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).

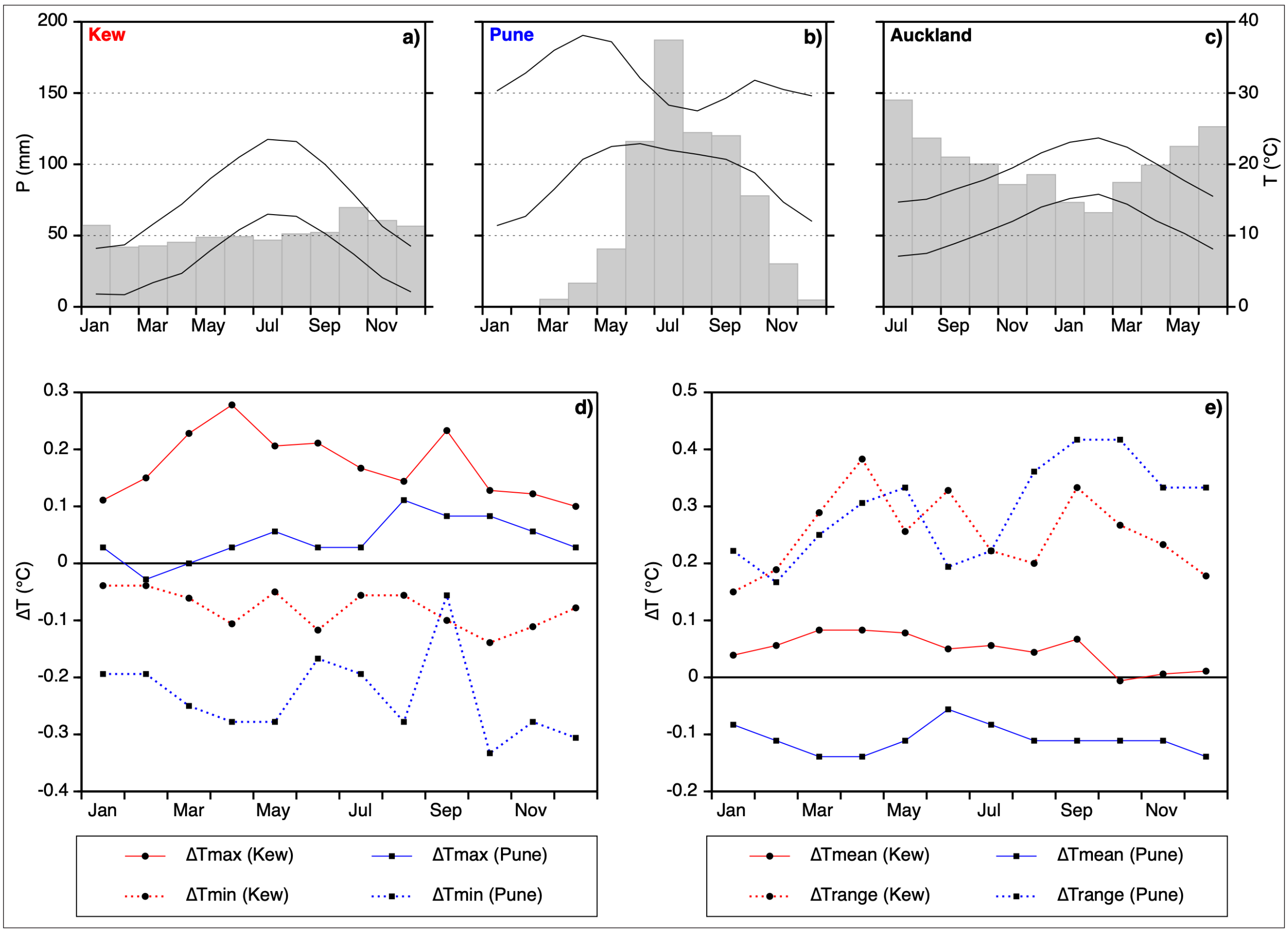

Figure 1: Stevenson Screen vs. Bilham Screen inter-comparison results for Kew Gardens, England (Bilham, 1937) and Pune, India (Gadre and Narayanan, 1939). Panels a–c: climographs for Kew (1981–2010, UK Met Office), Pune (1951–80, Indian Meteorological Department), and Auckland (1981–2010, NIWA). Panels d & e: monthly mean differences (Bilham minus Stevenson) in T_{min}, T_{max} T_{mean}, and T_{range} (ΔT_{min}, ΔT_{max}, ΔT_{mean}, ΔT_{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 not 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 ΔT_{min}, ΔT_{max}, and ΔT_{mean} statistics in Table 1b. ΔT_{range} is ΔT_{max} minus ΔT_{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 ΔT_{min} and ΔT_{max} largely cancel each other out, the impact on ΔT_{mean} is minor (-0.06°C). The results for ΔT_{min}, ΔT_{max}, and ΔT_{mean} differ substantially from Hessell's before-and-after study (Table 1a), but ΔT_{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 ΔT_{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		ΔT_{min}	ΔT_{max}	ΔT_{mean}[1]	ΔT_{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	“	-0.48	+0.38	-0.06	+0.86	Figure 2: reanalysis of Hessell (1980, Table 5)
c	Whenuapai[6]	5	-	15	-	-0.7	+0.6	-0.1	+1.3	Salinger (1981, Table AK.4, p. C9)
d	Oratia[6]	2	-	15	-	-0.2	+0.3	0.0	+0.5	“
e	Ruakura[6]	11	-	15	-	-0.4	+0.4	0.0	+0.8	“
f	Te Aroha[6]	15	-	5	-	-0.3	+0.7	+0.3	+1.0	“
g	Waihi[6]	4	-	15	-	-0.6	+0.6	0.0	+1.2	“
h	Tauranga[6]	10	-	10	-	-0.3	+0.8	+0.2	+1.1	“
i	Wellington[6]	15	-	15	-	-0.5	+0.8	+0.1	+1.3	“
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	“	-0.50	+0.32	-0.09	+0.82	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	+1.03	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. ΔT_{range} calculated here ($\Delta T_{max} - \Delta T_{min}$).

