Daily air temperature interpolated at high spatial resolution over a large mountainous region

Rusty Dodson*, Danny Marks**

US Environmental Protection Agency, Office of Research and Development, National Health and Environmental
Effects Research Laboratory, Western Ecology Division, 200 SW 35th Street, Corvallis, Oregon 97333, USA

ABSTRACT: Two methods are investigated for interpolating daily minimum and maximum air temperatures (T_{min} and T_{max}) at a 1 km spatial resolution over a large mountainous region (830 000 km^2) in the U.S. Pacific Northwest. The methods were selected because of their ability to (1) account for the effect of elevation on temperature and (2) efficiently handle large volumes of data. The first method, the neutral stability algorithm (NSA), used the hydrostatic and potential temperature equations to convert measured temperatures and elevations to sea-level potential temperatures. The potential temperatures were spatially interpolated using an inverse-squared-distance algorithm and then mapped to the elevation surface of a digital elevation model (DEM). The second method, linear lapse rate adjustment (LLRA), involved the same basic procedure as the NSA, but used a constant linear lapse rate instead of the potential temperature equation. Cross-validation analyses were performed using the NSA and LLRA methods to interpolate T_{min} and T_{max} each day for the 1990 water year, and the methods were evaluated based on mean annual interpolation error (IE). The NSA method showed considerable bias for sites associated with vertical extrapolation. A correction based on climate station/grid cell elevation differences was developed and found to successfully remove the bias. The LLRA method was tested using 3 lapse rates, none of which produced a serious extrapolation bias. The bias-adjusted NSA and the 3 LLRA methods produced almost identical levels of accuracy (mean absolute errors between 1.2 and 1.3°C), and produced very similar temperature surfaces based on image difference statistics. In terms of accuracy, speed, and ease of implementation, LLRA was chosen as the best of the methods tested.

KEY WORDS: Spatial interpolation · Potential temperature · Neutral stability

1. INTRODUCTION

Air temperature is an important input to a variety of spatially distributed hydrological and ecological models. These models use air temperature to drive processes such as evapotranspiration, snowmelt, soil decomposition, and plant productivity. Since most near-surface air-temperature data are collected at irregularly spaced point locations rather than over continuous surfaces, the point-based temperatures must be accurately distributed over the landscape in order to be useful in spatially distributed modeling. The objective of this study is to analyze and compare 2 methods for interpolating air temperature at high spatial (1 km grid) and temporal (daily) resolution over a large mountainous region (830 000 km^2). The methods were selected because of their ability to account for the relationship between elevation and temperature, and to handle large volumes of data.

Several methods exist for spatial interpolation of point-based data, including inverse-distance weighting, kriging, 2-dimensional splines, and trend-surface regression (Myers 1994). These methods often work well over relatively flat, homogeneous terrain. In mountainous terrain, however, the strong relationship between temperature and elevation precludes a sim-

---

*E-mail: dodson@mail.cor.epa.gov. Affiliated with Dynamac Corporation, 200 SW 35th St, Corvallis, Oregon 97333, USA. Formerly affiliated with Mantech Environmental Research Services Corp., PO Box 12313, Research Triangle Park, North Carolina 27709, USA
**Affiliated with US Department of the Interior, Geological Survey, Water Resources Division, 200 SW 35th St, Corvallis, Oregon 97333, USA. E-mail: danny@mail.cor.epa.gov

ple interpolation of point-based temperature observations. Unless the effect of elevation on temperature is explicitly accounted for, an interpolation of temperature can produce grossly inaccurate results. For example, in the case where a set of temperature observations exist around the base of a mountain, an interpolation which ignores elevation would seriously overestimate the temperature at the mountain top, as it would not account for the fact that temperature generally decreases with increasing elevation.

An additional problem with point-based temperature data is that the locations of meteorological stations tend to be biased toward lower elevations. High-elevation regions are represented poorly by the spatial distribution of most meteorological station networks (Robeson 1995). These 2 difficulties — the correlation between temperature and elevation, and the skewed elevational distribution of temperature-recording stations — present a considerable challenge to deriving accurate air-temperature surfaces over mountainous terrain.

The data generated in this study were one part of a larger database-production and ecological-modeling project. Our task was to derive daily surfaces for minimum and maximum air temperature at a 1 km grid resolution over a large mountainous region. Given the computational size of this task, we limited our choice of temperature-interpolation algorithms to a class of methods which are simple to implement and relatively fast. These algorithms have both been used in the literature but to our knowledge have not been subjected to any rigorous accuracy assessment at high spatial and temporal resolution. One algorithm assumes neutral atmospheric stability and the other assumes a constant linear lapse rate. Our objectives in this study were to assess the accuracy of each temperature-interpolation algorithm, and to determine, if possible, which algorithm is the better one to use based on simplicity, speed, and accuracy.

2. STUDY AREA

The study area encompassed the Columbia River Basin, an area of approximately 670 000 km² in the northwest U.S. and southwestern Canada, plus the coastal areas of Oregon and Washington (Fig 1), making a total area of about 830 000 km². The terrain of this region is heterogeneous and includes 3 mountain ranges (the Pacific Coast Range, the Cascade Range, and the Rocky Mountains) as well as relatively flat regions (the Willamette Valley, the Columbia Plateau, the Snake River Valley, and the Puget Trough).

The study area was represented by a 15-arc-second digital elevation model (DEM) (S. Jensen, US Geologi-

cal Survey, EROS Data Center, pers. comm. 1989). The DEM was projected to an Albers equal-area conic map projection and resampled to 1 km resolution using the Image Processing Workbench software package (Frew 1990, Longley et al. 1992).

2.1. Meteorological station data

The temperature data used in this analysis came from a combined set of SNOTEL, USDA Forest Service, and Canadian meteorological stations (USDA-SCS 1988, EarthInfo 1990, Environment Canada 1989). These stations provide daily measurements of minimum and maximum temperature (T_{min} and T_{max}, respectively) for sites specified by location and elevation. Most of the station measurements were made with high-quality electronic thermistors, which have a typical calibration accuracy of ±0.5°C, with a measurement precision of 0.1°C (USDA 1989, Marks et al. 1992). Some of the stations, however, were manually operated and provided measurement precisions as coarse as 0.28°C (0.5°F). The time period considered in this analysis is the 1990 water year (October 1, 1989 to September 30, 1990). 1990 was chosen because it represents a climatologically 'typical' year based on the historical climate record (Greenland 1994), and because it coincides with a comprehensive database of AVHRR satellite imagery (EROS Data Center 1991).

The temperature data were checked extensively for impossible and implausible values (Davidson 1996, p. 117) and for excessive amounts of missing values. Stations with more than 100 days of missing data were not considered in the analysis, nor were stations with reported elevations that differed from their corresponding DEM elevations by more than 500 m, making a total of 907 stations, 679 of which are within the study area boundaries. 1.37 % of the daily temperature observations consisted of missing values. The spatial distribution of meteorological stations is shown in Fig. 1. Fig. 2 shows the distribution of elevations for the station data and for the DEM cells within the study area. It is clear from this figure that there are proportionally more low-elevation stations than there are low-elevation DEM cells in the study area.

3. INCORPORATING ELEVATION EFFECTS ON AIR TEMPERATURE

The main difficulty in accurately interpolating temperature data in mountainous terrain is the effect of elevation on temperature. Mountains, acting as physical barriers, force air to move vertically, a process called *orographic uplift*. When an air parcel rises, it
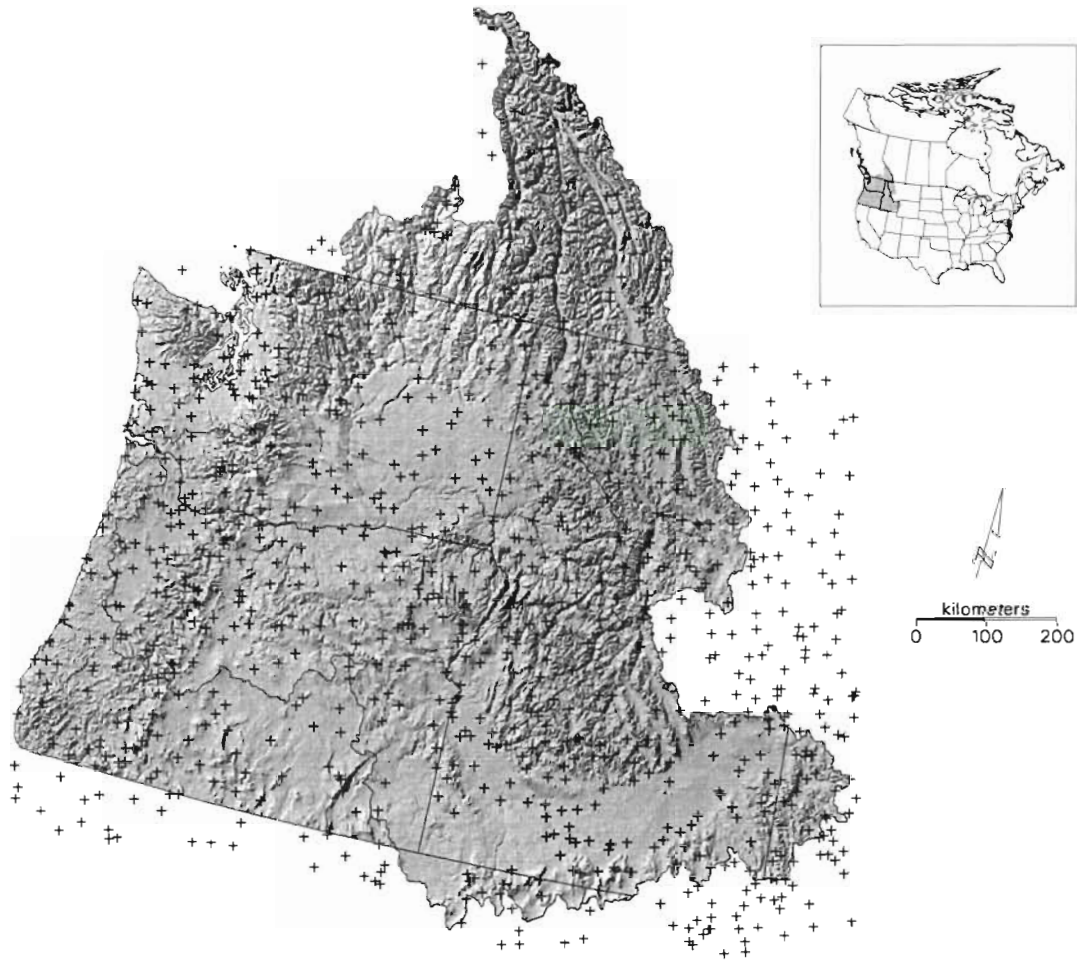
Fig. 1. Shaded-relief map of a 1 km DEM (digital elevation model) over the Columbia River Basin study area (828 595 km^2) north-western U.S. and Canada. Crosses denote meteorological station locations (907 total, 679 within the study area). Map projection is Albers equal-area conic

expands and cools. If no heat is exchanged with the outside system, this cooling is termed *adiabatic*. The rate at which air cools with elevation change, the *lapse rate*, varies from about $-9.8°C$ km^{-1} for dry air (the dry adiabatic lapse rate) to about $-4.0°C$ km^{-1} for very warm saturated air (the saturated adiabatic lapse rate) (Barry & Chorley 1987, p. 76). The lapse rate is seldom purely adiabatic due to outside heat exchange caused by radiational heating or cooling at the surface, horizontal mixing (advection) of air masses, and evaporation or condensation of moisture. The actual lapse rate at a given place and time is termed the *environmental lapse rate*. A typical value used for the global mean environmental lapse rate is $-6.5°C$ km^{-1} (Barry & Chorley 1987, p. 56).

