Evaluation of interpolated daily temperature data for high elevation areas in New Zealand

Andrew Tait and Gregor Macara

NIWA, Private Bag 14-901, Wellington, New Zealand

Correspondence: email andrew.tait@niwa.co.nz

Abstract

Accurate estimates of daily maximum and minimum temperature at locations where there are no actual observations are extremely useful for many purposes. The daily gridded Virtual Climate Station (VCS) estimates of these two variables (plus nine others), which are based on a thin plate smoothing spline model, have become a well-used dataset. However, there is some concern about the usefulness of the VCS temperature estimates in high elevations due to the possibility that the values are often too high (warm).

This study presents two alternate daily temperature interpolation methods, both based on the use of fixed lapse rates, and evaluates the accuracy of all three methods against two independent datasets; the first a set of mostly low elevation stations and the second a single high elevation site. Results of the comparison show that while the choice of interpolation method makes little difference in low elevation areas, the 'Norton' fixed lapse rate method is clearly the most accurate for the high elevation site.

This study has confirmed previous concerns about the significant warm bias (mostly in summer) of the operational VCS daily temperature estimates in high elevation locations. It also presents an improved methodology, which reduces this bias and is unlikely to have any negative impact on the accuracy of temperature estimates in low elevation locations. As a result, the operational VCS temperature interpolation methodology will be changed forthwith.

1. Introduction

Daily maximum and minimum temperature (Tmax and Tmin) observations from around 150 locations throughout New Zealand are operationally interpolated onto the Virtual Climate Station (VCS) grid (0.05 degrees lat/long covering all of New Zealand; 11491 grid points) every day and stored on the NIWA National Climate Database. The temperature interpolation methodology is described in Tait (2008) and presented schematically in Figure 1a. In brief, maximum and minimum temperatures are interpolated independently each day using a thin plate smoothing spline interpolation with two position variables (latitude and longitude) and a linear dependence on elevation which is determined from minimising the generalised cross validation error. The interpolation software used is ANUsplin v4.2 (Hutchinson, 2013).

Historic daily observations going back to January 1972 have been interpolated onto the VCS grid, yielding 40-year time series of uninterrupted daily Tmax and Tmin estimates that are updated every day. Daily gridded esti-

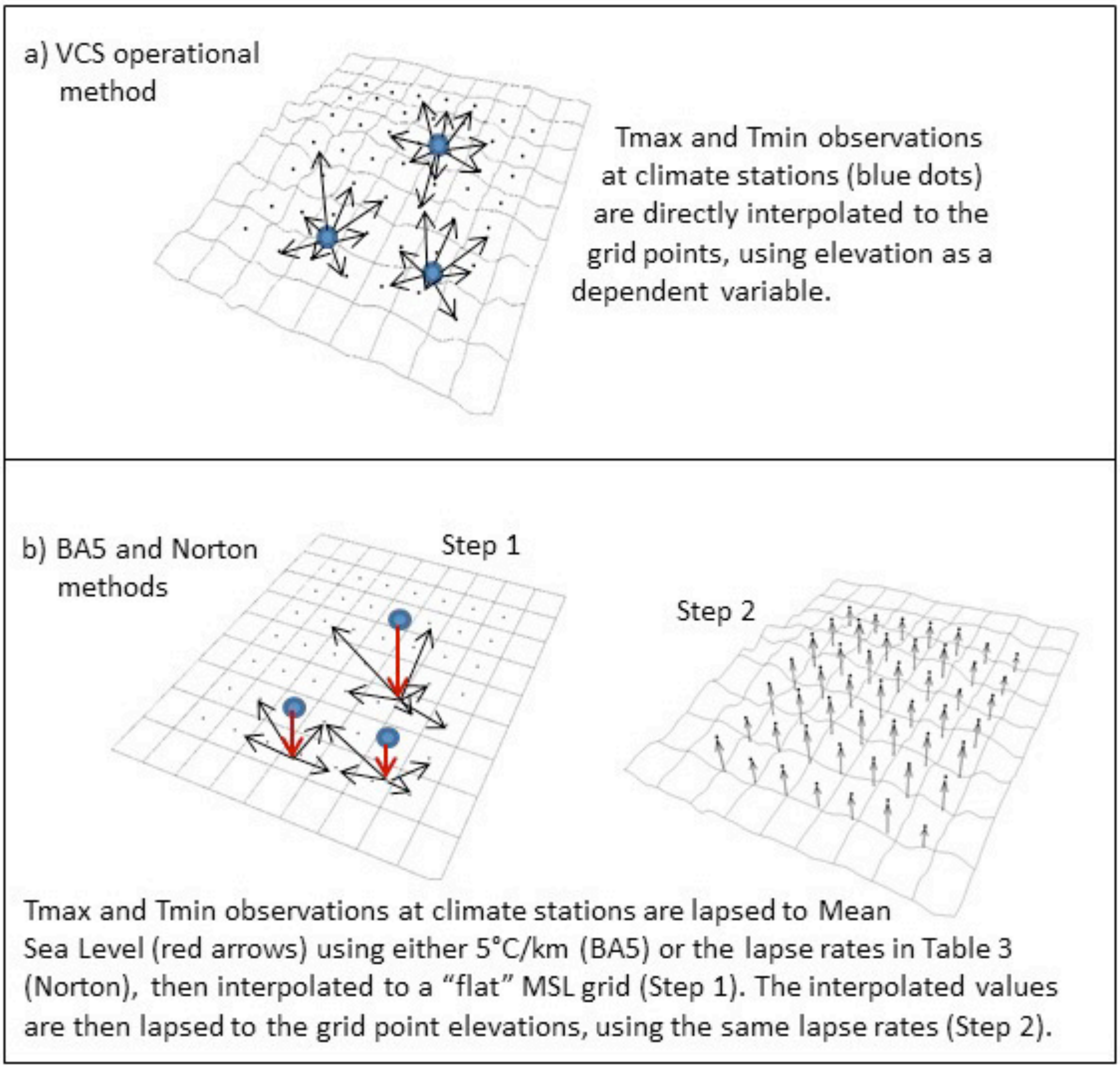

Figure 1: Schematic representations of the three temperature interpolation techniques compared in this study.

