ELSEVIER

Agricultural and Forest Meteorology 92 (1998) 211–225

AGRICULTURAL AND FOREST METEOROLOGY

Spatial interpolation of climatic Normals: test of a new method in the Canadian boreal forest

Ian A. Nalder*, Ross W. Wein

Department of Renewable Resources, University of Alberta, Edmonton, AB, Canada T6G 2E3

Received 20 April 1998; accepted 22 July 1998

Abstract

Spatial interpolation of climatic data is frequently required to provide input for plant growth models. As no single method is optimal for all regions, it is important to compare the results obtained using alternative methods applied to each data set. For estimating 30-year averages (Normals) of monthly temperature and precipitation at specific sites in western Canada, we examined four forms of kriging and three simple alternatives. One of the alternatives was a novel technique, termed 'gradient-plus-inverse distance squared' (GIDS), which combines multiple linear regression and distance-weighting. Based on the mean absolute errors from cross-validation tests, the methods were ranked GIDS > detrended kriging > nearest neighbour >co-kriging > inverse distance squared > universal kriging > ordinary kriging for interpolating monthly temperature, and GIDS > co-kriging>inverse distance squared > nearest neighbour > ordinary kriging > detrended kriging > universal kriging for interpolating monthly precipitation. GIDS gave the lowest errors, which averaged 0.5°C for monthly temperature and 3.6 mm, or 11%, for monthly precipitation. These errors were comparable with those from optimal methods in other studies. GIDS errors were also more consistent for a wide range of data variability than the other methods. The performance of kriging may have been constrained by the limited number of stations (32) in the study region, but if so, this is an unavoidable limitation in regions with sparse coverage of climate stations. Compared with kriging, GIDS was simple to apply and avoided the subjectivity involved in defining variogram models and neighbourhoods. We conclude that GIDS is a simple, robust and accurate interpolation method for use in our region, and that it should be applicable elsewhere, subject to careful comparison with other methods.

Keywords: Boreal forest; Climate interpolation; Temperature; Precipitation; Kriging; Canada

1. Introduction

The long-term dynamics of forests are commonly simulated to examine aspects such as growth and yield, succession, carbon dynamics and nutrient cycling. Climatic information is usually required for the sites being simulated, but long-term records are rarely available for specific sites, so as a surrogate, regional climate data from nearby climate stations are frequently utilised (Running et al., 1987; Bradley et al., 1995; Nalder and Merriam, 1995; Sykes and Prentice, 1995; Bugmann and Fischlin, 1996; Talkkari and Hypén, 1996; Liski and Westman, 1997; Peng et al., 1998). Either the assumption is made that the climate recorded at a given climate station is repre-

*Corresponding author. Tel.: +1-403-492-3242; fax: +1-403-492-1767; e-mail: inalder@gpu.srv.ualberta.ca

0168-1923/98/$ – see front matter
PII: S0168-1923(98)00102-6

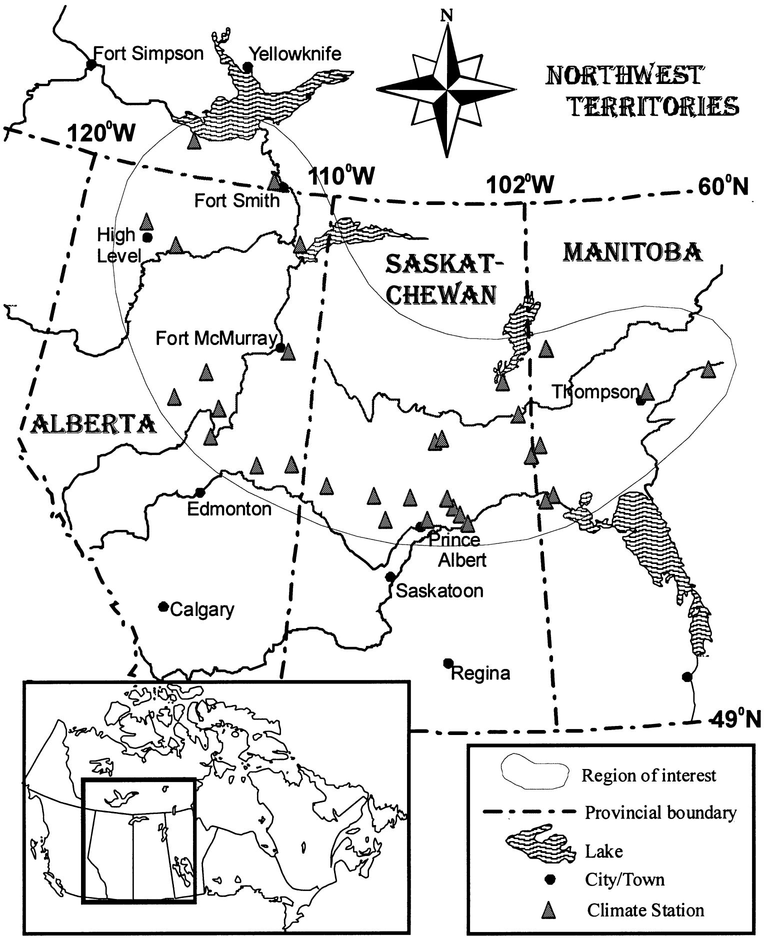

Fig. 1. Western Canada showing study region and climate stations.

sentative of the surrounding region, or else climate variables are spatially interpolated to generate data for a specific site or for a series of grid cells.

Spatial interpolation becomes particularly problematic when climate stations are few and widely separated. This is the case for much of the boreal forest of Canada, where climate stations can be separated by hundreds of kilometres. Our region of interest (Fig. 1) is defined by a separate study of forest floor dynamics for which 121 forest stands were sampled across northern Manitoba, Saskatchewan, Alberta and the southern Northwest Territories. This region, which we refer to as the western Canadian continental boreal forest (WCCBF), is relatively homogeneous, having low relief and broadly similar vegetative cover and land–water ratios. Given this uniformity, we were hopeful of finding a method that was simple to apply and would give good estimates for our study sites despite the limited number of climate stations.

Interpolation methods commonly applied for estimating temperature or precipitation include distance weighting (Tabios III and Salas, 1985; Eischeid et al., 1995; Lennon and Turner, 1995; Ashraf et al., 1997; Dodson and Marks, 1997), interpolating polynomials (Stewart and Cadou, 1981; Tabios III and Salas, 1985; Eischeid et al., 1995; Lennon and Turner, 1995), kriging (Tabios III and Salas, 1985; Phillips et al., 1992; Hammond and Yarie, 1996; Holdaway, 1996; Ashraf et al., 1997) and splines (Hulme et al., 1995; Lennon and Turner, 1995). Distance weighting, which estimates the variable of interest by assigning more weight to closer points, is the simplest technique. Interpolating polynomials fit a polynomial of an appropriate degree to the data points: higher degree polynomials provide a better fit, but may give totally unreasonable values between data points. Kriging, originally developed for mining ore estimation, assigns weights to minimise the variance and bias of the estimates. Spline methods, which are equivalent to kriging with a generalised covariance function (Cressie, 1986), fit polynomials to a restricted set of points to provide a smooth, minimum curvature surface passing through the points.

