Available online at www.sciencedirect.com
ScienceDirect
Agricultural and Forest Meteorology 139 (2006) 224–236
AGRICULTURAL AND FOREST METEOROLOGY
www.elsevier.com/locate/agrformet

Comparison of approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density

K. Stahl [a,*], R.D. Moore [a,b], J.A. Floyer [c], M.G. Asplin [a], I.G. McKendry [a]

[a] Department of Geography, The University of British Columbia, 1984 West Mall, Vancouver, BC, Canada V6T 1Z2
[b] Department of Forest Resources Management, The University of British Columbia, 2424 Main Mall, Vancouver, BC, Canada V6T 1Z4
[c] Department of Geology & Geophysics, The University of Calgary, Calgary, Alta., Canada T2N 1N4

Received 19 December 2005; accepted 12 July 2006

Abstract

This study compared 12 variations of regression-based and weighted-average approaches for interpolating daily maximum and minimum temperatures over British Columbia, Canada, a domain with complex topography and highly variable density and elevational distribution of climate stations. The approaches include simple extrapolation with elevation from the nearest climate station; nine variations of weighted-average methods employing three approaches to calculate lapse rates, two methods for station selection and three approaches for weight calculation; multiple linear regression using station coordinates as predictor variables; a method combining multiple regression and weighted averaging. Cross-validation for years with different densities and elevational distribution of climate stations showed varied mean prediction errors, which also depended on elevation and month. Methods that compute local lapse rates from the control points performed better for years for which there were a greater number of higher-elevation stations, which allowed for better estimation of lapse rates. The methods that involved specified lapse rates all performed similarly, indicating that the method for selecting control stations and for calculating weights have less effect on predictive accuracy than the method for accounting for elevation.

Keywords: Air temperature; Spatial interpolation; Lapse rates; Kriging; Trend surface

1. Introduction

Air temperature is one of the most important input variables in environmental models, in relation to both biological processes, such as phenological development of forest pests (e.g., Régnière, 1996; Logan and Powell, 2001) and physical processes, such as snow melt (e.g., Hock, 2003). The continuing development and application of spatially explicit environmental models has generated a need for methods for predicting spatial fields of weather variables, including air temperature, from irregularly spaced station data. While many studies have focused on the interpolation of climate normals (e.g., Holdaway, 1996; Nalder and Wein, 1998; Daly et al., 2002; Hamann and Wang, 2005), several have tested methods for spatial interpolation of daily maximum and minimum temperature (e.g., Dodson and Marks, 1997; Thornton et al., 1997; Bolstad et al., 1998; Couralt and Monestiez, 1999; Shen et al., 2001; Xia et al., 2001; Jarvis and Stuart, 2001; Hasenauer et al., 2003; Garen and Marks, 2005). In mountainous regions, the spatial variability of climate variables is modified by

* Corresponding author.
E-mail address: kstahl@geog.ubc.ca (K. Stahl).

0168-1923/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.agrformet.2006.07.004

physiographic features, the importance of which depends on the scale of interest (Daly, 2006): at larger scales, elevation is the most important factor; for medium to smaller scales additional terrain-related factors become relevant. For smaller regional scales with known physical influences, such as air flow barriers, water or ice bodies and urban areas, progress has also been made with the modelling at higher spatial and temporal resolutions through the inclusion of additional relevant guiding variables, such as incident solar radiation (Régnière, 1996; Choi et al., 2003; Chung and Yun, 2004). A general problem in mountainous regions is the lack of high-elevation observations to derive relations with elevation. Regional studies have supplemented high-elevation observations with data from other sources, such as upper-level soundings (Garen and Marks, 2005); at larger scales but high temporal resolution, first applications have started to test the inclusion of data from of meso-scale weather models (e.g., Liston and Elder, 2006). In this paper, we concentrate on the comparison of methods applicable over large, data sparse mountainous regions.

Schemes for spatial interpolation of air temperature vary in relation to three aspects: (1) the approach to adjusting for elevation, (2) the model used for characterising the spatial variation of air temperature and (3) the method of choosing prediction points. The simplest approach is to choose the nearest station and adjust for elevation according to some assumed lapse rate. Where more than one station is used in the prediction, a model is required to determine how to interpolate from them. Some models fit a surface to the data, including multiple regressions that use location variables as predictors (e.g., Bolstad et al., 1998; Xia et al., 2001; Jarvis and Stuart, 2001) and thin plate splines (Xia et al., 2001). Others can be framed as weighted-average approaches, most of which apply an equation similar to the following:

$$T_p = \frac{\sum_{i=1}^{n} w_i [T_i + f(h_p - h_i)]}{\sum_{i=1}^{n} w_i} \quad (1)$$

where T_p is the predicted temperature, T_i the observed temperature at control point i and w_i is the weight associated with control point i, and $f(h_p - h_i)$ is a function (usually linear) of the difference in elevation (h) between the prediction point and the control point.

Interpolation weights have been estimated using a range of approaches, including inverse-distance measures (Dodson and Marks, 1997; Shen et al., 2001), a truncated Gaussian filter (Thornton et al., 1997) and geostatistical methods based on kriging (Bolstad et al., 1998; Couralt and Monestiez, 1999; Garen and Marks, 2005). Jarvis and Stuart (2001) fitted a multiple regression model to account for effects of topographic and land-cover influences, then interpolated the residuals. Control points are typically chosen to lie within a defined cut-off distance, which has been specified a priori (e.g., Shen et al., 2001) or by using an iterative scheme based on station density (Thornton et al., 1997). A further variation for choosing control points is whether or not the chosen points are stratified by angular position relative to the prediction point (e.g., by taking the six closest points in each of four quadrants).

Adjustments for elevation have been made using specified lapse rates, typically a value, such as 6.0 or 6.5 °C/km (e.g., Running et al., 1987; Dodson and Marks, 1997), sometimes monthly varying values (e.g., Liston and Elder, 2006). To address the significant variations in lapse rates under different meteorological conditions and in different seasons, Couralt and Monestiez (1999) estimated daily lapse rates by regression for a region in southern France, then sorted these by synoptic-scale circulation patterns. However, they found that sorting lapse rates by synoptic type had only a marginal effect on prediction accuracy. Alternatively, lapse rates can be estimated from the observed station data for each time interval, for example, by using simple linear regression with elevation (Garen and Marks, 2005). Multiple regression and thin-plate spline models can explicitly include elevation effects within the fitted model (Bolstad et al., 1998; Xia et al., 2001). Thornton et al. (1997) used a weighted-linear regression approach to estimate lapse rates based on temperature differences between pairs of stations, with the weights based on distances between the prediction point and control points. Some studies adjusted temperatures to sea level using a specified lapse rate prior to the interpolation, then adjusted the predicted temperature to the elevation of the prediction point (Dodson and Marks, 1997; Couralt and Monestiez, 1999). Equivalently, Thornton et al. (1997) adjusted the observed data to the elevation of the prediction point prior to interpolation. Garen and Marks (2005) applied kriging to the residuals from a linear regression with elevation.

