Comment on “A Reanalysis of Long-Term Surface Air Temperature Trends in New Zealand”

Brett Mullan¹ · James Salinger² · James Renwick³ · David Wratt⁴

Received: 5 July 2017 / Accepted: 11 April 2018 / Published online: 25 April 2018
© The Author(s) 2018

Abstract

de Freitas et al. (2015) (henceforth dFDB) report a trend of 0.28 °C per century over the period 1909–2009 for New Zealand land surface temperatures, from their reanalysis of a composite of seven long-term records. This is much lower than the warming trend of about 0.9 °C per century reported previously by other researchers and much smaller than trends estimated from independent sea surface temperature data from the surrounding region. We show these differences result primarily from the way inhomogeneities in temperature time series at individual stations due to site or instrument changes are identified and adjusted for in the dFDB paper. The adjustments reported in that paper are based on a method designed by one of us (Salinger), but use only a short (1–2-year) overlap period with comparison stations and consider only inhomogeneities in monthly mean (rather than monthly maximum and minimum) temperatures. This leads to underestimates of the statistical significance of individual temperature discontinuities and hence rejection of many valid adjustments. Since there was a systematic tendency for the seven-station sites to be relocated to colder locations as the early half of the twentieth century progressed, this rejection of valid adjustments produces an artificially low rate of warming. We therefore disagree with the trend calculations in the dFDB paper and consider there is no reason to reject the previous estimates of around 0.9 °C warming per century.

Keywords Temperature time series · Trend analysis · Adjustment of climate records · Climate change · New Zealand

1 Introduction

A paper by de Freitas et al. ([1], hereafter referred to as dFDB) claims to find a very small warming rate for New Zealand

temperatures over the past century, a result very much at odds with all other published evidence (e.g., [2], Fig. SPM.1). The temperature records analyzed are collectively known as the New Zealand “seven-station series” (7SS), built up from temperature measurements made at seven locations across the country, which go back to the early years of the twentieth century or even before. There are three locations in the North Island: Auckland (in the north), Wellington (south-west), and Masterton (southeast); and four locations in the South Island: Nelson (north), Hokitika (west), Lincoln (east), and Dunedin (southeast). dFDB (their Fig. 1) provide a map of the site locations. All other locations mentioned in this paper can be found in the maps produced in a report ([3], hereafter known as M10, to follow dFDB’s terminology) produced by the New Zealand National Institute of Water and Atmospheric Research (NIWA).

A number of site moves and instrument changes have occurred over the century at the 7SS stations, necessitating the calculation of adjustments before joining sections of the record together. dFDB depend on an adjustment methodology they refer to as RS93, from a paper published in the early 1990s [4]. RS93 was developed for New Zealand data, and many publications followed from it (see Section 2). This

✉ Brett Mullan
brett.mullan@niwa.co.nz

James Salinger
jimboalinger09@gmail.com

James Renwick
james.renwick@vuw.ac.nz

David Wratt
david.wratt@niwa.co.nz

¹ National Institute of Water & Atmospheric Research Ltd (NIWA), Private Bag 14901, Kilbirnie, Wellington 6421, New Zealand

² Department of Geography, University of Otago, Dunedin, New Zealand

³ School of Geography, Environment and Earth Sciences, Victoria University of Wellington, Wellington, New Zealand

⁴ NIWA, and New Zealand Climate Change Research Institute, Victoria University, Wellington, New Zealand

Fig. 1 Mean temperatures (in °C) for the month of May, from 1910 to 1930, for the two Masterton sites (Workshop Road and Essex Street) and four comparison sites used in the RS93 calculation. The vertical line marks the site change, and the star the position of dFDB's estimate for the missing May 1920 Masterton mean temperature

contradicts the claim that the dFDB paper is the first to apply the RS93 technique to New Zealand temperature data. Moreover, we identify several methodological and calculation issues, including the approach taken in the dFDB paper to applying RS93 methodology (Section 3), which we argue lead to a substantial underestimate of New Zealand temperature trend in the dFDB paper.

2 Literature on New Zealand Temperature Series and Length of Comparison Period

In this section, we provide relevant background on the history of homogenizing New Zealand temperature records, based on the experience of climate scientists involved at first-hand in the research (especially our second author, Salinger).

The concept of New Zealand's 7SS was developed by Salinger in 1980 [5], as he was completing his PhD thesis ([6], hereafter S81 to use dFDB's terminology). In the thesis, temperature data from 77 climatological stations were checked for homogeneity, quality of observations, and changes in instrumentation or in the environs near the climate site. Where appropriate, the records were carefully adjusted for site changes and other disturbances by neighbor station comparisons before and after the period in question. Missing data were interpolated from the anomalies of nearby stations. Prior to 1908, fewer stations were available, although observations from Auckland and Dunedin go back to the 1850s. Adjustments were made, and trends analyzed for annual minimum, mean, and maximum temperature. Salinger reported in 1979 ([7]) that "... the longest instrumental records show a warming of over 1°C between the early 1860s and the present day." Seven of the longest records were combined to produce

the 7SS, where the published time series [5] used the homogeneity adjustments of S81.

In 1992, Salinger led a project to homogenize temperature and rainfall records for 24 reference climate stations, which included the 7SS as a subset. Four reports were produced, two describing station histories [8, 9] and two the temperature and rainfall trends [10, 11] after homogenization. The trend reports made explicit reference ([10], p. 3) to a new homogenization technique developed by Salinger in collaboration with statistician Rhoades. A paper [4] describing the technical aspects of this method was published the following year. This is the so-called RS93 paper referred to extensively by dFDB. The NZMS temperature trend report [10] noted that, between 1941–1950 and 1981–1990, "North Island mean temperatures increased by 0.8°C, which was uniform between east and west. In the South Island, the mean temperature increase was 0.7°C, ranging from 0.8°C in the west and south, to 0.5°C in inland areas." dFDB do not cite this NZMS report, which contains many tables and graphs of temperature time series for mean, minimum, and maximum temperatures, broken down by geographic region and season. They also do not cite the explicit statement that RS93 adjustments were calculated using comparison periods up to ± 4 years before and after a site change or discontinuity in the temperature series ([10], p. 3).

There were numerous subsequent papers in the international literature where Salinger exploited his 7SS homogenization research and periodically updated the 7SS. The numerical values of the offsets for all the site changes were not documented in the literature, as they had been for the S81 adjustments, but such a level of detail is rarely published. Instead, graphs and trend calculations of the homogenized records were published on numerous occasions, all the while attributing the adjustments to RS93. The 7SS trend in mean temperature over 1871–1993 was compared to trends in surrounding sea temperatures from data independently homogenized by the UK MetOffice [12]. The authors stated ([12], p. 1195) that: "We conclude that NZT [7SS] and NMAT [night marine air temperature] over the nearby ocean surface have both warmed by about 0.7°C since the beginning of the century, with a slightly smaller increase in SST [sea surface temperature]." This intercomparison study was later extended to the South Pacific, using 40 island stations and optimally interpolated sea surface and night marine air temperature data [13]. The overall conclusion ([13], p. 2871) was that "annual and seasonal surface ocean and island air temperatures have generally increased by 0.6° – 1.0°C since near 1910 throughout a large part of the South Pacific southwest of the SPCZ [South Pacific Convergence Zone]." Both these publications [12, 13] also graphed trends in maximum and minimum, as well as mean temperatures.

In 1995, Salinger [14] extended his high-quality temperature data set for New Zealand out into the Southwest Pacific,

focusing especially on the maximum and minimum temperatures and trends in their difference (i.e., trends in the diurnal cycle). In 1998, Salinger [15] analyzed his 7SS data for any abrupt changes in New Zealand temperature that coincided with major volcanic eruptions. In 2001, Salinger [16] described trends in daily temperature and rainfall extremes for New Zealand, examining both the 1951–1998 and the 1930–1998 periods. All these papers plus others (for example, [12] on p. 1196, [13] on p. 2860, [14] on p. 88, [15] on p. 12, and [16] on pp. 1439–1440) refer the reader to RS93 as the source of the adjustment methodology used to prepare the homogeneous records being analyzed. The dFDB paper refers to little of this literature, which contradicts the claim that dFDB are the first to publish New Zealand temperature trends derived by the RS93. Moreover, the dFDB paper does not indicate that the trend calculated therein of $0.28\text{ }^{\circ}\text{C}$ per century is very different from the 7SS trend calculated by the author of the methodology, and does not discuss reasons for the discrepancy. It also does not refer to a subsequent paper by Zheng et al. [17] who calculated a trend of $0.11 \pm 0.035\text{ }^{\circ}\text{C}$ per decade (i.e., $1.1\text{ }^{\circ}\text{C}$ per century) for 1896–1994 using Salinger's data set developed using the RS93 methodology.

