Applying the Rhoades and Salinger Method to New Zealand's "Seven-Station" Temperature Series.

A. Brett Mullan
National Institute of Water and Atmospheric Research (NIWA), Wellington, New Zealand.

Corresponding author address: A.B. Mullan, NIWA, Private Bag 14901, Kilbirnie, Wellington 6241, New Zealand. E-mail: Brett.Mullan@niwa.co.nz

Abstract
There is recent renewed interest in the accuracy of long-term trends in New Zealand temperature records. In December 2010, NIWA published a report documenting temperature adjustments for site changes at seven locations across the country, and derived an average "seven-station" New Zealand warming trend of 0.91°C per century over the period 1909 to 2009. This report updated temperature adjustments that had previously been based on a methodology published in Rhoades and Salinger (1993). Some controversy surrounding the effect of the different adjustment methods in 1993 and 2010 is the motivation of this paper. A number of examples are provided of the application of the Rhoades and Salinger method, and its strengths and weaknesses highlighted. A key conclusion is that four years before and after a site change is the minimum period that should be used when estimating an adjustment, in order to ensure both reasonable stability and statistical significance of the estimate.

1. Introduction

In December 2010, NIWA published a report (Mullan *et al.,* 2010) documenting revised adjustments to its so-called "seven-station" temperature series. This data set comprises long "reference" series for 7 locations (Auckland, Masterton, Wellington, Nelson, Hokitika, Lincoln and Dunedin (Figure 1)), plus a combined national composite series (NZT7). At each of the 7 locations, the reference series is pieced together using multiple records from near-neighbour sites. At this step, the raw data are adjusted for the effect of any site changes unrelated to broad-scale climate. For example, within the Wellington reference series there is a large difference between the baseline temperatures of Thorndon (near the Wellington waterfront) and Kelburn (at 125 m altitude in the Botanical Gardens) for the site change that occurred in January 1928. Site changes can also be said to occur even if there is not a physical relocation of the instruments, such as a change of instrumentation (e.g., mercury thermometers to thermocouples) or a change

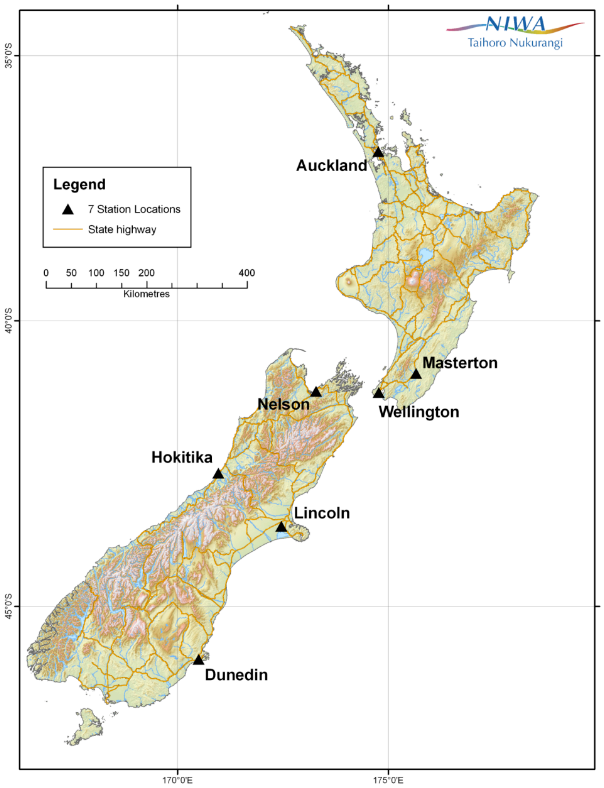

Figure 1: Location map of the reference sites comprising NIWA's "seven-station" temperature series.

of exposure (e.g., tree growth, building construction).

Altogether, 31 "site" adjustments were estimated to generate the December 2010 version of NZT7, which begins in the year 1909. Adjustments can sometimes be calculated by direct overlap of the near-neighbour sites (e.g., in the Auckland series, Albert Park and Mangere have a 30-year overlap). However, often there is no or insufficient overlap, and temperature records from more distant "comparison" sites must be used to estimate the reference series adjustment. In Mullan *et al.* (2010), these adjustments are calculated by taking annual temperatures for up to 10 years before and after a site change, and comparing changes at the reference site with those at multiple comparison sites. Comparison sites were selected from all available records after consideration of: length of overlap with the reference site change-point, absence of site changes in the comparison record, no evidence of quality control problems from the comparison site metadata, and high correlations between reference and comparison inter-annual temperature variations (in maximum and minimum temperature, as well as mean temperature).

The December 2010 NZT7 series is a revision of a long-term homogenised record originally developed by Salinger *et al.* (1992), who calculated the site change adjustments using an approach known as the Rhoades and Salinger (hereafter R&S) method (Rhoades and Salinger, 1993). The R&S method made use of monthly-average temperatures, whereas the Mullan *et al.* (2010) report used primarily annual averages. Nevertheless, the generic element common to both methods is the use of multiple comparison records which span across identified inhomogeneities at the reference site.

Rhoades and Salinger (1993) presented the mathematical formulation of their methodology for estimating the shift in a climatological time series at the time of a known site change. As an example of the approach, adjustments were estimated for minimum temperature and precipitation records at a number of sites in the vicinity of Lincoln, one of the seven-station reference sites (Figure 1). Salinger *et al.* (1992) applied the R&S method to estimate all necessary shifts to put together long-term composite temperature records at 24 selected reference stations spanning New Zealand, including three outlying islands (Raoul, Chatham and Campbell), and presented a complete set of graphs of the adjusted time series for seasonal and annual anomalies in mean, maximum and minimum temperature. The specific site shifts needed to derive the adjusted time series were not themselves documented in Salinger *et al.* (1992), but are noted in Mullan *et al.* (2010) for the subset of the seven-station reference sites as applied to mean temperatures.

The estimated site shifts for the seven-station reference sites are not identical between the Salinger *et al.* (1992) and Mullan *et al.* (2010) calculations, but tend to be quite similar. When the multiple adjustments are aggregated over time and over the seven reference sites, differences in the estimated shifts tend to balance out, and result in a New Zealand-average century-scale warming trend that is virtually the same regardless of the finer details of the adjustment method. However, there have been claims made outside the scientific literature (NZCSC, 2011) that the R&S methodology results in quite different adjustments to those presented in Mullan *et al.* (2010). Since Rhoades and Salinger (1993) provided just a 'single' example (Lincoln) of the adjustment calculations, and Salinger *et al.* (1992) only showed the resulting temperature time series and not the adjustments themselves, it seems timely to revisit the issue and provide a wider range of examples.