While previous studies have made significant progress in developing and testing methods for spatial interpolation of daily air temperature, there has been relatively limited attention to comparing them. Dodson and Marks (1997) and Couralt and Monestiez (1999) compared different methods for elevation adjustment, but did not vary the interpolation approach. Shen et al. (2001) compared different equations for calculating interpolation weights, but worked in an area of relatively low relief such that elevation adjustments

were not applied. Bolstad et al. (1998) compared three fundamentally different approaches (regional lapse, regional regression and kriging), while Xia et al. (2001) compared six methods, ranging from a simple nearest-neighbour with elevation adjustment to methods based on thin plate splines and kriging. One of the most comprehensive intercomparisons was reported by Jarvis and Stuart (2001), who modelled daily temperatures over England and Wales using trend surface analysis, thin-plate splines and interpolation based on both inverse-distance weighting and kriging.

In the context of weighted-average approaches, no studies appear to have compared approaches that combine different variations of the method for choosing prediction points, the method for computing weights and the method of adjusting for elevation. Furthermore, many studies have been conducted in regions with a relatively uniform station density and a reasonable number of high-elevation stations (e.g., Dodson and Marks, 1997; Thornton et al., 1997; Couralt and Monestiez, 1999; Hasenauer et al., 2003; Garen and Marks, 2005). However, many extensive mountainous regions have sparse station coverage, especially at higher elevations, and the ability to estimate daily weather data over such regions is required for the application of environmental simulation models, particularly in relation to pest outbreaks, such as the mountain pine beetle infestation that is currently spreading throughout western Canada.

The objective of this study is to test algorithms for spatial interpolation of daily maximum and minimum air temperature in British Columbia, Canada, an extensive region with complex topography and highly varying densities and elevational distribution of climate stations. Reflecting the vast and in part remote area, geostatistical interpolation algorithms based on linear regression and weighted averaging using elevation as the only guiding variable were used. In particular, attention is paid to the effects of station density and elevational distribution on the relative accuracy of the 12 methods.

2. Data sources

Station observations of daily maximum and minimum temperatures (T_{max} and T_{min}) were obtained from three different sources: Environment Canada's climate network (EC), The British Columbia Ministry of Forests Fire Weather Station Network (FW) and the British Columbia Ministry of Water, Land and Air Protection's Snow Pillow sites (SP) (Table 1). Stations are distributed unevenly across BC, with lower density towards the north (Fig. 1a). Many series are short: only 43 Environment Canada stations have a continuous record of temperature and precipitation data (with less than 5 years missing) covering the period from 1950 to 2003. The Fire Weather Network and the Snow Pillow stations were established relatively recently. At the same time, the EC network was reduced. This reduction, together with the gaps and seasonally limited observations of the two smaller datasets, results in an effective decrease of climate observations in recent years, despite the increasing number of stations (Fig. 1b). A further problem is that most stations are located at low elevations (Fig. 1c). The few high-elevation stations are mainly the Snow Pillow stations. Most Fire Weather Network stations and some EC stations only operate seasonally.

3. Prediction algorithms

Twelve variations of models based on linear regression and weighted-average interpolation are applied. They include simple extrapolation with elevation from the nearest climate station (NN), nine variations of weighted-average approaches (employing three approaches to calculate lapse rates, three approaches for weight calculation and two methods

Table 1
Sources of surface climate data

Source	Data set	No. of stations	Time period	Comments
Environment Canada	Canadian Climate CD West	1601 (BC) 21 (YT, NWT) 230 (AB)	Variable (longest since 1880)	
Ministry of Forests Protection Program	Fire Weather Station Network	283 (310)	Since 1990s	Summer data only (winter not maintained, no T below −20 °C)
Ministry of Water, Land and Air Protection	Archived Automatic Snow Pillow (ASP)	68	Since late 1980s or early 1990s	Station elevations are 1700 ± 1000 m

BC, British Columbia; YT, Yukon Territory; NWT, Northwest Territory; AB, Alberta.

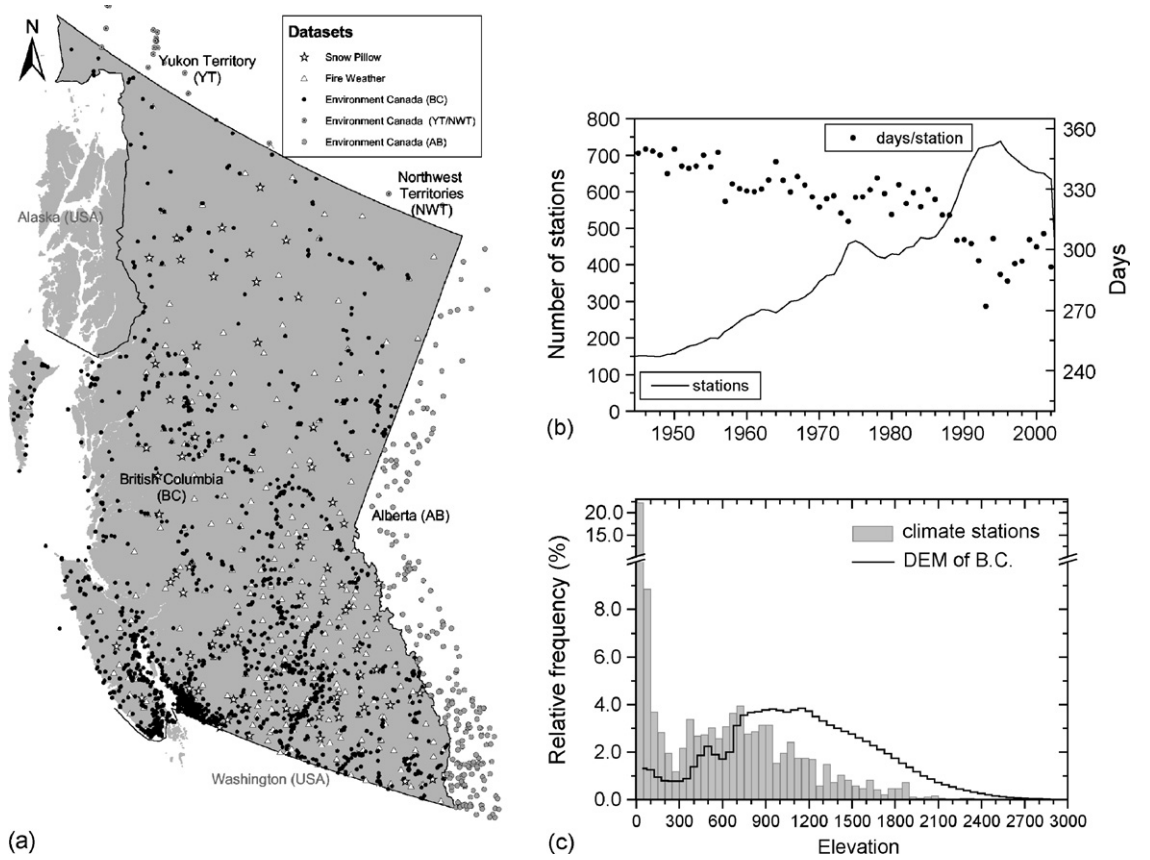

Fig. 1. Location of climate stations (a) and distribution of records with time (b) and elevation (c).