The dFDB paper follows parts of the RS93 strategy but also deviates from it in significant ways. We believe these differences result in the different trend estimates. Moreover, we consider that these departures from RS93 are not justified. The dFDB paper quotes extensively from RS93 ([4], at p. 900), beginning with: "The use of monthly differences means that the *t*-statistic has relatively high degrees of freedom even when computed from a short time interval of only 1 or 2 years before and after the site change. The period of comparison is kept relatively short in order to avoid contamination by gradual effects, or sudden but unrecognized effects, at one or more of the neighbouring stations." The emphasis here by RS93 on ± 1 - and ± 2 -year periods is an advertisement for the authors' new method. dFDB choose to interpret this sentence as "one must not use a comparison period longer than ± 2 years," but this is certainly not what RS93 are saying. In fact, the next sentence in the RS93 paper, which is not included in the quote by dFDB, states that "If no such effects are present it is optimal to use as long a period of comparison as possible." The statement is very relevant to dFDB's argument about length of comparison period and directly contradicts their assertion that RS93 are advocating a comparison period of ± 2 years or less. Moreover, it is stated explicitly in [10] (which is not referenced by dFDB) and in [18] (referenced by dFDB) that in practice, Salinger used up to ± 4 years in his adjustment calculations for a more stable result.

We also disagree with interpretations within the dFDB paper of some other relevant literature on the length of comparison periods. The dFDB paper refers to a seminal paper [19], claiming that on p. 1206 [sic], the authors "used ± 24 -month comparison periods by default for their algorithm based on

pairwise comparisons." However, from our examination of [19], we do not agree that ± 24 months should be interpreted as the default: [19] (p. 1706) clearly states that a ± 2 -year period is the *minimum* the authors will consider for adjustments; any shorter periods are combined to give a longer comparison period that is treated as one for adjustment purposes. There is therefore a clear difference between the practice outlined in [19] and that used in the dFDB paper: [19] do not estimate adjustments with less than ± 2 years of comparison data, whereas dFDB do not estimate adjustments using more than ± 2 years of data.

The dFDB paper references Hessell's analysis [20] of New Zealand temperature trends and accepts his opinion that changes in shelter and urbanization have tended to artificially increase the observed temperatures. But the dFDB paper does not incorporate Hessell's advice about the length of comparison periods for homogenizing records. Hessell used (Table 2, page 4 of [20]) "five year periods each side of the year of the change-over. It was necessary to confine the data to short periods to largely eliminate any longer-term trends." Thus, Hessell considered 5 years to be a "short period." Trewin [21] (p. 1523) also notes the use of a 5-year reference period when analyzing adjustments for the Australian ACORN-SAT data set.

The dFDB paper also does not cite or address points from a very relevant 2012 paper by Mullan [22], which sets out the RS93 methodology and applies it to 7SS data. This Mullan paper argues that 4 years before and after a site change is the minimum period that should be used when estimating statistical significance of homogeneity adjustments using the RS93 methodology, and demonstrates that shorter periods can lead to substantial errors. Mullan also demonstrates the importance of considering shifts in minimum and maximum temperatures, and not confining attention solely to the mean temperature adjustments. Some of the disagreements we have outlined in the next section with some particular adjustment calculations in the dFDB paper (next section) were also pointed out in Mullan [22].

3 Issues with Calculations in the dFDB Analysis

In this section, we examine in detail various further concerns we have with adjustment calculations in the dFDB paper. Table 1 summarizes the key results of our application of the RS93 methodology to the known 7SS site changes. Adjustments were calculated for mean, minimum, and maximum temperatures for up to ± 10 years either side of the site changes ($k = 10$ in the RS93 terminology) if the data for target and comparison sites allowed it. However, Table 1 is restricted to $k = 1$ and $k = 2$, largely for space reasons, but also to demonstrate our disagreements with dFDB's results even for

Table 1 Summary of RS93 adjustments for the 7SS sites (column 1) for selected site changes occurring in the named month (column 2), calculated for both ± 1 and ± 2 years of data either side of the site change. The last three columns show the adjustment and its 95% confidence interval for mean, minimum, and maximum temperatures; italicized values were statistically significant. Column 5, labeled “Mean Adj.,” gives the adjustment for the mean temperature as determined from

the average of minimum and maximum adjustments (see text), rather than directly from the mean temperature data (column 6). A negative adjustment means the new site is colder than the old site. The bracketed mean temperature adjustments (column 6, mean temperature only) are the RS93 $k = 2$ calculations using the $k = 1$ weighting of dFDB. Column 4 presents dFDB’s estimates using the same RS93 approach, but combining $k = 1$ and $k = 2$ estimates if both judged significant

