in three of the cases significantly so. Either positive height anomalies occurred in the south Tasman Sea or very strong negative 500 hPa anomalies occurred in the Tasman Sea or east of NZ. Neither the SOI nor the IPO show any clear tendency. Although an average of the analogues demonstrates a trend towards more southerly and westerly circulation, this was not dominant.

The warm year analogues changed for the NZEEZT-AGW series (Table 5, Figure 6 and Figure 7b), all being prior to 1972: 1916, 1917, 1962 and 1971. At the 500 hPa level warm years had consistent positive anomalies of 20 to 50 gpm

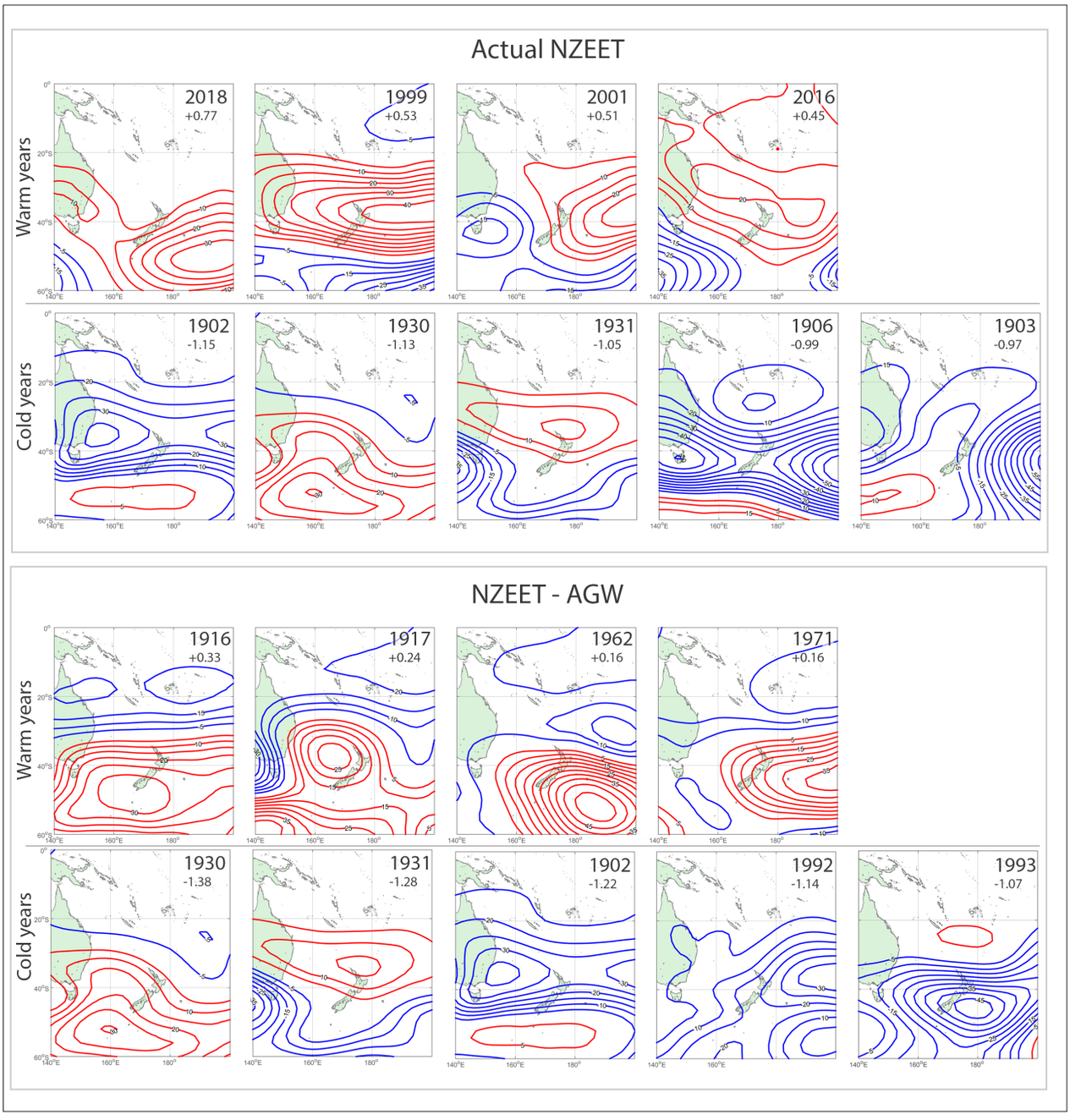



**Figure 7**: Warmest and coldest 500 hPa height anomalies. Top: NZEEZT years. where NZEEZT was > 0.45°C or < -0.84°C. Bottom: NZEEZT-AGW years where NZEEZT-AGW was > +0.15°C, or < -1.05°C. Source: 20th Century Reanalysis, Compo et al. (2011).

across or east of the South Island. In all cases both Z1 and M1 were negative, the latter strongly so indicating north to northeasterly flow anomalies over NZ. These years had lack of zonal and more frequent blocking circulation regimes. On interannual to interdecadal timescales both the negative phase of the IPO and positive phase of the SOI (La Niña phase) were important. The cold year analogues were similar to those for AGW, except 1992 was present instead of 1903. The 500 hPa patterns were similar. A strong feature was the positive M1 index, with southerly quarter airflow anomalies in the region with no dominance of any of the three circulation regimes. SAM was negative in all cases, the SOI generally neutral and IPO trending positive.

## 5. Discussion and conclusions

The New Zealand region has high-quality and detailed temperature records, with the introduction of Stevenson screens and precision sheathed thermometers in 1869 which have been operational since that time. Mullan (2012) did not homogenise station temperature series prior to 1909 despite calculating adjustments, comparisons with these with earlier (Salinger et al., 1992) and this work (Table 1) displayed very close agreement from 1870-1908. Such a high quality 148-year record provides a unique opportunity to investigate the climate teleconnections and atmospheric circulation for establishing climate trends and variability.