Several methods exist in the literature for dealing with elevation effects when interpolating temperature. One method is to compute temperature deviations, also called *anomalies*, by subtracting a monthly or annual mean from each temperature observation and then to interpolate the temperature anomaly data rather than the raw temperatures. This method does a good job at removing elevation effects in the investigation of temperature trends over time (Robeson 1993), however it is not well-suited to ecological or hydrological applications where the actual temperatures are of interest rather than the temperature anomalies.

Another method, climatologically aided interpolation (CAI), is related to the anomaly approach (Robeson 1993, Willmott & Robeson 1995). CAI involves computing temperature anomalies at each station, interpolating the anomalies, and using the interpolated anomaly surface to modulate a climatology (a pre-existing surface of long-term mean temperatures). Robeson (1993) employs CAI to create mean annual temperature surfaces for the land area of the globe, using the Legates & Willmott (1990) temperature data set as the climatology. Two disadvantages to this
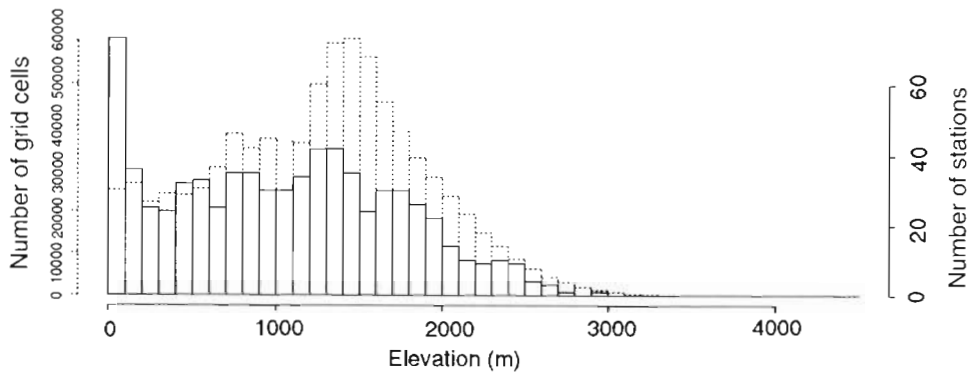
Fig. 2. Histograms of elevation for all meteorological stations (solid bars) and all DEM cells within the study area (dashed bars)

method are first, that the climatology surface often does not account for elevation, as is the case for the Legates & Willmott (1990) climatology, and second, that high-quality climatologies may not exist for a given region of interest, especially at higher spatial and temporal resolutions.

Hutchinson (1989) uses multi-dimensional thin-plate splines to fit temperature surfaces to 3 variables—latitude, longitude, and elevation—over the Australian continent. This approach creates a smooth surface of temperatures by minimizing the roughness of the interpolated surface. Hutchinson estimates a smoothing parameter by minimizing the generalized cross-validation errors of the fitted surface. The thin-plate spline method is reported to work as well as kriging while requiring less parameterization, however thin-plate splines are computationally demanding and complicated to implement.

The creators of the IIASA database (Leemans & Cramer 1991) use a simpler technique for incorporating elevation effects. Temperature values were first normalized to sea-level equivalents, using the station elevation and a constant linear lapse rate adjustment of $-6.0°C$ km^{-1}. The adjusted sea-level temperatures were then interpolated using a combination triangulation/smooth surface fitting approach. Finally, the interpolated sea-level temperatures were adjusted back to actual temperatures using the same lapse rate function and a surface of elevation values stored in a DEM. Based on visual inspection and comparison with the Legates & Willmott database (1990), Leemans & Cramer (1991) conclude that their temperature surfaces do a good job at resolving temperature features in mountainous regions. However, they also note that certain high-altitude and data-sparse regions (e.g. the Tibetan Plateau) are consistently colder than expected, possibly because of over-estimating the magnitude of the lapse rate in these regions. No rigorous accuracy assessment is performed.

Willmott & Matsuura (1995) explore 2 methods for incorporating elevation effects. Their first method,

'topographically informed interpolation', is essentially the same as the Leemans & Cramer (1991) method above, except for a different lapse rate ($-6.5°C$ km^{-1}) and spatial interpolation algorithm (a form of inverse-distance weighting using spherical geometry). The second method, 'topographically and climatologically informed interpolation', combines their first method with the CAI method described above (Robeson 1993, Willmott & Robeson 1995). Willmott & Matsuura perform a detailed accuracy assessment based on cross-validated interpolation errors for annual average air temperatures in the United States from 1920 to 1987. They find that their methods are considerably more accurate than simpler interpolation methods, with the topographically and climatologically informed interpolation method performing best.

Marks (1990) presents a method for elevation correction that is similar to, but slightly more complicated than, the linear lapse rate adjustment used by Leemans & Cramer (1991) and Willmott & Matsuura (1995). The basic procedure is the same: temperature observations are adjusted to sea-level equivalents; the sea-level temperatures are interpolated; and the surface of sea-level temperatures is converted to actual temperatures by mapping it to the elevations of a DEM surface. The difference here is in the way actual temperatures are adjusted to sea-level: instead of using a constant linear adjustment based on elevation and lapse rate, Marks employs the hydrostatic equation (Byers 1974, p. 82-85) to estimate air pressure at each station based on the station's elevation. The station temperature and pressure are then used, under an assumption of neutral atmospheric stability, in the potential temperature equation (Barry & Chorley 1987, p. 77) to compute sea-level potential temperatures. This approach has been implemented by Dolph et al. (1992), Marks et al. (1993), and Phillips & Marks (1996) to create temperature surfaces used in spatially distributed hydrological modeling, however these studies focus on modeling potential evapotranspiration and not specifically on the interpolation methods used to

create the model inputs. An in-depth assessment of the method's accuracy has not been done.

Given the size of our temperature interpolation task, we decided against a thin-plate spline approach due to its computational complexity. We rejected the CAI method due to the lack of a pre-existing high-resolution climatology suitable for use at a daily time step. We also rejected the anomaly approach because the actual temperature data, rather than the deviations, were of interest. We felt that the class of methods which convert temperatures to sea-level reference temperatures, interpolate, and then convert back to actual temperatures showed the most promise in terms of both computational feasibility and accounting for elevation effects. Therefore for this study we chose the constant linear lapse rate adjustment employed by Leemans & Cramer (1991) and Willmott & Matsuura (1995), and the neutral stability assumption procedure used by Marks (1990). We will refer to the former as Linear Lapse Rate Adjustment (LLRA), and the latter as the Neutral Stability Algorithm (NSA).

For the spatial interpolation step in the LLRA and NSA methods, we chose inverse-squared-distance interpolation (Isaaks & Srivastava 1989) for its speed, simplicity, and ease of implementation. Inverse-squared-distance interpolation is part of a general class of Inverse Distance Weighting interpolators, and will be referred to as IDW. Note that Leemans & Cramer did not use IDW in their implementation of LLRA, and that Willmott & Matsuura used a different, more sophisticated form of IDW in their work. We use simple IDW here in order to more easily compare results from the LLRA and NSA methods.

4. THE NEUTRAL STABILITY ALGORITHM (NSA)

The overall procedure of the NSA is: convert air temperature measurements (T_a) to sea-level potential temperatures (Θ_a), spatially interpolate Θ_a points to a grid surface, and use the inverse of the potential temperature function to map the Θ_a surface to DEM elevations. This procedure assumes that the atmosphere in the vicinity of a measurement site is neutrally stable. Neutral atmospheric stability implies that Θ_a is effectively the same for neighboring grid cells regardless of their elevations. Note that the atmosphere is, strictly speaking, not often neutrally stable. Neutral stability is a simplifying assumption which enables the processing of large volumes of data.

Air temperatures T_a (K) were converted to potential temperatures Θ_a (K):

$$\Theta_a = T_a(P_0/P_z)^{(R/m\,C_p)} \qquad (1)$$

where P_0 is 1.0×10^5 Pa (approximately sea-level pressure), P_z is the air pressure (Pa) at elevation z (m), R is

the gas constant (8.3143 J mol^{-1} K^{-1}), m is the molecular weight of dry air (0.02897 kg mol^{-1}), and C_p is the specific heat of dry air at constant pressure (1005 J kg^{-1} K^{-1}).

Measurement site elevations were used to derive the air pressures using a form of the hydrostatic equation (Byers 1974, Barry & Chorley 1987):

$$P_z = P_0[T_b/(T_b + \lambda z)]^{(mg)/(\lambda R)} \qquad (2)$$

where T_b is an assumed sea level temperature (300 K), λ is an assumed temperature lapse rate (–0.0065 K m^{-1}), z is the station elevation (m), and g is the acceleration due to gravity (9.80616 m s^{-2}). This form of the hydrostatic equation uses a constant lapse rate in order to compute pressure, however when combined with the potential temperature equation the effective lapse rate is variable and tends to decrease in magnitude with increasing elevation or decreasing temperature.

We performed a short sensitivity analysis of the NSA method to the parameters λ and T_b. Assuming a station at 1500 m elevation with T_a values of –10.0 and 10.0°C, we used NSA to derive neighboring T_a values at 0, 500, 1000, 2000, 2500, and 3000 m elevations. We used a range of values for λ (–0.002 to –0.010 K m^{-1}) and T_b (290 to 310 K), and compared the predicted T_a values to those using NSA with our assumed defaults of –0.0065 K m^{-1} for λ, and 300 K for T_b. For most combinations of λ and T_b, the final T_a values differed by no more than 0.5°C from the NSA using default parameters, and these discrepancies occurred at only the largest vertical extrapolations (station elevation ± 1500 m). Two combinations of parameters ($\lambda = -0.010$ K m^{-2}, $T_b = 290$ K) and ($\lambda = -0.002$ K m^{-2}, $T_b = 310$ K) produced larger discrepancies (up to 1.0°C), again only at (station elevation ± 1500 m). At elevations closer to station elevation, NSA was less sensitive to the λ and T_b parameters.