for station selection), linear regression (MLR) using station coordinates as predictor variables and a method combining multiple regression and weighted averaging. Table 2 gives an overview of the 12 models. The acronyms for the nine weighted-average approaches are derived from the algorithms with SL for specified lapse rate and LWR for lapse rate by weighted regression, ID for inverse distance weighting (IDS for squared

Table 2
Summary of model characteristics

Model number	Model abbreviation	Model	Lapse rate	Weight calculation	Station search
1	NN	n/a	Specified	n/a	Nearest neighbour
2	LWR-G	Weighted average	Computed by weighted regression	Gaussian Filter	Density-based
3	CLWR-G	Weighted average	Combination of weighted regression and specified	Gaussian Filter	Density-based
4	SLR-G	Weighted average	Specified	Gaussian filter	Density-based
5	SLR-K	Weighted average	Specified	Ordinary kriging	Density-based
6	SLR-KQ	Weighted average	Specified	Ordinary kriging	Quadrant
7	SLR-ID	Weighted average	Specified	Inverse distance	Density-based
8	SLR-IDQ	Weighted average	Specified	Inverse distance	Quadrant
9	SLR-IDS	Weighted average	Specified	Inverse distance squared	Density-based
10	SLR-IDSQ	Weighted average	Specified	Inverse distance squared	Quadrant
11	MLR	Multiple linear regression	Computed by regression	n/a	Density-based
12	GIDS	Multiple regression + weighted average	Computed by regression	Inverse distance squared	Density-based

weights), K for kriging, G for Gaussian-filter-based weighting. Note that, although we refer to lapse rates (which are positive for temperatures decreasing with elevation), we represent the temperature adjustments using temperature gradients (negative for temperatures decreasing with elevation) in the descriptions below.

3.1. Model 1: nearest-neighbour with elevation adjustment

In the nearest neighbour interpolation method (NN), the station closest to the prediction point is used to specify T_0, the nearest known temperature (°C). The temperature at the prediction point, T_p (°C), is estimated as:

$$T_p = T_0 + \lambda(h_p - h_0) \tag{2}$$

where λ is a specified temperature gradient (°C m^{-1}), h_p the elevation of the prediction point (m) and h_0 is the elevation of the predictor station (m).

3.2. Models 2–10: weighted-average approaches

These models involve different combinations of methods for: (a) adjusting for elevation, (b) calculating weights and (c) choosing prediction points. The combinations are summarized in Table 2.

Model 2 is based on descriptions of the DAYMET algorithms in Thornton et al. (1997) and Hasenauer et al. (2003). Because our implementation is likely not identical to the DAYMET model, we refer to it as the "lapse rate by weighted regression with Gaussian filter" (LWR-G) approach. The LWR-G method uses an iterative station density algorithm to determine a local set of predictor stations for each given prediction point, each with a weight that decreases with distance from the prediction point. Temperature gradients are calculated by summing the weighted contribution of the temperature gradient computed from every possible predictor station pair. The interpolation weights are assigned to all stations in the database by using a truncated Gaussian filter. From Thornton et al. (1997), this is given by:

$$w(r) = \begin{cases} 0; & r > R_p \\ \exp\left[-\left(\dfrac{r}{R_p}\right)^2 \alpha\right] - \mathrm{e}^{-\alpha}; & r \le R_p \end{cases} \tag{3}$$

where $w(r)$ is the weight associated with the radial distance r from the prediction point p, R_p the truncation distance from p and α is a dimensionless shape parameter. The truncation distance R_p is allowed to vary so that it may be reduced in areas of high station density and increased in areas of low station density. This is done by specifying a desired number of observations N, and using an iterative scheme to vary R_p such that the actual number of observations used in the interpolation converges towards this value of N. Following the truncation distance adjustment, a final set of interpolation weights is calculated using the new value of R_p; these weights are then used in the temperature interpolations as described below.

In the LWR-G model, T_p is interpolated using information from all stations with non-zero interpolation weights. A weighted least-squares regression is used to determine the local temperature gradient. Each unique pair of stations is assigned a regression weight equal to the product of the interpolation weights of the two stations in the pair. The regression model takes the form:

$$(T_1 - T_2) = \beta_0 + \beta_1(h_1 - h_2) + \varepsilon \tag{4}$$

where subscripts 1 and 2 refer to the two stations in the unique pair, β_0 and β_1 are coefficients to be estimated by weighted regression and ε is the prediction error. Once the coefficients are calculated, T_p is computed as:

$$T_p = \frac{\sum_{i=1}^{n} w_i[T_i + b_0 + b_1(h_p - h_i)]}{\sum_{i=1}^{n} w_i} \tag{5}$$

where b_0 and b_1 are the estimated regression coefficients and h_p is the elevation of the prediction point and the subscript i refers to the predictor stations.

Model 3 (CLWR-G) is a variation of the LWR-G model, motivated by the fact that in regions of mountainous terrain, there will often be prediction points that are either above or below the elevation range of the local weather station network. In order to prevent extrapolations from predicting unrealistic temperatures at high (or low) elevations, a third method, the constrained LWR option, computes the temperature gradient using a weighted average of the gradient computed by the LWR approach (b_{1D}) and a specified gradient (b_{1S}) for points that lie outside the elevation range of the control points. This option computes the gradient as

$$b_1 = w_D b_{1D} + w_S b_{1S} \tag{6}$$

where the weights are computed as

$$w_D = \begin{cases} (h_{\max} - h_{\min})/(h_p - h_{\min}), & h_p > h_{\max} \\ (h_{\max} - h_{\min})/(h_{\max} - h_p), & h_p < h_{\min} \end{cases} \tag{7}$$

and

$$w_S = 1 - w_D \tag{8}$$

where h_{max} and h_{min} are the maximum and minimum elevations of the predictor stations and h_p is the elevation of the prediction point.

Model 4 (SLR-G) incorporates the same method for selecting predictor points as used in models 2 and 3. However, observed temperatures are first adjusted to sea level using the specified monthly lapse rates prior to application of Eq. (1), with the weights computed using Eq. (3). The predicted temperature is then adjusted to the elevation of the prediction point using the specified lapse rate.

In models 5 and 6 (SLR-K and SLR-KQ), temperatures are first adjusted to sea level using specified monthly lapse rates prior to application of Eq. (1). The weights are computed using ordinary kriging. Following Garen and Marks (2005), a linear variogram is used to avoid the problems in determining the optimal semivariogram model for each prediction point and each day. In a small number of cases, the kriging solution yields negative weights; these were dealt with using the approach described by Deutsch (1996). In model 5 (SLR-K), control points are selected using the density-based approach used in models 2 and 3. In model 6 (SLR-KQ), only up to six of the closest points within each of four quadrants surrounding the prediction point are used, subject to the constraint that they lie within a specified truncation distance. If no neighbouring points fell within the truncation distance, only the nearest neighbour was used, following Shen et al. (2001).