The cooling impacts of major volcanic eruption events that resulted in depressed regional surface temperatures (0.3-0.5°C) along with strong southwesterly winds is clearly depicted both in annual values and centennial trends (Figures 3 and 4), consistent with the findings of Salinger (1998).

The warming trend in NZEEZT, broadly matches the CMIP5 simulations of AGW (Mullan et al., 2018). The trend from the CMIP5 simulations (Figure 3 CMIP5 modelling) from 1900–2015 is +0.7°C. The additional observed warming of 0.2°C could be a result of the positive SAM trend (Arblaster et al., 2011). This is compatible with Kidston et al. (2009) who showed positive correlations between the SAM index and temperatures over much of the country. The associated anomalous high pressure over the county would also be associated with increased solar radiation.

The IPO as noted by both Power et al. (1999) and Salinger and Mullan (1999), dominated variability on interdecadal timescales. During positive (negative) IPO phases, SST anomalies are negative (positive) in the NZ sector, and similarly with land surface air temperatures (Power et al., 1999). The two IPO phases induce interdecadal variations in NZ climate variability, including temperature changes, with warmer periods, or accelerated warming occurring during negative IPO phases, and slower warming during positive IPO phases.

ENSO is clearly important for NZEEZT on interannual timescales and confirms the conclusions of Gordon (1986) where La Niña years tend to exhibit above average temperatures with more northeasterlies, and El Niño years are cooler than average due to more frequent southwesterly winds.

Figure 8a schematic displays the various climate teleconnections and circulation in operation for warm and cold years. In terms of regional circulation, both negative (positive) values of Z1 and M1 were important in influencing very warm (cold) years. Thus, a more northeasterly circulation predominated in warm years, compared to southwesterly circulation in the cold years. For the extreme warm years, a positive SAM together with northerly quarter circulation were the most important factors. In cold years a negative SAM with westerly quarter circulation were the chief influences.