Several climate interpolations have been done for specific regions of western Canada (Longley, 1972; Longley and Janz, 1978; Stewart and Cadou, 1981; Olson, 1986; Hogg, 1994; Halsey et al., 1995) but none of these have attempted to define an optimal method of interpolation. There is little evidence that any one method is optimum across a range of conditions: rather, it is important to determine the best method for each circumstance (Lennon and Turner, 1995).

For our purposes, we wanted a simple, robust and flexible method of estimating temperature and precipitation at specific sites across the WCCBF. Kriging has the advantages that it is a well-proven technique, software is readily available, it provides optimal interpolation in the sense of providing the best linear unbiased estimate, it provides error estimates at the unknown points, and since it can return the original data values, it can be considered an exact technique. However, it assumes stationarity of data, which is almost never true. This assumption can be relaxed with specific forms of kriging, but still it has to be determined which form is applicable to a particular dataset. Additionally, definition of the required variogram models (see Section 2) is time consuming and somewhat subjective. Definition of neighbourhoods (the area outside which stations are not used in the kriging process) is also required which is difficult to do objectively. Consequently, we were interested in testing some simple alternatives to kriging, as well as testing the various forms of kriging, to determine the most suitable method for interpolation in our region.

2. Methods

2.1. Climatic data

We carried out spatial interpolation of mean monthly temperature (T) and mean monthly precipitation (P) using 30-year averages, termed Normals, for the period 1961–1990 (Environment Canada, 1994). The boundary of the study region (Fig. 1) was drawn primarily to encompass the 121 stands previously mentioned while avoiding agricultural land to the south, Hudson Bay to the east, the Rocky Mountains to the west and the transitional forest to the tundra in the north. Within this region, we identified 31 climate stations that had complete monthly temperature and precipitation Normals data for 1961–1990. Of these 31 stations, we excluded two (Hay River and Cree Lake) that were located on the shores of large lakes. These were excluded as none of our stands were located close to large lakes, and although lake influence can be considerable, it is generally localised (Holdaway, 1996). Three additional stations were available with precipitation Normals. The locations of the 32 stations used in this analysis are shown in Fig. 1. They range in elevation from 145 to 633 m a.m.s.l. with elevation generally increasing towards the southwest.

Latitude, longitude and elevation of each climate station were taken from Environment Canada (1994). Since the kriging software that we used calculates planar distances, we transformed latitude and longitude to a planar coordinate system following Ashraf et al. (1997). To do this, we calculated great circle distances (km) and bearings to each climate station from a central, reference point (57°N, 105°W), converted this polar representation to X–Y, then applied an X and Y offset of 750 and 500 km, respectively, so that all distances were positive.

Logarithmic transforms of precipitation can give a more normal distribution and/or improved predictions (Stewart and Cadou, 1981; Phillips et al., 1992; Daly et al., 1994). When we examined our monthly precipitation data, however, we found it was not significantly different from a normal distribution at the 5% level using the Kolgorov–Smirnov test for 9 of the 12 months and that log-transformed data made no difference to this result. As a further check, we calculated cross-validation errors using method 7 (see Section 2.9) with both raw data and log-transformed data:

the log transforms gave slightly higher errors. Consequently, untransformed data were used throughout the analysis.

2.2. Kriging – general

Four of the interpolation methods are variants of the basic kriging procedure. Cressie (1991) and Kitanidis (1997) have given detailed treatments of kriging. The basis of kriging is the semi-variogram, usually referred to as a variogram, which defines variation as a function of distance:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{k=1}^{N(h)} [Z(x_k + h) - Z(x_k)]^2$$

where $\gamma(h)$ is the semi-variance of variable Z at separation distance h and $N(h)$ is the number of pairs of points in the distance interval $h + \Delta h$. Values are calculated for each possible pair of stations and the mean values of the semi-variance are plotted for successive distance intervals to produce the experimental variogram. A model variogram is fitted to these points, and this model is used in generating an autocovariance matrix. Estimates for the test sites are calculated by summing weighted values for each climate station where the weights are determined so that $E\{Z^*(x_0) - Z(x_0)\} = 0$ and the estimation variance $\mathrm{Var}\{Z^*(x_0) - Z(x_0)\}$ is minimised. For co-kriging, the cross-semi-variance for the two variables of interest is calculated by

$$\gamma_{12}(h) = \frac{1}{2N(h)} \sum_{k=1}^{N(h)} [Z_1(x_k + h) - Z_1(x_k)][Z_2(x_k + h) - Z_2(x_k)]$$

where Z_1 is the first variable and Z_2 is the second variable.

We developed variogram models for each month for (1) temperature, (2) precipitation, (3) elevation, (4) residuals of temperature after detrending, (5) residuals of precipitation after detrending, and (6) cross-variograms of temperature–elevation and precipitation–elevation. The application of each is discussed below. In each case, the experimental variogram was examined and the most suitable type of model, e.g., Gaussian, was selected by eye, then the model was fitted to the data points using weighted least squares fit (Cressie, 1985). Experimental variograms were plotted using distance intervals of 100 km and a maximum distance of 1200 km.

We also investigated anisotropy by generating variograms and fitting models at 22.5° intervals with a tolerance of ±22.5°. To enable comparisons between directions, the same model form must be used throughout. We chose a linear model, as the limited number of data points in each arc made it unrealistic to define more complex models. Use of a linear model precludes using the conventional anisotropy ratio (the ratio of the range in the major direction of continuity to the range in a direction perpendicular to the major direction (Phillips et al., 1992; Pebesma, 1997)); therefore, we calculate an anisotropy ratio (AR) as

$$\mathrm{AR} = \max\left\{\frac{C_1(\theta)}{C_1(\theta + 90)}\right\},\ \theta = 0,\ 22.5,\ 45 \ldots 180$$

where C_1 is the slope coefficient of the fitted linear model and θ is the angle assessed with the directional variogram.

We used the GSTAT software package for generating experimental variograms, fitting variogram models and performing kriging estimates (Pebesma and Wesseling, 1998). Visual Basic$^{\mathrm{TM}}$ for Applications (VBA) programs were written to generate the required data and command files for GSTAT, as well as to read and summarise the output files.