Models 7 and 8 (SLR-ID and SLR-IDQ) are similar to models 5 and 6, except that the weights are computed using the following equation, with $c = 1$:

$$w_i = \frac{(1/d_i)^c}{\sum_{i=1}^{n} (1/d_i)^c} \tag{9}$$

where d_i is the distance between the prediction point and control point i and c is an exponent. Models 9 and 10 (SLR-IDS and SLR-IDSQ) are identical to models 7 and 8, except that the weights are computed using inverse-distance squared, that is, with $c = 2$ in Eq. (8).

3.3. Model 11: multiple linear regression approach

Model 11 (MLR) involves the fitting of a multiple linear regression using longitude (Long), latitude (Lat) and elevation (h) as predictor variables:

$$\hat{T} = b_0 + b_1\text{Long} + b_2\text{Lat} + b_3 h \tag{10}$$

where $\hat{T}$ is the fitted temperature, b_0 the intercept and b_1–b_3 are slope coefficients. Stations are selected using the density-based algorithm.

3.4. Model 12: gradient-plus-inverse-distance-squared (GIDS)

The GIDS model combines the multiple regression and inverse-distance-squared approaches (Nalder and Wein, 1998). It first involves fitting the multiple regression model Eq. (10), then using the coefficients b_1–b_3 to adjust the temperature at each control station to the location of the prediction point. These adjusted temperatures are then averaged using inverse-distance-squared weights to calculate the predicted temperature. The model can be expressed as follows:

$$T_p = \left[\sum_{i=1}^{n}\left(\frac{1}{d_i}\right)^2\right]^{-1} \times \sum_{i=1}^{n}\left\{[T_i + b_1(\text{Long}_p - \text{Long}_i) + b_2(\text{Lat}_p - \text{Lat}_i) + b_3(h_p - h_i)]\left(\frac{1}{d_i}\right)^2\right\} \tag{11}$$

where the subscripts p and i refer to the prediction and control points, respectively.

4. Model application

4.1. Estimation of specified lapse rates

Specified lapse rates were estimated on a monthly basis by identifying pairs of low-elevation/high-elevation stations that were reasonably close (within a few tens of km) and computing vertical temperature gradients from them. The model lapse rates were then chosen to follow the pattern shared by the majority of the stations. Fig. 2 shows the average monthly lapse rates and their standard deviations for T_{max} and T_{min} for a selection of six station pairs located on the west coast, in the Coast Mountains, on the Interior Plateau and in the Columbia Mountains in eastern BC. The figure also shows the representative monthly lapse rates chosen for use in the interpolation models. Throughout the year, gradients for T_{max} (0.2 to −9 °C/1000 m) vary more than those for T_{min} (−0.9 to −6 °C/1000 m). Gradients for T_{max} are larger and more stable in summer than in winter. The T_{max} gradient for January for one of the pairs shown is positive, indicating frequent inversions or cold air ponding at the climate station in Quesnel. The large standard deviation for this month and station

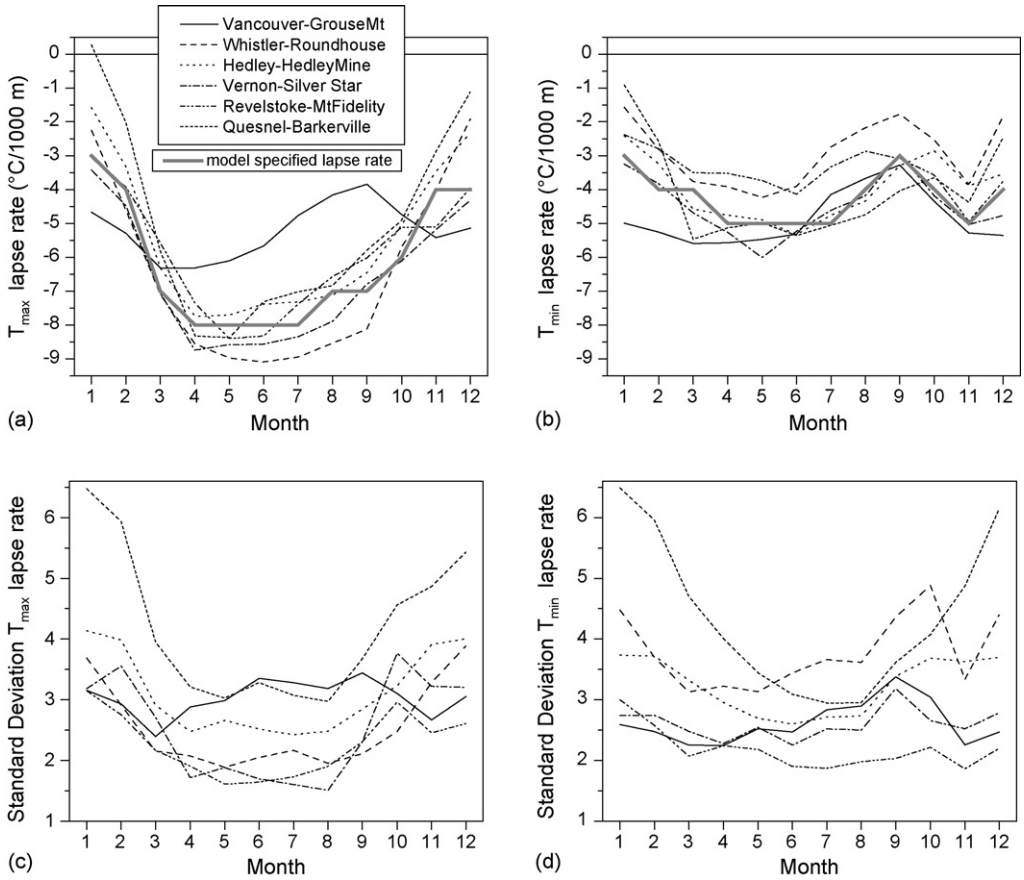

Fig. 2. Mean monthly lapse rates of (a) T_{max} and (b) T_{min} and their standard deviations (c and d) at station pairs across BC.

provides further evidence of this occasional phenomenon. The seasonal distribution of T_{min} lapse rates shows a more complex pattern. Gradients increase in absolute value in spring, become smaller in summer, particularly in September, then increase again during October and November before reaching their smallest gradients from December to February. The coastal station pair (Vancouver-Grouse Mountain) shows generally smaller annual variations than the other station pairs and has a different seasonal pattern. The difference between the T_{max} and T_{min} lapse rates for Vancouver-Grouse Mountain is also small compared to the other locations. Particularly in summer, however, day to day variations of lapse rates on the coast are higher than elsewhere.