Section 2 of this paper reprises the essence of the R&S methodology, and Section 3 gives a broad selection of results. The purpose is not to be fully comprehensive (that is, not to derive all the adjustments at all the seven reference sites) but to provide enough examples to demonstrate that detailed methodological differences between Salinger *et al.* (1992) and Mullan *et al.* (2010) have relatively little impact on the resulting estimated shifts, *provided* due regard is given to minimising short-term noise and errors in the underlying temperature data. Issues influencing the estimated adjustments that are addressed include:

- the effect of comparison site selection and relative weighting;
- the effect of bad data, missing data, and short anomalous periods of data;
- the effect of the length of the overlap comparison period;
- discussion of the circumstances under which adjustments should be accepted even where not statistically significant.

Section 4 of this paper contains further discussion and conclusions.

2. Data and Method

(a) Station temperature data

The data used in this study are monthly temperatures, which are all freely available from NIWA's Climate Database (http://cliflo.niwa.co.nz). Both monthly minimum and maximum temperatures are used on occasion, although most emphasis is placed on the monthly mean temperature (derived from the average of the daily minimum and daily maximum). Figure 1 shows the location of the seven reference stations. Similar figures are provided in Mullan *et al.* (2010) for each reference site separately, which also show all the comparison sites used in the NIWA report. By and large, this paper makes use of the same comparison sites and the same site change dates as in Mullan *et al.* (2010), unless otherwise noted. Information on site changes, and other data issues, is available from the NIWA climate database metadata, and from Fouhy *et al.* (1992) and Mullan *et al.* (2010).

(b) Rhoades and Salinger methodology

The R&S method calculates differences between monthly temperatures before and after a site change. Differences at the target site (where the series is to be adjusted) are compared with differences at multiple comparison sites, and the final adjustment is a weighted combination of the "Comparison – Target" differences. We will make one small change here, and express the result as the "Target – Comparison" differences, in order to have the same sign convention as in Mullan *et al.* (2010). Thus a negative temperature adjustment is to be interpreted as the new target site being colder than the previous one. Rhoades and Salinger (1993) suggest different possibilities for the weighting of separate adjustments from the multiple comparison sites. In their worked example, they specify weights that depend on the fourth power of the inter-station correlations.

Adjustments are calculated as a function of "k", the accumulating number of years before and after the site change. For k=1, temperatures for the 12 months (1 year) before the site change are subtracted from temperatures for the 12 months after the site change; for k=2, temperature differences are taken between the 24 months (2 years) before and after, etc. For convenience, the essential mathematics is now reproduced.

R&S estimate the adjustment for a 'target' station, using n comparison stations, i = 1,2,..,n, where i = 0 is used to denote the target site. If x_i^t is the temperature for station i and month t, then R&S compute the monthly difference series for k years before and after the month of the site change, as:

$$y_i^t = x_i^{\tau+t} - x_i^{\tau+t-12k}$$

for stations i = 0, 1, …, n,

where t = 1, 2, …, 12k; and τ+1 denotes the first month at the new site.

The weighted difference between the target y_0 series and comparison y_i series can then be expressed as:

$$z^t = \sum_{i=1}^{n} w_i \, (y_0^t - y_i^t)$$
$$= y_0^t - \sum_{i=1}^{n} w_i \, y_i^t \quad .... (1)$$

The simplification on the right-hand side of Equation (1) occurs because the weights w_i are normalized to 1. If there are missing months in the calculation, then the left-hand formulation must be used, retaining the paired differences between each comparison site and

the target site. The weights w_i then need to be renormalized to 1.
Finally, the R&S adjustment is the average of z^t over the 12k months (or 12k – m, for m missing months in the series). Thus, the shift due to the site change is

$$\bar{z} = \sum_{t=1}^{12k} \frac{z^t}{12k} \quad \text{....(2)}$$

The associated standard error is given by

$$s = \left\{ \frac{\sum_{t=1}^{12k} (z^t - \bar{z})^2}{12k(12k-1)} \right\}^{\frac{1}{2}} \quad \text{.... (3)}$$

The 95% confidence interval, approximately ± 2 standard errors, can then be calculated in the usual way from the Students' t-statistic (see R&S (1993)).

Various choices are possible for the weights w_i . If all comparison sites are treated equally, then $w_i = 1/n$. R&S suggest that the weights w_i be given by the 4th power of the correlations ρ_i between the ith comparison site and the target site difference series (y_i, y_0), normalised to unity:

$$w_i = \rho_i^4 / \sum_{j=1}^{n} \rho_j^4 \quad \text{...(4)}$$

Note that the correlations should be calculated from the inter-annual difference series y_i rather than the raw series, in order to minimise the effect of site changes at the target site. R&S appear to calculate the correlations ρ_i from just the first year of differences (i.e., k = 1). Since the correlations will change (sometimes substantially) as more years of data are considered, we use all 12*k months in estimating ρ_i as the 'standard' weighting assumption.

The R&S method is quite flexible and easy to apply, allowing a quick comparison of different options an analyst might wish to consider, such as: varying the length of the comparison period, filling in missing data or not, altering the mix of comparison sites, varying the way comparison sites are weighted, and diagnosing the cause of outliers. These and other issues are discussed in the following section.

3. Results

3.1 Example calculation for Masterton site change in May 1920

Figure 2 gives an example to introduce the application of the R&S methodology. There is a site change in the Masterton reference series in April/June 1920: the Worksop Road measurements in the town of Masterton ceased at the end of April 1920, and a nearby site in Essex Street commenced in June 1920. (In the NIWA climate database, the combined records are archived under the common name of Waingawa, database agent number 2473). There is thus one missing month in May 1920, and in the calculations of Figure 2 this month is assigned as the starting month of the new target site (Essex Street). The estimated shift in base temperature between the Worksop Road and Essex Street sites is shown as a function of "k", the number of years before and after the site change. Thus, k = 1 compares Masterton monthly temperature differences for May 1920 to April 1921 relative to May 1919 to April 1920; k = 2 compares the period May 1920 to April 1922 relative to May 1918 to April 1920, and so on. In all cases, the Masterton minus comparison sites differences are then combined with a weighting given by equation (4), where the correlations are calculated using all 12k months of data. The adjustments are all negative, indicating that the new Masterton site (Essex Street) is colder than the former one (Worksop Road).