2.3. Method 1 – ordinary kriging

Ordinary kriging estimates the value of a climatic variable at a point from its values at surrounding stations and a variogram model for that variable. The spatial vector is formed from X and Y coordinates. We assume second-order stationarity, i.e., the variogram function depends on the separation vector, not on location (Journel and Huijbregts, 1978) which requires that there is no trend in the data. This is seldom the case, but may be an acceptable assumption if the points used in the interpolation are restricted to a relatively small neighbourhood (Holdaway, 1996). In GSTAT, the neighbourhood can be defined by setting a maximum distance and/or minimum and maximum number of stations. To determine an optimum neighbourhood, we arbitrarily selected temperature data for the first month (January) and calculated cross-validation errors (see Section 2.10) for a range of neighbourhood constraints. A maximum distance of

300 km, which provided at least three stations, combined with a maximum limitation of eight stations, gave the lowest errors. This neighbourhood was also used for co-kriging and detrended kriging (Sections 2.4 and 2.5).

2.4. Method 2 – co-kriging

Co-kriging takes advantage of the correlation between two variables, in this case the climatic variable and elevation. Three variogram models were used for each run: (1) climatic variable, (2) elevation and (3) cross-variogram of the climatic variable and elevation. The spatial vector is formed from X and Y coordinates and elevation is a covariate.

2.5. Method 3 – detrended kriging

To better meet the assumptions of stationarity, we carried out a multiple linear regression of T and P against X, Y and elevation to remove first-order trends. The residuals were used to generate new variograms and then ordinary kriging was carried out on these residuals. The resulting estimates were added to the trend to give the T or P estimates. This method has been termed residual kriging (Holdaway, 1996), kriging with trend model (Journel and Huijbregts, 1978) and detrended kriging (Phillips et al., 1992). We use the latter term.

2.6. Method 4 – universal kriging

Universal kriging (Ver Hoef, 1993) also takes account of trends in the data by removing trends apparent within the kriging neighbourhood. We used the variogram model previously generated for T or P and specified a first-order trend model using X, Y and elevation as independent variables. The results remain sensitive to neighbourhood, although less than with ordinary kriging. Again, we used January temperature data to test for sensitivity and found that a radius of 600 km with a maximum of 10 stations gave the lowest errors. This neighbourhood was used for the remaining universal kriging runs.

2.7. Method 5 – nearest neighbour

The nearest neighbour method uses climatic data from the nearest climate station to estimate the value at a test site. It is not strictly interpolation, but essentially tests the assumption that a climate station is representative of an area, where the area is formed from all points that are closer to the station than to any other station. It is also known as the Thiessen polygon method (Tabios III and Salas, 1985). It is the simplest method and provided a baseline for testing improvements offered by the other methods. VBA programs were written to carry out the computations for this method as well as for methods 6 and 7.

2.8. Method 6 – inverse distance squared

The inverse distance squared method averages the climate variable from surrounding stations with more weight given to those that are closest. The weighting function is the inverse square of the distance, so that the predicted value for a site is given by

$$Z = \left[\sum_{i=1}^{N} \frac{Z_i}{d_i^2}\right] \Big/ \left[\sum_{i=1}^{N} \frac{1}{d_i^2}\right]$$

where Z is the estimated climatic variable (T or P), Z_i is the value at climate station 'i', d is the distance from the site to climate station 'i' and N is the number of climate stations used for the interpolation.

2.9. Method 7 – gradient plus inverse distance squared

The same multiple linear regressions developed for detrended kriging were also used to define climatic gradients as X, Y and elevation coefficients. These gradients were then used to predict each climate variable for each test site based on the remaining stations. The values so obtained ($N_T = 28$, $N_P = 31$) were averaged using inverse distance squared weighting

$$Z = \left[\sum_{i=1}^{N} \frac{Z_i + (X - X_i) \times C_x + (Y - Y_i) \times C_Y + (E - E_i) \times C_e}{d_i^2}\right] \Big/ \left[\sum_{i=1}^{N} \frac{1}{d_i^2}\right]$$

where X and X_i are the X coordinates of the test site and climate station 'i', respectively, Y and Y_i are the Y coordinates of the test site and climate station 'i', respectively, E and E_i are the elevations of the test site and climate station 'i', respectively, and C_x, C_y and C_e and regression coefficients for X, Y and elevation, respectively.

We refer to this method as 'gradient-plus-inverse distance squared' (GIDS), and to our knowledge it has not been used before. It was attractive as it seemed to explain much of the variation in the dataset while being easy to implement: the necessary calculations can be performed in a simple spreadsheet.

2.10. Assessment criteria

As a test of the accuracy of each method, we calculated T or P for each climate station after excluding that station from the input data. This is a common validation method in climate studies and has variously been termed as 'cross-validation' (Hulme et al., 1995; Holdaway, 1996), 'jack-knifing' (Phillips et al., 1992; Daly et al., 1994), and 'fictitious point' (Tabios III and Salas, 1985; Ashraf et al., 1997). We use the term cross-validation. A more rigorous test would be to reserve a certain number of stations for validation according to Hulme et al. (1995), but given the limited number of stations in our region, this was not practicable.

For each method and each month there were 29 observations for temperature and 32 for precipitation. We calculated error as 'actual minus predicted' and calculated the mean of these errors in three ways following Hulme et al. (1995): mean error (ME), mean absolute error (MAE) and root mean square error (RMSE). ME indicates the degree of bias, MAE provides a measure of how far the estimate can be in error, ignoring sign, and RMSE provides a measure that is sensitive to outliers. Differences among methods were tested using a two-tailed t-test for paired samples: MAE was used as the response variable as we were primarily interested in the absolute magnitude of likely errors for sensitivity testing of our forest growth model. The method that gave the lowest MAE was paired with each of the remaining methods.

3. Results

3.1. Variograms

For each month and each variable, we generated five variogram models for a total of 120 models. Ordinary and universal kriging utilised the same model. Detrended kriging required a separate model based on residuals. Co-kriging required a model for each variable plus a third model to characterise their covariance. Of the 120 models, 70 were linear, 38 were Gaussian, 10 were spherical and 2 were power. Model forms are defined in Pebesma (1997).

There was considerable variation in the variogram models among months (Fig. 2). For temperature, the models generally provided a good fit to the experimental variogram. The fit for co-kriging models was constrained by the need to use a common model-type and range for the three variograms, but in general, good fits were obtained. The models for precipitation provided poorer fits than for temperature, particularly for co-kriging where it was seldom possible to find a model type and range that provided good fits for the three variograms. Twelve models had negative slopes, e.g., January precipitation in Fig. 2.