7SS station	1st month of new site	k value	dFDB Adj.	Mean Adj.	RS93 adjustments		
					Mean temperature	Minimum temperature	Maximum temperature
Auckland	Nov 1950	$k = 1$		− 0.11	− 0.10 $\pm$ 0.24	− 0.80 $\pm$ 0.35	0.58 $\pm$ 0.25
		$k = 2$	0.00	− 0.00	0.00 $\pm$ 0.18 (− 0.01 $\pm$ 0.18)	− 0.57 $\pm$ 0.33	0.56 $\pm$ 0.11
Masterton	May 1920	$k = 1$		− 0.41	− 0.26 $\pm$ 0.26	0.15 $\pm$ 0.50	− 0.82 $\pm$ 0.28
		$k = 2$	0.00	− 0.37	− 0.31 $\pm$ 0.22 (− 0.32 $\pm$ 0.22)	0.15 $\pm$ 0.37	− 0.75 $\pm$ 0.25
	Oct 1942	$k = 1$		− 0.00	0.05 $\pm$ 0.34	0.29 $\pm$ 0.70	− 0.19 $\pm$ 0.56
		$k = 2$	0.00	− 0.20	− 0.10 $\pm$ 0.19 (− 0.11 $\pm$ 0.20)	0.22 $\pm$ 0.34	− 0.41 $\pm$ 0.29
Wellington	Jul 1912	$k = 1$		− 0.00	0.06 $\pm$ 0.28	0.08 $\pm$ 0.37	0.12 $\pm$ 0.46
		$k = 2$	0.17	0.17	0.16 $\pm$ 0.16 (0.19 $\pm$ 0.16)	0.05 $\pm$ 0.28	0.33 $\pm$ 0.28
	Jan 1928	$k = 1$		− 1.04	− 1.05 $\pm$ 0.29	− 0.79 $\pm$ 0.27	− 1.29 $\pm$ 0.35
		$k = 2$	− 1.00	− 1.11	− 1.04 $\pm$ 0.22 (− 1.11 $\pm$ 0.27)	− 0.73 $\pm$ 0.38	− 1.50 $\pm$ 0.26
Nelson	Dec 1920	$k = 1$		− 0.47	− 0.40 $\pm$ 0.23	− 0.95 $\pm$ 0.31	0.12 $\pm$ 0.43
		$k = 2$	− 0.40	− 0.60	− 0.66 $\pm$ 0.25 (− 0.69 $\pm$ 0.25)	− 0.76 $\pm$ 0.35	− 0.44 $\pm$ 0.40
	Jan 1932	$k = 1$		− 0.37	− 0.24 $\pm$ 0.31	0.03 $\pm$ 0.48	− 0.74 $\pm$ 0.32
		$k = 2$	0.00	− 0.23	− 0.26 $\pm$ 0.20 (− 0.24 $\pm$ 0.21)	0.02 $\pm$ 0.30	− 0.46 $\pm$ 0.24
	Jun 1997	$k = 1$		0.26	0.30 $\pm$ 0.19	0.52 $\pm$ 0.44	− 0.08 $\pm$ 0.32
		$k = 2$	0.05*	0.36	0.40 $\pm$ 0.19 (0.38 $\pm$ 0.18)	0.71 $\pm$ 0.24	0.03 $\pm$ 0.25
Hokitika	Nov 1928	$k = 1$		− 0.00	− 0.34 $\pm$ 0.39	− 0.39 $\pm$ 0.66	− 0.27 $\pm$ 0.42
		$k = 2$	0.00	− 0.00	− 0.20 $\pm$ 0.29 (− 0.17 $\pm$ 0.29)	0.05 $\pm$ 0.43	− 0.38 $\pm$ 0.39
	Jan 1945	$k = 1$		− 0.76	− 0.63 $\pm$ 0.23	− 0.52 $\pm$ 0.40	− 1.00 $\pm$ 0.24
		$k = 2$	− 0.11*	− 0.45	− 0.51 $\pm$ 0.23 (− 0.54 $\pm$ 0.26)	− 0.14 $\pm$ 0.30	− 0.89 $\pm$ 0.26
Lincoln	Dec 1915	$k = 1$		− 0.22	− 0.21 $\pm$ 0.23	0.04 $\pm$ 0.39	− 0.44 $\pm$ 0.30
		$k = 2$	− 0.45	− 0.43	− 0.39 $\pm$ 0.14 (− 0.37 $\pm$ 0.15)	− 0.52 $\pm$ 0.22	− 0.34 $\pm$ 0.18
	Nov 1923	$k = 1$		0.63	0.67 $\pm$ 0.48	0.48 $\pm$ 0.50	0.78 $\pm$ 0.73
		$k = 2$	0.59	0.32	0.48 $\pm$ 0.35 (0.49 $\pm$ 0.35)	0.63 $\pm$ 0.56	0.31 $\pm$ 0.43
	Jan 1926	$k = 1$		− 0.00	− 0.29 $\pm$ 0.70	− 0.64 $\pm$ 0.88	0.05 $\pm$ 0.80
		$k = 2$	− 0.51	− 0.29	− 0.43 $\pm$ 0.31 (− 0.58 $\pm$ 0.33)	− 0.58 $\pm$ 0.39	− 0.35 $\pm$ 0.48
	Jan 1944	$k = 1$		− 0.62	− 0.59 $\pm$ 0.21	− 0.85 $\pm$ 0.36	− 0.39 $\pm$ 0.14
		$k = 2$	− 0.60	− 0.64	− 0.61 $\pm$ 0.15 (− 0.61 $\pm$ 0.14)	− 0.87 $\pm$ 0.24	− 0.41 $\pm$ 0.16
	May 1964	$k = 1$		0.53	0.54 $\pm$ 0.29	0.59 $\pm$ 0.39	0.46 $\pm$ 0.30
		$k = 2$	+ 0.55	0.56	0.57 $\pm$ 0.15 (0.57 $\pm$ 0.15)	0.68 $\pm$ 0.24	0.44 $\pm$ 0.14
	Jan 1976	$k = 1$		− 0.00	− 0.04 $\pm$ 0.24	− 0.05 $\pm$ 0.47	− 0.08 $\pm$ 0.27
		$k = 2$	0.00	− 0.00	0.01 $\pm$ 0.16 (0.01 $\pm$ 0.16)	0.05 $\pm$ 0.22	− 0.05 $\pm$ 0.15
	Jun 1987	$k = 1$		0.16	0.17 $\pm$ 0.16	0.07 $\pm$ 0.20	0.32 $\pm$ 0.24
		$k = 2$	0.00	0.14	0.17 $\pm$ 0.11 (0.17 $\pm$ 0.11)	0.09 $\pm$ 0.15	0.28 $\pm$ 0.15
	Jan 2000	$k = 1$		0.22	0.28 $\pm$ 0.15	0.13 $\pm$ 0.17	0.43 $\pm$ 0.19
		$k = 2$	0.00	0.09	0.06 $\pm$ 0.10 (0.06 $\pm$ 0.10)	− 0.08 $\pm$ 0.13	0.18 $\pm$ 0.13
Dunedin	Dec 1942	$k = 1$		− 0.54	− 0.62 $\pm$ 0.39	− 0.05 $\pm$ 0.51	− 1.07 $\pm$ 0.60
		$k = 2$	− 0.69	− 0.80	− 0.83 $\pm$ 0.26 (− 0.84 $\pm$ 0.26)	− 0.43 $\pm$ 0.31	− 1.16 $\pm$ 0.32
	Jun 1947	$k = 1$		0.64	0.85 $\pm$ 0.37	0.46 $\pm$ 0.71	1.27 $\pm$ 0.63
		$k = 2$	0.85	0.90	0.83 $\pm$ 0.24 (0.83 $\pm$ 0.24)	0.50 $\pm$ 0.37	1.30 $\pm$ 0.41
	Nov 1960	$k = 1$		− 0.27	− 0.23 $\pm$ 0.27	− 0.13 $\pm$ 0.40	− 0.53 $\pm$ 0.34
		$k = 2$	0.00	− 0.25	− 0.25 $\pm$ 0.19 (− 0.24 $\pm$ 0.23)	− 0.04 $\pm$ 0.31	− 0.50 $\pm$ 0.25
	Sep 1997	$k = 1$		− 0.13	− 0.04 $\pm$ 0.20	− 0.26 $\pm$ 0.24	0.11 $\pm$ 0.31
		$k = 2$	0.00	− 0.00	− 0.03 $\pm$ 0.20 (− 0.04 $\pm$ 0.20)	− 0.17 $\pm$ 0.28	− 0.05 $\pm$ 0.19

*Combination of two adjustments within 2 years of each other (column 4)

comparison periods we consider too short. Comment is made in the text on larger k where appropriate. Table 1 is referred to throughout this section. For the convenience of the reader, we have added the dFDB results to this table. Discrepancies between our application of RS93 and that of dFDB are readily

apparent by comparing columns 4 and 5; discussion of reasons for the differences follows in Sections 3.1 to 3.7.

There are several points to note about the RS93 calculations in Table 1, before comment is made on dFDB’s specific errors. Adjustments are estimated between the two target sites

at one of the 7SS locations by comparing the changes with those at selected comparison sites. The comparison sites were selected by M10 after consideration of first-difference inter-site correlations and site differences, for mean, minimum, and maximum temperature records. dFDB have adopted M10's selection of comparison sites, with a couple of exceptions where dFDB's short comparison period makes them unsuitable. The adjustments for mean temperature in Table 1 are calculated in two ways: directly, using the RS93 methodology (column 6) and, indirectly, by averaging the adjustments for minimum and maximum temperatures (column 5). The consequences of the second approach are discussed below in Section 3.2.

When applying the RS93 fourth-power correlation weighting to combine the estimated shifts from the individual comparison sites (usually four), dFDB use the first-difference correlation derived from the $k=1$ data for *both* the $k=1$ and $k=2$ results. We disagree with this approach, which occasionally has a big impact on the results. It is true that RS93 (pp. 905–906) appears somewhat ambivalent on this, but we know that in practice (Salinger, pers. comm.), the correlations used all the data; in other words, the correlations changed with k . This is the approach taken in Table 1: for example, the $k=2$ results use inter-site correlation weights derived from 24 months of after-before difference, and not 12 months. However, in Table 1, additional calculations (in brackets) have been added to illustrate the consequence of applying dFDB's decision to use the $k=1$ correlation weights in the $k=2$ calculation.

Missing months occur from time to time. Our approach in Table 1, as discussed in Mullan [22], is to ignore the missing month entirely and adjust the degrees of freedom appropriately. Thus, if one month is missing for either a target or comparison site, then the $k=2$ calculation compares 23 months rather than 24 months before and after the site change. This is less time-consuming than estimating missing monthly values, and sensitivity tests show the approach has little effect on the answer (see Section 3.5 for further discussion).

Finally, while there are many numbers reported in the dFDB paper [1], it does not include sufficient specific information, in general, for us to reproduce the underlying calculations. To examine these calculations, we have referred back to unpublished reports produced in 2011 by the same authors, on behalf of the New Zealand Climate Science Coalition.¹ These reports, referred to here collectively as CSC11, challenged NIWA's calculations in M10. The numerical results described in CSC11 are identical to those in dFDB, so we feel

justified in consulting CSC11 for the technical details not provided by dFDB.

3.1 Using Too Short a Comparison Period

As the comparison period increases (denoted by increasing " k " in the RS93 formula, p. 403 in [1]), the 95% confidence interval progressively decreases. The estimated adjustment can also change substantially, particularly between $k=1$ and $k=2$. There are numerous examples of this behavior in Table 1, where the adjustment for small k is unstable. In most cases, the estimated adjustment shows little change beyond $k=4$, and Table 2 illustrates this for the October 1942 site change at Masterton.

From $k=2$ onwards, the maximum temperature adjustment is both statistically significant and stable at about -0.4 °C; the mean temperature adjustment of about -0.2 °C is also significant, whereas the minimum temperature adjustment never becomes significant (Table 2). And, of course, the confidence intervals continue to contract as k increases. Such sensitivity analyses are the reason that Mullan [22] recommended that $k \geq 4$ be used in applying the RS93 method. But the dFDB paper does not include any sensitivity analyses and hence misses the point that very short comparison periods are often not sufficient to achieve statistical significance.