The blocking in warm year analogues bear this out

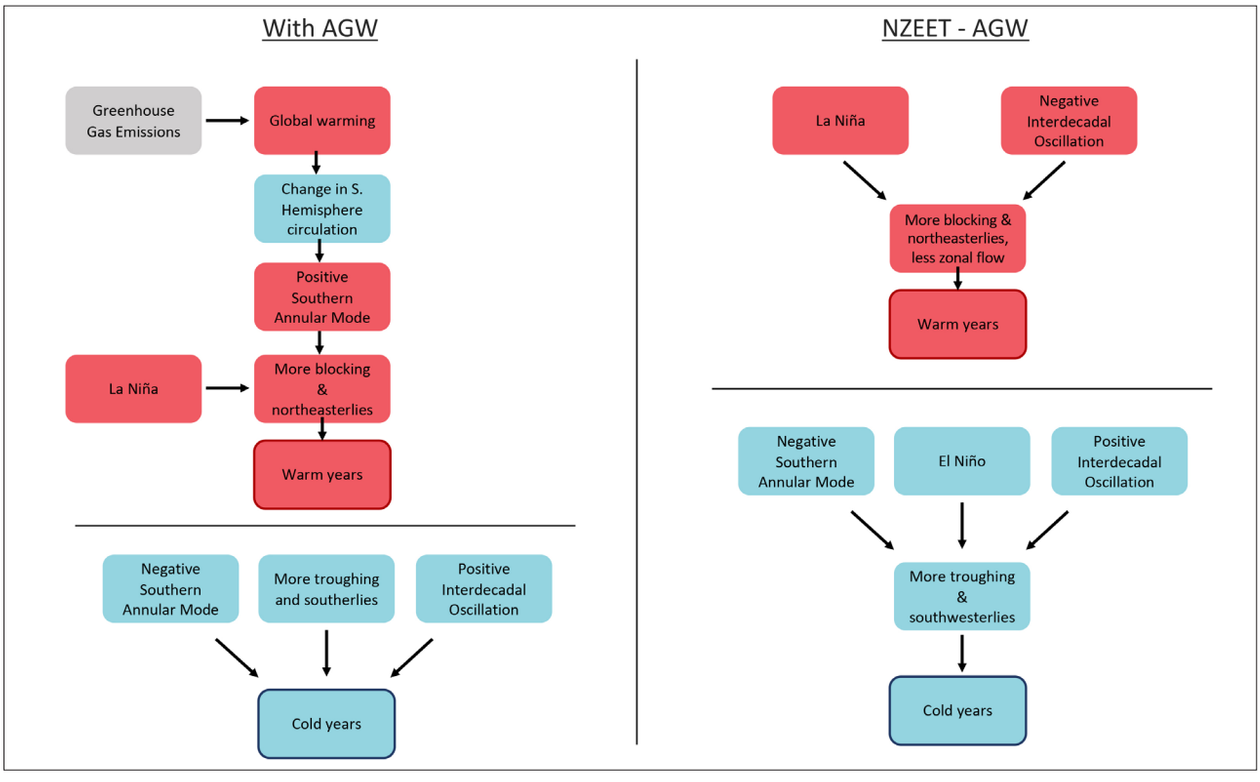



**Figure 8**: Schematic of likely driving mechanisms for warm and cold years. (a). With AGW, and (b) NZEEZT-AGW.

with 500 hPa anomalies showing very positive height anomalies over and to the east of NZ. These were characterised by a positive SAM and north easterly flow anomalies. The 500 hPa fields for cold analogues were less consistent but generally showed blocking patterns in the south Tasman Sea. There was a southerly quarter 500 hPa field, and more cyclonic type fields. Cold year analogues had strongly negative SAM values.