mates of nine other climate variables (pressure, rainfall, potential evapotranspiration, soil moisture, windspeed, solar radiation, earth temperature, vapour pressure and relative humidity) form the complete VCS dataset. Descriptions of the interpolation methodologies and error assessments of some of these other VCS variables have been performed (see Tait et al., 2006; Tait and Woods, 2007; Tait and Liley, 2009; and Tait et al., 2012). The VCS data can be accessed for free from http://cliflo.niwa.co.nz (select "Special Data Sets" > "Virtual Climate Network" when choosing Data Type).

The full VCS dataset has been used for multiple applications. VCS data are the base dataset used in New Zealand for empirical statistical downscaling of Global Climate Model projections (e.g. MfE, 2008) and are the primary dataset for most climate change impact model analyses. The data have been used to better understand the influences of Southern Hemispheric atmospheric circulation on New Zealand's rainfall (Ummenhofer et al., 2009) and are used as the basis for 15-day weather forecasts for the country (Renwick et al., 2009). The data are also used in physiological models, including pasture, crop, forest and animal production models, hydrological assessments, soil moisture and drought models, the development of a New Zealand snow model, ground water recharge models, renewable energy assessments, weed and pest habitat models, human disease models and water quality models (e.g. Clark et al., 2009; Elliott and Harper, 2010; Kirschbaum et al., 2012; Romera et al., 2010; Tait et al., 2008; McBride et al., 2014; Woods et al., 2006). Furthermore, several Regional Councils (RCs) use VCS data operationally for water allocation decisions (e.g. Environment Canterbury, Environment Waikato, and Hawke's Bay Regional Council). It is generally accepted that for most applications the VCS data are appropriate, but that the sparseness of the input data locations (i.e. actual climate stations) and the complexity of the terrain over which the data are interpolated affect its accuracy.

In particular, there is some concern that the accuracy of the VCS Tmax and Tmin estimates in regions of high mountainous terrain in New Zealand is poor, and may not be acceptable for applications such as snow water equivalence and glacier mass balance modelling (*pers comm*, Brian Anderson, Antarctic Research Centre, Victoria University). For background, Table 1 shows the percent of New Zealand's land area in certain altitude bands. This shows that there is

Altitude band (m above sea level)	Percent of land area (%)
0 - 300	45.1
300 - 500	17.3
500 - 1000	23.9
1000 - 2000	13.4
>2000	0.3

Table 1: Percent of New Zealand's land area in the given altitude bands.

a significant area above 1000m (13.7%) which could be classed 'high mountainous terrain'. The concern over the validity of using VCS temperature estimates for these parts of the country is supported by a recent snowline study (Clark et al., 2009), glacier mass balance analyses (Anderson et al., 2006; Anderson and MacIntosh, 2012) and an unpublished New Zealand treeline mapping study using a relationship with mean temperatures in the warmest month of the year (a methodology first described in Köppen, 1931), which all show that the VCS Tmax and Tmin estimates in high elevations are frequently too high (warm). In each of the above studies, Tmax and Tmin estimates using a fixed lapse rate of 5°C/km yielded improved results. This paper will compare results using this fixed lapse rate method plus another method which varies the lapse rate by season and data type (Norton, 1985) against the operational VCS Tmax and Tmin estimates for low and high elevation locations throughout the country. A decision on which method should be used operationally for the VCS Tmax and Tmin variables, based on this comparison, will be made.

2. Previous estimation of interpolation error

Tait (2008) analysed the operational VCS Tmax and Tmin interpolation error using data from 20 climate stations left out of the daily interpolations (but re-introduced for the operational interpolations) (Table 2; see also Figure 1 in Tait, 2008 for a map of the sites). The 2008 analysis showed that the average root mean square error (RMSE) varied between 0.7 and 2.0 °C for daily maximum air temperature, with a mean over the 20 sites of 1.2 °C, and between 1.2 and 3.2 °C, with a mean of 1.6 °C, for daily minimum air temperature (Table 2). The study concluded that the slightly higher error associated with the minimum air temperatures was related to lower interpolation accuracy during the winter months caused by localized temperature inversions, and that in both the maximum and minimum air temperature interpolations, higher RMSEs tended to be related to higher elevation and/or more remote locations.

Station ID	Lat (°S)	Long (°E)	Height (m)	No. Obs (days)	Mean Max Temp (°C)	RMSE Max Temp (°C)	Mean Min Temp (°C)	RMSE Min Temp (°C)
1129	35.4	173.8	204	4754	18.4	1.4	11.0	1.4
1434	36.9	174.8	75	6907	18.6	0.7	12.6	2.1
1550	37.4	175.8	91	6243	18.5	1.2	9.5	1.7
1645	37.8	176.3	91	6060	18.8	1.3	9.0	1.2
1841	38.7	176.1	376	8017	17.1	0.9	7.0	1.3
2005	37.2	174.9	82	6168	18.4	0.7	10.2	1.1
2103	37.8	175.1	104	11623	18.4	0.9	9.3	1.5
3014	39.7	176.9	9	2626	19.1	1.2	7.1	1.3
3232	40.3	175.5	15	6832	17.6	0.6	8.1	1.5
3549	39.5	174.2	98	6389	16.7	1.4	8.7	1.7
3925	42.1	171.9	198	11582	16.6	2.0	6.1	1.7
4241	41.3	173.2	2	11099	17.4	1.1	7.7	1.3
4310	41.5	174.0	4	3947	18.1	1.2	7.6	1.3
4458	42.5	172.9	387	8225	16.5	1.9	4.0	2.4
4566	42.8	173.3	85	4667	17.2	1.8	5.5	3.2
4764	43.8	171.8	160	12193	16.4	1.1	5.8	1.1
5280	45.1	170.1	427	4690	15.1	1.7	2.8	1.7
5577	45.2	169.3	171	4822	16.7	0.8	4.2	1.2
5823	46.6	168.4	5	11996	14.0	1.0	7.1	2.0
17029	41.1	175.1	56	2357	17.5	1.1	8.1	1.4
Average				7060	17.3	1.2	7.6	1.6