3.2 Not Considering Minimum and Maximum Temperature

Table 1 shows numerous examples where either the minimum or maximum temperature adjustment is significantly different from zero at the 95% confidence level, but the mean temperature adjustment is not. Maximum and minimum temperatures may respond differently with site moves, for a variety of reasons. One of the most common in New Zealand, a mountainous country, is a change in sheltering (e.g., a hilltop to a valley). A more sheltered environment, other factors being equal, typically has slightly higher maximum and much lower minimum temperatures. In this situation, the adjustments for these two records would partially compensate, but usually leave a residual contribution to the mean temperature adjustment.

Consider the November 1960 site change at Dunedin, which is one of the first examples worked through in the dFDB paper. This claims that the RS93 shift for $k=1$ is -0.23 ± 0.27 °C and for $k=2$ is -0.24 ± 0.24 °C, and sets the shift to zero because neither of these estimates differ from zero by more than their respective 95% confidence intervals. While dFDB's $k=2$ calculation does not quite agree with ours (see Section 3.3 below), the real issue is that the RS93 shift for the maximum temperature is significant for both $k=1$ and $k=2$. So for $k=1$, the RS93 shifts for mean, minimum, and maximum temperature are -0.23 ± 0.27 , -0.13 ± 0.40 , and -0.53 ± 0.34 °C, respectively. Since the RS93 shift for maximum

¹ NZ Climate Science Coalition (2011). *Statistical Audit of the NIWA 7-Station Review*. Two documents, one entitled *Supplementary Information*. Last accessed 03 April 2018.

<https://www.climateconversation.org.nz/docs/Statistical%20Audit%20of%20the%20NIWA%207-Station%20Review%20Aug%202011.pdf>

<https://www.climateconversation.org.nz/docs/Statistical%20Audit%20of%20the%20NIWA%207-Station%20Review%20Aug%202011%20SI.pdf>

temperature *is* significant, an adjustment to the temperature record should be made at this point, using the best estimate (i.e., $-0.53\text{ }^{\circ}\text{C}$). If the adjustment for the mean temperature is set to zero as in the dFDB paper, then the adjustment for the minimum temperature should be $+0.53\text{ }^{\circ}\text{C}$, in order for the mean temperature after the site change to remain the average of minimum and maximum. However, a shift of $+0.53\text{ }^{\circ}\text{C}$ for the minimum temperature record is not allowed, since this lies outside the RS93 confidence interval. Thus, the dFDB approach of not adjusting the mean temperature when non-significant has introduced a mathematical contradiction. This issue is well understood by climate scientists who homogenize data time series, and the standard solution which avoids any internal contradiction is to work from the minimum and maximum temperature records. Thus, in this example, the maximum temperature shift is accepted to be $-0.53\text{ }^{\circ}\text{C}$, the minimum shift to be $0.00\text{ }^{\circ}\text{C}$ (because it is not significant), and therefore the mean shift to be $-0.27\text{ }^{\circ}\text{C}$ (as the average of the maximum and minimum shifts). In Table 1, the mean temperature adjustments in the fourth column have all been calculated in this way. Note that the RS93 methodology is non-linear, because of the fourth power correlation weighting, and thus the mean temperature shift is not numerically identical to the average of the minimum and maximum shifts. (The raw temperature data will not match identically to two decimal places, either, because the monthly temperatures on the NIWA Climate Database are all rounded to one decimal place.)

There are a number of examples of non-zero adjustments for the mean temperature in Table 1, which dFDB consider to be zero, in particular, for the RS93 $k=2$ estimates: Dunedin November 1960 shift of $-0.25\text{ }^{\circ}\text{C}$, Masterton May 1920 shift of $-0.37\text{ }^{\circ}\text{C}$, Masterton October 1942 shift of $-0.20\text{ }^{\circ}\text{C}$, Nelson January 1932 shift of $-0.23\text{ }^{\circ}\text{C}$, and Lincoln June 1987 shift of $+0.14\text{ }^{\circ}\text{C}$. Four of these five shifts are negative, meaning that the new site is colder than the earlier one, and thus, the earlier temperatures need to be reduced. Failure to apply the correct adjustment will result in too small a warming trend. In fact, it is apparent from Table 1 that the majority of the adjustments for mean temperature are negative. This systematic bias in the early site changes—that is, the tendency for the earliest sites to be located in warmer locations than later ones—means that if valid adjustments are rejected, the calculated warming trend will be erroneously low.

It is well recognized that “discontinuities often affect the maximum and minimum temperatures differently” ([18], p. 1504), so that partial cancelation between minimum and maximum temperature adjustments occurs frequently (e.g., see Table 1). Obviously, one of the adjustments to either minimum or maximum temperature will be larger than the adjustment to the mean temperature (apart from the rare special case of all three adjustments being identical). That is, the mean temperature adjustment will never be the largest of the three adjustments. To avoid any ambiguity, a common international

practice is to start by homogenizing minimum and maximum temperatures, and only calculate the homogenized mean temperature at the end by averaging the minimum and maximum [23, 24]. Unfortunately, dFDB have chosen to focus solely on the mean temperature, without realizing that the statistical significance of the mean temperature adjustment is not the only constraint: there are also the significance tests on the minimum and maximum temperatures, and the mathematical relationship between the three temperature records. In practice, all these constraints must be satisfied simultaneously.

3.3 Using the $k=1$ Correlation in $k=2$ Calculations

Consider again the November 1960 site change at Dunedin (Table 1), which is one of the examples dFDB work through. They calculate that the RS93 shift for $k=1$ is $-0.23 \pm 0.27\text{ }^{\circ}\text{C}$ and for $k=2$ is $-0.24 \pm 0.24\text{ }^{\circ}\text{C}$. We get the same answer for $k=1$, but for $k=2$, we find $-0.24 \pm 0.23\text{ }^{\circ}\text{C}$ (actually $-0.235 \pm 0.226\text{ }^{\circ}\text{C}$ to three decimal places) using dFDB's $k=1$ correlation weighting. If the more sensible $k=2$ weighting is used, then the mean temperature shift is estimated to be $-0.25 \pm 0.19\text{ }^{\circ}\text{C}$, which is statistically significant in its own right. As pointed out above, the maximum temperature shift is clearly significant, anyway. Table 1 demonstrates that, for the most part, there is little difference between using the $k=1$ or $k=2$ correlations in the $k=2$ adjustments; however, Dunedin November 1960 is an exception, and incidentally one of the first worked examples in CSC11. The reason the choice of correlation period makes so much difference in this case is that the correlations for two of the comparison-target site pairs change substantially between 12 and 24 months. dFDB state that the $k=1$ correlations are 0.41, 0.82, 0.85, and 0.55, for comparison sites Kelburn, Adair, Waimate, and Invercargill Aero, respectively. They note the fact that two of the correlations are low (Kelburn and Invercargill). The $k=1$ correlations that we calculate are 0.407, 0.824, 0.846, and 0.551, for comparison sites in the same order. However, the $k=2$ correlations are 0.803, 0.916, 0.919, and 0.895, respectively. This leads to a substantial change in the weighting of the four comparison sites, from 0.025, 0.422, 0.469, and 0.084 to 0.168, 0.285, 0.288, and 0.259, respectively. At $k=10$, the correlations have only changed slightly to 0.806, 0.880, 0.891, and 0.878, respectively.

Clearly, it makes more sense to calculate correlations over the same period as the RS93 shift is estimated. In any case, we have demonstrated in Section 3.2 that the dFDB result of a zero adjustment for Dunedin's November 1960 site change (their Table 3) should be replaced by our Table 1 value of $-0.25\text{ }^{\circ}\text{C}$. This increases the Dunedin trend from $0.30\text{ }^{\circ}\text{C}/\text{century}$ (dFDB, Table 4) to $0.66\text{ }^{\circ}\text{C}/\text{century}$.