4.2. Model parameters

Some of the models have parameters that can be adjusted. The models that use the density-based station search require as input the number of iterations, the initial truncation radius and the desired number of stations N. The number of iterations generally showed no significant performance improvement above 3 and was therefore set to that value. The initial truncation radius also showed no major influence and was set to 140 km. This radius was also used to set the boundaries in the Quadrant search station selection.

Models using Gaussian filter weighting require the input of a shape parameter α. Thornton et al. (1997) optimized the best combination of the shape parameter and the number of desired stations by creating a response surface of the mean absolute error (MAE) of a cross-validation for a range of shape parameter values and desired number of stations (Thornton et al., 1997). Similar to applications of the DAYMET algorithm (Hasenauer et al., 2003), such a surface for BC also shows a trough of lowest MAE values from lower station numbers and lower shape parameters to higher station numbers and higher shape parameters. The interpolations for the best overall results were carried out with a shape parameter of 3.5 and a desired station

number of 30. This combination resulted in large differences in the final truncation radii across BC, which was found to vary from 46 to 862 km for 1965 and 43–509 km for 2000. The largest radii occurred in areas of lower station density, particularly in the north of BC. Choosing a lower desired number of control points would have resulted in smaller radii and a more uniform model performance across BC but higher MAE.

4.3. Testing of interpolation algorithms

To assess the model performance we carried out a cross-validation. For each day, each station was, in turn, withheld from the set of control points and was predicted by interpolating from the other stations. The prediction error was aggregated over all stations and was expressed as both mean absolute error (MAE) and root mean squared error (RMSE). For two selected years with differing data availability, 1965 and 2000, error statistics for all 12 models were analysed in detail concerning elevational, seasonal and spatial variations. Then, year to year (1965–2000) variations in the error statistics were assessed for five selected models: the three models using a truncated Gaussian filter weighting scheme but different lapse rate estimations (LWR-G, CLWR-G and SLR-G), the multiple linear regression (MLR) and the gradient-plus-inverse-distance-squared (GIDS) model.

5. Results

5.1. Validation for 1965 and 2000

Table 3 summarizes the MAE and RMSE for all models and for BC stations only. Stations in Alberta (AB) and the Yukon (YT) and Northwest Territories (NWT) were used for the interpolation to reduce edge effects. Values of MAE range from 1.22 to 1.99 while RMSE varies from 1.83 to 2.83 depending on the year and model. All models predicted T_{max} more accurately than T_{min}. Validation errors are lower for the year 2000 than for 1965. The overall MAE for the models based on specified lapse rates are similar with the exception of the inverse distance (SLR-ID), which especially for 1965 T_{max} does not perform as well. Among the models that calculate lapse rates, GIDS performs best and in the year 2000 shows the lowest MAEs of all models. The other models based on calculated lapse rates perform less well than the models using specified lapse rates. These differences between the models overall MAE are larger for the low-density year 1965. The nearest neighbour model shows the poorest performance.

The validation errors generally increase with elevation (Fig. 3). The highest MAE are found above 1500 m for 1965, when fewer high-elevation stations were available. Errors for the LWR-G, MLR and GIDS models are particularly high for 1965, because the methods calculate inaccurate temperature gradients in the absence of high-elevation control points. Better results are found for the year 2000, due to the presence of the generally higher-elevation Snow Pillow and Fire Weather stations. In fact, for the highest elevation band in the year 2000, the MLR model performs best among all models, closely followed by GIDS, LWR-G and the CLWR-G model, which all have an MAE of 0.2–0.5 °C lower than the models based on specified lapse rates.

Validation errors also vary seasonally and they vary similarly for all models (Fig. 4). Temperatures from December to February are less well predicted than for

Table 3
Mean absolute error (°C) aggregated for all BC stations for 1965 and 2000

No.	Model	MAE 2000		MAE 1965		RMS 2000		RMS 1965	
		T_{max}	T_{min}	T_{max}	T_{min}	T_{max}	T_{min}	T_{max}	T_{min}
1	NN	1.53	1.93	1.57	1.99	2.26	2.78	2.37	2.83
2	LWR-G	1.42	1.72	1.59	1.92	2.06	2.40	2.30	2.69
3	CLWR-G	1.40	1.71	1.57	1.89	1.98	2.37	2.26	2.61
4	SLR-G	1.35	1.64	1.47	1.76	1.91	2.25	2.10	2.42
5	SLR-K	1.27	1.63	1.31	1.66	1.85	2.30	1.91	2.29
6	SLR-KQ	1.28	1.64	1.35	1.72	1.87	2.33	2.02	2.45
7	SLR-ID	1.34	1.62	1.52	1.77	1.91	2.24	2.22	2.47
8	SLR-IDQ	1.29	1.60	1.40	1.73	1.86	2.23	2.06	2.45
9	SLR-IDS	1.29	1.61	1.39	1.69	1.87	2.27	2.05	2.36
10	SLR-IDSQ	1.28	1.62	1.36	1.72	1.88	2.30	2.04	2.45
11	MLR	1.34	1.62	1.50	1.81	1.93	2.24	2.11	2.52
12	GIDS	1.22	1.55	1.31	1.68	1.83	2.19	1.92	2.39

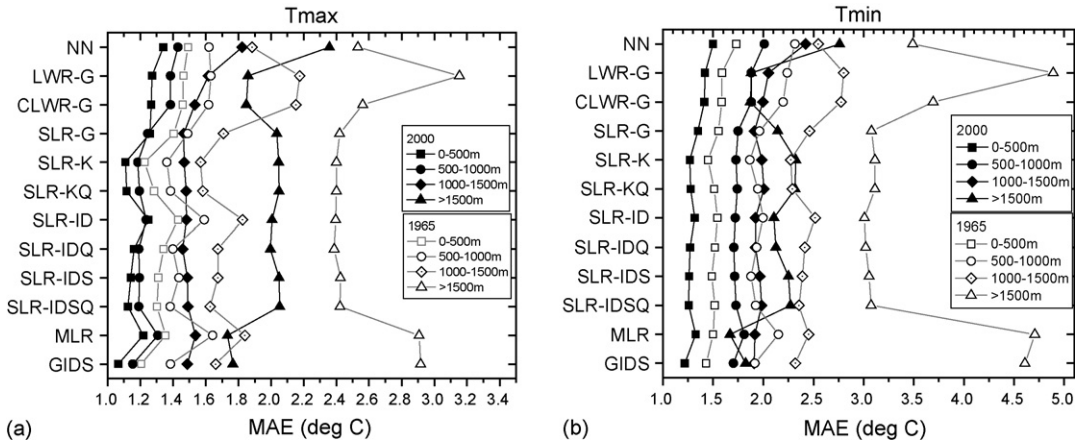

Fig. 3. Mean absolute error for different elevation bands for: (a) T_{max} and (b) T_{min}.

other months. The annual variation is stronger in 2000 than in 1965, which indicates that the seasonal differences may in fact partly be due to the lower relative number of stations available during winter 2000 than summer 2000. Although 1965 has fewer records altogether they are all fairly complete, which contrasts to the data record for 2000, where many stations have incomplete records as they operated only seasonally (Fig. 1).