3.2. Anisotropy

Strong anisotropy was evident in both temperature and precipitation data with T having consistently higher anisotropy ratios (AR) than P (Table 1). AR values above 1.0 indicate some degree of anisotropy. For both T and P, anisotropy varied considerably over the year. For T, the direction of major variance was NE–SW from April to August and NNE–SSW for the remaining months, but for P it was more variable, from N–S to ESE–WNW. The directional variograms also showed a strong seasonal trend in variance, with T variance being very low in summer, but P variance being low in winter months.

3.3. Gradients

The multiple linear regressions carried out for GIDS and detrended kriging methods revealed strong gradients in the climatic data which varied by month in a fairly consistent manner. Fig. 3 plots the T and P coefficients for X, Y and elevation, together with r^2

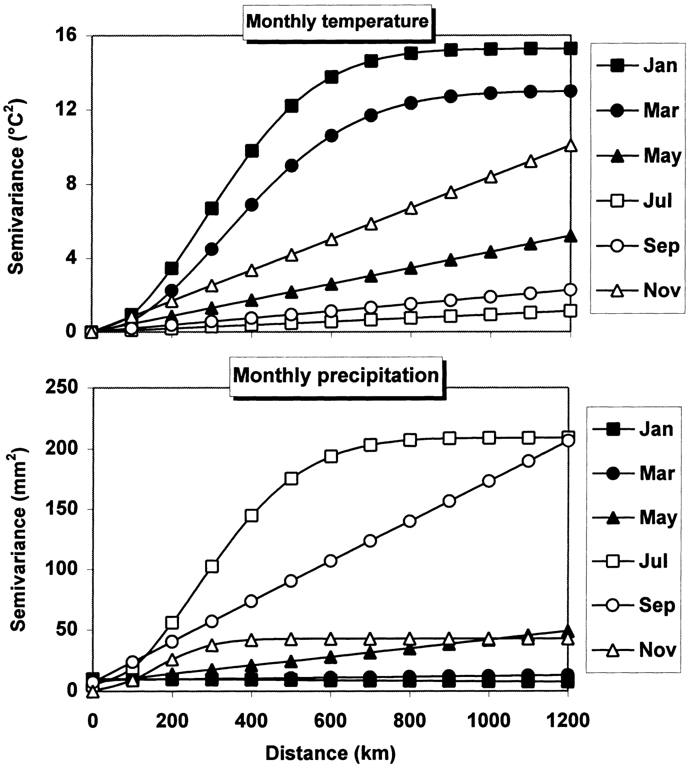

Fig. 2. Monthly temperature and precipitation variogram models used for ordinary and universal kriging (for clarity, only alternate months are shown).

Table 1
T and *P* anisotropy data

	AR_T	AR_P	θ_T	θ_P
January	5.9	1.7	22.5	22.5
February	5.9	2.3	22.5	22.5
March	6.1	5.4	22.5	22.5
April	8.1	1.8	45.0	135.0
May	10.5	4.2	45.0	0.0
June	9.7	4.2	45.0	0.0
July	2.8	3.6	45.0	0.0
August	2.1	2.3	45.04	5.0
September	6.4	1.7	22.56	7.5
October	6.1	3.8	22.56	7.5
November	6.3	3.7	22.54	5.0
December	6.1	2.3	22.52	2.5

AR – Anisotropy ratio.
θ – Direction of major variance.
Subscripts denote *T* or *P* data.

values, with all stations included in the regression. All *T* coefficients, except those for elevation in March, October and November, were significant at the 5% level. For P, the gradients were neither as strong nor as significant: 16 of the 36 coefficients were not significant at the 5% level. Consistent seasonal patterns, however, were still evident. In contrast to *T*, the highest r^2 occurred during the summer. Second-order terms were tested but the improvements in r^2 were only slight.

3.4. Cross-validation errors

A summary of the errors obtained from the cross-validation tests is presented in Table 2. Mean error is relatively low for all methods, but is generally lowest for kriging. For both *T* and *P*, the GIDS method gave the lowest MAE and RMSE. The lowest ME was obtained with detrended kriging for *T* and universal kriging for *P*. Ordinary kriging gave consistently poor performance. When methods were ranked, the ranking was the same for both MAE and RMSE. With the

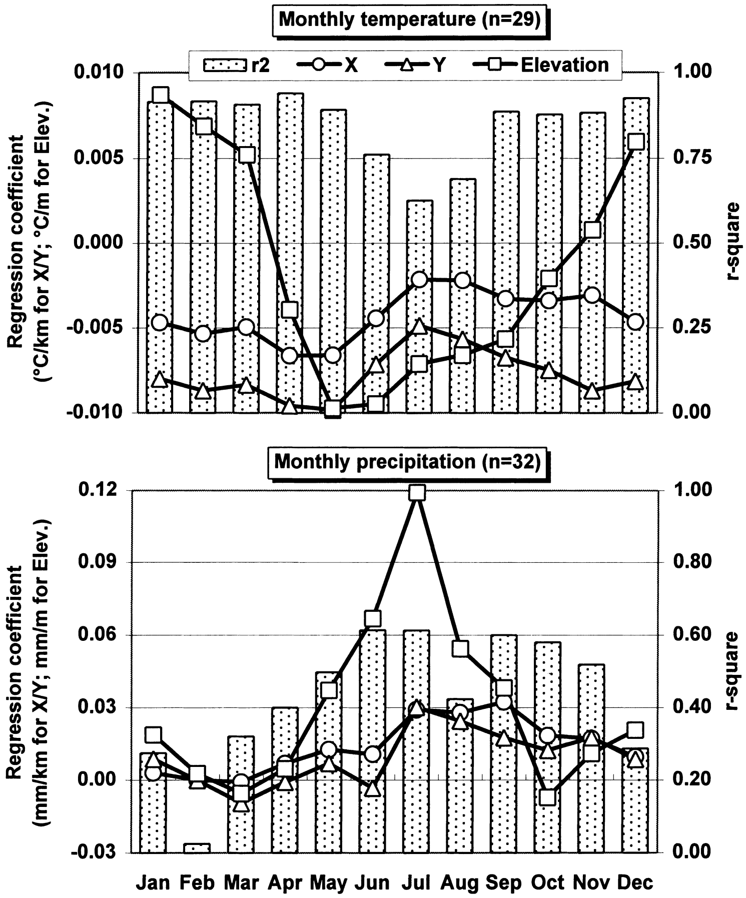

Fig. 3. Coefficients and r^2 values from a multiple, linear regression of monthly temperature and precipitation against X, Y and elevation.