Table 2 As in Table 1, but expanding on the Masterton site change in October 1942. RS93 adjustments are calculated for increasing overlap periods before and after the change-point, using the six comparison

sites of Bagshot (Masterton), Kelburn (Wellington), Pahiatua, Taihape, Appleby (Nelson), and Albert Park (Auckland) (see M10 for location of comparison sites). Italicized values were statistically significant

7SS station	1st month of "new" site	<i>k</i> value	Mean temperature adjustment	RS93 adjustments		
				Mean temperature	Minimum temperature	Maximum temperature
Masterton	Oct 1942	<i>k</i> = 1	0.00	+ 0.05 ± 0.34	+ 0.29 ± 0.70	− 0.19 ± 0.56
		<i>k</i> = 2	− 0.20	− 0.10 ± 0.18 (− 0.11 ± 0.20)	+ 0.22 ± 0.34	− 0.41 ± 0.29
		<i>k</i> = 4	− 0.17	− 0.18 ± 0.16	− 0.01 ± 0.27	− 0.34 ± 0.26
		<i>k</i> = 6	− 0.23	− 0.25 ± 0.13	− 0.07 ± 0.22	− 0.46 ± 0.20
		<i>k</i> = 8	− 0.22	− 0.28 ± 0.10	− 0.10 ± 0.19	− 0.44 ± 0.18
		<i>k</i> = 10	− 0.23	− 0.27 ± 0.09	− 0.05 ± 0.17	− 0.46 ± 0.15

3.4 Averaging *k* = 1 and *k* = 2 Results Incorrectly

dFDB comment that: "The convention has been to use the mean of both the significant results, for *k*=1 and *k*=2, when making an adjustment." In our view, this is not an appropriate "convention" and is inconsistent with RS93 since it unevenly weights the years about the site change. The *k*=2 calculations include all the *k*=1 data, and so the *k*=1 and *k*=2 calculations are not independent. Indeed, if the *k*=1 and *k*=2 results are averaged, this gives three times the weighting to the difference of the first years either side of the site change, relative to the difference of the second years either side of the site change. The RS93 methodology gives uneven weighting to comparison sites, but there is no intention to unevenly weight different years going into the calculation. Note that the RS93 difference series equation (for $y_t^{(i)}$, see [1]) "crosses over" in the way the years are differenced. Thus, if years after a site change are numbered A1, A2, etc., and years before a site change as B1, B2, etc., then the *k*=1 calculation uses the difference (A1 − B1), and the *k*=2 calculation uses the difference [(A1 − B2) + (A2 − B1)]. The differences in these bracketed terms are averaged over all months in the named years. However, the *k*=2 calculation is identical to [(A1 − B1) + (A2 − B2)], in which the first term is the *k*=1 difference.

This would be a relatively trivial objection to the analysis in dFDB were it not for the fact that this averaging, which we argue is incorrect, leads to a smaller warming rate in the composite series. This can be seen by an inspection of the *k*=1 and *k*=2 RS93 shifts in Table 1, focusing on those instances when both the *k*=1 and *k*=2 shifts are statistically significant.

A more robust approach (Mullan [22]) is to calculate adjustments over progressively longer periods and select the adjustment once it stabilizes and no longer changes substantially with *k*. However, with only *k*=1 and 2 being in the dFDB paper, it would have been more defensible to simply take the adjustment for the *k*=2 case. Sometimes the shift

changes little with *k*, but focusing on those cases where the change is substantial (mean temperature only), we have:

- Dunedin in December 1942 has a *k*=1 significant shift of − 0.62 °C and a *k*=2 shift of − 0.83 °C. dFDB should take − 0.83 °C as the shift, not the average of − 0.73 °C. Actually, dFDB quote a value of − 0.69 °C, but this is because they have used a different site change date (October 1940, we believe) to M10 (see M10 Dunedin table, p. 156).
- Lincoln in November 1923 has a *k*=1 shift of + 0.67 °C and a *k*=2 shift of + 0.48 °C.
- Nelson in December 1920 has a *k*=1 shift of − 0.40 °C and a *k*=2 shift of − 0.66 °C.

In these cases, the average of *k*=1 and 2 is numerically more positive than *k*=2 alone, leading to a smaller warming trend. Other examples, not shown in Table 1, occur for Hokitika in September 1912 and in August 1943 (Section 3.7).

3.5 Estimating Missing Data from Climatological Values (Masterton 1920)

From time to time, missing data occur in the temperature records of the 7SS series or the comparison sites used to calculate the RS93 adjustments for site changes. One option in these cases is the suggestion of Mullan [22], where the missing months are simply skipped over; an alternative is to estimate the temperature for the missing month. The dFDB paper [1] does not document how missing data have been addressed, but we are aware from an examination of CSC11 and other correspondence that dFDB have replaced missing data by a climatological value. We disagree with this practice, since an individual month could have a temperature very different from the climatological average. Using a climatological value rather

than an interpolated value also adds noise and, therefore, makes it less likely for the adjustment to pass the significance test that dFDB require. A better approach is to interpolate from anomalies of comparison sites. However, had dFDB used a longer comparison period, then the occasional bad estimate for missing data would have little consequence on the calculated adjustment.

A particularly notable case is the dFDB adjustment for the Masterton record in 1920, where the daily observations are available for only the first half of May (and therefore the monthly average temperature for May 1920 is missing), and a new site begins in June 1920. Our Table 1 shows that the RS93 shift for $k = 1$ or 2 is about -0.4 °C, meaning the earlier record at the Workop Road site up to April 1920 is too warm by 0.4 °C relative to the new Essex Street site from June 1920 onwards. dFDB's Table 3 shows an adjustment of 0.00 °C, and not applying any offset to the earlier data has a major impact on the Masterton trend. Figure 1 provides insight into what is happening, showing the mean temperatures for the month of May over the period 1910–1930, for the two Masterton sites and the four comparison sites recommended by M10 and used by [1] and CSC11. At the four comparison sites, May 1920 is substantially colder than the May of the year either side (1919, 1921), but dFDB's estimate for the May 1920 temperature at "Masterton" of 10.0 °C is higher than those of the year either side. The dFDB estimate is not only too warm by about 1 °C, but is also assigned to the new and colder site at Essex Street.

We carried out sensitivity tests to determine how the RS93 adjustment varied with the estimate for the missing May 1920. If May 1920 is taken as the first (and missing) month at the new site, then Table 1 shows a RS93 $k = 2$ shift of -0.31 ± 0.22 °C for the mean temperature. Table 1 result is for the default calculation where the missing value is skipped over. Using our best estimate for the missing May 1920 temperature of 8.9 °C, the $k = 2$ shift for mean temperature is -0.30 ± 0.22 °C, and for $k = 4$ it is -0.32 ± 0.16 °C. As the estimated May 1920 temperature increases towards dFDB's estimate of 10.0 °C, the RS93 shift gets smaller and the 95% confidence interval gets larger. Only for an estimate of 9.9 °C or larger for the May temperature does the RS93 calculation become non-significant (-0.25 ± 0.25 °C, for 9.9 °C estimate), but of course the maximum temperature adjustment remains significant throughout.

In summary, we are able to reproduce the dFDB result of a non-significant site change in May or June 1920 only by a combination of assumptions: an erroneously high estimated value for the missing May 1920, starting the new site in May 1920 instead of June 1920 (i.e., associating the high estimate with the colder of the two sites), restricting the RS93 calculations to $k = 2$ or less, and ignoring the much larger shift in maximum temperature. For any other

combination of assumptions, there is a statistically significant difference between the Masterton Workop Road and Essex Street mean temperatures of about -0.3 °C. The dFDB paper ends up making no adjustments to the raw data for the Masterton sites ([1], Table 3). However, applying a shift of -0.31 °C (our Table 1, column 5) increases dFDB's 1912–2009 Masterton trend from 0.36 ([1], Table 4) to 0.51 °C/century. Applying both the adjustments from column 4 of Table 1, for May 1920 and October 1942, increases the Masterton trend to 0.80 °C/century, which matches exactly with NIWA's RS93 calculation (M10, p. 57).

3.6 Arithmetic Errors (Masterton 1991, Nelson 1932)

In our opinion, the dFDB paper includes at least two simple arithmetic errors, for the Masterton January 1991 and Nelson January 1932 mean temperature adjustments. For the Masterton site change in January 1991, there is an overlap in records between the earlier Waingawa site and the later East Taratahi site. dFDB (in CSC11) claim that East Taratahi is warmer than Waingawa, and calculate from monthly data covering the 1982–1990 site overlap that the change is $+0.059 \pm 0.062$ °C and therefore not statistically significant; their Table 3 thus sets the adjustment to zero. However, our calculation gives a change of -0.074 ± 0.060 °C, which shows the new site (East Taratahi) is significantly colder than Waingawa. It is obvious from a graph of the overlapping data (M10, page 51, Figure 4) that East Taratahi is indeed the colder of the two sites, so dFDB are incorrect.