Although the mean absolute errors are not extreme, individual days can have large errors. The largest prediction errors occur for the lowest and highest temperatures, which demonstrates that the models are unable to predict the full temperature range (Fig. 5). All

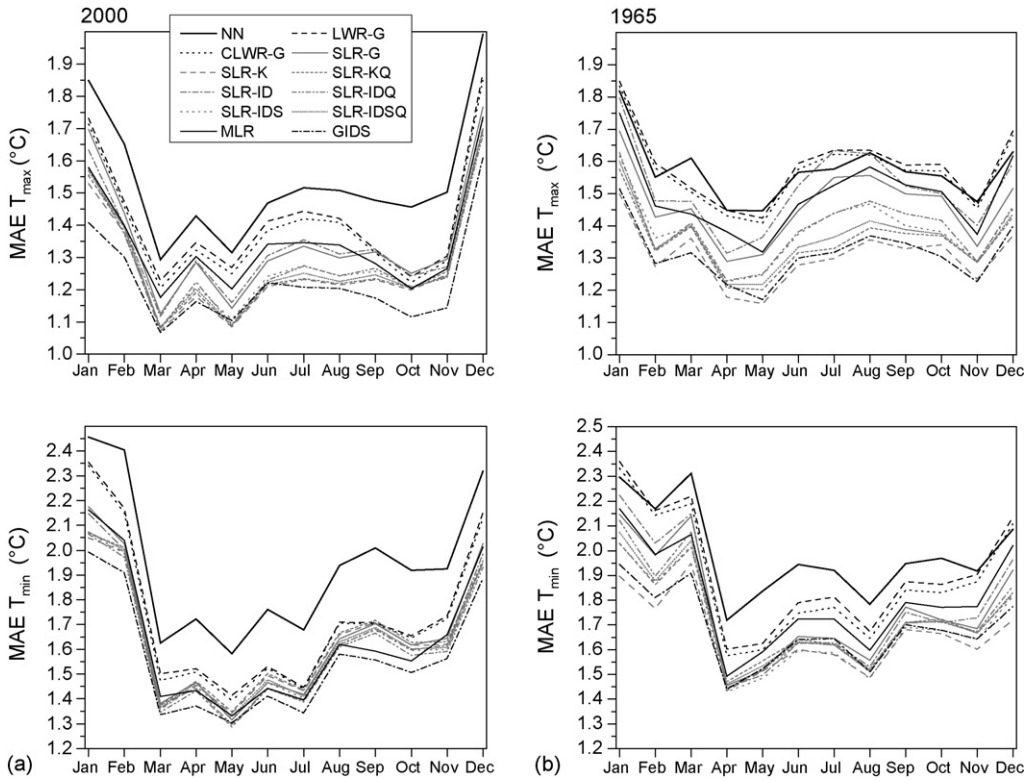

Fig. 4. Mean absolute error by month for the year 2000 (left panels) and 1965 (right panels).

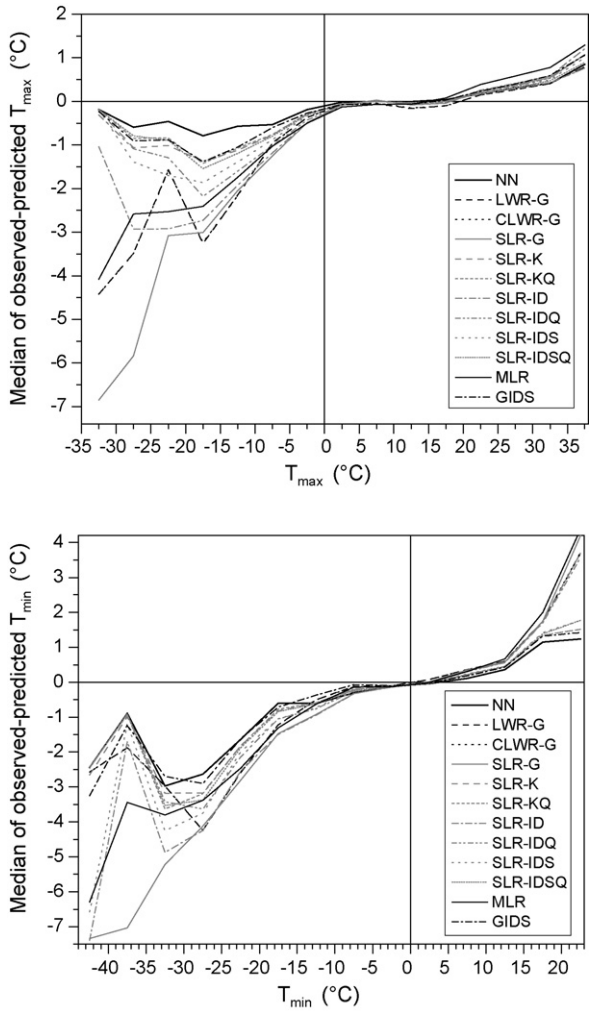

Fig. 5. Median of the residuals (observed–predicted) of daily T_{max} and T_{min} for the year 2000, aggregated within 5 °C intervals.

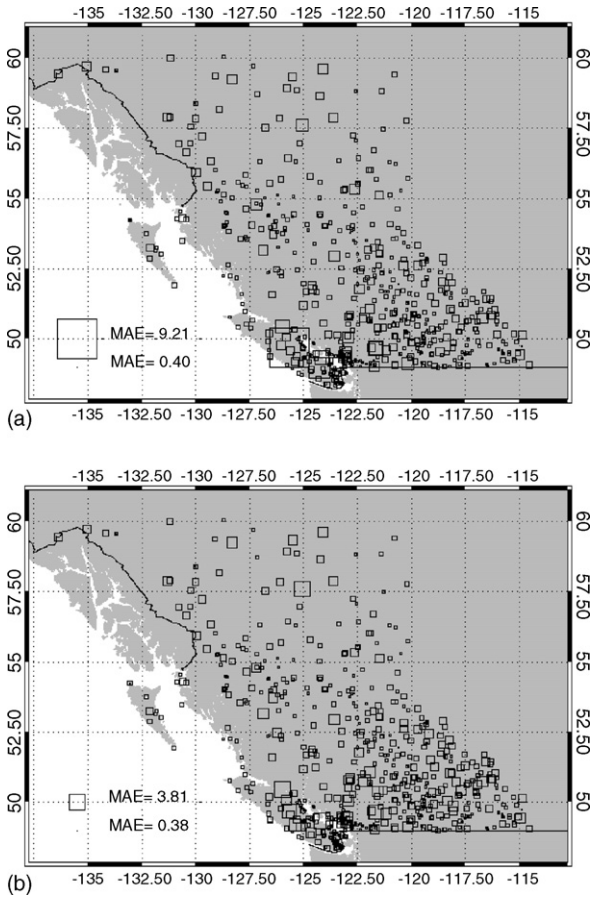

Fig. 6. Mean absolute error at each station for the validation of T_{max} in 2000 for (a) the GIDS model and (b) the SLR-K model. Symbols are scaled to MAE.

models systematically underestimate high and overestimate low temperatures.