When NZEEZT values are detrended (Figure 8b), relationships with Z1 are unchanged whereas the relationship with M1 become stronger. The correlation with SAM weakens which supports the link (Arblaster et al., 2011) between trends in AGW and the SAM (Figure 8b). The importance of ENSO in determining cold or warm years increases, whilst IPO associations remain the same for interdecadal variability. Detrended warm year analogues differed somewhat in that 500 hPa blocking area was located further south across the South Island of NZ and to the east. La Niña years were particularly prominent in the warm sample, as was the IPO, with northerly flow strengthened. Lack of a zonal flow and a more prevalent blocking pattern occurred. There was only one common year (1971) between the actual and detrended analogues. In contrast, the actual and detrended cold years are similar, with three being the same: 500 hPa fields of blocking patterns in the south Tasman Sea with a southerly quarter 500-hPa field, or more cyclonic in nature near NZ. This was consistent with the negative phase of the SAM and southerly quarter meridional flow and a tendency towards El Niño episodes.

This study has shown the importance of AGW forcing as depicted by CMIP5 simulations and the SAM in determining long-term warming in the New Zealand region. It is of note that all the warm years >+0.45°C) occurred from 1998 onwards, and all the cold years (<-0.84°C) prior to 1933. Detrending emphasises the

role of modes of interannual to decadal variability (SOI and IPO). The cold years (anomaly<-1.05°C) are similar, whereas the warm years (anomaly>+0.15°C) are quite different, apart from one. The consequences of atmospheric circulation in this mode (lack of zonal and more frequent blocking) in the current climate (AGW) would be an extremely warm year.

## Acknowledgements

Dr Brett Mullan provided the average results data from the 13 CMIP5 model simulations of climate which had both AGW and natural variability (NAT) over the 1900–2015 period. The differences between AGW and NAT were used to calculate the AGW signal. We also thank the detailed comments of three anonymous reviewers which have significantly improved this paper.

## Dedication

Our colleague and good friend Brett Mullan died of cancer on 22 April whilst this manuscript was under review. Brett has been a mainstay of the Meteorological Society of New Zealand contributing strongly over the last 38 years. From 1983 – 1985 he was editor of *Weather & Climate* and the MSNZ quarterly newsletter, and 1996 to 1998 president, as well as serving for many years on the committee. Brett has made incredibly significant contributions and authored seminal papers in meteorology. These include the analysis of Southern Hemisphere climate and circulation variability over interannual (El Niño-Southern Oscillation) to interdecadal (Interdecadal Pacific Oscillation) timescales. The development of relationships with climate variability has been a basis for seasonal climate prediction for New Zealand commencing in the 1990s. He has carried out research into climate change and climate modelling, with particular emphasis on Southern Hemisphere and New Zealand regional effects (Southern Oscillation, greenhouse warming, ocean-atmosphere coupled models and, decadal variability, integrated climate impact models). In particular he has been at the forefront of development of climate change scenarios for New Zealand, leading and completing two major reports for the Ministry for the Environment published in 2008 and 2016 which are widely used as the basis for climate change adaptation and mitigation planning in New Zealand. More recently Brett re-evaluated the calculation of the 7SS, which had been under scrutiny from climate sceptics in New Zealand. Brett led a seminal paper which robustly defended established temperature trends and addressed the criticism. He has made research visits to the UK Hadley Centre for Climate Prediction and Research, and CSIRO Division of Atmospheric Research in Australia. Over the 40-year period Brett's contribution to New Zealand meteorology and climate science and beyond has been substantial, including mentoring to many scientists. He will be missed.

## References

Arblaster J.M., Meehl, G.A., Karoly D.J., 2011. Future climate change in the Southern Hemisphere: Competing effects of ozone and greenhouse gases. *Geophysical Research Letters*, Volume 38 (2), L02701 pp. 1-6. https://doi.org/ 10.1029/2010GL045384.

Bottomley, M., Folland, C.K., Hsiung, J., Newell, R.E., and Parker, D. E. 1990. *Global Ocean Surface Temperature Atlas*, 20 + iv pp. and 313 color plates, Her Majesty's Stn. Off., Norwich, UK, 1990.