Table 2: Root mean square error (RMSE) and mean maximum and minimum air temperature at the 20 validation sites used in Tait (2008).

3. Discussion on two alternate interpolation methodologies

One alternate method for interpolating daily temperature uses a set lapse rate of 5°C/km to lapse the daily maximum and minimum temperature observations down to mean sea level (MSL), then spatially interpolate the MSL temperatures using a bilinear (e.g. inverse distance) interpolation onto the VCS (0.05 degree lat/long) grid, and finally lapse these gridded MSL temperatures up to the gridpoint elevations using the same 5°C/km lapse rate (Figure 1b). The evidence from the above mentioned glaciological, snow water equivalence and tree line studies suggests that this alternate method (which is coded here as BA5 – for Brian Anderson: 5°C/km) performs better at estimating daily maximum and minimum temperatures at high elevations.

A second alternate method is similar to the BA5 approach, but uses lapse rates which vary seasonally and by data type (Figure 1b). The lapse rates used here were derived from a study by Norton (1985), which used Tmax and Tmin monthly normals data calculated over the period 195180 at around 300 climate stations from throughout New Zealand. Norton also calculated lapse rates for mean daily temperature and for two elevation bands (below and above 300m). For the purposes of this study, only the lapse rates for Tmax and Tmin were used and these were averaged over all elevations (Table 3).

Other alternate interpolation methods exist (such as linear and polynomial regression and kriging, for example; see Matheron, 1981, Philips *et al*., 1992) however for the purposes of this study we have limited ourselves to a

	Maximum Temperature (°C / km)	Minimum Temperature (°C / km)
Summer (DJF)	6.1	4.1
Autumn (MAM)	6.3	3.2
Winter (JJA)	6.4	3.0
Spring (SON)	6.6	4.2

Table 3: Lapse rates used in this study, from Table 1 of Norton (1985).

comparison of the operational VCS method (based on spline interpolation) and the two lapse rate methods (BA5 and Norton).

4. Comparison of interpolated datasets

The independent data validation exercise for the VCS operational method using 20 validation sites (Table 2; Tait, 2008) was repeated for the BA5 and Norton interpolation methods. The statistics compared are the root mean square error (RMSE), mean absolute error (MAE), percent of positive differences (as a measure of bias), and percent of differences greater than 2 °C (as a measure of poor performance), where the differences here are the estimated daily Tmax and Tmin from each model minus actual daily Tmax and Tmin at the 20 validation sites (Tables 4 and 5).

Note: the values presented as the average of the 'percent of positive differences' for each model in Tables 4 and 5 are not the simple average of the percent of positive difference values at the 20 validation sites. This is because both high and low values (i.e. values significantly different from 50%) depict a model bias, and simple averaging would mask these biases. Hence, the values at the bottom of these columns are the average over the 20 sites of the absolute difference from 50% of the percent of positive differences.

4.1 Error comparison

The RMSE and MAE statistics in Tables 4 and 5 show that the three interpolation methods are very similar in their error characteristics. The 20-station averages indicate that the BA5 method is slightly more accurate (both RMSE and MAE) than the other two methods for Tmax, but the Norton method is slightly superior for Tmin. On a station-by-station basis the model with the lowest error (represented by the number of best performances, NBPs) is almost evenly split for Tmax, but dominated by the Norton method for Tmin.

4.2 Bias comparison

The percent of positive differences shows that all three models are frequently biased in their temperature interpolations, sometimes too high and sometimes too low depending upon location. Taking 40–60% as being reasonable (i.e. ±10% from a nil-bias), most stations (14–15 out of 20 for Tmax and 12–15 out of 20 for Tmin) have a percent of positive differences outside this range, across all three models. Overall, the 20-station average of the absolute difference from 50% of the percent of positive differences shows that the VCS method is slightly less biased for maximum temperature and similarly, the Norton method for minimum temperature. The model with the least bias differs from station to station, with the VCS method having the highest (Tmax) or equal-highest with Norton (Tmin) number of NBPs.

4.3 Poor performance comparison

The last three columns in Tables 4 and 5 compare the percent of absolute differences (absolute value of estimate minus actual) greater than 2°C. This threshold was chosen as an indicator of poor model performance. While what is generally considered a poor estimate is dependent upon the application of the data, for model comparison purposes this threshold will suffice. In general, based on the 20-station average, the BA5 method has slightly fewer large errors the other two methods for Tmax while the Norton method slightly out-performs the others for Tmin (which is consistent with the corresponding RMSE results). Station-by-station, the BA5 model has slightly more NBPs for Tmax while the Norton model is a clear favourite forTmin.