4.1. The spatial interpolation algorithm

Potential temperatures at the meteorological station locations were interpolated to the geographic grid spacing of the DEM data using a simple inverse distance weighting (IDW) approach. Within the general class of IDW methods, we used an inverse-squared-distance weighting function and a neighborhood size of 8 (i.e. the 8 nearest neighbors). We decided that 8 was a reasonable number of neighbors to use in terms of reducing computation time while maintaining a smooth surface. From our experience with inverse-squared-distance interpolation, moving beyond 5 or 6 neighbors adds little additional information.

Due to the uneven density of stations across the study area, a maximum neighbor distance threshold

was not used. In the case where some neighbors are very far from the interpolation point, the $1/d^2$ weighting function ensures that distant stations receive proportionally little weight. When a missing value was encountered during the interpolation, that station was ignored and the nearest 8 stations with valid data were used.

The IDW interpolation is:

$$V_{gc} = [\Sigma_i(v_i/d_i^2)]/[\Sigma_i(1.0/d_i^2)] \qquad (3)$$

where V_{gc} is the interpolated value at a grid cell, v_i is the measured value of the ith nearest neighbor, and d_i is the distance to the ith nearest neighbor. In the case where $d = 0.0$ for a particular v_i, V_{gc} is assigned the value of v_i, making IDW an exact interpolator.

The IDW interpolation algorithm was chosen because it is simple, relatively fast, and easy to implement. Robeson (1994) investigated 3 methods of spatially interpolating temperature anomaly data and found that, on average, the inverse-distance method gave about the same results as the other 2 methods, triangulated surface patches and thin-plate splines.

Using a subset of the Columbia River Basin data set and a 10 km grid cell resolution, we compared the results of interpolating potential temperatures by kriging (Phillips & Marks 1996) versus IDW (Table 1). The kriging method tended to smooth the data slightly by reducing the extreme values, however the surfaces interpolated by kriging and IDW were very similar, with mean absolute pixel differences of 0.54°C or less between predicted Θ_a for each of the days during which the comparison was made. The similarity between the kriging and IDW results is a compelling argument in favor of the simpler and faster IDW method.

4.2. The cold bias

The interpolated surfaces of Θ_a were converted back to air temperature (T_a) by inverting Eq. (1) and using the DEM surface to map Θ_a to the elevations of the DEM grid. A typical air temperature surface made by NSA (minimum temperature on January 1, 1990) is shown in Fig. 3. The general spatial pattern of temperatures is intuitively consistent. Low elevations (coastal areas, interior valleys) tend to be warmer than high elevations, and inland areas have cooler minimum temperature values than coastal areas. However, closer examination revealed that the estimated temperatures for some mountain cells were far cooler than expected. Daily temperatures for selected cells were plotted for the entire year and indicated a possible bias in the interpolated surfaces. For example, a cell in the Olympic Peninsula at 2080 m elevation had an average annual T_{max} of –3.9°C (median = –4.5°C). The average

Table 1. Comparison of potential temperature interpolation using kriging and inverse-distance weighting (IDW) on 743 station values to estimate 5693 ten km grid cells for the U.S. portion of the Columbia River Basin. Data are from Phillips & Marks (1996). Units are °C. MPD: mean pixel difference; MAPD: mean absolute pixel difference. Pixel differences were computed as (IDW surface minus kriging surface)

	Kriging		IDW
January 10, 1990			
Min.	2.94		–3.69
Max.	21.97		24.84
Median	11.97		11.82
Mean	11.91		11.87
SD	4.36		4.47
MPD		–0.04	
MAPD		0.54	
April 5, 1990			
Min.	8.65		4.76
Max.	27.68		29.78
Median	18.80		18.87
Mean	19.07		19.05
SD	3.67		3.77
MPD		–0.02	
MAPD		0.50	
August 3, 1990			
Min.	16.75		15.56
Max.	39.84		41.77
Median	30.51		30.52
Mean	30.74		30.72
SD	3.84		3.93
MPD		–0.02	
MAPD		0.49	

summer T_{max} (June, July, August) was 3.2°C (median = 2.8°C). Over the entire year, 256 days had a T_{max} less than 0.0°C. For comparison purposes, we looked at T_{max} statistics for the nearest comparable station. The station with the most similar elevation and latitude to the Olympic Peninsula grid cell was in western Montana at an elevation of 2103 m and a latitude 35 km south of the grid cell's latitude. For this station, the mean T_{max} was 7.2°C; the mean summer T_{max} was 17.1°C, and the number of sub-zero T_{max} days was 102. The Olympic Peninsula grid cell was rather extreme, but many other high-elevation cells exhibited similar behavior and led us to suspect some sort of bias in the NSA temperature surfaces.

5. CROSS-VALIDATION ANALYSIS OF THE NSA

In order to assess the accuracy of the NSA, we performed a cross-validation analysis using all stations within the study area (n = 679), and interpolating both T_{min} and T_{max} each day for the entire year. The cross-validation procedure was as follows: For each meteoro-
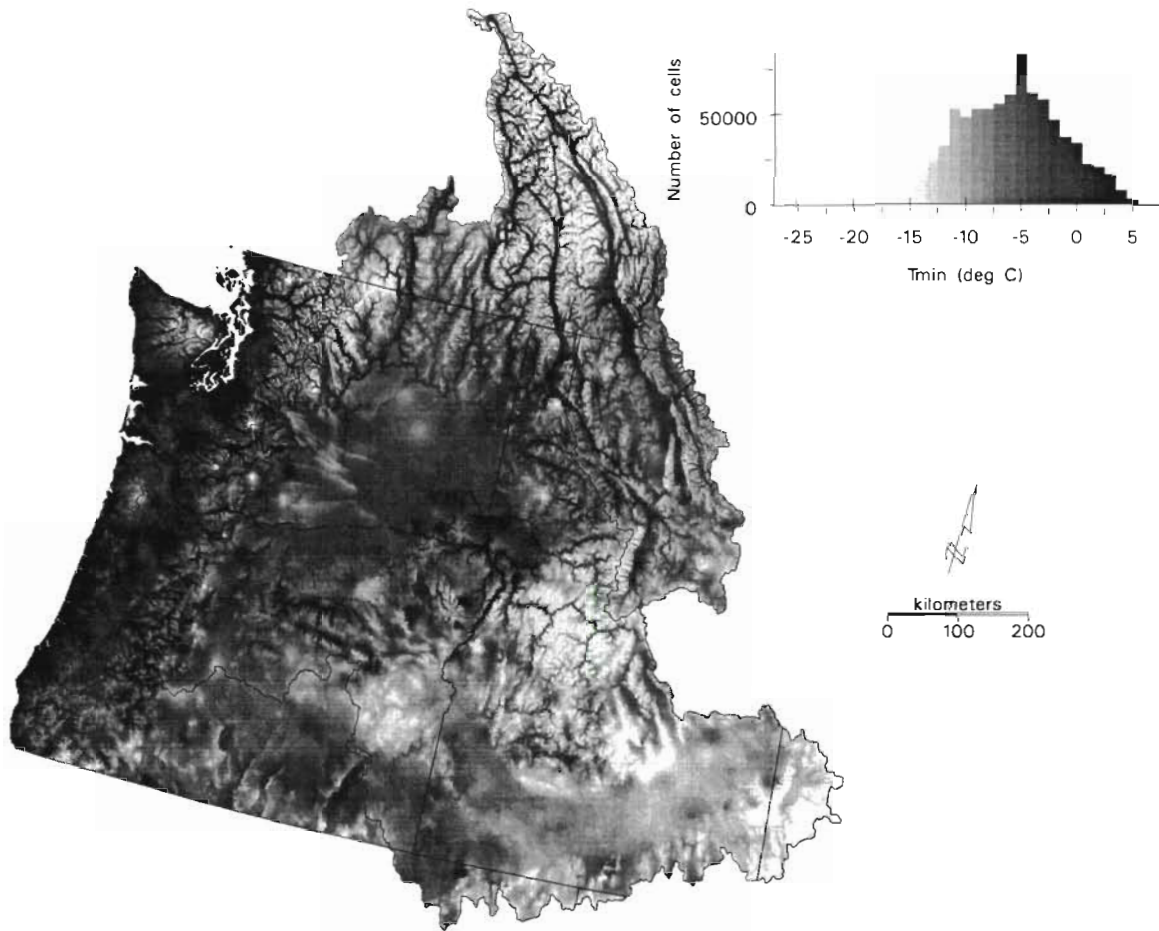
Fig. 3. Map of minimum temperature on January 1, 1990, interpolated with the neutral stability algorithm

logical station point, find the DEM grid cell that contains the station location and estimate T_{min} and T_{max} for each day of the year from the nearest 8 stations which are outside the grid cell. Compute the interpolation error (IE) as the observed temperature (at the station inside the grid cell) minus the interpolated temperature (interpolated from the 8 stations outside the grid cell). Note that the DEM cell and the station it contains do not necessarily have identical elevations. The mean elevation difference (DEM – station) is 32 m (standard deviation = 98 m).

The cross-validation analysis produced, for each of the 679 stations, 365 interpolation errors for both T_{min} and T_{max}. Fig. 4a shows the cross-validation results for a high-elevation (2286 m) station in southern Idaho. The interpolated value (dashed line) is consistently lower than the known temperature (solid line) throughout the year. The interpolation error (observed minus interpolated value) for this point is graphed in Fig. 4b. While there is considerable noise in the signal of daily interpolation error, it is clear that this point is, in general, being predicted too cold. Rather than deal

with the noisy daily errors, in the following analyses we use the mean annual interpolation errors for T_{min} and T_{max} (2 sets of 679 values). These 2 sets of mean annual errors were combined into a single data set (n = 1358).

A histogram of the mean annual interpolation errors (Fig. 5) shows a relatively symmetrical distribution centered near zero with a mean value of 0.22°C (i.e. on average, each point was predicted 0.22°C too cold). A plot of station elevation versus interpolation error (Fig. 6) shows a very weak positive relationship for elevation values above about 1000 m, but in general the relationship between elevation and interpolation error is nonexistent.