For all models, large errors occur for individual stations in the northern part of the province, where station density is low, and on Vancouver Island and in the Coast Mountains (Fig. 6). Although it is the model with the best overall performance, the GIDS model produced a large MAE of 9.21 °C for one particular high-elevation station on Vancouver Island (Fig. 6a), which is surrounded by only low elevation stations. Here, strong temperature gradients between stations located directly on the coast and stations located along an inlet further inland, but only marginally higher in elevation, resulted in excessive model-calculated lapse rates. The same large error was found for all models that calculate lapse rates (MLR, LWR-G). The model with the constrained lapse rate and the models with specified lapse rates also generally perform less well in this region than elsewhere; however, their MAE's for the same station were considerably better.

5.2. Temporal variation of cross-validation errors

Cross-validation errors for all years from 1965 to 2000 for five selected models confirm the patterns found for 1965 and 2000. The GIDS model performs best in all years, followed by the model using the specified lapse rate (SLR-G), the multiple linear regression (MLR) and then the constrained and unconstrained lapse rate weighted regression models (CLWR-G and LWR-G). The ranking of the models by their annual error statistics remains relatively unchanged with time. However, annual MAEs for each model vary through time. While the MAE for a particular model can vary by 0.2–0.3 °C from year to year, throughout the study period overall performance differences between the models narrow and all models generally improve with

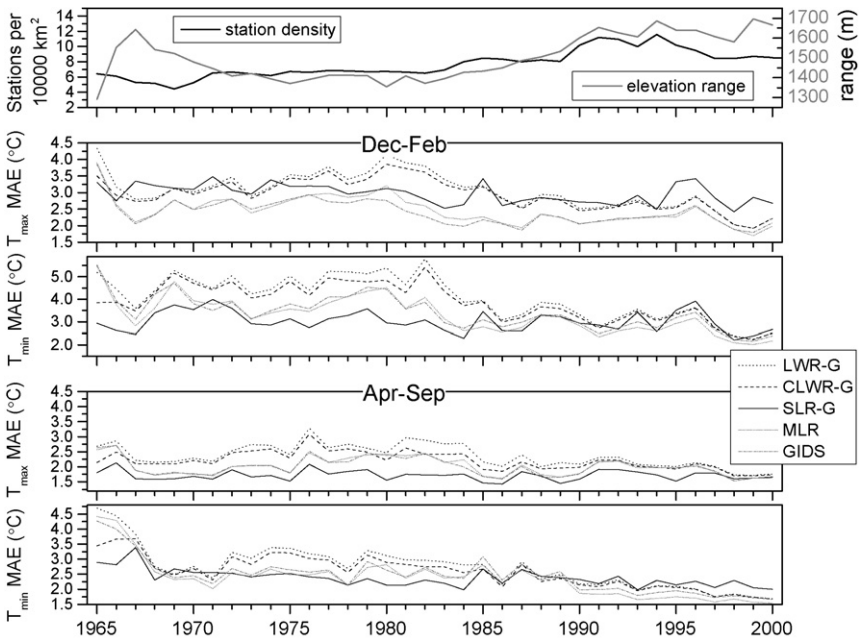

Fig. 7. Mean station density and annual mean cross-validation errors of five models for all stations above 1500 m.

time. Seasonal differences are also similar in all years, with high MAE from December to January, in particular for T_{min}, and lower MAE in summer. While the highest MAE for winter T_{min} did not occur in the warmest years, there is no general statistically significant relation between seasonal temperature and model performance.

Illustrative, however, are the seasonal results for highest elevation band >1500 m (Fig. 7). While most models show a clear improvement with increasing station density and elevation range within the dataset, the cross-validation error of the only model based on specified monthly lapse rates (SLR-G) remains relative constant through time. Until around 1985, however, this is the best model for high-elevation temperature interpolation, with the exception of summer T_{max}, which in most years is better predicted by GIDS or MLR. After 1985, consistent with the availability of more high-elevation stations, the other models (which all calculate lapse rates from surrounding stations) perform similarly well or even better. With a substantial drop in MAE of 2–4 °C depending on the model from the early 1980s to the late 1990s, the improvement in model performance is greatest for the cross-validation of T_{min}, particularly in winter.

6. Discussion

All of the models performed better in more recent years due to the greater availability of data, especially for higher elevations. Differences among model performance were greatest for the highest elevation band (>1500 m), almost certainly due to differences in the ability of the different approaches to account for vertical temperature gradients. The methods that compute local lapse rates for each prediction point (LWR-G, CLWR-G, MLR, GIDS) outperformed the other methods for the highest stations in recent years. However, LWR-G, MLR and GIDS, which depend entirely on the locally calculated lapse rates, performed substantially worse than the other models in 1965, even the nearest neighbour approach.

The models employing the Gaussian filter differ only in the way that lapse rates are calculated. Focusing on the results for the highest elevation stations (>1500 m), the model performance for 1965–1985 decreased from SLR-G to CLWR-G to LWR-G, probably as a result of the lack of higher-elevation stations, which would introduce greater error into the calculation of lapse rates. This error would be greatest for the LWR-G model. The performance difference narrowed during the last 5–10 years of the record, and finally in the year 2000 the LWR-G and CLWR-G models performed similarly or better than the SLR-G approach, likely due to the increased accuracy in calculating local, time-varying lapse rates with higher-elevation data. MLR and GIDS, which use the same station selection approach but a regression model (with inverse distance squared weighting in GIDS) show a

similar pattern hence supporting the strong influence of high-elevation stations. The regression-based approaches, particularly GIDS, however, always perform better than all other approaches, which calculate lapse rates. Event in the early years of the study periodm GIDS outperformed other algorithms for the interpolation of T_{max} in winter.

Of the methods involving specified lapse rates, the SLR-K, SLR-KQ and the SLR-IDSQ methods provided the best performance in terms of both lower MAE and reduced bias. Kriging did not provide a large improvement over inverse-distance weighting, likely because the assumed linear variogram is inappropriate. Quadrant searching slightly improved the performance of the inverse-distance methods, probably because it results in a more even distribution of control points and reduces the potential for bias in the presence of regional trends. The use of quadrant search with kriging did not perform better than the density-based algorithm, possibly because kriging inherently accounts for the arrangement of stations.

Averaged over several years, lapse rates in BC show a clear seasonal pattern, the relative variability of which is consistent with seasonal gradients found in other mountainous regions of the world (Rolland, 2003). The absolute lapse rates, however, vary regionally and temporally, which presents a source of error in methods that use a specified lapse rate, especially for predicting the extreme temperatures. In particular, winter temperatures can be influenced by cold air drainage and ponding during cold periods. These phenomena can invert vertical temperature profiles and thus complicate the extrapolation of air temperature to higher elevations from low-elevation stations. The high variability of daily lapse rates during winter (Fig. 2c and d) reflects temperature inversions as well as rapidly changing circulation patterns during winter in BC. Further research should build on the work of Jarvis and Stuart (2001) and focus on methods for temperature interpolation and extrapolation that can account for differences in topographic influence under different seasons and synoptic situations. A challenge will be to develop automated methods that can be applied over large domains like BC, where the surface climate response to synoptic-scale weather systems can exhibit significant spatial variability particularly during winter (Stahl et al., 2006a). A further challenge will be to develop approaches for distinguishing local-scale effects, such as cold-air ponding at individual stations from regional-scale air mass controls on vertical temperature gradients.