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R.J., Yin, X., . . . Worley, S. J. (2011). The Twentieth Century reanalysis project. Q*uarterly Journal of the Royal Meteorological Society,* 137(654): 1-28. https://doiorg/10.1002/qj.776

Dean, S.M., Stott, P.A., 2009. The effect of local circulation variability on the detection and attribution of New Zealand temperature trends. *Journal of Climate* 22(23), 6217–6229. https://doi.org: 10.1175/2009jcli2715.1

Dee D.P, Uppala S.M, Simmons A.J, Berrisford P., Poli P.,

Kobayashi S., Andrae U., Balmaseda M.A., Balsamo G., Bauer P., Bechtold P., Beljaars A.C.M., van de Berg L., Bidlot J., Bormann N., Delsol C., Dragani R., Fuentes M., Geer A.J., Haimberger L., Healy S.B., Hersbach H., Hólm E.V., Isaksen L., Kållberg P., Köhler M., Matricardi M., McNally A.P., Monge-Sanz B.M., Morcrette J-J, Park B-K, Peubey C., de Rosnay P., Tavolato C., Thépaut J-N, Vitart F. 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system *Q. J. R. Meteorol. Soc.* 137: 553–597.

Flato G., Marotzke J., Abiodun B., Braconnot P., Chou S.C., Collins W., Cox P., Driouech F., Emori S., Eyring V., Forest C., Gleckler P., Guilyardi E., Jakob C., Kattsov V., Reason C. Rummukainen M.. 2013. Evaluation of Climate Models. In: T.F. Stocker, D. Qin, G.K . Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, P.M. Midgley (eds). *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge: Cambridge University Press. http://cmip-pcmdi.llnl.gov/cmip5/

Folland, C.K. and Parker, D.E. 1995. Correction of instrumental biases in historical sea surface temperature data, *Q. J R. Meteorological. Soc*., 121, 319-367. https://doi.org/10.1002/qj.49712152206

Folland, C.K., Salinger, M.J, 1995. Trends in New Zealand and surrounding ocean surface temperature, 1871 - 1993. *International Journal of Climatology* 15, 1195 - 1218.https://doi.org/10.1002/joc.3370151103

Folland C.K.; Salinger, M.J, and Rayner, N. 1997. A comparison of annual South Pacific island and ocean surface temperatures. *Weather and Climate* 17(1): 23-42.

Folland, C.K., Salinger, M.J., Jiang, N., and Rayner, N., 2003. Trends and variations in South Pacific island and ocean surface temperatures. *Journal of Climate*, Volume 16 (17), pp. 2859-2874. https://doi.org/10.1175/1520-0442(2003)016<2859:TAVISP>2.0.CO;2.

Fouhy, E., Coutts, L; McGann, R.P, Collen, B and Salinger, M.J., 1992. South Pacific Historic Climatological Network Climate Station Histories. Part 2: New Zealand and Offshore Islands. NZ Meteorological Service, Wellington, ISBN 0-477-01583-2, 216 p.

Gong, D. and Wang, S., 1999. Definition of Antarctic oscillation index, *Geophys. Res. Lett.*, 26, pp. 459–462. https://doiorg/10.1029/1999GL900003

Gordon, N.D., 1986. The Southern Oscillation and New Zealand weather, *Monthly Weather Review*, 114, 371 – 387. https://doi.org/10.1175/1520-0493(1986)114<0371:TSOANZ>2.0.CO;2

Hector, J., 1869. Meteorological Report, 1868. New Zealand Government Printer, 1869, Wellington, 60pp.

Henley, B.J., Gergis, J., Karoly, D.J., Power, S.B., Kennedy, J., and Folland, C.K. 2015. A Tripole Index for the Interdecadal Pacific Oscillation. *Climate Dynamics,* 45(11-12), 3077-3090. https://doi.org/10.1007/s00382-015-2525-1

Huang, B., Thorne,P.W., Smith, T.M., Liu,W., Lawrimore, J., Banzon,V.F., Zhang, H.M., Peterson, T.C., and Menne, M. 2016. Further Exploring and Quantifying Uncertainties for Extended Reconstructed Sea Surface Temperature (ERSST) Version 4 (v4). *J. Climate*, 29 (9): 3119–3142, https://doi.org/10.1175/JCLI-D-15-0430.1