5.1. The ΔZ measure

A map of the T_{min} interpolation errors (Fig. 7) shows a few spatial patterns of interest. A band of relatively high-magnitude errors occurs along the Cascade Range, while the interior valleys have mostly low-mag-
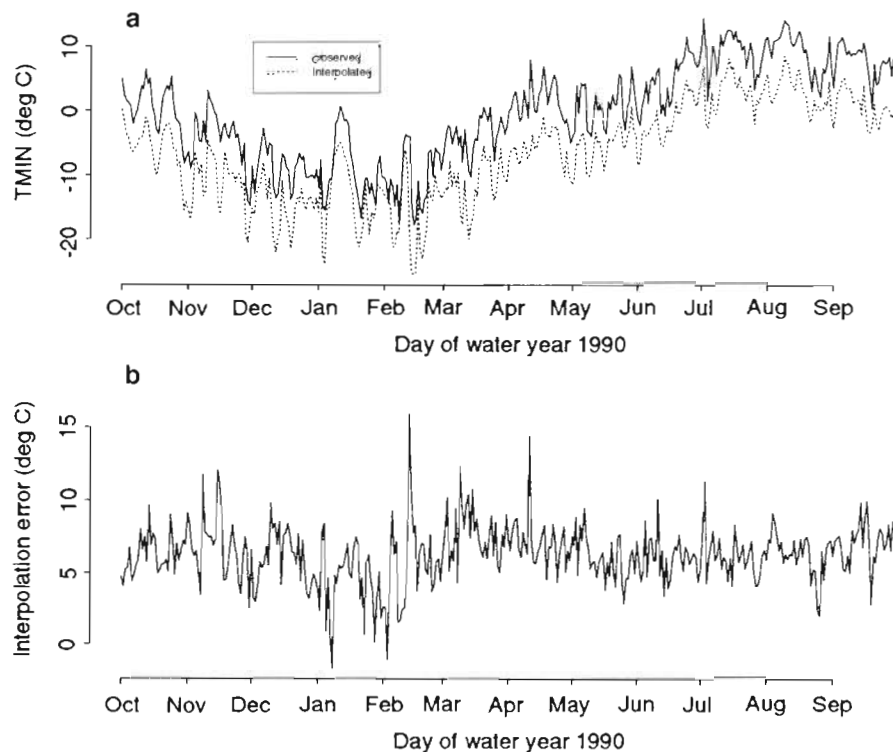
Fig. 4. (a) Daily minimum temperature for a representative high-elevation meteorological station in southern Idaho (2286 m). The solid line is the observed temperature and the dashed line is the cross-validated interpolated temperature using the NSA method. (b) Daily interpolation error (observed minus cross-validated interpolated temperature) for the station depicted in (a)
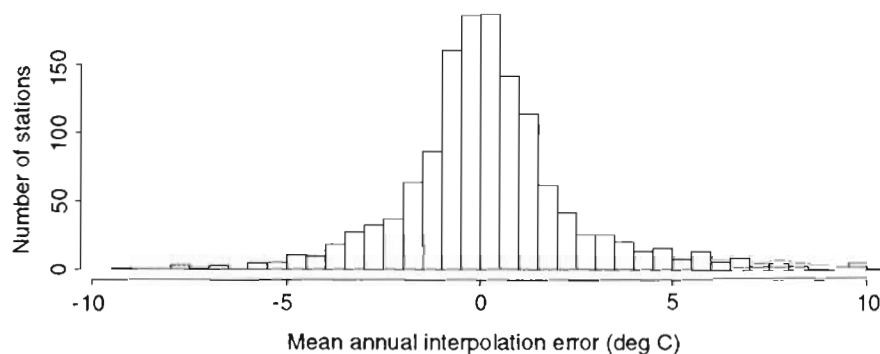

Fig. 5. Histogram of NSA mean annual interpolation error (observed minus cross-validated interpolated temperature) for combined T_{min} and T_{max} over the 679 stations within the study area

nitude errors. The mountains of central Idaho tend to have errors at about the same magnitude as those in the Cascades, even though the Cascades have lower elevations. It appears that the interpolation errors are related to changes in elevation across the landscape rather than to absolute magnitudes of elevation.

These spatial patterns of interpolation error led to the development of a measure we call ΔZ, which is the difference between a DEM grid-cell elevation and the weighted mean of the DEM elevations of a neighboring set of stations:

$$\Delta Z = Z_{gc} - \{[\Sigma_i(z_i/d_i^2)] / [\Sigma_i(1.0/d_i^2)]\} \quad (4)$$

where Z_{gc} is the DEM elevation (m) of the grid cell under consideration, z_i is the DEM elevation of the ith nearest station ($i = 1$ to 8), and d_i is the distance to the ith nearest station. The purpose of the ΔZ measure is to quantitatively flag regions of the study area which are likely to be susceptible to interpolation error. Grid cells with high positive (negative) values of ΔZ are cells which are much higher (lower) in elevation than the neighboring stations used for interpolating tempera-
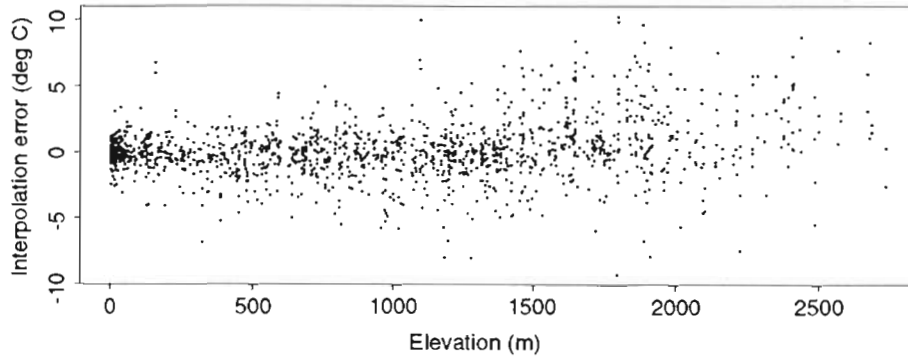
Fig. 6. Station elevation versus mean annual interpolation error for combined T_{min} and T_{max} over the 679 study area stations
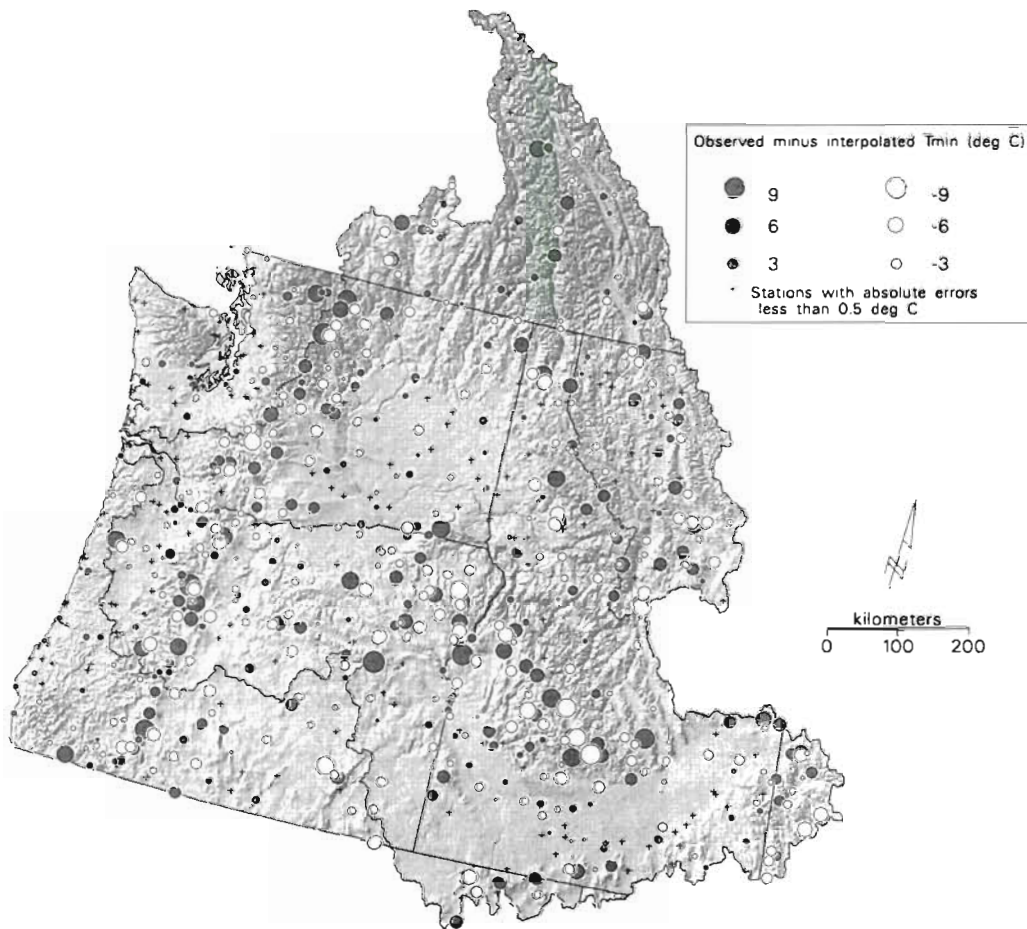

Fig. 7. Map of mean annual interpolation error for T_{min}. Circle diameters are proportional to error magnitude. Errors less than ±0.5°C are denoted by a cross

ture, and therefore represent areas where significant vertical extrapolation occurs when potential temperatures are converted to DEM-elevation temperatures. We suspect that it is this type of extrapolation that is responsible for the very cold temperatures observed in the initial set of interpolated temperature surfaces.

Cells with low-magnitude values of ΔZ are at about the same elevation as neighboring stations and are not expected to have serious problems associated with extrapolation.

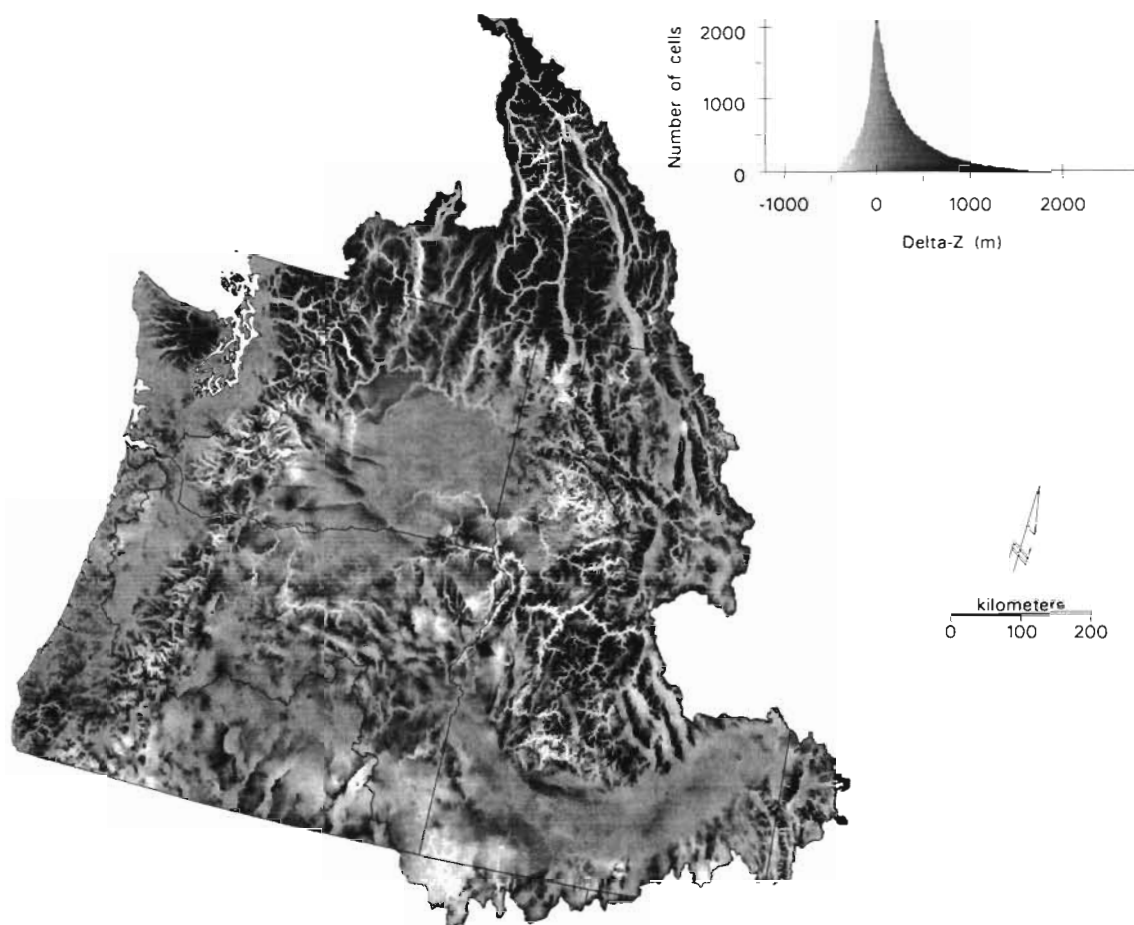Fig. 8 shows the spatial distribution of ΔZ across the study area. As expected, we see high positive ΔZ