The three adjustment estimates of Figure 2 show the consequence of different ways of handling the missing May 1920 value. At the Masterton sites, the May temperature at Worksop Road in 1919 is 9.4°C and the May temperature at Essex Street in 1921 is 10.0°C. At all four comparison sites (the same as used in Mullan *et al.,* 2010, namely Albert Park (Auckland), Thorndon (Wellington), Christchurch Gardens and Taihape), May is colder in 1920 than in either 1919 or 1921. This is entirely consistent with the Trenberth (1976) M1 pressure indices, which show May

1920 as a month of strong southerly flow over New Zealand (the M1 index, or Hobart minus Chatham Island anomalous pressure difference, being +0.2, +11.0 and -2.7 hPa for the three Mays 1919-1921, respectively, where a positive index implies more southerly flow than the 1971-2000 normal).

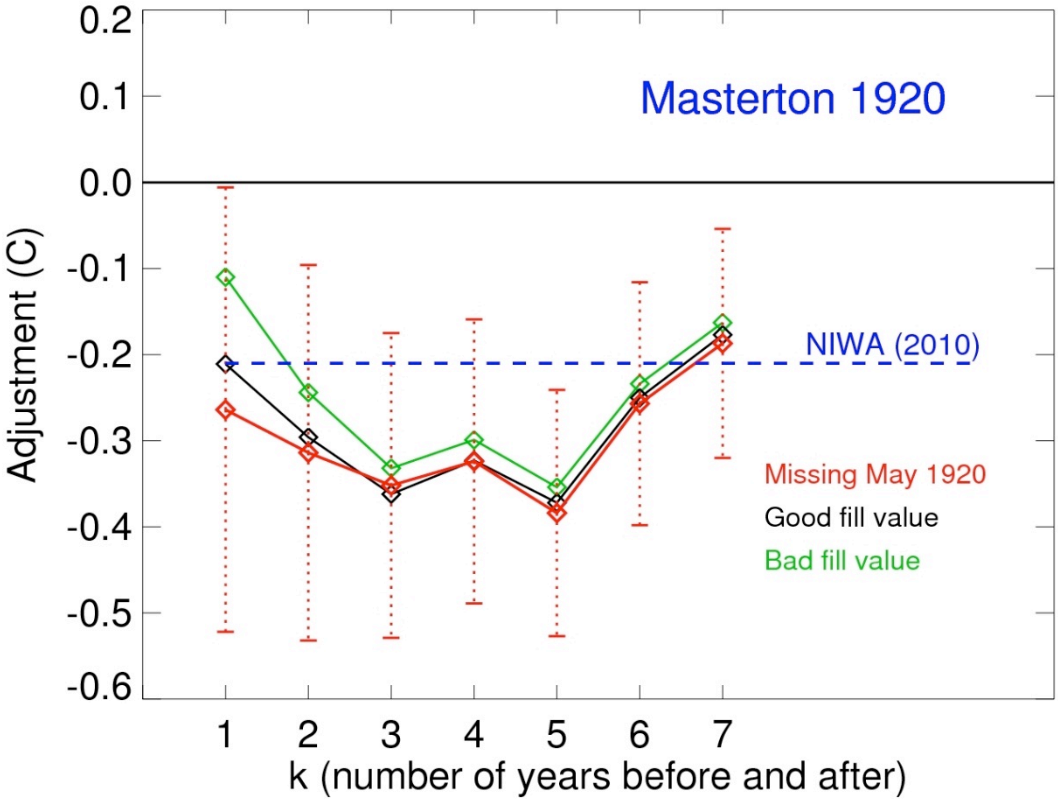

Figure 2: R&S temperature adjustments as a function of k for the May 1920 site change in the Masterton reference series, using four comparison sites (Albert Park, Thorndon, Christchurch Gardens and Taihape). The three lines show the effect of different ways of handling the missing value at May 1920: leave May 1920 missing (red diamonds and line), estimate the missing value (black), apply a naïve estimate for the missing value (green). The vertical red dotted lines denote the 95% confidence intervals about the estimates (red diamonds). The horizontal dashed blue line is the NIWA 2010 adjustment (of -0.21°C) based on comparing annual temperatures 1912-1919 with 1921-1927

One option for May 1920 is to leave the temperature as missing, and evaluate equation (1) with the May 1920 data for all comparison sites also omitted. This result is given by the red line in Figure 2, where the 95% confidence limits have been added. A second alternative (black line) is to estimate the missing May 1920 for Essex Street (none of the comparison sites have any missing months within the May 1913 to April 1927 period of the figure). A value of 8.9°C has been estimated by interpolation from the comparison sites. The final alternative, to illustrate the sensitivity of the method, is to assume a very naïve estimate of 10.0°C for the May 1920 mean temperature. This value is the same as the May of the following year at the site, and thus completely ignores the effect of inter-annual circulation variations. (One can also arrive at a value of 10.0°C by naively replacing the missing value by a 'climatology' over 1912-1927, which averages across both the warmer Worksop Road site and the colder Essex Street site).

For this site change, the estimated adjustments are not particularly stable and show considerable variation as more years of data are added. However, the adjustment is clearly negative (Essex Street colder), and for k > 2 the three estimates for a given k are all within 0.03°C of each other. The Mullan *et al.* (2010) adjustment was estimated to be -0.21°C in this case. For k = 1 or 2, there is quite a divergence of answers. If a bad estimate of 10.0°C is used to fill the missing May 1920 value, then the 95% confidence interval is increased substantially such that neither the k=1 or k=2 shift differs significantly from zero.

The variation in the adjustment with the length of data period is a consequence of many factors that could be affecting either the target Masterton sites or the comparison sites: possible errors in the measurements, systematic deviations from the climatological target-comparison site difference because of persistent circulation anomalies such as an El Niño period, and short or longer-term trends due to exposure changes. In Figure 2, the target-comparison site correlations are different for each value of k: at k=1, the correlations are 0.67, 0.79, 0.97 and 0.78 (for Albert Park, Thorndon, Christchurch and Taihape, respectively), resulting in weights for equation (4) of 0.11, 0.21, 0.48, and 0.20; at k=7, the correlations are 0.83, 0.78, 0.91 and 0.83, and the weights are 0.24, 0.18, 0.34 and 0.24.

A strength of the R&S methodology is that comparison sites can be quickly exchanged in order to assess what is affecting the shift estimates. For example, the Albert Park site may be undergoing non-climatic warming during this early period of its record (Mullan *et al*., 2010). If Albert Park is removed as a comparison site, the shift estimates reduce by about 0.05°C (to a less negative value). This change is, however, almost invariant with k whereas we would expect it to show a progressive change with the length of data period if there was systematic warming at the Auckland site. Experiments like this establish that the temporal variation in the estimated Masterton adjustment of Figure 2 is determined by the Masterton data and not by the comparison sites.