Station ID (agent #)	RMSE			MAE			% of Positive Differences			% of Absolute Differences > 2°C		
	VCS	BA5	N	VCS	BA5	N	VCS	BA5	N	VCS	BA5	N
1129	1.39	1.37	1.38	1.24	1.21	1.23	91.4	91.4	91.5	9.4	8.7	9.0
1434	0.69	0.70	0.69	0.50	0.51	0.50	58.8	63.5	63.6	1.6	1.8	1.6
1550	1.18	1.11	1.34	0.87	0.78	0.79	46.4	41.7	30.4	6.7	5.5	6.2
1645	1.27	1.19	1.31	1.08	1.01	1.14	7.6	7.0	4.4	9.1	6.2	9.2
1841	0.87	0.91	0.86	0.62	0.65	0.61	43.5	35.6	38.8	3.4	3.9	3.1
2005	0.66	0.70	0.69	0.51	0.54	0.54	68.1	76.0	76.1	0.8	0.9	0.9
2103	0.92	0.89	0.91	0.68	0.66	0.66	43.5	53.5	44.7	3.3	2.8	3.3
3014	1.22	0.92	1.07	1.06	0.80	0.96	4.3	6.1	3.6	6.5	1.7	2.8
3232	0.62	0.57	0.53	0.45	0.40	0.38	34.8	44.6	42.8	0.7	0.7	0.4
3549	1.40	1.05	1.04	1.10	0.78	0.77	64.6	60.3	70.8	14.1	6.0	6.3
3925	1.96	2.21	2.33	1.58	1.86	2.00	26.3	13.5	11.2	30.5	41.3	47.2
4241	1.08	1.10	1.19	0.88	0.90	0.99	30.8	22.5	17.8	5.0	5.2	7.2
4310	1.18	1.14	1.21	0.90	0.86	0.94	19.9	19.0	15.4	6.3	6.1	7.2
4458	1.91	1.53	1.52	1.51	1.16	1.15	44.3	54.6	54.9	28.8	16.0	15.6
4566	1.75	1.64	1.63	1.28	1.18	1.17	29.7	36.5	37.4	20.1	16.0	15.6
4764	1.14	1.18	1.16	0.86	0.90	0.89	75.9	78.9	77.8	6.8	7.7	7.4
5280	1.69	1.73	1.74	1.16	1.12	1.12	61.8	51.5	55.7	13.7	13.1	13.3
5577	0.80	0.88	0.94	0.56	0.64	0.71	62.0	24.7	19.0	2.0	2.7	3.4
5823	0.96	0.96	0.95	0.70	0.69	0.68	44.2	46.7	45.4	4.2	4.2	4.1
17029	1.10	1.32	1.60	0.90	1.14	1.44	15.1	6.2	2.1	5.9	9.9	17.5
Average	1.19	1.16	1.20	0.92	0.89	0.93	19.6	21.3	23.9	9.0	8.0	9.1
NBPs	7.5	6	6.5	6.5	6	7.5	11.5	7.5	1	6.5	8	5.5

Table 4: Tmax error statistics at the 20 validation sites for the three interpolation methods. The best performing interpolation method is highlighted for each station and statistic, with the number of best performances (NBPs) presented in the final row.

Station ID (agent #)	RMSE			MAE			% of Positive Differences			% of Absolute Differences > 2°C		
	VCS	BA5	N	VCS	BA5	N	VCS	BA5	N	VCS	BA5	N
1129	1.45	1.51	1.50	1.11	1.16	1.15	30.1	28.4	28.9	16.3	18.2	17.8
1434	2.06	2.04	2.00	1.67	1.64	1.59	5.8	7.0	8.2	37.4	36.5	35.5
1550	1.68	1.70	1.64	1.29	1.30	1.25	41.2	34.6	39.6	20.9	20.9	20.2
1645	1.21	1.24	1.18	0.93	0.93	0.88	42.4	32.3	36.9	8.9	9.7	8.6
1841	1.26	1.22	1.28	0.91	0.88	0.94	35.6	35.6	30.2	10.1	9.3	10.5
2005	1.12	1.14	1.14	0.82	0.83	0.83	52.4	52.4	54.1	6.8	7.2	7.3
2103	1.49	1.48	1.43	1.10	1.09	1.05	38.2	37.3	42.2	14.9	14.5	13.5
3014	1.34	1.20	1.27	1.02	0.90	0.96	68.8	61.3	69.2	12.6	8.8	10.6
3232	1.47	1.49	1.43	1.10	1.13	1.08	74.8	78.7	77.0	16.7	17.4	16.1
3549	1.74	1.60	1.50	1.32	1.20	1.12	46.2	68.4	58.4	22.0	18.1	14.8
3925	1.73	1.77	1.68	1.32	1.43	1.33	45.5	27.9	33.4	21.3	24.0	20.7
4241	1.31	1.49	1.41	1.06	1.22	1.15	29.8	21.2	24.5	11.6	17.2	14.5
4310	1.25	1.23	1.23	0.94	0.92	0.92	51.2	48.7	50.7	10.5	9.6	9.9
4458	2.38	1.94	1.91	1.91	1.43	1.41	79.0	60.2	59.7	40.4	25.2	24.4
4566	3.23	2.92	2.90	2.53	2.24	2.24	77.9	76.5	76.2	52.1	46.8	46.9
4764	1.09	1.09	1.09	0.76	0.76	0.76	54.2	54.0	53.8	6.4	6.4	6.3
5280	1.68	1.72	1.72	1.17	1.20	1.20	61.3	56.3	52.8	15.8	16.9	16.6
5577	1.19	1.14	1.14	0.93	0.82	0.83	69.0	45.7	53.8	7.8	7.5	7.2
5823	2.02	1.99	1.98	1.63	1.60	1.60	11.3	12.0	12.2	31.8	31.0	30.7
17029	1.38	1.21	1.14	1.00	0.95	0.84	52.9	28.3	38.9	13.2	7.4	7.9
Average	1.60	1.56	1.53	1.23	1.18	1.16	15.8	17.4	15.5	18.9	17.6	17.0
NBPs	4.3	3.3	12.3	5.3	4.8	9.8	9	2	9	4	5	11

Table 5: Tmin error statistics at the 20 validation sites for the three interpolation methods. The best performing interpolation method is highlighted for each station and statistic, with the number of best performances (NBPs) presented in the final row

	Latitude (°S)	Longitude (°E)	Elevation (m)	Distance apart (km)
Brewster AWS	44.084	169.429	1650	1.3
P069071	44.075	169.425	1876	

Table 6: Site details of the high elevation temperature recording site and the nearest 0.05°lat/long grid point used in this study.