Huang, B, Angel, W., Boyer, T., Cheng, L., Chepurin, G., Freeman, E., Liu, C., Zhang, H-M, 2018. Evaluating SST Analyses with Independent Ocean Profile Observations. *Journal of Climate*, 31 (13): 5015–5030. https://doi.org/10.1175/JCLI-D-17-0824.1

Huang, B., Menne, M.J., Boyer, T., Freeman, E., Gleason, B.E., Lawrence, J. H., Liu, C. Renne, J., Schreck, C.J., Sun F., Vose, R., Williams, C.N., Yin, X. , and Zhang, H-M. 2019. Uncertainty Estimates for Sea Surface Temperature and Land Surface Air Temperature in NOAA Global Temp Version 5. *Journal of Climate* 33, 1351-1379, https://doi.org/ 10.1175/JCLI-D-19-0395.1

Huang B., Thorne P. W., Banzon V.F., Boyer T., Chepurin

G., Lawrimore J. W., Menne M.J., Smith T.M., Vose R.S., Zhang H-M., 2017. Extended Reconstructed Sea Surface Temperature version 5 (ERSSTv5), Upgrades, validations, and intercomparisons. *Journal of Climate*, Volume 30, pp. 8179-8205. https://doiorg/ 10.1175/JCLI-D-16-0836.1

Huang, B, Angel,W., Boyer, T., Cheng, L., Chepurin, G. Freeman, E., Liu,C., Zhang, H-M. 2018. Evaluating SST Analyses with Independent Ocean Profile Observations. *Journal of Climate*, 31 (13): 5015–5030. https://doi.org/10.1175/JCLI-D-17-0824.1

Jones, P. D., Groisman, M, Coughlan, M., Plummer, N., Wang, W. C. and Karl, T. R. 1990. Assessment of urbanisation effects in time series of surface air temperature over land, *Nature,* 347, 169—172. https://doi.org/10.1038/347169a0

Karl, T.R., Jones, P.D., Knight, N., Plummer, N., Razuvayev, V., Gallo, J., Lindesay, J., Charlson, R.J. and Peterson, T.C., 1993. A new perspective on recent global warming: Asymmetric trends of daily maximum and minimum temperature, *Bulletin of the American Meteorological Society*, Volume 74, pp. 1007-1023.

Karl, T.R., Williams, C.N., Young, P.J., 1986. A model to estimate the time of observation bias associated with mean monthly maximum, minimum and mean temperatures, *Journal of Applied Meteorology and Climatology*., Volume 25, pp. 145-159. https://doi.org/10.1175/1520-0450(1986)0252.0.CO;2

Kelly, P.M., Jones, P.D. and Pengqun, J., 1996. The spatial response of the climate system to explosive volcanic eruptions. *International Journal of Climatology* 16(5), 537- 550.

Kent, E.C., Rayner, N.A., Berry, D.I., Saunby, M., Moat, B.I., Kennedy, J.J. and Parker, D.E. 2017. A call for new approaches to quantifying biases in observations of sea surface temperature. *Bull. Amer. Meteor. Soc.,* 98, 1601–1616, https://doi.org/10.1175/BAMS-D-15-00251.1.

Kidson J.W., 2000. An analysis of New Zealand synoptic types and their use in defining weather regimes. I*nternational Journal of Climatology* 20 (3): 299-315. https://doi.org/10.1002/(SICI)1097-0088(20000315)20:3<299::AID-JOC474>3.0.CO;2-B

Kidston, J., Renwick, J.A., and McGregor, J., 2009. Hemispheric-scale seasonality of the Southern Annular Mode and impacts on the climate of New Zealand. J*ournal of Climate*, 22, pp. 4759-4770. https://doi.org/ 10.1175/JCLI-D-11-00474.1