3.2 Stability of adjustment estimate

Adjustment estimates are sometimes quite stable as further years are added to the comparison period (see later examples), but it is also quite common for the estimate to require several years of data (k > 3) to settle down to a reasonably stable answer. Figure 3 investigates this issue for the 1942 Masterton site adjustment: the Essex Street record ends November 1942, and the Waingawa electricity substation record starts in October 1942, giving a two month overlap. The bottom panel of Figure 3 shows the R&S adjustment for k=1 through to k=8 (after which there is a break in the Taihape record). Two examples are shown to assess the sensitivity to different assumptions:

- joining the Masterton and Waingawa data at October 1942 (start of Waingawa record), with the 'standard' weighting that varies with k;
- joining the data at December 1942 (after end of Masterton record), with a weighting fixed by the k=1 inter-station correlations. Note that the Waingawa mean temperatures for October and November 1942 are 13.2°C and 14.2°C, respectively; the Masterton temperatures, not in the electronic database but recoverable from the original paper records, are 13.3°C and 14.2°C, respectively (after converting from Fahrenheit and rounding to one decimal place).

The two estimated adjustments differ by 0.05°C at k=1 and 2, but thereafter the differences are minimal. Beyond k=4 the adjustment asymptotes to the longer-term estimate of -0.26°C found by Mullan *et al.* (2010).

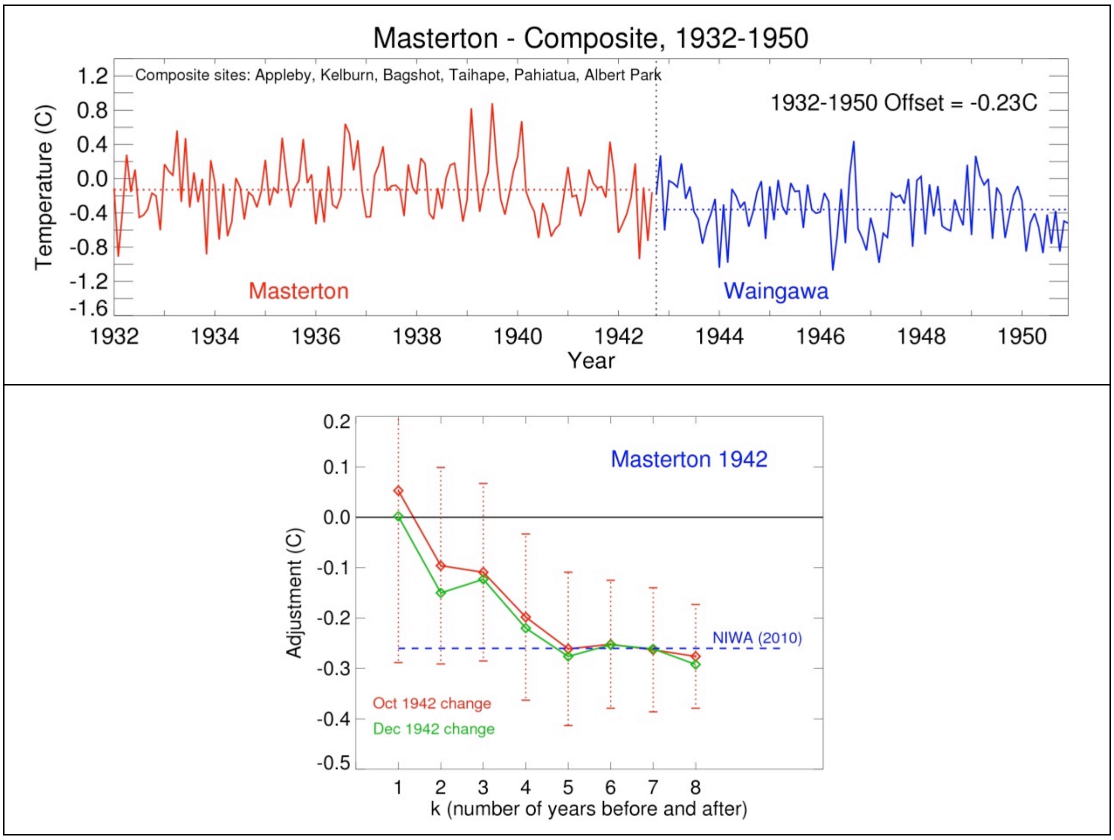

Figure 3: Anomaly time series and R&S adjustments for the 1942 site change from Masterton (Essex Street) to Waingawa electricity substation: (top) monthly time series of temperature anomalies, shown as the difference between the reference site (Masterton in red, Waingawa in blue) and the average over six comparison sites (Appleby, Kelburn, Bagshot, Taihape, Pahiatua and Albert Park); (bottom) R&S adjustment calculated for two cases: starting the Waingawa data in October 1942 and recalculating station weighting as each new year of data is added (red line and diamonds, with vertical dotted lines denoting 95% confidence interval); and starting the Waingawa data in December 1942 and using station weights determined from just one year (k=1) of data (green line and diamonds). See text for further details.

The top panel of Figure 3 provides a clue as to why the adjustment at k=1 is so much smaller than for k>3. This panel displays the time series of the difference in monthly mean temperature between the Masterton/Waingawa site and a composite derived by averaging the records from six comparison sites. Before differencing, each series was first normalised by subtracting the monthly 'climatology' for the respective periods as plotted (i.e., Jan 1932 to September 1942, and October 1942 to December 1950), but then adding back in the annual mean climatology before plotting to retain the long-term offset. This normalisation has minimal effect with monthly mean temperature, but is helpful when visualising monthly maximum or minimum temperatures where there are often distinct annual cycles in the target-comparison site differences.

The mean shift between the 1932-1942 and 1942-1950 periods, with the break at September/October 1942, is -0.23°C. The two time series segments, for Essex Street and the Waingawa substation, appear fairly stable about their long-term levels. However, there are occasional short-term deviations, one of which appears immediately after the October 1942 site change: for about the first eight months the relative Waingawa temperatures are at a very similar level to the previous Essex Street values (and successive removal of the comparison sites shows this is a feature of the Masterton data and not the comparison sites). Why this deviation occurred is not apparent from surviving metadata, and there was no substantial circulation anomaly at this time since the long-running 1941-42 El Niño had finished. However, a similarly anomalous period, but in this case a deviation *below* the long-term level, occurs after October 1946. Had the 1942-43 and 1946-47 periods been transposed, the R&S calculation would have given a much larger negative adjustment at k=1 than the longer-term k=4-7 estimates.

It would appear that short-term anomalous periods are relatively common in the New Zealand temperature data sets, and are not worth correcting (although Rhoades and Neill, 1995, tried to). The straightforward solution is to take a long enough comparison period so the effects of short-term anomalies are minimised. As a final comment on this example, it should be noted that removing the Auckland Albert Park data makes no difference to the results: all the numbers discussed above are the same to two decimal places, apart from the k<3 R&S adjustments.