5. Evaluation using independent high elevation temperature data

The above comparison has shown that the three methods are very similar in terms of the average error and bias over the 20 validation sites. At the three sites with the highest elevation (Stations 1841 (376m), 4458 (387m) and 5280 (427m)) the results were mixed, however it must be noted that none of these sites is particularly high. The next stage of this study compares daily Tmax and Tmin estimates from the three interpolation methods with independent (i.e. not used in any of the interpolated datasets) air temperature data from a high elevation site. The site is Brewster AWS (Automatic Weather Station) located near Lake Ohau (Table 6). The temperature data from this station were kindly provided by Dr. Nicolas Cullen, Geography Department, University of Otago.

Daily Tmax and Tmin for the period 1/6/2004 – 18/11/2009 were extracted from the hourly temperature data provided by the University of Otago for the high elevation automatic weather station at Brewster Glacier. Of the possible 1997 daily values, 145 days were coded as "missing" due to data gaps. The largest data gaps are from 19/8/2004 – 3/10/2004 and from 15/4/2006 – 14/5/2006. The 24-hour period from which the max and min temperatures were extracted was chosen to match the period used for the interpolated data (i.e. the 24-hours up to 9am each day).

Prior to any comparisons with the observational data, the interpolated daily temperatures at the nearest grid point to the Brewster AWS (grid point ID = P069071) were adjusted to take into account the elevation difference between itself and the AWS site (see Table 6). A fixed lapse rate of 5 °C/km was used for these adjustments.

5.1 Maximum daily temperature comparison

Figure 2 shows a time series plot of the daily Tmax difference (estimates minus observations) over the nearly five and a half year period for which data were available. from the AWS. The thick lines are the 30-day moving averages of the difference between the VCS (blue), BA5 (red) and Norton (green) Tmax estimates and the Brewster AWS observations. It can be seen that the VCS method clearly overestimates Tmax more than the BA5 method, particularly during the months from September – May (spring, summer and autumn; peaking in late summer). The differences for the VCS and BA5 methods are consistently least in the winter months. The amplitude of the "seasonality" of the differences is approximately 8 °C for the VCS method and approximately 5 °C for the BA5 method. Therefore, the BA5 method is clearly an improvement over the VCS method for estimating maximum temperature at this location, yet it still overestimates the observed Tmax, particularly in summer.

Using the Norton method it is apparent that the bias of the difference has been reduced, compared with both the VCS and BA5 methods, though there is still a mostly positive difference. The "seasonality" of the differences is similar to the BA5 method. Table 7 shows a statistical comparison between the three methods. This table clearly shows that the Norton method out-performs both the VCS and BA5 methods for estimating maximum daily temperature at this site. The Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are much lower and there is a near-doubling in the number of days when the difference is plus or minus 2°C. The bias is reduced, but still positive.

5.2 Minimum daily temperature comparison

Figure 3 shows a time series plot of daily Tmin difference (estimates minus observations) and Table 8 presents a summary of the errors and biases. The Norton method has a lower standard deviation (SD), MAE and RMSE, and a higher percentage of days with the difference within 2 °C, compared