Fig. 8. Map of ΔZ (grid cell elevation minus weighted mean station elevation). Positive ΔZ values denote grid cells which are higher in elevation than the nearest 8 meteorological stations; negative ΔZ denotes cells which are lower than the 8 neighboring stations.

values associated with mountain tops, e.g. Mount Rainier (DEM elevation = 4210 m; ΔZ = 2739 m) and Mount Adams (DEM elevation 3650 m; ΔZ = 2446 m). A more interesting, and less obvious, pattern is the large group of very high ΔZ values in the mountains of Washington's Olympic Peninsula. Only a few grid cells in the Olympic Mountains have elevations above 2000 m, yet most of the ΔZ values for these cells are extreme (greater than 1000 m). This is due to the fact that these mountains are devoid of any meteorological stations and are surrounded by stations very close to sea level (the highest neighboring station is at 134 m). As for areas which are lower than neighboring stations, the high-magnitude negative ΔZ values in Fig. 8 are fewer in number and are mostly confined to valley bottoms and narrow canyons.

If we compute a ΔZ value for each grid cell that contains a meteorological station, while ignoring the station within that grid cell, then the resulting set of ΔZ

values corresponds to the situation in which we ignored a station during the cross-validation of temperature interpolations. We will denote these values as $CV\Delta Z$. When the $CV\Delta Z$ are plotted against the interpolation errors, we see an obvious and relatively strong positive relationship (Fig. 9a).

The relationship depicted in Fig. 9a confirms our suspicion of a bias: high positive values of ΔZ tend to have high positive interpolation errors (i.e. tend to be predicted too cold). Fig. 9a also shows that the converse is true: high negative values of ΔZ tend to be predicted too warm. A simple linear regression of $CV\Delta Z$ on interpolation error (IE) yields a meaningful fit (R^2 = 0.43) and the following equation: IE = 0.297 + 0.004449($CV\Delta Z$) (n = 1358) Thus the interpolation errors tend to increase 4.449°C for each 1000 m increase in ΔZ. Note that spatial autocorrelation in the temperature data prevents the unbiased assessment of statistical significance in the regression analysis.
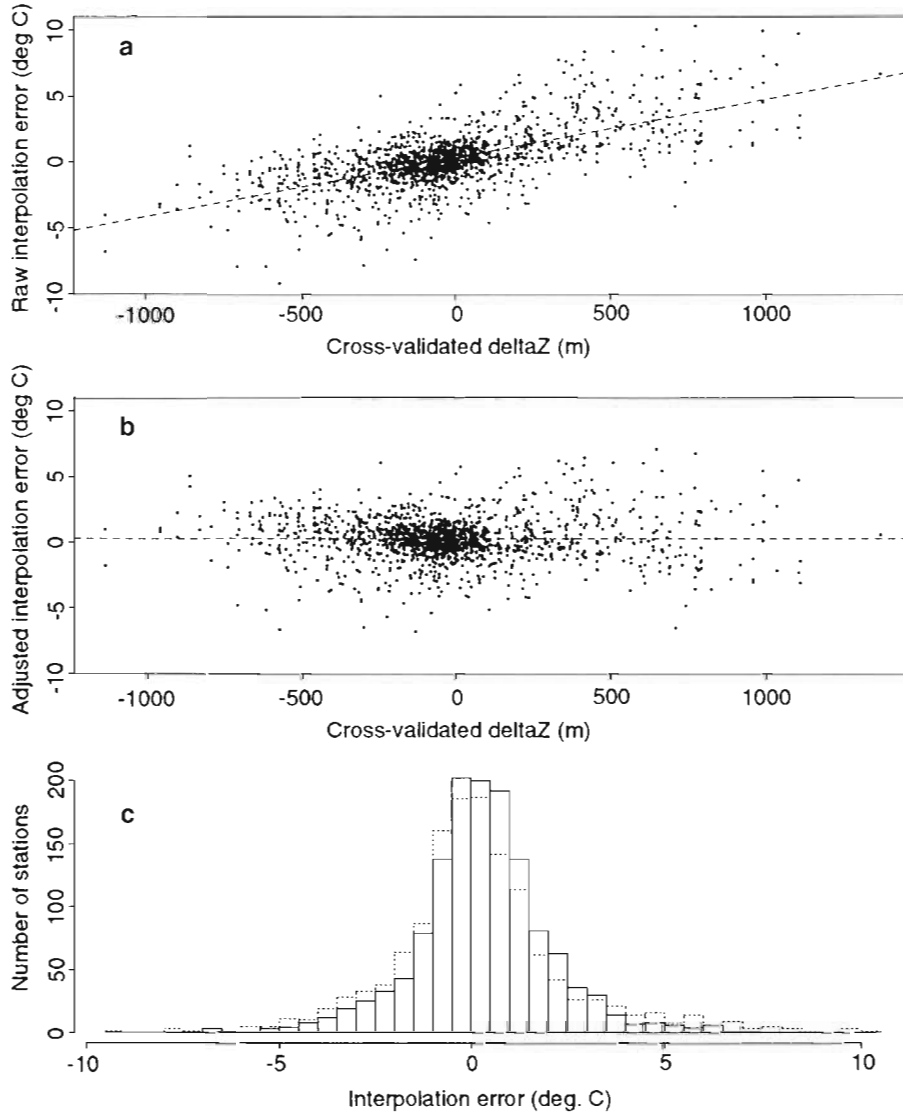
Fig. 9. (a) CVΔZ versus mean annual interpolation error for the unadjusted NSA method, with a linear regression line (dashed). Data are combined T_{min} and T_{max} from 679 stations (n = 1358). The regression equation was: IE = 0.297 + 0.004449 CVΔZ (R^2 = 0.43). (b) CVΔZ versus mean annual interpolation error for the bias-corrected NSA method, with a linear regression line (dashed). The regression equation was: IE = 0.297 − 0.0000003 CVΔZ (R^2 = 0.00). (c) Histograms of mean annual interpolation error for NSA: uncorrected (dashed bars, mean = 0.21, MAE = 1.58) and bias-corrected (solid bars, mean = 0.30, MAE = 1.27)

To remove the interpolation bias, we used the slope of the regression above[1] as a bias correction factor (BCF). The cross-validated temperature estimates were adjusted by 4.449°C for each km of CVΔZ. These adjusted temperatures were subtracted from the observed data to obtain adjusted interpolation errors, which were plotted against CVΔZ in Fig. 9b. A visual inspection of this figure reveals no apparent relationship, and a regression of CVΔZ on adjusted interpola-

tion error yields an R^2 of 0.0. The histogram of adjusted interpolation errors (Fig. 9c) shows that the bias correction resulted in a modest increase of interpolation accuracy. When compared to the raw, unadjusted errors, the solid histogram of Fig. 9c has smaller tails (fewer large errors) and larger central bars (more small errors) than the dashed histogram. Mean absolute error (MAE) decreased from 1.58 to 1.27°C, and the standard deviation decreased from 2.28 to 1.73°C).

The above BCF (4.449°C km^{-1}) was computed using temperatures over an entire year and over a large geographic region. To assess the stability of the correction factor in space and time, it was re-computed using

___
[1]The intercept is assumed zero so that known data values (where ΔZ is zero) are not altered during the bias correction procedure described in Section 5.2

Table 2. Regression and interpolation error (IE) statistics for spatial and temporal subsets of the study area, for each interpolation method. The 'Int.', 'Slope', and 'R^2' columns pertain to the intercept, slope, and fit of the regression where cross-validated ΔZ is used to predict IE. IE pertains to mean interpolation error for pooled daily T_{min} and T_{max}. 'IE$_{MAE}$' and 'IE$_{stdev}$' are the mean absolute error and standard deviation of IE, respectively. Error standard deviation is also known as root mean square error (RMSE). Slope units are (°C km^{-1}); Int. and IE units are (°C)