3.3 Further examples of sensitivity of R&S adjustment

Figure 4 shows further examples of temperature adjustments associated with site changes. Except for the 1928 Kelburn site change when k=5 and k=6, the Mullan *et al.* (2010) estimates are within the 95% uncertainty band of the R&S estimates when a comparison period of about 4 years or more about the site change is considered. Unless otherwise noted below, the comparison sites are the same as selected by Mullan *et al.* (2010).

- <u>Kelburn site change, January 1928</u>: this example (Figure 4, top left) demonstrates that the hilltop Kelburn site is about 1°C colder than the previous harbour-side Thorndon site, and the R&S estimate is stable as further years are added successively to the comparison period. An interesting feature, which is quite commonly observed, is the very different adjustments required for the monthly maximum and minimum temperatures. In this case, it is apparent that the more sheltered Thorndon site allows an enhanced difference in daily maximum temperatures relative to Kelburn, but reduces the difference in overnight minimum temperatures, possibly due to ponding of cooler air.

- <u>Nelson site change, December 1920</u>: this example calculates the shift between the Nelson Town sites in Nile Street and at the Cawthron Institute. Estimation of site shifts are more problematic in New Zealand prior to about 1930 because of the reduced number of available comparison sites. There are further complications in this instance with poor exposure at Nile Street and with bad and missing temperature data. The Nelson Nile Street site has a period of missing data one year prior to the site change (May to September 1919); and Dunedin Botanical Gardens, one of the few suitable comparison sites, has a period of dubious data in 1921: Fouhy *et al.* (1992, p. 183) comment that "Maximum temperatures were wrong from April to May and possibly in October also". A visual inspection of the Dunedin data against other sites suggests that the intermediate months June-September 1921 may also be wrong. This bad Dunedin data has been retained in the NIWA climate database for future analysts to use at their discretion.

Figure 4 (top right) shows the R&S adjustments for the December 1920 Nelson site change with and without the April-October 1921 Dunedin data. There is a large difference in the estimates for k=1, but surprisingly little with a longer comparison period. There is a substantial change in the estimated shift between k=3 and k=4, which is not affected by a different

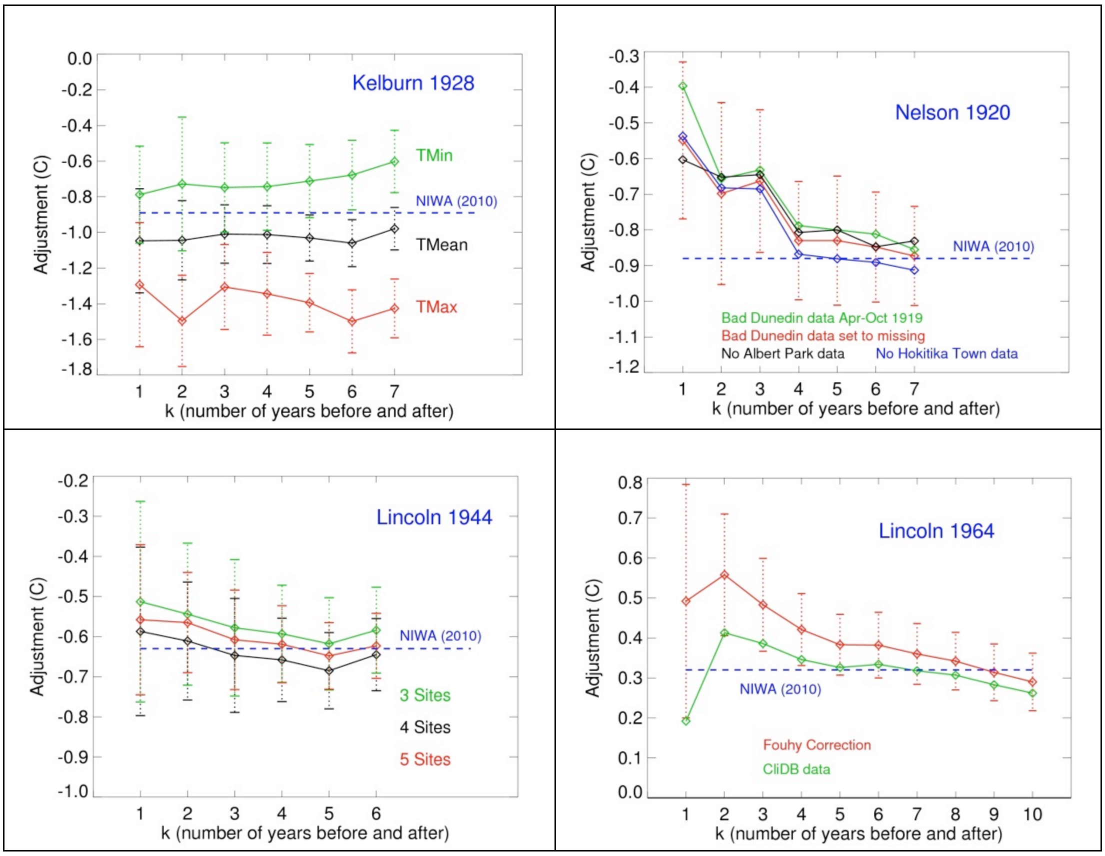

Figure 4: R&S temperature adjustments as a function of k for four site changes in the seven-station temperature series: (top left) Kelburn Wellington in January 1928, using the comparison sites Albert Park, Taihape, Waingawa and Christchurch Gardens; (top right) Nelson in December 1920, using a selection of comparison sites from Albert Park, Thorndon, Hokitika Town and Dunedin Botanical Gardens; (bottom left) Lincoln in January 1944, using a selection of comparison sites from Christchurch Gardens, Ashburton Council, Onawe, Waimate and Wigram; (bottom right) Lincoln in May 1964, using the comparison sites Christchurch Gardens, Christchurch Aero, Ashburton Council and Darfield. Further details are discussed in the text.

choice of comparison site (Figure 4 illustrates the result of removing either Albert Park or Hokitika Town from the set of four comparison sites). The Fouhy *et al.* (1992) discussion of the Nelson climate stations indicates on-going problems with the site exposure and measurement quality. For example, an April 1920 site inspection noted that "The thermometer screen was 'almost under' a tall hedge and surrounded by tall trees and it was not known how long conditions had been like that" (Fouhy *et al.*, 1992, p. 127). The R&S analysis in Figure 4 suggests a possible undocumented 'site change' 3-4 years prior to December 1920, such that the 1918-1920 Nile Street temperatures are about 0.65°C warmer than at Cawthron, but the temperatures prior to December 1917 are at least 0.85°C warmer (remembering that the k > 3 R&S results include all data for k ≤ 3 as well).