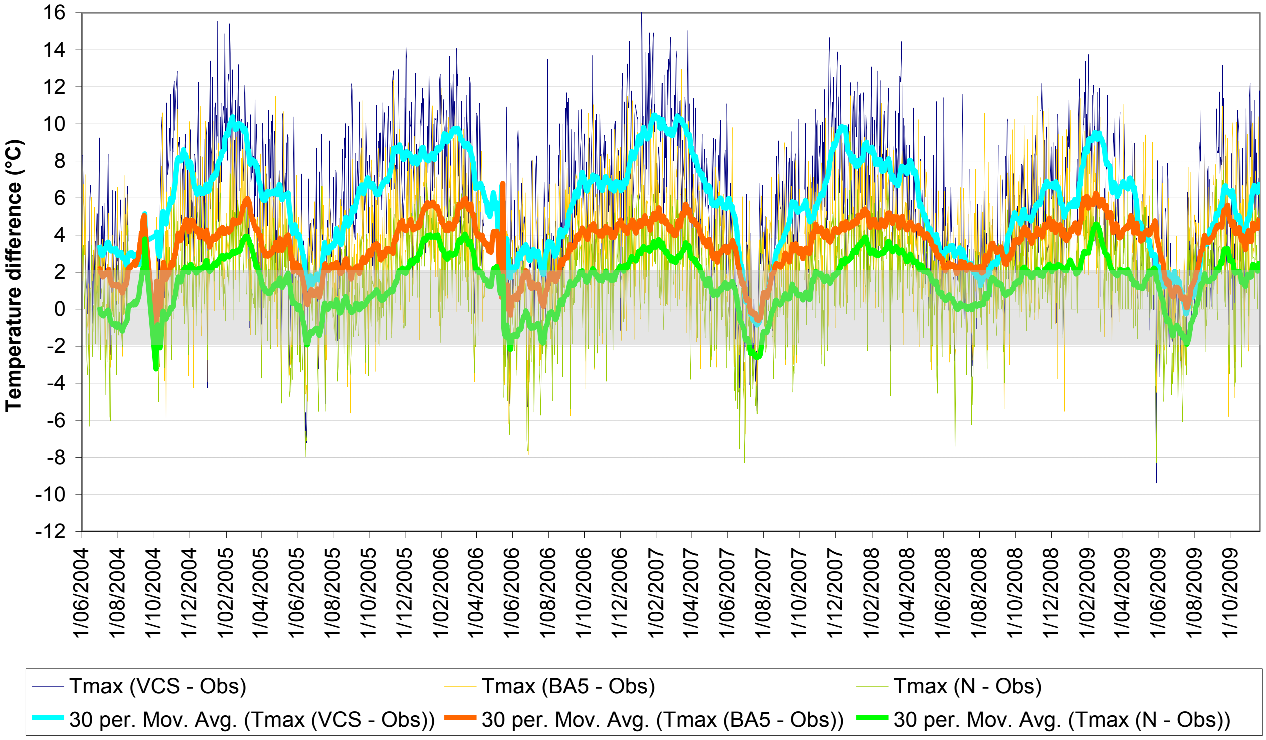

Figure 2: Plot of the maximum daily temperature differences at Brewster Glacier for the period 1/6/2004 – 18/11/2009 comparing observations with estimates from three interpolation methods: VCS, BA5 and Norton (N). The thin lines are the daily differences; the thick lines are 30-day moving averages of the differences; and the shaded area is ±2 °C.

	VCS	BA5	N
Mean diff	5.74	3.49	1.52
Median diff	5.73	3.5	1.77
SD of diffs	3.98	3.33	2.67
MAE	6.04	4.03	2.56
RMSE	6.98	4.82	3.07
% positive diffs (bias)	91.85	85.69	74.57
% diffs within ± 2 °C	14.58	26.62	42.12

Table 7: Statistics describing the differences between maximum daily temperature observations at Brewster Glacier for the period 1/6/2004 – 18/11/2009 and estimates from three interpolation methods: VCS, BA5 and Norton (N). The best performing interpolation method is highlighted for each statistic

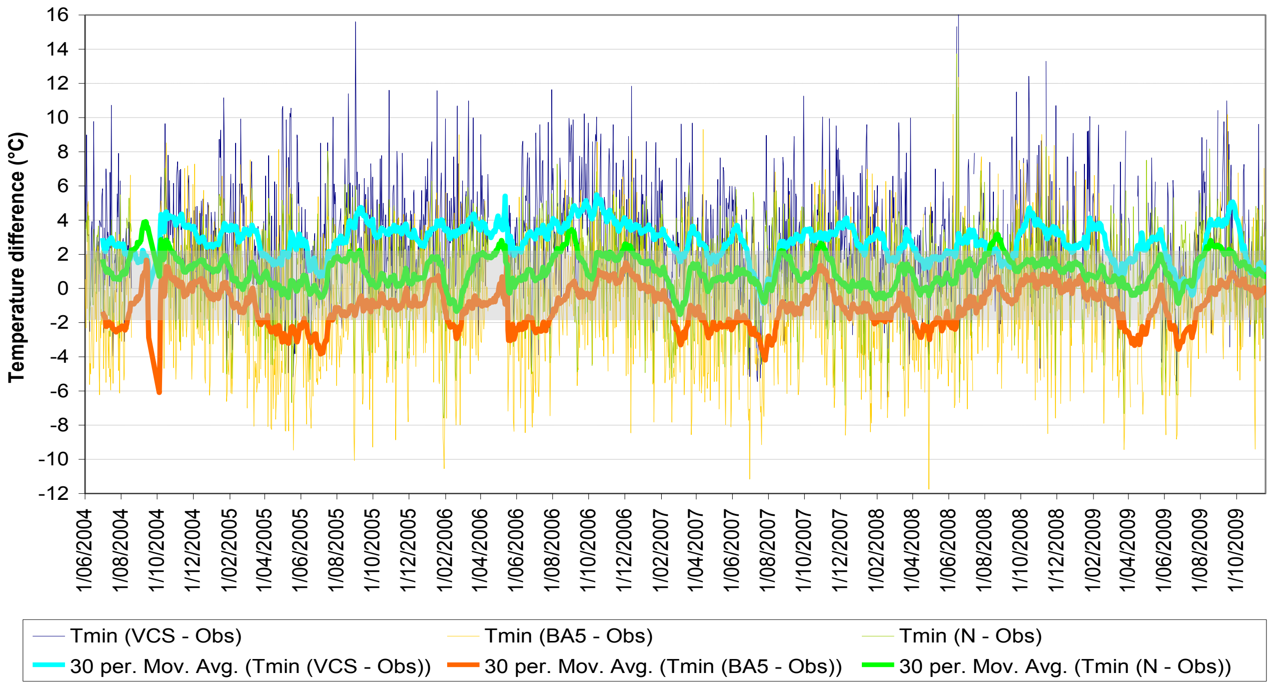

Figure 3: Plot of the minimum daily temperature differences at Brewster Glacier for the period 1/6/2004 – 18/11/2009 comparing observations with estimates from three interpolation methods: VCS, BA5 and Norton (N). The thin lines are the daily differences; the thick lines are 30-day moving averages of the differences; and the shaded area is ±2 °C.