The January 1932 site change at Nelson is reported by dFDB as a 0.00 °C shift ([1], Table 3), but there appears to be an arithmetic error in the dFDB calculation. M10 use three comparison sites to estimate the adjustment: Albert Park (Auckland), Hokitika, and Dunedin Botanical Gardens. For $k = 1$, the correlation between Nelson and Dunedin is very small, so dFDB quite reasonably replace it with Kelburn (Wellington) data. Our RS93 calculation in Table 1 uses the same comparison sites as dFDB (i.e., Wellington in place of Dunedin). For $k = 1$, our estimated adjustment is -0.24 ± 0.31 °C, exactly the same as dFDB (in CSC11), and not significant. However, for $k = 2$, our adjustments are -0.26 ± 0.20 °C using correlations based on 2 years of data before and after the site change, or -0.24 ± 0.21 °C using correlations based on 1 year of data as dFDB have elected to do. Neither of these estimated (and statistically significant) adjustments match dFDB's result (in CSC11) of -0.21 ± 0.25 °C (not significant). It would appear that dFDB could have inadvertently transposed the adjustment and its 95% confidence interval, the discrepancy between 0.24 and 0.25 being likely round-off by dFDB in calculating the correlation weighting. The correct calculation gives a statistically significant offset between the two Nelson sites at January 1932, requiring an adjustment slightly larger than -0.2 °C. In any case, Table 1

Table 3 As Table 1, but only for the Hokitika Town record. RS93 adjustments are calculated for overlapping consecutive ± 1 year periods leading up to August 1944, using four comparison sites of Westport Aero,

Appleby (Nelson), Kelburn (Wellington), and New Plymouth. Italicized values were statistically significant

7SS station	1st month of "new" site	<i>k</i> value	Mean temperature adjustment	RS93 adjustments		
				Mean temperature	Minimum temperature	Maximum temperature
Hokitika Town	Aug 1940	<i>k</i> = 1	0.00	-0.07 ± 0.35	-0.09 ± 0.46	-0.03 ± 0.33
	Aug 1941	<i>k</i> = 1	0.00	$+0.11 \pm 0.30$	$+0.03 \pm 0.47$	$+0.23 \pm 0.29$
	Aug 1942	<i>k</i> = 1	-0.57	-0.53 ± 0.24	-0.55 ± 0.36	-0.59 ± 0.28
	Aug 1943	<i>k</i> = 1	$+0.66$	$+0.73 \pm 0.26$	$+0.71 \pm 0.45$	$+0.60 \pm 0.34$
	Aug 1944	<i>k</i> = 1	0.00	-0.01 ± 0.20	$+0.15 \pm 0.29$	-0.25 ± 0.31

shows a very large shift in the maximum temperature at January 1932, requiring the mean temperature to also be adjusted.

Clearly, the Nelson record should be adjusted in 1932. Note that again the adjustment is negative, and failing to accept an adjustment will reduce the warming trend at the Nelson location. dFDB (Table 4) claim a 1909–2009 trend for Nelson of $+0.27$ °C/century. Applying this single correction to dFDB's Table 3 (e.g., using -0.23 °C for the January 1932 shift) almost doubles the Nelson trend to 0.53 °C/century. Accepting all the *k* = 2 adjustments from column 4 of our Table 1 (i.e., -0.60 °C for December 1920, -0.23 °C for January 1932, and $+0.36$ °C for June 1997) further increases the Nelson trend to 0.60 °C/century.

3.7 Interpretation Error for Hokitika 1943

The analysis of the Hokitika record is an example where a short intercomparison period can be misleading. While the use in the dFDB paper of a very short intercomparison period avoids problems of long-term non-climatic drifts in the data, the calculated shift is then highly sensitive to short-term noise in the temperature record. Figure 2 presents an example for the Hokitika site change in the 1940s to illustrate this. The Hokitika temperature records for the early 1940s are somewhat untidy, as shown in Fig. 2, where the two Hokitika sites have been compared against the average of two long-term comparison sites (Nelson and Wellington). Two further sites (Westport Aero and New Plymouth) with shorter records are also used in Table 1 RS93 calculation. Measurements were made at Hokitika Town up to January 1946 (M10, p. 110), and there is about 2.5 years of overlap with the replacement site at Hokitika Southside, whose record runs from August 1943 to December 1964.

In late July 1943, heavy rains on the west coast of New Zealand's South Island caused flooding in the Hokitika meteorological enclosure such that the rain gauge needed to be moved to higher ground ([10], p. 113). At the same

time, the NZMS inspector noted that: "The [Stevenson] screen was in very poor condition. Most of the louvres had fallen off, the roof was rotten and it needed painting." Repairs were made, and Fig. 2 shows that, in the August 1943 to January 1946 overlap between Hokitika Town and Hokitika Southside, the Town side is 0.68 °C warmer than the Southside site ([1], Table 3, and M10, p. 117). Prior to 1942, there is a long period of stable temperatures (relative to the Nelson/Wellington composite) at the Town site, but between early 1942 and July 1943, the Hokitika Town temperatures (especially the minimum) are consistently lower than periods before or after. It seems reasonable to assume that the Stevenson screen fell into disrepair at about this time, and resulted in recorded temperatures being lower than they should.

dFDB (Table 3) have accepted the M10 overlap offset of -0.68 °C (Southside colder than Town site), but have been misled by the short period of bad data at Hokitika Town prior to August 1943 to erroneously reduce the Hokitika trend. Their shift of $+0.57$ °C in August 1943 is determined (CSC11, p. 29, Supplementary) by averaging *k* = 1 and *k* = 2 RS93 results for years before and after July/August 1943. They are therefore adjusting the "good" Hokitika Town data post-July 1943 back to the "bad" data when the screen was in disrepair. The dashed horizontal line in Fig. 2 shows the level of the Town temperature relative to Southside reported in dFDB; this is clearly incorrect, being well below the actual observed temperatures prior to 1942.

Table 3 demonstrates how it is possible to apply the RS93 method to this period of data, where *k* = 1 estimates are made for a sequence of overlapping comparison periods, each starting in the month of August. The fourth column shows the final adjustment for mean temperature, as calculated by averaging those adjustments of minimum and maximum temperature which are statistically significant. Beginning back in August 1940, the RS93 adjustment for ± 1 year (August 1939–July 1940 versus August 1940–July 1941) is clearly zero. Stepping forward 1 year, and comparing August 1940–July 1941 to August 1941–

July 1942, the adjustment is again zero, as might be expected from a visual inspection of Fig. 2. However, in the following 2 years, the Hokitika Town temperatures drop by 0.57°C (August 1942 “shift”) and then rise again by 0.66°C (August 1943 “shift”), almost compensating each other. Thus, this alternative calculation would indicate that the earlier (pre-1942) Hokitika Town mean temperature is 0.09°C ($0.66-0.57$) cooler than the post-August 1943 Town record and therefore 0.59°C warmer ($0.68-0.09$) than the Hokitika Southside record. So we argue that by recognizing only one of the two inhomogeneities in the 1942–1944 Hokitika Town record, the dFDB paper erroneously reduces the Hokitika warming trend by more than $0.5^{\circ}\text{C}/\text{century}$.

3.8 Ad Hoc Detrending of Wellington and Auckland Temperatures

The examples above describe what we consider to be significant errors in adjustments in the dFDB paper for site changes and other inhomogeneities in the 7SS record. Most of the other discrepancies between [1] and M10, not commented on above, are due to the short comparison period used by dFDB. Indeed, had dFDB used $k=4$ (± 4 years either side of a site change) as Salinger did originally [10], they could have avoided most of the problems we have described, and obtained adjustments close to those of M10.

However, as a result of applying adjustments we consider to be incorrect, the dFDB paper calculates what we consider to be artificially low trends at five of the 7SS sites which are very different from those at the other two sites, Wellington and Auckland. The Wellington Kelburn site, used from 1927 up to the present, is well maintained and in a very open location at the top of a hill. The Auckland measurements were taken at Albert Park from 1909 to 1989. M10 carefully examined the Albert Park record and found that some tree growth and increased urbanization may have led to about 0.3°C of excess warming between about 1928 and 1960. M10 found that a correction of this magnitude would reduce the apparent Auckland 1909–2009 trend from 1.53 to 1.15°C and the overall New Zealand 1909–2009 7SS trend from 0.91 to 0.85°C .

But the dFDB paper argues that the trends at Auckland and Wellington should be the same as the average of the trends they calculate for the other five sites ([1], Fig. 4). We disagree with this statement: it is legitimate in homogenization analysis to investigate spatially anomalous trends, but this does not equate to enforcing all sites to have the same trend. The evidence we have presented in this paper suggests that the dFDB 5-site trend is much too low, and this conclusion is supported by the trends in observed sea surface temperatures and from an independent 11-site series (see Section 4).

Further, in the case of the Wellington record, the adjustments shown in dFDB’s Table 1 lead to a 1909–2009 trend of $+0.95^{\circ}\text{C}/\text{century}$, which is very different to the trend reported in dFDB Table 4 of $+0.43^{\circ}\text{C}/\text{century}$. No details of this discrepancy are given in [1] at all, but an examination of CSC11 indicates that dFDB have compared the trend differential between Kelburn (Wellington) and Appleby (Nelson) over 1932–1996 and find Kelburn warming at a rate of $0.004^{\circ}\text{C}/\text{year}$ faster than Appleby. The dFDB paper does not examine whether this is a genuine difference between these sites, or if it is Appleby that has a spurious warming rate. Instead, the Kelburn temperatures are progressively adjusted from 2005 back to 1928, at a rate of $0.4^{\circ}\text{C}/\text{century}$, so that at the January 1928, changeover from Thorndon to Kelburn the offset between the two sites is not -1.00°C (dFDB’s Table 1) but -0.69°C (CSC11), and the slope of the Kelburn warming between 1928 and 2005 is also reduced.