- Lincoln site change, January 1944: Figure 4 (bottom left) shows the R&S estimated adjustments for the January 1944 site change at Lincoln from south of the main Lincoln College buildings to a field north of the campus; this was the site change (but for minimum temperature) that Rhoades and Salinger (1993) gave a worked example for. This is an example similar to Kelburn 1928 for which the adjustments are fairly stable as the length of comparison period is extended. Figure 4 also demonstrates a small dependence on the comparison sites used: the 5-site curve uses the four comparison sites of Mullan *et al.* (2010) plus Wigram Aero, whose record starts in September 1937. The 4-site curve drops the Wigram data, and the 3-site curve additionally drops Christchurch Gardens. (The Gardens site has a considerably larger warming trend than Lincoln (Wratt *et al.,* 2003, Fig. 3.8), suspected as being due to progressive sheltering with the growth of large trees). However, regardless of the mix of comparison sites, the estimated adjustments are significantly different from zero and close to the Mullan *et al.* (2010) value, and show the new site to be colder than the previous one. In contrast to the Kelburn 1928 site change, the Lincoln 1944 shift is much larger for minimum temperature than for maximum temperature (not shown).

- <u>Lincoln site change, May 1964</u>: Figure 4 (bottom right) shows the R&S estimated adjustments for the subsequent May 1964 site change at Lincoln to a research farm 0.5 km west of the College campus. This new site is significantly warmer, by about 0.3°C, than the earlier site. Figure 4 also illustrates the influence of potentially bad data during 1964-65. One curve shows estimated adjustments using data as in the NIWA climate database; and the second curve uses the correction of +0.55°C to the August 1964 to January 1965 mean temperatures recommended by Fouhy *et al.* (1992, p. 142). There is big difference between the estimated temperature shifts for $k \leq 3$. This example reinforces the message of previous figures that the R&S adjustments become relatively insensitive to data errors for a comparison period of 8 years or more ($k \geq 4$).

3.4 Statistical significance of R&S adjustment

The R&S method has the advantage of providing a simple expression for the standard error of the estimated adjustments (equation (3) above), although it should be noted that it makes the assumption that the target-comparison site monthly differences are independent, identically distributed normal random variables. This is probably not true, given that we might expect some serial correlation in site-to-site differences when there are persistent circulation anomalies, such as during El Niño periods. A correction to equation (3) could be made by assuming the number of independent variables is less than 12k. Nevertheless, proceeding with the R&S formulation for the purposes of this paper, we note that the 95% confidence interval is typically in the range ±0.3 to 0.4°C for $k = 1$, reducing to about ±0.10 to 0.15°C by $k = 4$ or 5.

Thus, a modest temperature shift of 0.2-0.3°C could be overlooked if the R&S method is applied only up to $k = 2$ (±2 years either side of a site change), and one insisted on the adjustment being significantly different from zero. If there happened to be more than one such shift in the reference series, and all were ignored, this could lead to a serious overestimate or underestimate of the long-term temperature trend at the reference site. Salinger et al. (1992) avoided this error by focussing on $k = 2$ *and* $k = 4$, as well as comparing annual differences over longer periods still, in order to ensure a stable estimate had been reached. An insistence on any adjustment being significantly different from zero (i.e., the 95% confidence band about the estimate does not overlap zero) would seem to be a justifiable strategy in the case of an undocumented site change. However, if the station metadata very clearly identify a site change, then it seems

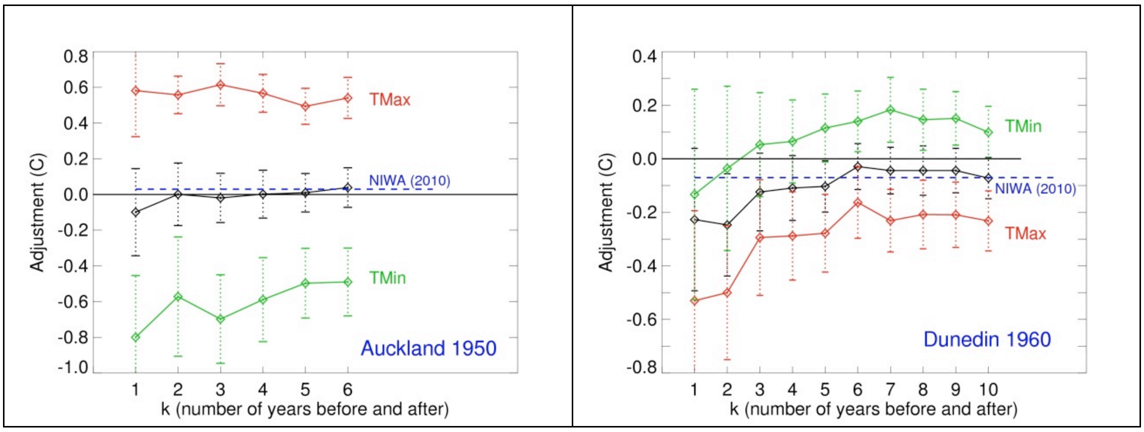

Figure 5: R&S temperature adjustments as a function of k for two site changes in the seven-station temperature series: (left) Albert Park (Auckland) in November 1950, using the comparison sites Waipoua, Riverhead Forest, Waiuku Forest, Te Aroha and Ruakura; (right) Musselburgh (Dunedin) in November 1960, using the comparison sites Kelburn, Waimate, Adair and Invercargill Aero. Further details are discussed in the text.

reasonable to apply the 'best' estimate of the shift to the raw data series, even in the absence of statistical significance.

Often a non-significant adjustment can be supported from consideration of other data. Figure 5 provides two examples where a site shift in mean temperature is not significantly different from zero, following the R&S approach, but there are clearly significant adjustments to both the minimum and maximum temperatures separately. Thus, it is mandatory to adjust the mean temperature for mathematical consistency in the records before and after the site shift. The temperature adjustment at Auckland (Albert Park) is a consequence of replacing the Stevenson screen housing the thermometers in November 1950 (Fouhy *et al.,* 1992, p. 29) by a smaller version, sometimes called a Bilham screen after the British scientist who studied its design characteristics (Bilham, 1937). This screen replacement occurred at many New Zealand climate stations through the 1940s and 1950s; Figure 5 (left) shows the consequence of effectively greater sheltering was to increase the daily maximum temperature and reduce the daily minimum, but with only a small change to the daily mean temperature.