7. Conclusions

This study compared twelve methods for interpolating daily maximum and minimum temperatures over British Columbia, Canada, a region with complex topography and highly variable station density and elevational distribution. The simplest method, which uses the nearest station with an adjustment for elevation, generally had the greatest errors. All models performed better for years with greater station density, particularly in relation to higher-elevation stations. Prediction errors also depended on elevation and month. Methods that compute local lapse rates from the control points performed significantly more poorly for years for which there were a smaller number of higher-elevation stations, which introduced greater error into the estimation of lapse rates. These methods should therefore not be applied in the absence of sufficient high-elevation data. Specified lapse rates, on the other hand, cannot account for the large variability in daily lapse rates in winter. The study illustrates that even if further research makes progress in including effects of topography and synoptic-scale weather systems on temperature patterns at different spatial scales, from local to regional, an appropriate observation network that includes high-elevation stations is indispensable for spatial interpolation of climate variables. This is of particular interest as global warming so far has been affecting the cold season most strongly, when high-elevation temperatures can play an important role, for example, in Mountain Pine Beetle mortality (Stahl et al., 2006b) and the occurrence of mid-winter rain-on-snow events.

Acknowledgments

We gratefully acknowledge financial support from the Ministry of Water, Land and Air Protection through Contribution Agreement TBIO4025 and through financial support from the Government of Canada's Climate Change Impacts and Adaptation Program via Grant A676, as well as through a Natural Sciences and Engineering Research Council Discovery Grant to R.D.M. Eric Meyer of BC Ministry of Forests provided access to the Fire Weather climate data, while Dave Hutchinson of Environment Canada assisted with accessing daily climate data from that agency.

References

Bolstad, P.V., Swift, L., Collins, F., Régnière, J., 1998. Measured and predicted air temperatures at basin to regional scales in the

southern Appalachian mountains. Agric. Forest Meteorol. 91, 161–178.

Choi, J., Chung, U., Yun, J.I., 2003. Urban effect correction to improve accuracy of spatially interpolated temperature estimates in Korea. J. Appl. Meteorol. 42 (12), 1711–1719.

Chung, U., Yun, J.I., 2004. Solar irradiance—corrected spatial interpolation of hourly air temperature in complex terrain. Agric. Forest Meteorol. 126, 129–139.

Couralt, D., Monestiez, P., 1999. Spatial interpolation of air temperature according to atmospheric circulation patterns in southeast France. Int. J. Climatol. 19, 365–378.

Daly, C., Gibson, W.P., Taylor, G.H., Johnson, G.L., Pasteris, P., 2002. A knowledge-based approach to the statistical mapping of climate. Climate Res. 22, 99–113.

Daly, C., 2006. Guidelines for assessing the suitability of spatial climate data sets. Int. J. Climatol. 26, 707–721, doi:10.1002/joc.1322.

Deutsch, C.V., 1996. Correcting for negative weights in ordinary kriging. Comput. Geosci. 22, 765–773.

Dodson, Marks, 1997. Daily air temperature interpolated at high spatial resolution over a large mountainous region. Climate Res. 8, 1–20.

Garen, D.C., Marks, D., 2005. Spatially distributed energy balance snowmelt modelling in a mountainous river basin: estimation of meteorological inputs and verification of model results. J. Hydrol. 315, 126–153.

Hamann, A., Wang, T., 2005. Models of climatic normals for genecology and climate change studies in British Columbia. Agric. Forest Meteorol. 128, 211–221.

Hasenauer, H., Merganicova, K., Petritsch, R., Pietsch, S.A., Thornton, P.E., 2003. Validation of daily climate interpolations over complex terrain in Austria. Agric. Forest Meteorol. 119, 87–107.

Hock, R., 2003. Temperature index melt modelling in mountain areas. J. Hydrol. 282, 104–115.

Holdaway, M.R., 1996. Spatial modeling and interpolation of monthly temperature using kriging. Climate Res. 6, 215–225.

Jarvis, C.H., Stuart, N., 2001. A comparison among strategies for interpolating maximum and minimum daily air temperatures. Part II: the interaction between number of guiding variables and the type of interpolation method. J. Appl. Meteorol. 40, 1075–1084.

Logan, J.A., Powell, J.A., 2001. Ghost forests, global warming, and the mountain pine beetle (Coleoptera: Scolytidae). Am. Entomol. 47, 160–173.

Liston, G.E., Elder, K., 2006. A meteorological distribution system for high-resolution terrestrial modeling (MicroMet). J. Hydrometeorol. 7, 217–234.

Nalder, I.A., Wein, R.W., 1998. Spatial interpolation of climatic normals: test of a new method in the Canadian boreal forest. Agric. Forest Meteorol. 92, 211–225.

Régnière, J., 1996. A generalized approach to landscape-wide seasonal forecasting with temperature-driven simulation models. Environ. Entomol. 25, 869–881.

Rolland, C., 2003. Spatial and seasonal variations of air temperature lapse rates in alpine regions. J. Climate 16 (7), 1032–1046, doi:10.1175/1520-0442.

Running, S.W., Nemani, R.R., Hungerford, R.D., 1987. Extrapolation of synoptic meteorological data in mountainous terrain and its use for simulating forest evapotranspiration and photosynthesis. Can. J. Forest Res. 17, 472–483.

Shen, S.S.P., Dzikowski, P., Li, G.L., Griffith, D., 2001. Interpolation of 1961–1997 daily temperature and precipitation data onto Alberta polygons of ecodistrict and soil landscapes of Canada. J. Appl. Meteorol. 40, 2162–2177.

Stahl, K., Moore, R.D., McKendry, I.G., 2006a. The role of synoptic-scale circulation in the linkage between large-scale ocean-atmosphere indices and winter surface climate in British Columbia, Canada. Int. J. Climatol. 26, 541–560, doi:10.1002/joc.1268.

Stahl, K., Moore, R.D., McKendry, I.G., 2006b. Climatology of winter cold spells in relation to mountain pine beetle mortality in British Columbia, Canada. Climate Research 32, 13–23.

Thornton, P.E., Running, S.W., White, M.A., 1997. Generating surfaces of daily meteorological variables over large regions of complex terrain. J. Hydrol. 190, 214–251.

Xia, Y., Fabian, P., Winterhalter, M., Zhao, M., 2001. Forest climatology: estimation and use of daily climatological data for Bavaria, Germany. Agric. Forest Meteorol. 106, 87–103.