	VCS	BA5	N
Mean diff	2.78	-1.03	1.03
Median diff	2.55	-1.13	1.34
SD of diffs	3.06	3.49	2.59
MAE	3.28	2.93	2.31
RMSE	4.13	3.64	2.79
% positive diffs (bias)	83.69	37.37	67.44
% diffs within ± 2 °C	37.58	41.41	47.79

Table 8: Statistics describing the differences between minimum daily temperature observations at Brewster Glacier for the period 1/6/2004 – 18/11/2009 and estimates from three interpolation methods: VCS, BA5 and Norton (N). The best performing interpolation method is highlighted for each statistic.

with the other two methods. However, the BA5 method for Tmin has a mean and median difference nearest zero (the absolute value of the mean difference is actually equal for BA5 and N), and has the least bias. Figure 3 shows that there is noticeably less seasonality in the Tmin differences, compared with the Tmax differences.

Note: temperature data from another high elevation site (Glenmary AWS, lat = 43.997°S, lon = 169.883°E, height = 2032m) were also compared with the interpolated Tmax and Tmin values from the station's nearest grid point, using the three methods. Results were highly comparable to the above analysis, although the data period was much shorter (22/3/2005 – 2/10/2007 with 431 days coded as "missing" due to data gaps).6.

6.Summary and Conclusions

All three interpolation methods (VCS, BA5 and Norton) perform very similarly at estimating daily Tmax and Tmin at the 20 validation sites selected. On average, the BA5 method has *slightly* lower average and extreme errors and the VCS method has *slightly* lower bias for Tmax, while the Norton method clearly has the lowest errors and bias for Tmin.

However, from a comparison with independent temperature data from Brewster Glacier, the Norton method is clearly superior at estimating Tmax for this site, while the BA5 and Norton methods have similar skill at estimating Tmin. It should be noted that the Norton method consistently outperformed the BA5 and VCS methods when comparing the SD, MAE and RMSE This suggests that the Norton method is more able to capture day-to-day variations in the temperature Based on these results, the Norton method is the preferred temperature interpolation method for this high elevation site.

We now more broadly consider the three methods presented in this paper, and their applicability for temperature estimation in areas of high mountainous terrain in New Zealand. Each of the methods produce temperature estimates using point-based temperature data and a linear dependence on elevation. This presents two main challenges to deriving accurate air temperature estimates for New Zealand's mountainous terrain: 1) temperature observation locations are heavily biased towards lower elevations, so mountainous terrain is poorly represented by the spatial distribution of New Zealand's climate station network; and 2) the correlation between temperature and elevation, i.e. the lapse rate, varies in time and space (Dodson and Marks, 1997).

For the operational VCS method, it seems to us that using a model which calculates an optimal smoothing parameter (and hence lapse rate) based on all the available data on a given day is appropriate for addressing point 2 above, but is very likely to be problematic based on point 1. This is because the determination of the daily lapse rate is dominated by the temperature/altitude relationship at low elevation, mostly coastal, sites where there is little altitudinal gradient. Hence, the altitudinal "forcing" of the relationship is weak and there is the potential for the calculated lapse rate to be quite different from that present in high elevation areas.

As demonstrated in this paper and in previous studies, the use of a prescribed fixed lapse rate (e.g. the BA5 method) has yielded improved estimations of temperature in New Zealand's mountainous terrain. This is probably because the lapse rate calculation is not biased by the location and density of the observation sites (point 1, above). However, the use of a fixed lapse rate can be problematic as it may be an unreliable representation of the environmental (actual) lapse rate at a given location on a given day (point 2 above; Minder et al., 2010).

Seasonally varying lapse rates have been demonstrated to improve temperature estimates (Rolland, 2003; Huang et al., 2008), and in addition, studies have also shown that

Tmax and Tmin lapse rates can be quite different (Bolstad et al., 1998; Rolland, 2003; Blandford et al., 2008). This is most likely also the case for New Zealand as demonstrated by the results presented in this paper, with the lapse rates provided by Norton (1985), which were derived from a reasonably large dataset with good spatial coverage, emerging as the most appropriate to use in New Zealand.

Of course, given that quite large temperature variations can result from fine-scale variation in insolation (related to slope and aspect), wind movement patterns, vegetation cover and surface albedo, we fully acknowledge that there is considerable complexity to interpolating temperatures in New Zealand's mountainous terrain. As such, this is a topic of which further research could provide valuable insight towards future improvements of temperature interpolation procedures for New Zealand.

As mentioned in the introduction, there are multiple users and uses of the VCS dataset, so the authors are reluctant to make significant changes to the operational interpolation methodology. Such changes will require regenerating the entire VCS data record (in this case, for Tmax and Tmin), which may lead to issues where users of the data have done extensive modelling work based on the previous dataset. Still, based on the results of this comparison study (and the other high elevation studies mentioned above), it appears that changing the operational interpolation methodology for daily Tmax and Tmin to the Norton method will not result in major differences to low elevation areas while at the same time significantly improve the data accuracy at higher elevations. It is therefore recommended that this change is made and clearly noted on the Cliflo metadata webpage associated with the VCS data.

Acknowledgements

The authors would like to thank Brian Anderson, Victoria University Wellington; Nicolas Cullen, University of Otago; and James Sturman, NIWA for their assistance with this study. This work has been supported by NIWA as part of its Government-funded, core research.

References

Anderson, B.M., Lawson, W.J. Owens, I.F. and Goodsell, B. 2006. Past and Future Mass Balance of Ka Roimata o Hine Hukatere (Franz Josef Glacier). *Journal of Glaciology*,52: 597–607.

Anderson, B. and Mackintosh, A. 2012. Controls on Mass Balance Sensitivity of Maritime Glaciers in the Southern Alps, New Zealand: The Role of Debris Cover. *Journal of Geophysical Research: Earth Surface*, 117 (F1): n/a–n/a. doi:10.1029/2011JF002064.