Figure 5 (right) shows a similar example for the site shift in November 1960 from the south to the northwest side of the Musselburgh (Dunedin) pumping station. Fouhy *et al.* (1992, p. 179) suggest that the new site was more exposed than the old one, which is supported by the R&S finding of lower maximum temperatures and higher minimum temperatures at the northwest site. There is minimal effect on the daily mean temperature, but although the mean temperature shift is small (-0.07°C according to the Mullan *et al.* (2010) estimate, and -0.09°C at k = 4 by the R&S method), once again it is mandatory to apply this adjustment when considered in the broader perspective of trends in maximum and minimum temperatures as well as mean temperature.

3.5 Assessment of Hokitika 1943 shift

Figure 6 provides the final example in this paper of the application of the R&S methodology, in this case to the August 1943 site change at Hokitika from the Town to the Southside aerodrome south of the Hokitika river (Mullan *et al*., 2010). The example is a salutary one of the poor R&S estimate that can result if the period of comparison is too short. Figure 6 (top panel) shows the normalised time series (as in Figure 3) of monthly mean temperature differences between the Hokitika sites and the composite of two comparison sites, Appleby and Kelburn. The period of comparison is made as long as possible, starting in January 1932 with

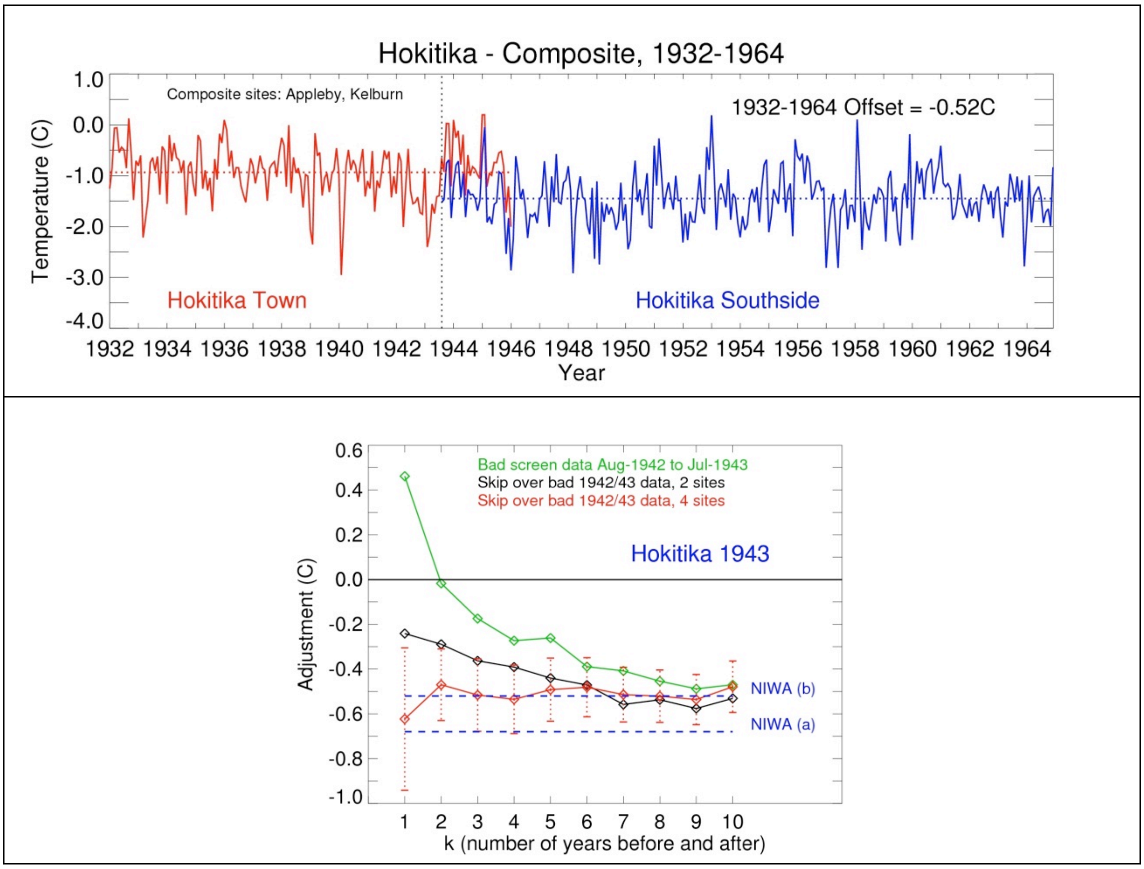

Figure 6: Anomaly time series and R&S adjustments for the August 1943 site change from Hokitika Town to Hokitika Southside: (top) monthly time series of temperature anomalies, shown as the difference between the reference site (Town in red, Southside in blue) and the average over two comparison sites (Appleby and Kelburn); (bottom) R&S adjustments and NIWA 2010 adjustments for various combinations of data and comparison sites as discussed in the text. The 95% confidence intervals (red vertical dotted lines) are shown for the case of four comparison sites (Appleby, Kelburn, Westport Aero and New Plymouth), where the offset is estimated between Hokitika Town data up to July 1942 and Hokitika Southside data from August 1943 onwards (note there is a year of omitted data, August 1942 to July 1943). See text for further details.

the opening of the Appleby climate station, and ending in December 1964 when the Hokitika Southside site closed. As the figure shows, there is an overlap period, August 1943 to January 1946, when both Hokitika sites were operating. With respect to Appleby and Kelburn, there is a long-term stable offset between the Hokitika Town site (January 1932 to January 1946) and the Hokitika Southside site (August 1943 to December 1964), with Southside colder by 0.52°C.

A closer inspection of the monthly time series suggests there is some anomalous behaviour in the Hokitika Town temperatures (expressed as deviations from the average of the Appleby and Kelburn temperatures) immediately before and after the August 1943 date when the Southside station opened. Drawing on the station metadata, Fouhy *et al.* (1992, p. 111) comment on a site inspection visit in July 1943 after heavy rainfall caused local flooding and problems with the rain gauge. The visit found the thermometer screen to be in very bad order: “Most of the louvres had fallen off, the roof was rotten and it needed painting. The necessary repairs were made.” Thus, the repaired Stevenson screen at the Town site went into operation at the same time as the Southside measurements began. Figure 6 (top) indicates that temperatures post-August 1943 at the Town site were mostly above the long-term level of the Town record, whereas temperatures for about 9 months prior to the July 1943 repairs were substantially below the long-term level.

This combination of circumstances has a major influence on the R&S adjustments calculated from a short overlap period (Figure 6, bottom panel). The green line in Figure 6 shows the R&S estimate for the shift between the Hokitika Town record up to July 1943 and the Hokitika Southside record from August 1943 onwards, using the same comparison sites as in the top panel (i.e., Appleby and Kelburn). The k = 1 adjustment is +0.46 ±0.42°C, implying a statistically significant adjustment with the Southside site *warmer* than the Town site, which is obviously wrong. As extra comparison years are added, the R&S adjustment quickly reverses sign and becomes significantly negative (i.e., Southside colder than the Town site).