Table 2
Cross-validation errors averaged across all test sites and months for the seven interpolation methods

	1-OK	2-CK	3-DK	4-UK	5-NN	6-IDS	7-GIDS
Temperature errors (°C)							
ME	−0.19	−0.13	**0.03**	0.09	−0.18	−0.39	−0.06
MAE	1.14	0.79	0.54	0.89	0.75	0.87	**0.51**
RMSE	2.27	1.21	0.72	1.44	0.98	1.25	**0.66**
Precipitation errors (mm)							
ME	0.16	0.10	0.23	**0.03**	0.17	−0.23	−0.28
MAE	3.89	3.62	4.27	4.96	3.69	3.67	**3.59**
RMSE	5.92	5.00	5.84	10.47	5.15	4.96	**4.93**

OK – Ordinary kriging; CK – Co-kriging; DK – Detrended kriging; UK – Universal kriging; NN – Nearest neighbour; IDS – Inverse distance squared.
Lowest error values in each row are in bold.

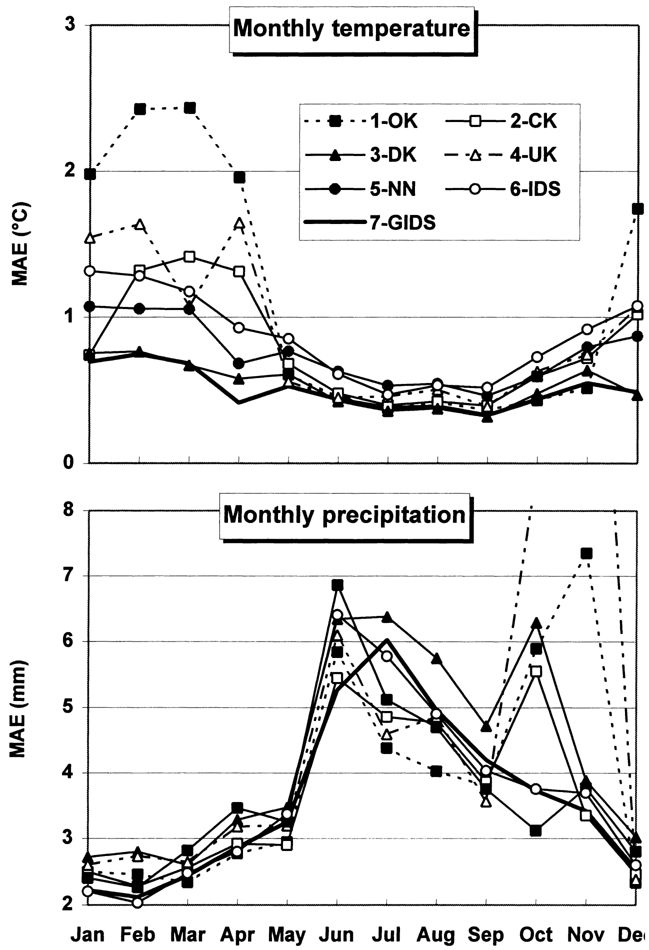

Fig. 4. Cross-validation MAE for mean monthly temperature and precipitation averaged across all test sites for seven interpolation methods. See Table 2 for abbreviations.

exception of universal kriging of *P*, RMSE was only slightly higher than MAE, indicating that the methods were not susceptible to extreme errors.

For all methods, there was substantial variation in error across the year (Fig. 4). For temperature, the highest errors occurred in winter and the lowest values from June to October, which probably reflects the greater temperature differences across the region in winter, e.g., the differences between the hottest and coldest stations are 11.2°C in February compared with 1.4°C in July. In contrast, the highest *P* errors occurred from December to May.

The GIDS method was used as the baseline when testing for significant differences among the methods. For *T* interpolation, GIDS was different from all other methods, except detrended kriging, at the 1% level of significance. For *P*, however, GIDS was only significantly different from detrended kriging (Prob. = 0.0004) and universal kriging (Prob. = 0.01).

3.5. Latitude–longitude vs. X–Y

A comparison of errors using the GIDS method showed that the use of latitude and longitude rather than *X* and *Y* coordinates increased the MAE from 0.51°C to 0.53°C for temperature and from 3.68 to 3.82 mm for precipitation. The differences in temperature MAE were significant at the 5% level using a two-tailed *t*-test for paired means.

4. Discussion

4.1. Gradients

In the WCCBF, there are strong gradients in *T* such that much of the variation can be explained by a simple multiple linear regression against station location and elevation. For most months, r^2 was about 0.9 (Fig. 3) and nearly all terms were significant. Similar high r^2 have been reported in other studies in continental North America (Stewart and Cadou, 1981; Hogg, 1994; Holdaway, 1996).

In all months, temperatures decreased with increasing *X* and *Y*, i.e., to the NE. The effect of elevation was more complex: from May to October, the elevation coefficient ranged from -0.005 to $-0.01°C\ m^{-1}$, which brackets the standard adiabatic lapse rate of $-0.006°C\ m^{-1}$ (Lennon and Turner, 1995), but in winter the coefficient became positive, presumably due to frequent inversions in the region (Longley and Janz, 1978; Olson, 1986). This is in sharp contrast, for instance, to Great Britain, where lapse rates for *T* range from -0.0059 to $-0.0074°C\ m^{-1}$ (Lennon and Turner, 1995). Some studies have used the standard adiabatic lapse rate in predicting monthly temperatures (Leemans and Cramer, 1991; Hammond and Yarie, 1996), but it is clear that would not be appropriate in the WCCBF. Holdaway (1996) noted that elevation is not an important factor in determining temperature outside mountainous areas, but our results contradict this. Elevation was a significant regression term for *T* in 9 months and for *P* in 7 months. Additionally, co-kriging using elevation as a covariate consistently gave lower errors than ordinary kriging (Table 2). Elevation may have a small range in the WCCBF, but it was an important predictor. It is also worth noting that the *T*-coefficient curves for elevation, *X* and *Y* approximate sine waves with a period of 1 year (Fig. 3). This smooth transition from month to month suggests that temporal interpolation to submonthly intervals may be productive.

For precipitation, the regressions explain much less of the variation (Fig. 3), with an average r^2 of 0.43 and only 55% of the coefficients being significant. In general, *P* increases towards the NE and with increasing elevation. During the winter months, r^2 is lower and coefficients are close to zero, indicating little predictive power in the regressions.

The gradients in both *T* and *P* were confirmed by the directional variograms, which indicated strong temperature trends towards the NE or NNE, and weaker precipitation trends which were somewhat variable in direction but mostly oriented N to NE (Table 1). These gradients have important implications for the performance of the various interpolation methods, particularly kriging.