7. ΔT_{mean} from Mullan et al. (2010, Figure 6, p. 27).

8. ΔT_{max} from Mullan et al. (2010, Figure A3.2, p. 35). ΔT_{min} & ΔT_{range} derived from ΔT_{mean} & ΔT_{max}.

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

10. 6 x 12 months (Nov–Oct).

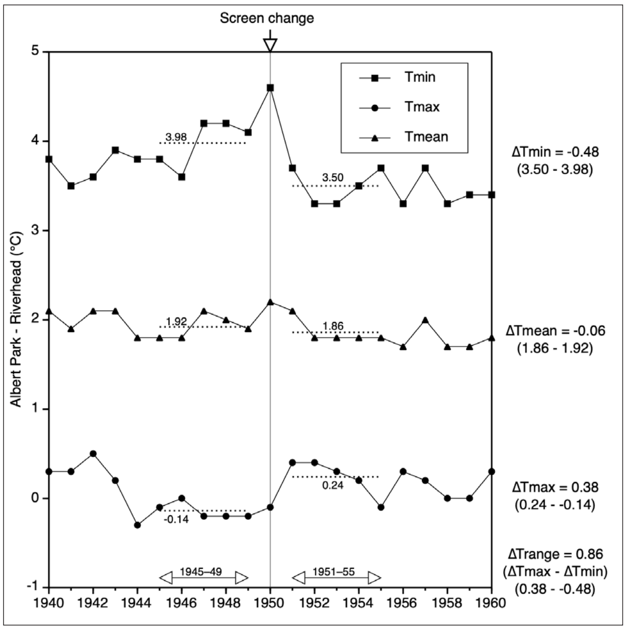

Figure 2: Albert Park – Riverhead paired-site analysis of the impact of the 1950 screen change. Solid lines show annual mean differences for T_{min}, T_{max}, and T_{mean}, reworked from data presented in Hessell's (1980) Table 5 (see Section 3.1 for details). Dotted lines show means over 1945–49 and 1951–55. Derived values for ΔT_{min}, ΔT_{max}, and ΔT_{mean} are the differences across these five-year blocks. ΔT_{range} is ΔT_{max} minus ΔT_{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 ΔT_{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 ΔT_{min}, ΔT_{max}, and ΔT_{mean} (ΔT_{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: ΔT_{min} negative; ΔT_{max} positive; ΔT_{mean} small (because ΔT_{min} and ΔT_{max} largely cancel); and ΔT_{range} positive. This is similar to Riverhead Forest (Table 1b). Inter-site spreads for ΔT_{min} and ΔT_{max} (both ~0.5°C) are similar in magnitude to the respective means (-0.4, +0.6). Only Whenuapai has a negative ΔT_{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 ΔT_{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 ΔT_{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. ΔT_{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) ΔT_{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 ΔT_{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 ΔT_{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 ΔT_{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.