Subset/Method	Int.	Slope	R^2	IE$_{MAE}$	IE$_{stdev}$
NSA-raw					
Daily, entire study area	0.30	4.45	0.43	1.58	2.28
Monthly, entire study area	0.29	4.51	0.43	1.59	2.30
Daily, Jan	0.31	3.42	0.29	1.51	2.14
Daily, Aug	0.29	5.10	0.45	1.75	2.52
Daily, Oct–Mar	0.30	3.78	0.36	1.45	2.10
Daily, Sep–Apr	0.29	5.12	0.46	1.76	2.53
Daily, NW quadrant	0.25	4.62	0.52	1.40	2.05
Daily, NE quadrant	0.34	3.79	0.46	1.46	2.08
Daily, SW quadrant	0.37	4.90	0.44	1.67	2.39
Daily, SE quadrant	0.25	4.57	0.35	1.78	2.55
NSA-adj					
Daily, entire study area	0.30	0.00	0.00	1.27	1.73
Monthly, entire study area	0.29	0.00	0.00	1.28	1.73
Daily, Jan	0.31	−1.03	0.03	1.41	1.84
Daily, Aug	0.29	0.65	0.01	1.33	1.88
Daily, Oct–Mar	0.30	−0.67	0.02	1.26	1.69
Daily, Sep–Apr	0.29	0.67	0.01	1.36	1.87
Daily, NW quadrant	0.25	0.17	0.00	1.02	1.42
Daily, NE quadrant	0.34	−0.65	0.02	1.21	1.55
Daily, SW quadrant	0.37	0.45	0.01	1.35	1.81
Daily, SE quadrant	0.25	0.12	0.00	1.51	2.05
LLRA3.9					
Daily, entire study area	0.13	−0.96	0.03	1.30	1.75
Monthly, entire study area	0.14	−0.91	0.03	1.31	1.75
Daily, Jan	0.15	−2.02	0.12	1.47	1.93
Daily, Aug	0.13	−0.06	0.00	1.31	1.83
Daily, Oct–Mar	0.14	−1.55	0.09	1.30	1.75
Daily, Sep–Apr	0.13	−0.37	0.00	1.37	1.85
Daily, NW quadrant	0.08	−0.72	0.03	1.02	1.41
Daily, NE quadrant	0.13	−1.61	0.13	1.26	1.64
Daily, SW quadrant	0.22	−0.52	0.01	1.38	1.81
Daily, SE quadrant	0.13	−0.90	0.02	1.56	2.07
LLRA4.8					
Daily, entire study area	0.17	0.00	0.00	1.26	1.72
Monthly, entire study area	0.17	0.06	0.00	1.26	1.73
Daily, Jan	0.18	−1.04	0.04	1.40	1.85
Daily, Aug	0.16	0.84	0.02	1.30	1.86
Daily, Oct–Mar	0.17	−0.59	0.01	1.23	1.68
Daily, Sep–Apr	0.16	0.59	0.01	1.34	1.86
Daily, NW quadrant	0.11	0.24	0.00	1.00	1.40
Daily, NE quadrant	0.17	−0.64	0.02	1.17	1.55
Daily, SW quadrant	0.25	0.44	0.01	1.34	1.81
Daily, SE quadrant	0.15	0.05	0.00	1.51	2.05
LLRA6.5					
Daily, entire study area	0.22	1.61	0.09	1.26	1.81
Monthly, entire study area	0.22	1.67	0.09	1.27	1.82
Daily, Jan	0.23	0.57	0.01	1.36	1.82
Daily, Aug	0.21	2.45	0.16	1.38	2.01
Daily, Oct–Mar	0.23	1.02	0.04	1.20	1.71
Daily, Sep–Apr	0.21	2.20	0.14	1.39	2.00
Daily, NW quadrant	0.17	1.84	0.15	1.05	1.53
Daily, NE quadrant	0.24	0.98	0.05	1.13	1.58
Daily, SW quadrant	0.30	2.05	0.12	1.36	1.92
Daily, SE quadrant	0.20	1.67	0.07	1.50	2.13

smaller time periods and spatial subsets of the study area (Table 2, 'NSA-raw'). The BCF did not change by more than about 1.0°C km^{-1} when computed over just the warm months (April to September), the cold months (October to March), the warmest month (August), and the coolest month (January). When computed using mean monthly temperatures instead of daily temperatures (reducing the computation time by a factor of 30), the BCF differed by only 0.064°C km^{-1}. When the study area was partitioned spatially into quadrants (using the median of the X,Y station coordinates to define the quadrants), the BCF was stable to within 0.65°C km^{-1}.

5.2. Bias-correcting the interpolated NSA temperatures

Since the bias correction of the cross-validated temperature estimates was successful and the correction factor appeared to be fairly stable over space and time, we applied the procedure to the daily temperature surfaces. The bias correction of interpolated temperature for a given grid cell is:

$$T_{adj} = T_{raw} + BCF(\Delta Z) \qquad (5)$$

where T_{adj} is the temperature (°C) adjusted for bias correction, T_{raw} is the raw temperature (°C) interpolated with NSA, BCF is the bias correction factor (0.004449°C m^{-1}) and ΔZ is the ΔZ value (m) of the grid cell. The ΔZ measure works well as a basis for adjusting temperatures for 2 reasons. First, the ΔZ surface is relatively smooth and continuous, thus a correction based on ΔZ will tend to produce a surface free of sharp breaks and discontinuities. This is desirable since air temperature varies smoothly over space. The second reason that ΔZ works well is that the correction will not alter known data points. We ensured this by setting the intercept value from the regression in Fig. 9a (0.297°C) to zero. Grid cells that contain a station point will have a ΔZ value near zero, since the contained station will dominate the $1/d^2$ weight computation. Therefore the bias correction at these cells (i.e. the value added to T_{raw}) will be near zero.

6. THE LINEAR LAPSE RATE ADJUSTMENT (LLRA) METHOD

The LLRA analysis was done in order to answer 2 questions. First, is the NSA method, with its additional amount of computational complexity and its bias, any more accurate than the simpler LLRA method and therefore worth the trouble of bias correction? Second, will the bias occur with the LLRA method or is it strictly a problem with NSA?

The LLRA method involves the same procedure as the NSA, except that instead of using Eqs. (1) and (2) to compute potential temperatures, the following equation is used to compute sea-level reference temperatures (T_{sl}):

$$T_{sl} = T_a - \lambda z \qquad (6)$$

where T_{sl} is the sea-level reference temperature (°C), T_a is the measured air temperature (°C), λ is the lapse rate (°C km^{-1}), and z is the station elevation (km). The only parameter that needs to be determined is λ.

6.1. Choosing the lapse rate

The observed bias in the NSA method occurs presumably because the NSA lapse rate is too steep. The effective lapse rate used by NSA varies with temperature and elevation, but is usually within about a degree of the dry adiabatic lapse rate (DALR) of –9.8°C km^{-1}.

Thus it would seem that an optimal lapse rate exists, less steep than the DALR, which would not cause a bias in the interpolated temperature surfaces. Since there are many ways to empirically derive a lapse rate from a set of observed air temperatures and elevations, we initially used 2 approaches for deriving λ and performed a cross-validation analysis on each approach.

In the first approach, T_{min} and T_{max} were used to compute mean annual temperature (T_{ann}) for all stations which had no missing values for T_{min} or T_{max} (n = 669). Station elevation (Z, in meters) was regressed on T_{ann}, resulting in a reasonably strong fit (R^2 = 0.72) and the following equation: T_{ann} = 12.18 – 0.003931Z (Fig. 10). The regression slope (–3.931°C km^{-1}) was taken as the lapse rate.

The T_{ann}/elevation regression approach aggregates the data spatially and temporally and computes a single lapse rate. The second method used a less aggre-

gated approach to compute a large number of regional monthly lapse rates. For the same set of 669 stations, elevation was regressed on monthly mean T_{min} and T_{max} using a moving spatial window which considered all stations within a 100 km radius. The 100 km threshold was chosen as a distance small enough to represent a relatively local area yet large enough to produce a reasonable sample size for bivariate regression. The sample sizes ranged from 1 to 41, with a mean of 19.5. For a particular regression to be included in the calculation of the overall lapse rate, we considered only those with sample sizes of at least 10 and with R^2 values of at least 0.7, where 0.7 was chosen as a threshold for a meaningful relationship. The constraint on sample size brings the total number of stations available for this analysis down to 605.

The local regressions were run on each valid station, for each month, for both T_{min} and T_{max}. The slopes (lapse rates) for all regressions which satisfied the 0.7 R^2 threshold were averaged in order to produce a single lapse rate estimate for the region of –6.506°C km^{-1}. The T_{max} data produced a larger number of valid regressions (where R^2 > 0.7) than did the T_{min} data. The mean number of valid regressions per month (out of a possible 605) was 167 for T_{min} and 390 for T_{max}. These numbers suggest that, at a region size of 100 km, the T_{max} lapse rate is more stable over space than that of T_{min}. One possible explanation for this is that T_{min} is more susceptible to cold air drainage effects which would tend to create local temperature inversions and confound the elevation/temperature relationship within the moving spatial window.

6.2. Cross-validation analysis of LLRA

Given the 2 derived lapse rates of –3.931 and –6.506°C km^{-1}, we performed cross-validation analyses on the LLRA method just as was done with NSA.
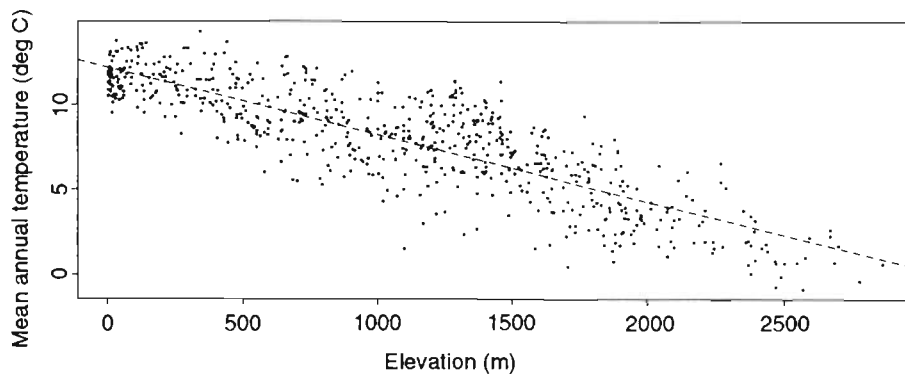

Fig. 10. Elevation versus mean annual temperature for all stations with no missing values (n = 669), with a linear regression line (dashed). The regression equation was: T_{ann} = 12.18 – 0.003931Z (R^2 = 0.72)
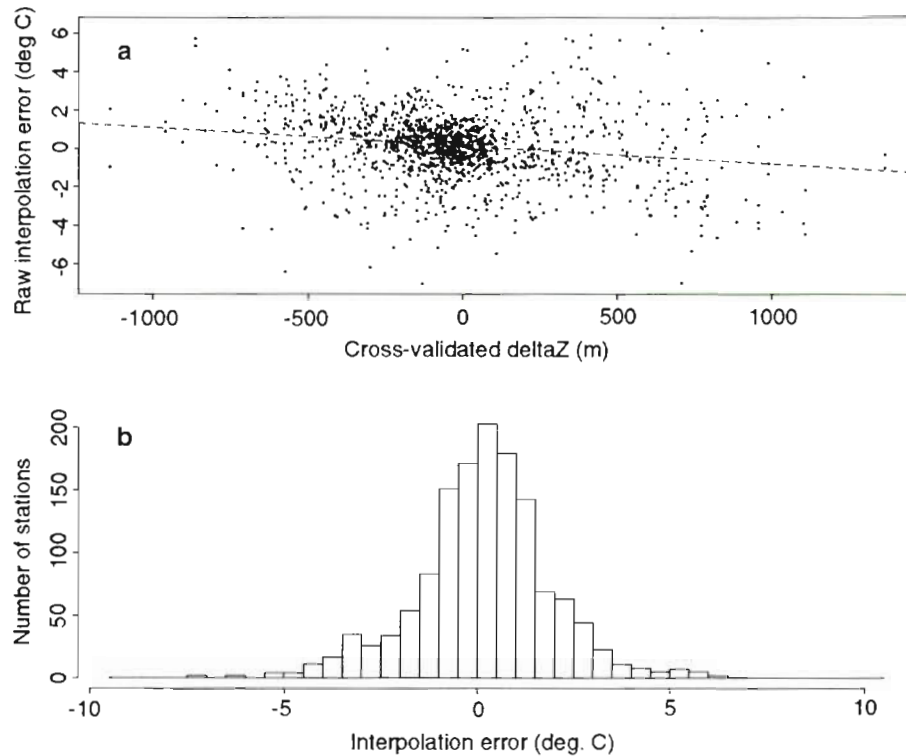
Fig. 11 (a) CVΔZ versus mean annual interpolation error for the LLRA3.9 method, with a linear regression line (dashed). Data are combined T_{min} and T_{max} from 679 stations (n = 1358). The regression equation was: IE = 0.135 – 0.00096 CVΔZ (R^2 = 0.03). (b) Histogram of mean annual interpolation error for LLRA3.9 (mean = 0.15, MAE = 1.30)

These 2 analyses will be denoted LLRA3.9 and LLRA6.5. The LLRA3.9 method (Fig. 11) produced a mean interpolation error of 0.15°C and a standard deviation of 1.75. A regression of CVΔZ on IE produced the equation: IE = 0.135 – 0.00096 CVΔZ (R^2 = 0.03). For the LLRA6.5 method (Fig. 12), the mean IE was 0.19°C with a standard deviation of 1.81. The regression equation was: IE = 0.220 + 0.001611 CVΔZ (R^2 = 0.09). Neither of the regressions showed a strong relationship between CVΔZ and IE, however both had non-zero slopes, and the NSA bias correction procedure resulted in a very slight reduction in the interpolation errors for both methods. The bias-corrected LLRA data were generated only to test whether LLRA can benefit from bias-correction. The following analyses and discussion pertain to uncorrected LLRA data.