In the case of the Auckland record, the dFDB paper indicates that the averaged trend of the other five sites (Masterton, Nelson, Hokitika, Lincoln, and Dunedin) “was used to reduce the Albert Park trend over 1916–1975” ([1], page 404). This produces a “counter-trend” of -0.54°C by the end of the Albert Park record at March 1976 and largely negates the large temperature difference of 0.66°C between the Auckland sites of Albert Park and Mangere as calculated from the direct overlap 1962–1986 (M10, p. 23) to give a net adjustment of -0.12°C ($-0.66+0.54$, dFDB Table 3). Such a small temperature difference (of 0.12°C) between Mangere and Albert Park is not tenable according to evidence provided in M10. Mangere lies to the southwest of Auckland City and is subject to Tasman sea temperatures, which are at least 0.5°C colder than those to the east of the Auckland isthmus (near Albert Park) (M10, Fig. A6.1, page 42), under the influence of the warm southward-flowing Auckland current. Hessel’s [25] analysis of spatial temperature variations across the Auckland region, reproduced in M10, provides further confirmation.

The dFDB paper presents further arguments in their Table 2 where trends are calculated between pairwise site differences. The paper states, in the case of the Auckland Aero versus Whenuapai comparison, for example, that: “The station histories for these two stations reveal no significant site changes during 1976–2009, and screening disclosed no abrupt shifts.” This is not correct: both abrupt shifts and (possibly spurious) differential trends occur in all the temperature series dFDB consider. Figure 3 displays a major problem that can occur if differential trends are calculated without careful inspection of the underlying data. Site changes, flagged in the station histories on the NIWA Climate Database (but ignored by dFDB), are marked by the vertical lines.

Mangere warms with respect to Auckland Aero from about 1966 to 1975, which is why [10] and M10 [3] did not use Mangere data until after its site change in April 1976. Prior

Fig. 2 Time series of deseasonalized differences in mean temperature between two Hokitika records, Hokitika Town (January 1932 to January 1946) and Hokitika Southside (August 1943 to December 1964), and a composite formed by averaging the monthly mean temperatures at Appleby (Nelson) and Kelburn (Wellington). The vertical line marks the start of the Hokitika Southside record in August

1943. The horizontal solid and dotted lines mark the mean levels of the Hokitika minus composite differences over the Town and Southside records respectively. The heavy horizontal dashed line marks dFDB's calculation of the offset of the Hokitika Town record relative to the Hokitika Southside record

to the Whenuapai site change in July 1968, that rural site was warming relative to urban Auckland Aero (from comparison with other sites, not shown). When the Whenuapai climate site was opened in 1943, the enclosure was in a "well-exposed position" on the southwest border of the airfield. However, by 1968, the enclosure had "buildings in vicinity." After the 1968 Whenuapai site change, there is little differential trend relative to Auckland Aero for the next 20 years, but at the Auckland Aero site move in January 1989, there is a large discontinuity. A naïve differential trend calculation gives a relative warming of Auckland Aero versus Whenuapai of $+0.93\text{ }^{\circ}\text{C}/\text{century}$ over 1962–1993 (cf., dFDB Table 2), but this is largely artificial and greatly exaggerated by the 1989 site change at Auckland Aero, and the period selected by dFDB for their trend calculation.

The other Auckland site pairs (Mangere and Ardmore) in dFDB Table 2 also have issues with site changes or other

inhomogeneities. We consider it is not good climatological practice to calculate trend differentials between temperature data (as done here in the dFDB paper), without a careful inspection of the time series, which must include multiple graphical intercomparisons of both maximum and minimum temperatures.

4 Other Evidence for Warming in NZ Region

Several lines of evidence point to a New Zealand land temperature trend much larger than the $0.28\text{ }^{\circ}\text{C}$ per century calculated by dFDB.

Trends from an Independent Analysis of New Zealand Temperature Records "Berkeley Earth" was a project conceived in 2010 to reanalyze the Earth's surface air

Fig. 3 Time series of monthly mean temperature differences between two pairs of Auckland sites: (upper) Mangere minus Auckland Aero, both considered urban by dFDB, and (lower) Auckland Aero minus Whenuapai (rural), relative to a 1972–1988 climatology. Vertical lines mark site changes, as documented in the NIWA Climate Database. Data were first deseasonalized to remove any bias in the trend caused by starting and ending the trend at different points in the annual cycle. The linear trend imposed on the Auckland Aero minus Whenuapai difference is clearly not appropriate

Fig. 4 New Zealand annual mean temperature anomalies (in °C), relative to 1981–2010, over the period 1909–2012, comparing the NIWA 7-station temperature series (in M10) and the Berkeley Earth NZ region series. Linear trends are inset. Data available from <https://www.niwa.co.nz/our-science/climate/information-and-resources/nz-temp-record/seven-station-series-temperature-data> and <http://berkeleyearth.lbl.gov/regions/new-zealand>

temperature record, and developed a new mathematical framework for incorporating short and discontinuous temperature records [26]. Composite temperature trends were calculated globally and by region. The Berkeley Earth temperature series for the “New Zealand region” is shown in Fig. 4, along with the 7SS series of M10, over the period 1909 (when the 7SS starts) until 2012 (when the Berkeley Earth regional analysis currently ends). The warming trend over 1909–2012 is calculated to be 0.890 °C per century, identical to three decimal places between the two independent analyses. The Berkeley analysis includes the 7SS stations, as well as others; the two pieces of evidence below are completely independent of the New Zealand land temperature record.

Trends from an Independent Further Set of 11 Stations² In 2010, a set of 11 stations spanning New Zealand at which there had been no significant site moves for many decades was identified from the NIWA climate archive. These were all different stations from the ones used in the 7SS, and for six of them, the records went back to at least the 1930s. The temperature trend from this “11SS,” with no homogenization, was 1.0 °C for 1931–2008. The sites include two offshore islands, and two higher elevation sites, one in each island. Without the offshore islands, the trend was 1.1 °C for the 77-year period. The correlation between the time series of the 11SS and 7SS is 0.956. This similarity between the independent 11 and 7 station series indicates that the 7SS is a large enough sample to reliably capture trends and variability for New Zealand.

² <https://niwa.co.nz/our-science/climate/information-and-resources/nz-temp-record/temperature-trends-from-raw-data>

Retreat of NZ Glaciers Chinn [27] reported on a study of 127 Southern Alps glaciers, which indicated an average shortening by 38% and loss of 25% in area in the century up to 1990. He concluded that the upward shift of glacier mean elevation with this century of change is approximately equivalent to a temperature rise of 0.6 °C. He noted that extensive debris cover on many glaciers was significant in damping the climate signal. Subsequently, Chinn et al. [28] estimated ice volume changes (in water equivalents) in the Southern Alps for the period 1976–2008; these results have recently been updated by Willsman [29] and indicate a decrease in ice volume from 54.0 km³ in 1976 to 40.7 km³ by the end of the 2016 glacier year. Of that reduction, 48% was from the 12 largest glaciers, and 52% from the small- to medium-sized glaciers. These papers conclude that the largest glaciers were still adjusting to temperature increases that occurred during the twentieth century.

Sea Surface Temperature Trends in the Ocean Surrounding New Zealand The latest version of the UK MetOffice Hadley Centre global sea surface temperature product [30] has been analyzed for trends in the period 1909–2009, as shown for the New Zealand region in Fig. 5. The data set comprises ten ensemble member realizations to cover the uncertainty range of the input observations. Trends were calculated separately for each of the ten ensemble members and then averaged. Figure 5 shows the ensemble mean trend values (there was only a small variation in trend statistics across the individual ensemble members). Near the New Zealand coast, century trends are typically around 0.8 °C, though above 1 °C in places and below 0.5 °C to the southeast of the country. Consistent with these trends, [12] concluded that nighttime marine air temperatures in the New Zealand region had increased by about 0.7 °C since the beginning of the twentieth century (to 1993) and sea surface temperatures by slightly less than this.