The black and red lines in Figure 6 (bottom panel) show the results when the year of data with the rotten Stevenson screen is excised (i.e., August 1942 to July 1943), and the shift calculated between Town data ending July 1942 and Southside data starting August 1943. The R&S estimates are shown for the same two comparison sites as previously (Appleby and Kelburn), and also with two additional sites as used in Mullan *et al.* (2010). These latter R&S estimates are very stable, especially the four-site one, and agree closely with the long-term Mullan *et al.* (2010) estimate of -0.52°C, labelled as NIWA (b) in Figure 6. The NIWA (a) estimate of -0.68°C was the one actually adopted in the Mullan *et al.* (2010) study, and corresponds to the offset between Hokitika Town and Southside during the overlap period (Figure 6, top panel).

The current NIWA adjustment for this Hokitika site change is to start the Southside record at January 1945 and adjust the earlier Town record down by -0.68°C. However, in view of the analysis presented here, a better solution is to join the records at August 1943 and adjust the earlier Town record down by -0.52°C. This would have the effect of reducing the 100-year warming trend at Hokitika by about 0.2°C. The bad Town data during August 1942 to July 1943 should also be corrected, although this will not affect the trend in any substantial way.

4. Discussion and Conclusions

This paper has provided a number of examples of applying the Rhoades and Salinger (1993) site adjustment methodology to temperature records making up the so-called "seven-station" series. This methodology is just one of many that have been suggested internationally (Peterson *et al.*, 1998), but it is of particular interest in the New Zealand context because it was the approach used in the NZ Meteorological Service study which formally developed the seven-station temperature series (Salinger *et al.*, 1992). This series has recently been updated by NIWA, in a report available on the web (Mullan, *et al.,* 2010). The two approaches produce estimates of site shifts that are not exactly the same, but are only slightly different in a random way such that Mullan *et al.* (2010) found the 100-year seven-station composite trend to be the same to two decimal places (although not for the seven reference series separately).

The most important conclusion in this paper relates to the preferred length of comparison period over which the site adjustment is assessed. Statistical confidence intervals shrink as k increases. For modest adjustments on the order of 0.2 to 0.3°C, one usually requires $k \geq 3$ to obtain an adjustment where the 95% confidence interval is clear of zero. However, it is not adequate to stop as soon as a significant result is found. Adjustments for site changes are often applied to several decades of record. It is therefore essential for the estimate to be stable and robust, and not unduly influenced by short-term effects such as bad, missing or anomalous data. The results presented here suggest at least 4 years before and after a site change should be compared ($k \geq 4$). This is consistent with much other work, such as Karl and Williams' (1987) recommendation of at least 5 years before and after a site change. More modern techniques for assessing site adjustments (e.g., Wang *et al.,* 2007) also tend to use several decades of record where possible.

Working with observational data, particularly with historical measurements where metadata are sparse, will always be fraught with difficulties, and controversy is likely to remain about interpretations of some site adjustments. This is the advantage of tackling the site change issue in a number of different ways. The Rhoades and Salinger method has one great advantage in that it can be easily automated and is very flexible. The method can readily assess the consequence of variations in: the length of comparison period (k years), the mix of comparison sites, the weighting of comparison sites, and the inclusion or exclusion of certain periods of data. Filling in missing data values can be a very labour-intensive job in climatological analyses, and the R&S method can simply skip over missing months with very little effect on the final adjustment.

Examples of all these issues have been presented. In practice, the way comparison sites are weighted in the R&S scheme has the least effect, although this conclusion is in the context of comparison sites being pre-selected on the basis of high correlations with the target site inter-annual temperature variation. Typically, we expect correlations of about 0.8 or higher. The examples also show that the specific comparison sites used do influence the estimated adjustment, and this 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. However, short-term variability in the target site temperatures is by far the dominating factor in the 'shape' of the adjustment curve as a function of years (k) of comparison data before and after the site change. This variability may be due to errors in the measurements (e.g., mis-reading the thermometer if the mercury column is broken), to exposure changes (e.g., growth of surrounding trees, damaged thermometer screen), or to persistent circulation anomalies (e.g., El Niño events). The R&S methodology could probably be adapted as an aid to identifying unknown site changes, and diagnosing outliers and periods of bad data.

Acknowledgements

This work was funded by NIWA's National Climate Centre Core Funding, allocated by the New Zealand Government. The author thanks Stephen Stuart (NIWA) for a careful review of the manuscript, and James Sturman (NIWA) for providing Figure 1.

References

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

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.

Karl TR, Williams CN. 1987. An approach to adjusting climatological time series for discontinuous inhomogeneities. *J. Clim. Appl. Meteorology*, **26**, 1744-1763.

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, ISSN 1174-264X, 175p., http://www.niwa.co.nz/our-science/climate/information-and-resources/nz-temp-record/seven-station-series-temperature-data).

NZCSC. 2011. New Zealand – Unaffected by global warming. New Zealand Climate Science Coalition, http://nzclimatescience.net/index.php.

Peterson TC, Easterling DR, Karl TR, 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 EJ, 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.

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

Rhoades DA, Neill WA. 1995. The effect of station drift on the estimation of shifts in meteorological records due to site changes. *International Journal of Climatology*, **15**, 207-220.

Salinger MJ, McGann RP, Coutts L, Collen B, Fouhy E. 1992. South Pacific Historical Climate Network. Temperature Trends in New Zealand and Outlying Islands, 1920–1990. NZ Meteorological Service, Wellington, ISBN 0-477-01598-0, 46 p.

Trenberth KE. 1976. Fluctuations and trends in indices of the Southern Hemisphere circulation. *Quart. J. Royal Meteor. Soc.*, **102**, 65-75.

Wang X L, Wen Q H, Wu Y. 2007. Penalized maximal *t* test for detecting undocumented mean change in climate data series. *Journal of Applied Meteorology and Climatology*, **46**, 916–931.

Wratt D, Mullan B, Salinger J, Allan S, Morgan T, Kenny, G. 2003. Overview of Climate Change Effects and Impacts Assessments – A guidance manual for local government in New Zealand (1st Edition). NIWA Client Report WLG2003/44, prepared for Climate Change Office, Ministry for the Environment, 141p. (available from website: http://www.climatechange.govt.nz/resources/local-govt/effects-impacts-may04/index.html

Submitted to *Weather and Climate* May 2012, revised June 2012.