4.2. Kriging methods

Ordinary kriging assumes second-order stationarity and isotropy, which was clearly not the case with our data. As noted earlier, they may be acceptable assumptions if the neighbourhood is sufficiently restricted, but even with the relatively small neighbourhood used in this analysis (3–8 stations), the performance of ordinary kriging was generally the worst of the seven methods (Table 2). This was particularly so with *T* where the gradients were strongest. Holdaway (1996) obtained much lower errors for ordinary kriging of *T* in Minnesota, reporting a mean square error (MSE) of $0.77°C^2$ vs. $5.16°C^2$ from this study. The lower error may be attributable to the smaller area, higher density of stations and availability of 90 years of data in the Holdaway study. However, detrended kriging reduced our MSE to $0.51°C^2$ while making virtually no difference in Minnesota; therefore, the lower errors in Minnesota are more likely due to the smaller gradients, particularly for elevation. Trends are less of a factor for precipitation and our errors were closer to other studies. For example, in the central United States, Tabios III and Salas (1985) found ordinary kriging of annual precipitation gave an MAE of 0.96 inches (24.4 mm) which is slightly higher than the value of 21.4 mm that we obtained by summing monthly errors.

The performance of kriging in our study may be constrained by the limited number of stations. Insufficient data points lead to unstable variograms and may give inappropriate models. Bilonik (1983) suggested 50 as a minimum to obtain a stable variogram. Webster and Oliver (1992) recommended 150–200 points where variation is isotropic and larger numbers for anisotropic cases. The number of stations used in this analysis (29 for T and 32 for P) would appear sub-optimal, and may explain why we could not achieve good fits with many variograms, particularly for precipitation and directional models. If so, this is a limitation of kriging in regions with a sparse coverage of climate stations. It is not clear, however, that poor variogram models will lead to large prediction errors, e.g., Holdaway (1996) noted that temperature errors can be insensitive to the variogram model. We investigated model sensitivity with our data set and found that cross-validation error estimates were fairly stable over a range of variogram models, from those that provided a good fit to those that provided a poor fit. Consequently, we doubt that the number of stations formed a significant limitation on kriging performance in our study.

Of more concern, in our experience, was the selection of the neighbourhood because cross-validation errors were sensitive to this factor. The neighbourhoods used, which were based on a study of January temperatures, are not likely to be optimal for all runs. Hence the errors may be higher than what we would have obtained if the neighbourhoods for each variable had been optimised. It is possible to optimise neighbourhoods (and variogram models) by selecting for minimum cross-validation error (Phillips et al., 1992). The difficulty with this approach is that it fits the model to the data set and removes the possibility of using cross-validation error as an independent test of methods. In the absence of sufficient stations to set aside as test sites, cross-validation was our only feasible method of comparison, therefore, optimisation through cross-validation was ruled out. It should be noted, however, that the results presented for kriging already partially fitted the data as January T data were used to determine the neighbourhoods and data from all stations were used to develop the variogram models. In this sense, the kriging results reported here enjoy an advantage over the other methods that require no a priori selection of model parameters.

As with ordinary kriging, co-kriging assumes stationarity, so its performance is similarly hampered by the presence of strong gradients. The inclusion of elevation as a covariate, however, was expected to reduce errors as elevation was an important predictor in this region. This turned out to be the case: errors were reduced, particularly for T. The relatively slight improvement for P may be due to either the poor fit of precipitation variogram models for co-kriging or to the weaker elevational trends.

When spatial temperature trends were taken into account, as with detrended kriging, T errors were further reduced (Table 2) and the MAE was not significantly different from the method with the lowest overall errors. Precipitation errors, however, were higher than for both ordinary kriging and co-kriging. This result was somewhat surprising, and counter to that of Daly et al. (1994) who reported that detrended kriging of annual precipitation in Oregon gave lower MAE than either co-kriging or ordinary kriging. We suspect it is a consequence of the anomalous behaviour of the residual variograms: 5 months had high nuggets with negative slopes.

Universal kriging also takes trends into account, but gave higher errors than detrended kriging (Table 2). MAE and RMSE for temperature were the second highest of all methods, and those for precipitation were the highest. This poor performance, particularly for precipitation, was unexpected given that Tabios III and Salas (1985) found that universal kriging of annual precipitation gave substantially lower MAE than ordinary kriging in the central United States. Our relatively high precipitation RMSE indicates the presence of some extreme values, which can be seen to occur in October and November (Fig. 4). Within these 2 months, inclusion of two stations (Lynn Lake and Whitesands Dam) produced extremely high errors. Removal of these data points reduced MAE and RMSE to a more reasonable 4.31 and 5.74 mm respectively. Both stations are on the northern periphery of our region and are relatively close together, so it is likely that one (or both) is anomalous resulting in unrealistic weights within the neighbourhood.

4.3. Non-kriging methods

Compared with kriging, the simpler methods performed surprisingly well. For both MAE and RMSE, the nearest neighbour method ranked third for temperature and fourth for precipitation (Table 2). Interestingly, it gave lower precipitation errors than three forms of kriging, lending support to previous observations (Lennon and Turner, 1995) that more complex methods are not necessarily more accurate.

Inverse distance squared gave slightly lower MAE and RMSE for precipitation than the nearest neighbour method. The errors for temperature, however, were higher and it was the one method where bias (ME) was large enough to be of concern. Apart from bias, it ranked ahead of ordinary kriging and universal kriging for both temperature and precipitation, but this superiority is not reflected in other data sets. In the central United States, inverse distance squared gives higher RMSE than ordinary kriging and co-kriging for interpolation of evapotranspiration (Ashraf et al., 1997) and higher MAE than ordinary kriging and universal kriging for interpolation of annual precipitation (Tabios III and Salas, 1985).

GIDS provided the lowest MAE and RMSE of all methods for both temperature and precipitation. In terms of mean error, it ranked second for temperature and last for precipitation, but these biases were in any case very low.

4.4. Seasonal variation

The overall errors, however, hide a much more complex seasonal picture (Fig. 4). For instance, ordinary kriging, which gave poor overall performance, gave the lowest errors for November T and for P in March, July and August. Across the year, no one method is consistently best. GIDS is the most consistent, although its performance for P from July to September is poor. This might suggest that methods should be chosen on a month-by-month basis, but we reject this approach for two reasons. First, at the monthly level of aggregation, we are less likely to find significant differences that will clearly establish the preferred approach. Second, the objective of our study was to find a single method of interpolating climate across the WCCBF: mixing and matching methods would seriously complicate the analysis.