In the regressions of CVΔZ and IE, the negative slope of the LLRA3.9 method (–0.960°C km^{-1}) and the positive slope of the LLRA6.5 method (1.611°C km^{-1}) suggest that a lapse rate between –3.931 and –6.506°C km^{-1} exists where the associated regression slope is zero. For a final LLRA analysis, we assumed a linear relationship between the lapse rates input to LLRA (i.e. –3.931 and –6.506°C km^{-1}) and the associated regression slopes (i.e. –0.960 and 1.611°C km^{-1}). Solving for

(slope = 0.0) gave an 'optimal' lapse rate of –4.893°C km^{-1}, where the R^2 relationship between CVΔZ and IE should be exactly zero. We ran another cross-validation analysis using this optimal lapse rate (Fig. 13), and the resulting interpolation errors had a mean of 0.17°C and a standard deviation of 1.72. The regression equation was: IE = 0.167 – 0.0000015 CVΔZ (R^2 = 0.00). This method will be denoted LLRA4.8 in the discussion below. Table 2 shows the spatial and temporal sensitivity of the CVΔZ/IE regressions and summary statistics of IE for each of the LLRA methods.

7. DISCUSSION

The NSA method uses standard principles and formulas from the meteorological literature (Byers 1974, Barry & Chorley 1987) combined with a commonly-used spatial interpolation routine (IDW) to distribute point-based measurements of daily air temperature. When a single station is used to predict a nearby grid cell's temperature, the effective lapse rate of the NSA can be back-calculated from the elevations and temperatures of the station and grid cell [(T_{grid} – $T_{station}$)/ (Z_{grid} – $Z_{station}$)]. To get at the root of the ΔZ bias, we
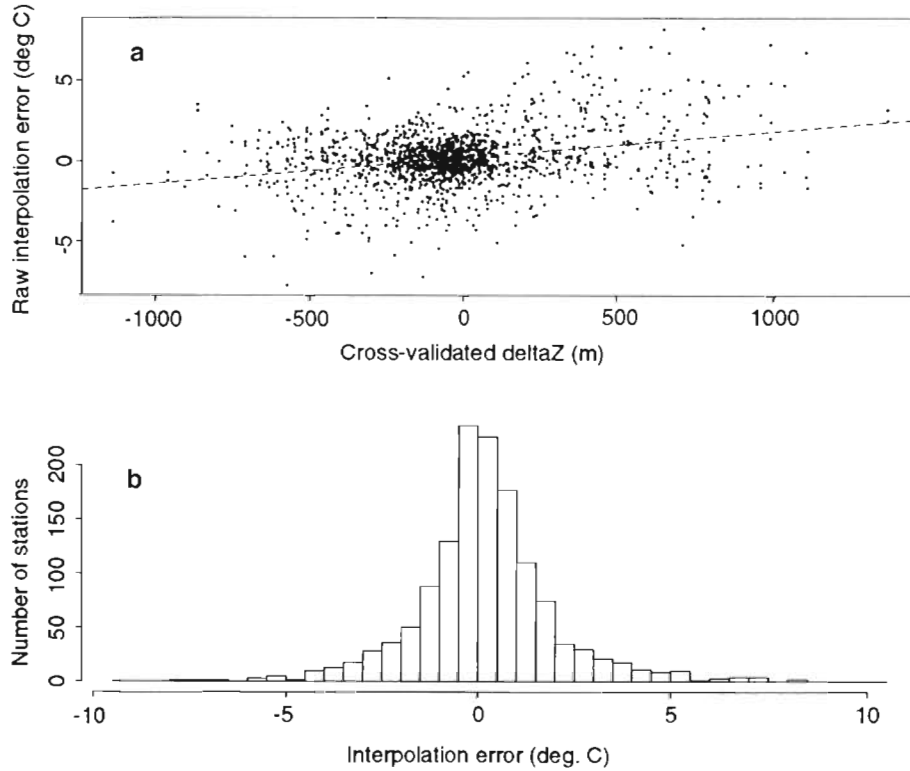
Fig. 12. (a) CVΔZ versus mean annual interpolation error for the LLRA6.5 method, with a linear regression line (dashed). Data are combined T_{min} and T_{max} from 679 stations (n = 1358). The regression equation was: IE = 0.220 + 0.001611 CVΔZ (R^2 = 0.09). (b) Histogram of mean annual interpolation error for LLRA6.5 (mean = 0.19, MAE = 1.26)

computed the effective NSA lapse rate, using a wide range of station temperature (–40 to 40°C), elevation (0 to 4000 m), λ (–2.0 to –10.0°C km^{-1}), and T_b (290 to 310 K) values. In most cases the effective NSA lapse rate was steeper than –8.0°C km^{-1}. In the extreme case where a sea-level station temperature of –40°C is used to predict a DEM cell at 4000 m (with λ = –2.0 and T_b = 310), the effective lapse rate goes as shallow as –7.0°C km^{-1}. The consistently-steep lapse rates imposed by NSA are the source of the ΔZ bias. The physics of the hydrostatic and potential temperature equations fails to account for the non-adiabatic processes at work in the environment, such as horizontal advection or inversions caused by cold air drainage.

The LLRA methods, on the other hand, make use of a lapse rate derived explicitly from the observed data which better reflects the physical realities of the study area. The LLRA results confirm the relationship between lapse rate and ΔZ bias. In Table 2, the slope and R^2 columns reflect the strength of the ΔZ bias. At the shallow lapse rate of LLRA3.9, the slope is always negative, indicating that high-elevation cells tend to be predicted slightly too warm. At the optimal LLRA4.8 lapse rate, the slope wavers around zero, indicating no

systematic relationship with ΔZ. And at the steeper LLRA6.5 lapse rate, the slope is consistently positive, indicating that high-elevation cells tend to be predicted slightly too cold. Note that in all cases the R^2 value denotes a very poor or non-existent fit, however it tends to stay closer to zero at LLRA4.8 and to rise slightly when the lapse rate changes to LLRA3.9 and LLRA6.5. At the very high lapse rates used by the NSA-raw method, the slope is much higher and the R^2 rises enough to indicate signs of a meaningful relationship.

Table 3. Interpolation error (IE) statistics for various interpolation methods (n = 1358 stations). Units are °C. MAE: mean absolute error; Med: median. The mean IE is also known as mean bias error (MBE). Error standard deviation is also known as root mean square error (RMSE)

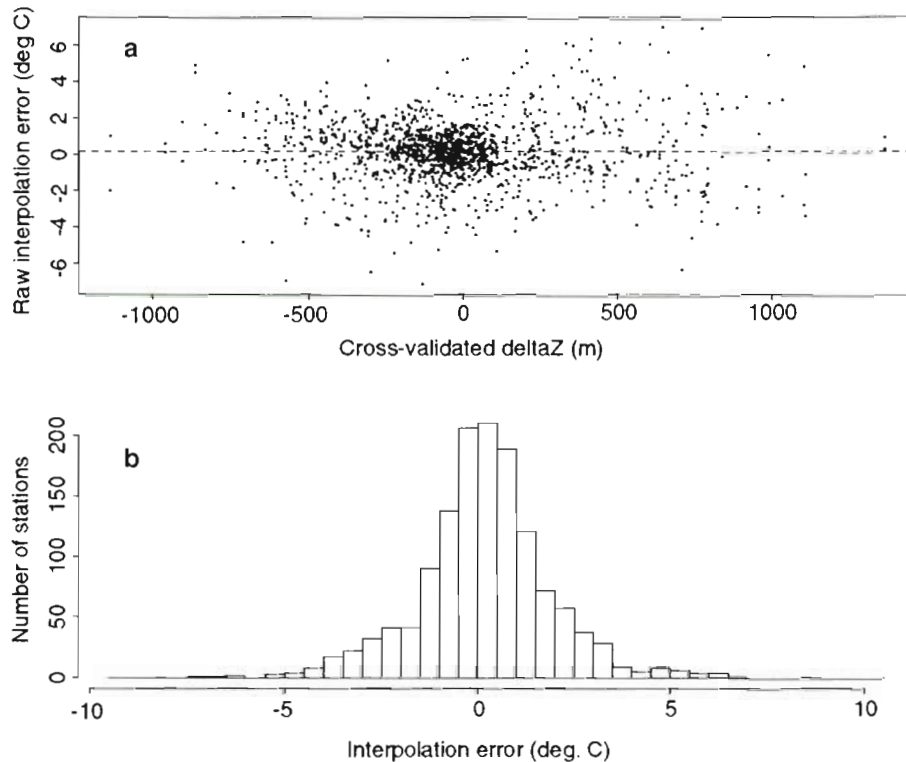| Method	Min.	Max.	Mean	Med.	SD	MAE
LLRA3.9	–7.05	6.28	0.15	0.22	1.75	1.30
LLRA4.8	–7.10	6.99	0.17	0.16	1.72	1.26
LLRA6.5	–7.74	8.21	0.19	0.13	1.81	1.26
NSA-raw	–9.23	10.25	0.21	0.06	2.28	1.58
NSA-adj	–6.81	7.10	0.30	0.26	1.73	1.27

Fig. 13. (a) CVΔZ versus mean annual interpolation error for the LLRA4.8 method, with a linear regression line (dashed). Data are combined T_{min} and T_{max} from 679 stations (n = 1358). The regression equation was: IE = 0.167 – 0.0000015 CVΔZ (R^2 = 0.00). (b) Histogram of mean annual interpolation error for LLRA4.8 (mean = 0.17, MAE = 1.26)

error for each of the methods under investigation. The errors for the NSA method are reported both before and after the ΔZ bias adjustment, and are denoted 'NSA-raw' and 'NSA-adj'. The NSA bias adjustment corrected the larger interpolation errors and slightly improved the overall accuracy of the method by lowering the MAE and standard deviation. However, the bias adjustment slightly raised the overall bias of the method by moving the mean interpolation error farther from zero (from 0.21 to 0.30°C). The overall bias indicates that, on average, each cross-validated station was predicted 0.30°C too cold. This is different from the ΔZ bias, which indicates a systematic relationship between vertical extrapolation (ΔZ) and interpolation error. The bias correction procedure effectively removes the ΔZ bias. The 3 LLRA methods in Table 3 show very similar results. LLRA4.8 has the lowest overall error in terms of MAE and standard deviation, while LLRA3.9 has lower-magnitude extrema and LLRA6.5 has a median error closer to zero. The accuracy of all methods (except NSA-raw), as measured by MAE and standard deviation, is essentially the same to within 0.1°C.