Marshall, G.J. 2003. Trends in the Southern Annular Mode from observations and reanalyses. *Journal of Climate,* 16: 4134–4143 https://doi.org/10.1175/1520-0442(2003)016<4134:TITSAM>2.0.CO;2

Mullan, A.B., 2012. Applying the Rhoades and Salinger Method to New Zealand's "Seven-Station" Temperature Series, *Weather and Climate*, Volume 32(1), pp. 23-37. https://doi.org/ 10.2307/26169723

Mullan, A. B., Stuart. S. J., Hadfield, M. G., Smith, M.J., 2010. Report on the Review of NIWA's 'Seven-Station' Temperature Series NIWA Information Series No. 78. 175 pp.

Mullan, A. B., Sood, A., and Stuart, S., 2018. Climate Change Projections for New Zealand: Atmosphere Projections Based on Simulations from the IPCC Fifth Assessment Wellington: Ministry for the Environment.https://www.mfe.govt.nz/sites/default/files/media/Climate%20Change/Climate-change-projections-2nd-edition-final.pdf

Parker, D.E., 1994. Effects of changing exposure of thermometers at land stations, *International Journal of Climatology*, 14, pp. 1-31. https://doi.org/ 10.1002/joc.3370140102

Parker, D. E., Folland C. K., Scaife,A.A., Colman, A., Knight, J., Fereday, D., Baines, P. and Smith, D. 2007. Decadal to interdecadal climate variability and predictability and the background of climate change. *Journal of Geophysical Research: Atmospheres*, 112 D18115, pp. 1-18. https://doi.org/10.1029/2007JD008411

Power, S., Casey, T., Folland, C. K., Colman, A., and Mehta,

V. 1999. Inter-decadal modulation of the impact of ENSO on Australia. *Climate Dynamics*, 15, pp. 319-323. https://doi.org/ 10.1007/s003820050284

Rhoades, D. A. and Salinger, M. J., 1993. Adjustment of temperature and rainfall records for site changes, *International Journal of Climatology*., 13, pp. 899-913. https://doiorg? 10.1002/joc.3370130807

Robock, A., 2003. Introduction: Mount Pinatubo as a test of climate feedback mechanisms, in Volcanism and the Earth's Atmosphere, *Geophys. Monogr. Ser*., 139, edited by A. Robock, and C. Oppenheimer, pp. 1–8, AGU, Washington, D. C.

Robock, A. and Free, M.P. 1995. Ice cores as an index of global volcanism from 1850 to the present. J*ournal of Geophysical Research: Atmospheres,* 100 (D6), 11549-11567. https://doi.org/10.1029/95JD00825

Salinger, M. J., 1980. The New Zealand temperature series, *Climate Monitor*, 9, pp. 112-118.

Salinger, M. J. 1981. New Zealand climate. The instrumental temperature record. Unpublished Ph.D thesis, Victoria University of Wellington, 357pp.

Salinger, M. J.; 1998. New Zealand climate: The impacts of major volcanic eruptions. *Weather and Climate*, 18(1) 11-20.

Salinger, M.J., McGann, R.P., Coutts, L., Collen, B., and Fouhy, E. 1992.

South Pacific historical climate network. Temperature trends in New Zealand and outlying islands, 1920-1990. New Zealand Meteorological Service, 46pp, Wellington, New Zealand.

Salinger, M. J., and Mullan, A. B., 1999. New Zealand climate: temperature and precipitation variations and their links with atmospheric circulation 1930–1994, *International Journal of Climatology,* 19, pp. 1049–1071. doi: 10.1002/(SICI)1097-0088(199908)19:10<1049::AID-JOC417>3.0.CO;2-Z

Salinger, M. J., Renwick, J. A., and Mullan, A. B., 2001. Interdecadal Pacific Oscillation and South Pacific climate. *International of Climatology,* 21, 1705–1721. https://doi.org/10.1002/joc.691

Salinger, M.J., Renwick, J., Behrens, E., Mullan, A.B., Diamond,H.J., Sirguey, P., Smith, R.O., Trought, M.C.T., Alexander, L.V., Cullen,N.J., Blair Fitzharris, B., Hepburn, C.D., Parker, A.K. and Sutton, P.J. 2019. The Unprecedented Coupled Ocean-Atmosphere Summer Heatwave in the New Zealand Region 2017/18: Drivers, Mechanisms and Impacts, *Environ. Res. Lett*, 14 (2019) 044023 https://doi.org/10.1088/1748-9326/ab012a.