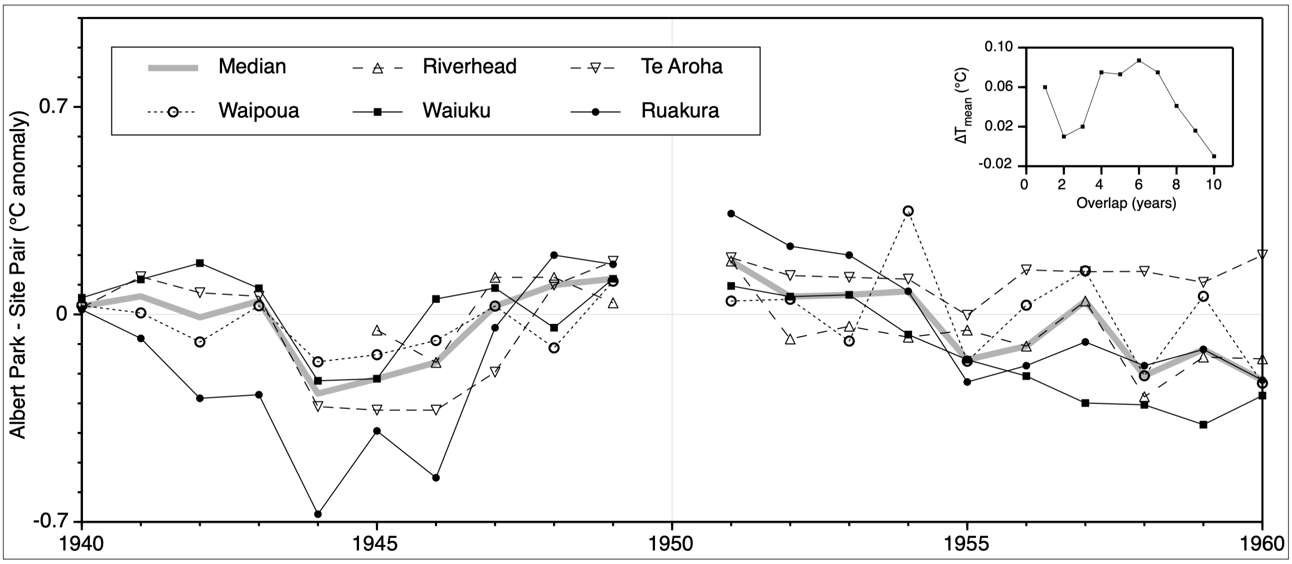

Figure 3: Reworking of the Mullan et al. (2010) paired-site analyses for Albert Park (results extracted from their Figure 6). For comparison purposes, annual T_{mean} differences for each comparison site (Albert Park minus that site) are shown as departures from their respective means for the common period 1945–1955, excluding 1950. The thick grey line is the median across available sites. The inset plot shows the impact on ΔT_{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 ΔT_{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 ΔT_{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 ΔT_{min} and ΔT_{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, ΔT_{min} and ΔT_{max} are both large and of opposite sign. Because they are of similar magnitude the net effect is a very small increase in ΔT_{mean}. Error estimates for ΔT_{min}, ΔT_{max}, and ΔT_{mean} are respectively ±0.19°C, ±0.11°C, and ±0.11°C. In addition to the larger error estimates, ΔT_{min} shows greater sensitivity to the overlap period, increasing from -0.80°C at one year to -0.49°C at six. Because ΔT_{max} is relatively stable, it is evolving ΔT_{min} that is responsible for ΔT_{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 ΔT_{range} results are quite consistent. This suggests that ΔT_{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 ΔT_{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 ΔT_{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 ΔT_{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 ΔT_{range}, the Hessell (1980) before-and-after study (row a) is a clear outlier in Table 1. It has the only positive ΔT_{min} and the largest values for ΔT_{max} and, especially, ΔT_{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 ΔT_{min}, ΔT_{max}, and ΔT_{mean} in Table 1a are invalid estimates of screen change impacts. ΔT_{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 ΔT_{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 ΔT_{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 ΔT_{mean} estimate by about 0.1°C.

4.5 Implications for climate change analyses

Excluding Hessell's (1980) estimates of ΔT_{min}, $\Delta T_{max,}$ and ΔT_{mean} in Table 1a, because they are significantly impacted by secular temperature increase (see Section 4.3), leaves broadly consistent results. ΔT_{min} is consistently negative by a few tenths of a degree, ΔT_{max} is consistently positive by a similar amount, and ΔT_{mean} is relatively small (because ΔT_{min} and ΔT_{max} cancel). The sign of ΔT_{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) ΔT_{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*International 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.