4.5. Selected method

The GIDS method gave the lowest MAE and RMSE while providing low MEs. It was robust in that it provided the lowest MAE and RMSE for two data sets with very different characteristics (T and P), and from month to month it gave a more consistent performance than any other method. It was not, however, significantly better than detrended kriging for temperature, or better than ordinary kriging, co-kriging, nearest neighbour, and inverse distance squared for precipitation. Nevertheless, we consider that GIDS is preferable for our application because it is simple to apply while providing robust performance and low errors for both temperature and precipitation. Although specific forms of kriging (detrended kriging for T and co-kriging for P) have been shown to provide comparable performance, we rule out these kriging methods because: (1) they are not as objective as GIDS: predictions depend on subjectively choosing the variogram model and the neighbourhood, (2) fitting variogram models is very time-consuming, (3) the mathematics are complicated which means the results can be difficult to check and interpret, and (4) the kriging errors presented here are probably optimistic, since the cross-validation was not truly independent of the input data.

GIDS provides cross-validation errors comparable with those from other climate interpolation studies. Detrended kriging of monthly temperature in northern Minnesota gives a MSE of 0.734°C^2 (Holdaway, 1996) which is higher than the GIDS MSE of 0.44°C^2 from this study. An optimum combination of multiple regression and thin plate splines for interpolation of T in Great Britain achieves an r^2 (actual vs. predicted) between 0.89 and 0.94 (Lennon and Turner, 1995): the comparable r^2 from our data were between 0.83 and 0.96. Hulme et al. (1995) used a mixed spline-regression model to interpolate climate for 'greater Europe' and give MAEs between 0.5°C and 0.8°C for January and July T_{max} and T_{min}, which brackets our MAEs for January and July T_{mean} of 0.7°C and 0.4°C. The same study also found MAEs for January and July P (expressed as percentage of means) were 11.9% and 9.1% which are remarkably close to the values of 11.1% and 8.1%, respectively, that we obtain using GIDS. In a comparison of methods for estimating annual precipitation in the central United States, universal kriging gave the lowest MAE of 0.96 in. (24.3 mm) (Tabios III and Salas, 1985). In comparison, we ran GIDS for annual precipitation in our region and obtained a very similar MAE (26.8 mm). Due to the sparse network of climate stations across the WCCBF, we had expected that our errors would be substantially greater than for other regions, but in fact they appear very similar to other studies.

Given the low errors and robust performance, we believe GIDS can be successfully applied in other regions. It has already been used at the national scale to generate gridded climate for Canada (Price et al., 1998). There is potential for reducing GIDS errors further by including additional explanatory variables in the regression (according to Lennon and Turner, 1995). In the WCCBF region, we believe that any improvements would be small compared to the errors presented in this study and have chosen not to pursue this. Log-transforms of precipitation may be also beneficial in other regions, although they proved not to be warranted in the WCCBF. It should also be noted that our cross-validation errors were based on a planar *X–Y* coordinate system, but GIDS can just as easily use latitude and longitude. Although *X–Y* gave slightly lower errors than latitude–longitude, possibly because of the non-linearity inherent in longitude at high latitudes, this may not be the case in other regions.

4.6. Other techniques

The method of thin plate splines is becoming popular for climate interpolation, and may have application in our region. We doubt, however, that splines would give lower errors since comparisons of which we are aware show that other methods, particularly kriging, can perform as well or better. In a comparison of two sets of data, Laslett (1994) concluded that "kriging sometimes outperforms splines by a considerable margin, and never performs worse than splines". Hutchinson and Gessler (1994) re-analysed the data of Dubrule (1984) and concluded that Dubrule's kriging analysis had essentially the same predictive error as their optimum spline analysis. Lennon and Turner (1995) found that splines and regression performed equally when based on the same number of variables. Given the lack of evidence for superiority of splines, GIDS is particularly attractive as it is intuitive and relatively easy to grasp. It is also robust and simple to apply which suggests that it may be useful as a benchmark for evaluating other techniques.

4.7. Applying GIDS

What accuracy will the GIDS method provide when interpolating to actual forest stands? This can never be answered quantitatively, since it would require long term climate data for each stand, but we can draw some general conclusions. On one hand, errors may be higher than presented here because these interpolations were tested against climate stations where conditions are fairly uniform (open, level ground, and frequently at airports), while the variations in topography and forest structure will probably lead to greater variability. On the other hand, errors will tend to be lower for two reasons. First, interpolation distances from the nearest climate station are generally shorter which should decrease errors. For example, the average distance from each cross-validation test station to the nearest climate station was 91 km, but the average distance from each of our 121 stands to the nearest climate station is only 56 km. Second, the peripheral effect tends to give larger errors near the edge of the study region (Phillips et al., 1992). In this study, about half of the climate stations were on the periphery, but none of our 121 stands are peripheral; therefore, interpolation to these stands should give better results. On balance it is likely that predictions for our stands will have less error than those listed in Table 2 and Fig. 4. It must be emphasised, however, that this statement applies to meso-climate, which may not be well correlated with the micro-climate at a specific site.

The strong seasonal trends in errors in this region must be considered in tree growth modelling. Temperature errors, and in fact variations across the region, are relatively small from May to October which brackets the growing season. During this period, absolute interpolation errors are likely to average about 0.4°C. Errors in winter, however, are likely to be nearly double. Whilst this may be immaterial for tree growth, since trees are not photosynthesising at this time, it is important for determining soil temperature (Bonan, 1989) and the length of the growing season (Burton and Cumming, 1995; Sykes and Prentice, 1995), as well as the survival range of a species. In contrast to temperature errors, precipitation errors are low in winter and much higher in summer. Both are important for growth: summer precipitation limits moisture deficits while winter precipitation (snow) provides a moisture reservoir and affects soil temperature due to its insulating properties. Clearly, it is important to assess potential errors in all months. Error estimates from this study will be useful in defining limits for sensitivity testing of tree growth models.

5. Conclusion

We have demonstrated a simple and effective method, which we term GIDS, for spatial interpolation of climatic Normals across the western Canadian continental boreal forest. This method provides cross-validation accuracies at least as good as established kriging techniques without the complexity and subjectivity of kriging. The technique should also be applicable in other areas, although this would need to be tested, and may have potential for interpolation at increased temporal resolution. We have chosen the GIDS method to generate climatic data for 121 stands in the WCCBF to investigate long-term forest floor dynamics.

Acknowledgements

Funding came from an NSERC67 Scholarship from the Natural Sciences and Engineering Research Council of Canada and a Walter H. Johns Graduate Fellowship from University of Alberta, both to I.A. Nalder. We thank D.T. Price and E.H. Hogg for comments on an earlier version of the manuscript.

References

Ashraf, M., Loftis, J.C., Hubbard, K.G., 1997. Application of geostatistics to evaluate partial weather station networks. Agric. For. Meteorol. 84, 225–271.

Bilonik, R.A., 1983. Risk qualified maps of hydrogen ion concentration for the New York state area for 1966–1978. Atmos. Environ. 17, 2513–2524.