5 Summary and Conclusion

In this paper, we identify what we consider to be several methodological flaws in the dFDB paper. We conclude that, as a consequence, the temperature trend of an increase of 0.28 °C per century for the period 1909–2009 for New Zealand land surface temperatures derived in the dFDB paper is substantially too low, and that no need has been established for significant downward revision of the trend of around 0.9 °C per century found in previous studies. In particular:

- We disagree with the statement in the dFDB paper that analyses using the RS93 methodology have not

Fig. 5 Trends in sea surface temperature, in °C/century over the period 1909–2009, calculated from the new HadISST.2.1.0.0 data set (provided by John Kennedy, UK MetOffice Hadley Centre, UK). The data values are on a $1^{\circ} \times 1^{\circ}$ latitude-longitude grid

previously been published. Such work was published by Salinger et al. in 1992 [10], Folland and Salinger in 1995 [12], and a number of other analyses referred to in the discussion above. These studies all derived substantially higher New Zealand temperature trends than those obtained in the dFDB paper, with its interpretation of the RS93 methodology.

- We dispute dFDB's representation of our earlier RS93 results as coming from Salinger's 1981 thesis. Thus, in dFDB's Table 5 where they compare trends from different adjustment methods, the columns they attribute to "7SS (S81)" and "7SS (RS93)" should instead be attributed to "RS93 (NIWA)" and "RS93 (dFDB)." The trends which dFDB attribute to S81 are copied from M10, who clearly state that these trends are derived from RS93.
- The application of the RS93 methodology for identifying and adjusting for temperature inhomogeneities in the dFDB paper used overlap periods with comparison stations which were too short. This, together with considering only changes in mean temperatures, and not requiring consistency with changes in maximum and minimum temperatures, led to many valid adjustments being rejected because they were incorrectly identified as not statistically significant.
- We have also identified some problems in dealing with missing data and calculations of actual temperature adjustments, and disagree with the detrending of Auckland and Wellington data.
- The net result of the issues with the methodology we have outlined above is a substantial underestimate of New Zealand temperature trends in the dFDB paper.
- Other lines of evidence, not referenced in the dFDB paper, point to values in the order of an increase of $0.7\text{--}1.0^{\circ}\text{C}$ per century for New Zealand temperatures. These include surface temperature analyses by other investigators using time series adjusted with the RS93 methodology [10, 12, 13, 17], analyses of trends by the Berkeley Earth project using an entirely independent homogenization methodology [26], the observed decline of New Zealand glaciers [27–29], and analyses of trends in sea surface temperature in the region surrounding New Zealand [12, 13, 17], and this paper].

Acknowledgments This analysis (by BM) was supported by core funding NIWA receives from the New Zealand government Ministry of Business, Innovation and Employment, as part of the "Climate Past and Present" project. John Kennedy of the UK MetOffice is thanked for providing the HadISST2.1 sea surface temperature data.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- de Freitas, C. R., Dedekind, M. O., & Brill, B. E. (2015). A reanalysis of long-term surface air temperature trends in New Zealand. *Environmental Modelling and Assessment*, 20(4), 399–410. <https://doi.org/10.1007/s10666-014-9429-z>.
- IPCC. (2013). Summary for policymakers. 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, United Kingdom and New York, NY, USA: Cambridge University Press.
- 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, ISSN 1174-264X, 175p., <http://www.niwa.co.nz/our-science/climate/information-and-resources/nz-temp-record/seven-station-series-temperature-data>). Accessed 26 January 2017.
- Rhoades, D. A., & Salinger, M. J. (1993). Adjustment of temperature and rainfall records for site changes. *International Journal of Climatology*, 13, 899–913.
- Salinger, M. J. (1980). The New Zealand temperature series. *Climate Monitor*, 9(4), 112–118.
- Salinger, M.J. (1981). New Zealand climate: the instrumental record. PhD thesis, Victoria University of Wellington, Wellington, New Zealand.
- Salinger, M. J. (1979). New Zealand climate: the temperature record, historical data and some agricultural implications. *Climatic Change*, 2, 109–126.
- Collen, B. (1992). *South Pacific historical climate network, Climate station histories: part 1, Southwest Pacific region*. Wellington: New Zealand Meteorological Service ISBN 0-477-01582-4, 72p.
- Fouhy, E., Coutts, L., McGann, R., Collen, B., & Salinger, M. J. (1992). *South Pacific historical climate network, Climate station histories: part 2, New Zealand and offshore islands*. Wellington: NZ Meteorological Service ISBN 0-477-01583-2, 216p.
- Salinger, J., McGann, R., Coutts, L., Collen, B., & Fouhy, E. (1992). *South Pacific historical climate network, Temperature trends in New Zealand and outlying islands, 1920–1990*. Wellington: NZ Meteorological Service ISBN 0-477-01598-0, 46 p.
- Salinger, J., McGann, R., Coutts, L., Collen, B., & Fouhy, E. (1992). *South Pacific historical climate network, Rainfall trends in New Zealand and outlying islands, 1920–1990*. Wellington: NZ Meteorological Service ISBN 0-477-01597-2, 33 p.
- Folland, C. K., & Salinger, M. J. (1995). Surface temperature trends in New Zealand and the surrounding ocean, 1871–1993. *International Journal of Climatology*, 15, 1195–1218.
- Folland, C. K., Salinger, M. J., Jiang, N., & Rayner, N. A. (2003). Trends and variation in South Pacific Island and ocean surface temperatures. *Journal of Climate*, 16, 2859–2874.
- Salinger, M. J. (1995). Southwest Pacific temperatures: trends in maximum and minimum temperatures. *Atmospheric Research*, 37, 87–99.
- Salinger, M. J. (1998). New Zealand climate: the impact of major volcanic eruptions. *Weather and Climate*, 18, 11–20.
- Salinger, M. J., & Griffiths, G. M. (2001). Trends in New Zealand daily temperature and rainfall extremes. *International Journal of Climatology*, 21, 1437–1452.
- Zheng, X., Basher, R. E., & Thompson, C. S. (1997). Trend detection in regional-mean temperature series: maximum, minimum, mean, diurnal range, and SST. *Journal of Climate*, 10, 317–326.
- Peterson, T. C., Easterling, D. R., Karl, T. R., Groisman, P., Nicholls, N., Plummer, N., Torok, S., Auer, I., Boehm, R., Gullett, D., Vincent, L., Heino, R., Tuomenvirta, H., Mestre, O., Szentimrey, T., Salinger, J., Forland, E. J., Hanssen-Bauer, I., Alexandersson, H., Jones, P., & Parker, D. (1998). Homogeneity adjustments of in situ atmospheric climate data: a review. *International Journal of Climatology*, 18, 1493–1517.
- Menne, M. J., & Williams, C. N. (2009). Homogenization of temperature series via pairwise comparisons. *Journal of Climate*, 22, 1700–1717.
- Hessell, J. W. D. (1980). Apparent trends of mean temperature in New Zealand since 1930. *New Zealand Journal of Science*, 23, 1–9.
- Trewin, B. (2013). A daily homogenized temperature data set for Australia. *International Journal of Climatology*, 33, 1510–1529.
- Mullan, A. B. (2012). Applying the Rhoades and Salinger method to New Zealand's "seven-station" temperature series. *Weather and Climate*, 32, 23–37.
- Torok, S. J., & Nicholls, N. (1996). A historical annual temperature dataset for Australia. *Australian Meteorological Magazine*, 45, 251–260.
- Vincent, L. A., & Gullett, D. W. (1999). Canadian historical and homogeneous temperature datasets for climate change analyses. *International Journal of Climatology*, 19, 1375–1388.
- Hessell, J.W.D. (1988). The climate and weather of Auckland. New Zealand Meteorological Service Miscellaneous Publication 115 (20), Wellington, ISSN 0110-6937, 34p.
- Rohde, R., Muller, R., Jacobsen, R., Perlmuter, S., Rosenfeld, A., Wurtele, J., Curry, J., Wickham, C., & Mosher, S. (2013). Berkeley Earth temperature averaging process. Geoinformatics & geostatistics: an overview, 1:2, <https://doi.org/10.4172/gigs.1000103>.
- Chinn, T. (1996). New Zealand glacier responses to climate change of the past century. *New Zealand Journal of Geology and Geophysics*, 39, 415–428.
- Chinn, T., Fitzharris, B. B., Willsman, A., & Salinger, M. J. (2012). Annual ice volume changes 1976–2008 for the New Zealand Southern Alps. *Global and Planetary Change*, 92, 105–118. <https://doi.org/10.1016/j.gloplacha.2012.03.002>.
- Willsman, A. (2017). Annual glacier ice volumes 1977–2016. Report prepared for the Ministry for the Environment, 20 p. <http://www.mfe.govt.nz/publications/environmental-reporting/annual-glacier-ice-volumes-1977-2016>
- Kennedy, J.J., Rayner, N.A., Millington, S.C., & Saunby, M. (2015). The Met Office Hadley Centre Sea Ice and Sea-Surface Temperature data set, version 2, part 2: sea surface temperature analysis. (In preparation).