Blandford, T.R., Humes, K.S., Harshburger, B.J., Moore, B.C. and Walden, V.P. 2008. Seasonal and Synoptic variations in Near-Surface Air Temperature Lapse Rates in a Mountainous Basin. *Journal of Applied Meteorology and Climatology*, 47: 249-261.

Bolstad, P.V., Swift, L., Collins, F. and Règniére, J. 1998. Measured and predicted air temperatures at basin to regional scales in the southern Appalachian mountains. *Agricultural and Forest Meteorology*, 91: 161-176.

Dodson, R. and Marks, D. 1997. Daily air temperature interpolated at high spatial resolution over a large mountainous region. *Climate Research*, 8: 1-20.

Elliott, S. and Harper, S. 2010. Catchment models for Campylobacter: Detailed Dynamic Model and Associated Risk Model. *NIWA Client Report HAM2010-069*.

Huang, S., Rich, P.M., Crabtree, R.L., Potter, C.S. and Fu, P. 2008. Modelling monthly near-surface air temperature from solar radiation and lapse rate: application over complex terrain in Yellowstone National Park. *Physical Geography*, 29: 158-178.

Hutchinson, M.F. 2013. *ANUSPLIN*, http://cres.anu.edu.au/outputs/anusplin.php.

Kirschbaum, M., Watt, M., Tait, A., Ausseil, A-G., Palmer, D. 2012. Future productivity of Pinus radiata in New Zealand under expected climatic changes. *Global Change Biology*, 18, 1342–1356, doi: 10.1111/j.1365-2486.2011.02625.x.

Köppen, W. 1931. *Grundriss der Klimakunde*. Berlin: De Gruyer.

Matheron, G. 1981. Splines and kriging: their formal equivalence. *Syracuse University Geology Contribution* **8**, 77–95.

McBride, G., Tait, A. and Slaney, D. 2014. Projected changes in reported campylobacteriosis and cryptosporidiosis rates as a function of climate change: a New Zealand study. *Stochastic Environmental Research and Risk Assessment*, doi: 10.1007/s00477-014-0920-5.

MfE, 2008. Climate change effects and impacts assessment. A Guidance Manual for Local Government in New Zealand. 2nd Edition. Prepared by Mullan, B; Wratt, D; Dean, S; Hollis, M. (NIWA); Allan, S; Williams, T. (MWH NZ Ltd), and Kenny, G. (Earthwise Consulting Ltd), in consultation with Ministry for the Environment. NIWA Client Report WLG2007/62.

Minder, J.R., Mote, P.W. and Lundquist, J.D. 2010. Surface temperature lapse rates over complex terrain: Lessons from the Cascade Mountains. *Journal of Geophysical Research*, 115: doi: 10.1029/2009JD013493.

Norton, D. 1985. A multivariate technique for estimating New Zealand temperature normal. *Weather and Climate*, 5, 64-74.

Philips, D.L., Dolph, J. and Marks, D. 1992. A comparison of geostatistical procedures for spatial analysis of precipitation in mountainous terrain. *Agricultural and Forest Meteorology* **58**, 119–141.

Renwick, J., Mullan, B. and Porteous, A. 2009. Statistical Downscaling of New Zealand Climate. *Weather and Climate*, 29, (2009), 24-44.

Rolland, C. 2003. Spatial and Seasonal Variations of Air Temperature Lapse Rates in Alpine Regions. Journal of Climate, 16: 1032-1046.Clark M, Hreinsson E, Martinez G, Tait A, Slater A, Hendrikx J, Owens I, Gupta H, Schmidt J, Woods R, 2009. Simulations of seasonal snow for the South Island, New Zealand. *Journal of Hydrology (NZ)*, 48(2), 41-58.

Romera, A., Beukes, P., Clark, C., Clark, D., Levy, H. and Tait A. 2010. Use of a pasture growth model to estimate herbage mass at a paddock scale and assist management on dairy farms. *Computers and Electronics in Agriculture*, 74(1), 66–72.

Tait, A.B., Henderson, R.D., Turner, R.W., and Zheng, X., 2006. Thin plate smoothing spline interpolation of daily rainfall for New Zealand using a climatological rainfall surface.

International Journal of Climatology **26**, 2097–2115.

Tait, A.B. and Woods, R.A., 2007. Spatial interpolation of daily potential evapotranspiration for New Zealand using a spline model. *Journal of Hydrometeorology* **8**, 430–438.

Tait, A.B. 2008. Future projections of growing degree days and frost in New Zealand and some implications for grape growing. *Weather and Climate*, 28, 17–36.

Tait, A., Baisden, T., Wratt, D., Mullan, B. and Stroombergen, A. 2008. An initial assessment of the potential effects of climate change on New Zealand agriculture. *New Zealand Science Review*, 65(3), 50–56.

Tait, A. and Liley, B. 2009. Interpolation of daily solar radiation for New Zealand using a satellite-derived cloud cover surface. *Weather and Climate*, 29, 70–88.

Tait, A., Sturman, J. and Clark, M. 2012. An assessment of the accuracy of interpolated daily rainfall for New Zealand. *Journal of Hydrology (NZ)*, 51(1), 25-44.

Ummenhofer, C., Sen Gupta, A. and England, M. 2009. Causes of Late Twentieth-Century Trends in New Zealand Precipitation. *Journal of Climate*, 22(1), 3-19.

Woods, R., Hendrikx, J., Henderson, R. and Tait, A. 2006. Estimating mean flow of New Zealand rivers. *Journal of Hydrology (NZ)*, 45 (2), pp 95-110.

Submitted to *Weather and Climate* November 2013, Revised August 2014