Bonan, G.B., 1989. A computer model of the solar radiation, soil moisture, and soil thermal regimes in boreal forests. Ecol. Modelling 45, 275–306.

Bradley, P., Gaston, G., Kolchugina, T., Vinson, T.S., 1995. Simulating carbon storage in forests of eastern Russia. Water Air Soil Pollut. 82, 299–308.

Bugmann, H., Fischlin, A., 1996. Simulating forest dynamics in a complex topography using gridded climatic data. Clim. Change 34(2), 201–211.

Burton, P.J., Cumming, S.G., 1995. Potential effects of climatic change on some western Canadian forests, based on phenological enhancements to a patch model of forest succession. Water Air Soil Pollut. 82, 401–414.

Cressie, N.A.C., 1985. Fitting variogram models by weighted least squares. Math. Geol. 17, 563–586.

Cressie, N.A.C., 1986. Kriging nonstationary data. J. Am. Stat. Assoc. 81(395), 625–634.

Cressie, N.A.C., 1991. Statistics for Spatial Data. Wiley, New York, 900 pp.

Daly, C., Neilson, R.P., Phillips, D.L., 1994. A statistical–topographic model for mapping climatological precipitation over mountainous terrain. J. Appl. Meteorol. 33, 140–158.

Dodson, R., Marks, D., 1997. Daily air temperature interpolated at high spatial resolution over a large mountainous region. Clim. Res. 8(1), 1–20.

Dubrule, O., 1984. Comparing splines with kriging. Comput. GeoSci. 10, 327–338.

Eischeid, J.K., Baker, F.B., Karl, T.R., Diaz, H.F., 1995. The quality control of long-term climatological data using objective data analysis. J. Appl. Meteorol. 34, 2787–2795.

Environment Canada, 1994. Canadian Monthly Climate Data and 1961–1990 Normals on CD-ROM, Version 3.0E. Environment Canada, Ottawa, Canada.

Halsey, L.A., Vitt, D.H., Zoltai, S.C., 1995. Disequilibrium response of permafrost in boreal continental western Canada to climate change. Clim. Change 30, 57–73.

Hammond, T., Yarie, J., 1996. Spatial prediction of climatic state factor regions in Alaska. Ecoscience 3(4), 490–501.

Hogg, T.H., 1994. Climate and the southern limit of the western Canadian boreal forest. Can. J. For. Res. 24, 1835–1845.

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

Hulme, M., Conway, D., Jones, P.D., Jiang, T., Barrow, E.M., Turney, C., 1995. Construction of a 1961–1990 European climatology for climate change modelling and impact applications. Int. J. Climatol. 15, 1333–1363.

Hutchinson, M.F., Gessler, P.E., 1994. Splines – more than just a smooth interpolator. Geoderma 62, 45–67.

Journel, A.G., Huijbregts, C.J., 1978. Mining Geostatistics. Academic Press, London, UK, 600 pp.

Kitanidis, P.K., 1997. Introduction to Geostatistics: Applications in Hydrogeology. Cambridge University Press, Cambridge, 249 pp.

Laslett, G.M., 1994. Kriging nonstationary data. J. Am. Stat. Assoc. 89(426), 391–400.

Leemans, R., Cramer, W.P., 1991. The IIASA Database for Mean Monthly Values of Temperature, Precipitation, and Cloudiness on a Global Terrestrial Grid. Working paper No. WP-90-41, Biosphere Dynamics Project. IIASA, Laxenburg, Austria.

Lennon, J.J., Turner, J.R.G., 1995. Predicting the spatial distribution of climate: temperature in Great Britain. J. Anim. Ecol. 64, 370–392.

Liski, J., Westman, C.J., 1997. Carbon storage in forest soil of Finland 1. Effect of thermocline. Biogeochem. 36, 239–260.

Longley, R.W., 1972. Climate of the Prairie Provinces. Climatological Studies Number 13. Environment Canada, Ottawa, Canada.

Longley, R.W., Janz, B., 1978. The Climatology of the Alberta Oil Sands Environmental Research Program study area. AOSERP Report 39. Fisheries and Environment Canada, Edmonton, Canada.

Nalder, I.A., Merriam, H.G., 1995. Simulating carbon dynamics of the boreal forest in Pukaskwa National Park. Water Air Soil Pollut. 82, 283–298.

Olson, R., 1986. Climate of Wood Buffalo National Park. Environment Canada, Edmonton, Canada.

Pebesma, E.J., 1997. GSTAT User's Manual. Available from http://www.frw.uva.nl/~pebesma/gstat/, 51 pp.

Pebesma, E.J., Wesseling, C.G., 1998. GSTAT: A program for geostatistical modelling, prediction and simulation. Comput. GeoSci. 24(1), 17–31.

Peng, C., Apps, M.J., Price, D.T., Nalder, I.A., Halliwell, D., 1998. Simulating carbon dynamics along the Boreal Forest Transect Case Study (BFTCS) in central Canada 1. Model testing. Biogeochem. 12(2), 381–392.

Phillips, D.L., Dolph, J., Marks, D., 1992. A comparison of geostatistical procedures for spatial analysis of precipitation in mountainous terrain. Agric. For. Meteorol. 58, 119–141.

Price, D.T., Nalder, I.A., Siltanen, R.M., 1998. A 10 km national climate surface for Canadian global change studies. Proc. Int. Workshop Scaling and Modelling in Forestry: Applications in Remote Sensing and GIS, 19–21 March 1988. Université de Québec à Montreal, Canada, in press.

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. For. Res. 17, 472–483.

Stewart, R.B., Cadou, C.F., 1981. Spatial estimates of temperature and precipitation normals for the Canadian Great Plains. Research Branch, Agriculture Canada, Ottawa, Canada.

Sykes, M.T., Prentice, I.C., 1995. Boreal forest futures: modelling the controls on tree species range limits and transient responses to climate change. Water Air Soil Pollut. 82, 401–414.

Tabios III, G.Q., Salas, J.D., 1985. A comparative analysis of techniques for spatial interpolation of precipitation. Water Resour. Bull. 21(3), 365–380.

Talkkari, A., Hypén, H., 1996. Development and assessment of a gap-type model to predict the effects of climate change on forests based on spatial forest data. For. Ecol. Manage. 83, 217–228.

Ver Hoef, J.M., 1993. Universal kriging for ecological data. In: Goodchild, M.F., Parks, B.O., Steyaert, L.T. (Eds.), Environmental Modeling with GIS. Oxford University Press, New York, pp. 447–453.

Webster, R., Oliver, M.A., 1992. Sample adequately to estimate variograms of soil properties. J. Soil Sci. 43, 177–192.