Smith, T. M., and Reynolds, R. W. , 2005. A global merged land–air–sea surface temperature reconstruction based on historical observations (1880–1997). *J. Climate*, 18, 2021–2036. https://doi.org/10.1175/JCLI3362.1

Smith, T.M., Reynolds, R.W., Peterson, T.C., and Lawrimore, J. 2008. Improvements to NOAA's historical Merged Land–Ocean Surface Temperature analysis (1880–2005). *J. Climate*, 21, 2283–2296. https://doi.org/10.1175/2007JCLI2100.1

Tiemann, T.K. and Mahbobi, M., 2015. *Introduction to Business Statistics with interactive spreadsheets.* 1st Canadian Edition. 124pp. https://open.umn.edu/opentextbooks/textbooks/introductory-business-statistics-with-interactive-spreadsheets-1st-canadian-edition

Trenberth, K. E., 1976. Fluctuations and trends in indices of the southern hemispheric circulation, *Quarterly Journal of the Royal Meteorological Society*, Volume 102 (431), pp. 65-75. https://doi.org/ 10.1002/qj.49710243106.

Troup, A. J., 1965. The 'southern oscillation'. *Quarterly Journal of the Royal Meteorological Society*, 91 (390), pp. 490-506. https://doi.org/10.1002/qj.49709139009

Vose, R. S., and Coauthors, 2012. NOAA's Merged Land–Ocean Surface Temperature Analysis. *Bulletin of the American Meteorological Society*, 93, 1677–1685. https://doi.org/10.1175/BAMS-D-11-00241.1

# Honour Roll for the Meteorological Society of New Zealand

K. Richards

The highest level of recognition awarded by the Meteorological Society of New Zealand is Honorary Membership. This recognises outstanding contributions made by individual members to meteorology or climatology. Recognition is given to outstanding researchers, leaders in forecasting or applied climatology, and exceptional communicators, who foster Governmental and public understanding of issues in weather and climate.

The Meteorological Society has awarded ten Honorary Memberships since its inauguration. At the Society's Annual General Meeting in Wellington, New Zealand in 2019, one more member joined this select group.

**Dr Antony Brett Mullan**

Dr Brett Mullan was awarded Honorary Life Membership of 'MetSoc' at the 2019 Annual General Meeting. This is our highest recognition and Brett was the 11th person to be admitted to our Honour Roll. He had been a 'MetSoc' member since 1980, a long-serving committee member, Editor of Weather and Climate (1983–1985) and President of the Society (1996-1998).

Brett's research over more than 40 years included seminal papers on the analysis of Southern Hemisphere climate and circulation variability over interannual (El Niño-Southern Oscillation) to interdecadal (Interdecadal Pacific Oscillation) timescales, research into Southern Hemisphere and New Zealand climate change and climate modelling, and the re-evaluation of New Zealand's temperature series (NZT). He had been at the forefront in the development of climate change scenarios for New Zealand; reports he led and completed for the Ministry for the Environment (2008, 2016) underpin climate change adaptation and mitigation planning in New Zealand.

Until recently, Brett was a Principal Scientist (Climate Variability and Change) at NIWA, Wellington. He began his working career as a Trainee Meteorologist and Operational weather forecaster at the Meteorological Service of New Zealand, gained a Sc.D. (Meteorology) from Massachusetts Institute of Technology, USA, and had been a Visiting Scientist at the UK Met. Office Hadley Centre for Climate Prediction and Research, Bureau of Meteorology Research Centre (Melbourne) and C.S.I.R.O. Division of Atmospheric Research in Australia. Brett passed away in April 2020.