The cross-validation technique used for assessing interpolation accuracy is useful and informative, how-

ever its value is limited as it only provides information on areas associated with observed data (Robeson 1994). Cross-validation tells us nothing about areas between measurement stations, especially those areas which are beyond the elevational range of the stations. While it is impossible to assess accuracy in areas without measured data, these areas should at least be examined and evaluated with exploratory data analysis techniques (Tukey 1977). To this end, we picked January 1 and July 1 as examples of cold and warm days and examined the entire distribution of interpolated T_{min} and T_{max} using all NSA and LLRA methods. Fig. 14 shows boxplots of these temperature surfaces as well as the station data used to derive the surfaces. The discrepancy between the minimum station values and the minimum interpolated values in Fig. 14 reflects the elevational bias in the station sample (illustrated in Fig. 2). We know that the study area contains higher elevations than the station data, therefore we expect the interpolated data to have lower minima. However, there is little basis for determining how much lower the interpolated data should go. It is probably safe to dismiss the NSA-raw method because of its relatively high interpolation errors and its extremely low minima (e.g. the minimum July T_{max} of –9.7°C seen in Fig. 14d,
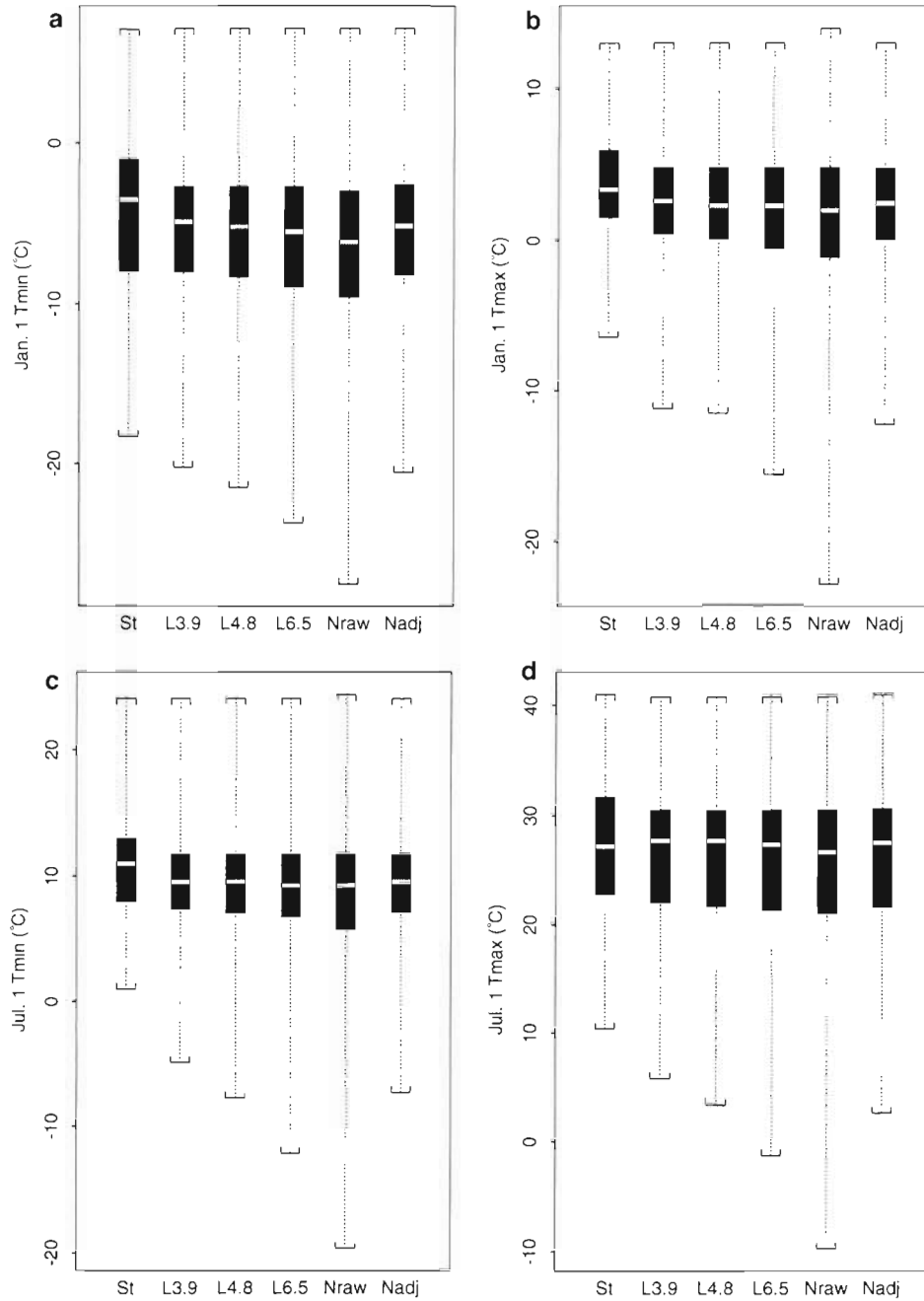
Fig. 14. Boxplots of temperature distributions for stations and interpolated surfaces for (a) January 1 T_{min}, (b) January 1 T_{max}, (c) July 1 T_{min}, and (d) July 1 T_{max}. The shaded portion of a boxplot denotes data values between the first and third quartiles. The white line in the shaded portion corresponds to the median value (the second quartile). The dashed lines extend to the minimum and maximum values. St: station data; L3.9: surface data using LLRA3.9; L4.8: surface data using LLRA4.8; L6.5: surface data using LLRA6.5; Nraw: surface data using NSA-raw; Nadj: surface data using NSA-adj. n = 679 for station data, 828 595 for surface data

which is nearly 20°C lower than the station minimum). But of the remaining methods, it is difficult to make decisions about which performs best in the 'unknown' regions of the study area, i.e. the high mountain areas where no observed data exist.

As a final measure of the discrepancy of the predicted temperatures for each method, we computed pixel-by-pixel image differences of January 1 and July 1 temperature surfaces using pairwise combinations of the interpolation methods under investigation

Table 4. Image difference statistics for January 1 and July 1 (n = 828 595 pixels). Units are °C. MAD: mean absolute difference; Med.: median

Methods	Min.	Max.	Med.	Mean	SD	MAD
T_{min}, Jan 1						
NSA-adj minus LLRA3.9	−5.63	0.99	−0.07	−0.12	0.31	0.22
NSA-adj minus LLRA4.8	−4.38	1.39	0.07	0.08	0.24	0.18
NSA-adj minus LLRA6.5	−4.34	6.04	0.23	0.41	0.76	0.59
LLRA4.8 minus LLRA3.9	−2.82	0.94	0.00	−0.20	0.38	0.28
LLRA4.8 minus LLRA6.5	−1.89	4.69	0.32	0.33	0.62	0.46
LLRA3.9 minus LLRA6.5	−2.83	7.53	0.34	0.53	0.97	0.74
T_{max}, Jan 1						
NSA-adj minus LLRA3.9	−5.63	1.26	−0.10	−0.15	0.36	0.27
NSA-adj minus LLRA4.8	−4.43	1.28	0.05	0.04	0.22	0.15
NSA-adj minus LLRA6.5	−4.33	5.73	0.21	0.37	0.69	0.53
LLRA4.8 minus LLRA3.9	−2.80	1.28	−0.13	−0.20	0.39	0.31
LLRA4.8 minus LLRA6.5	−1.83	4.48	0.17	0.33	0.62	0.48
LLRA3.9 minus LLRA6.5	−2.83	7.20	0.34	0.53	0.97	0.74
T_{min}, Jul 1						
NSA-adj minus LLRA3.9	−2.39	4.18	−0.11	−0.17	0.41	0.32
NSA-adj minus LLRA4.8	−1.67	3.70	0.02	0.03	0.18	0.13
NSA-adj minus LLRA6.5	−1.95	5.10	0.20	0.34	0.61	0.48
LLRA4.8 minus LLRA3.9	−2.82	1.26	0.00	−0.19	0.38	0.27
LLRA4.8 minus LLRA6.5	−2.20	4.38	0.32	0.32	0.61	0.46
LLRA3.9 minus LLRA6.5	−3.12	7.20	0.34	0.51	0.96	0.73
T_{max}, Jul 1						
NSA-adj minus LLRA3.9	−3.23	4.69	−0.18	−0.27	0.57	0.46
NSA-adj minus LLRA4.8	−1.89	4.17	−0.07	−0.08	0.28	0.22
NSA-adj minus LLRA6.5	−1.74	4.16	0.13	0.24	0.47	0.35
LLRA4.8 minus LLRA3.9	−2.47	1.41	0.01	−0.19	0.39	0.28
LLRA4.8 minus LLRA6.5	−2.12	4.59	0.36	0.33	0.61	0.46
LLRA3.9 minus LLRA6.5	−3.19	7.08	0.35	0.51	0.96	0.73

(Table 4). Of all the methods compared in Table 4, the pair which differed the most was LLRA3.9 and LLRA6.5, with a mean absolute difference (MAD) of 0.73 to 0.74°C. The NSA-adj and LLRA4.8 methods were the most similar, with MADs of 0.13 to 0.22°C.

Based on the error statistics (Table 3) and the image difference statistics (Table 4, Fig. 14), all methods (aside from NSA-raw) perform about the same. The rationale for choosing the 'best' method, then, becomes a question of simplicity. A single cross-validation analysis, where T_{min} and T_{max} are computed for 365 d over 679 stations, running simultaneously on four SPARC 20 and two SPARC 10 CPUs, takes approximately 48 h to complete[2] Cross-validation is desirable as a means for assessing accuracy, but it is not required by every interpolation method considered in this study. The LLRA3.9 and LLRA6.5 methods rely only on the calculation of a lapse rate. For LLRA6.5 this involved a rather complicated moving-window regression proce-

dure which depended on some subjective parameters (i.e. minimum regression sample size, distance radius, and threshold R^2 value). The LLRA3.9 lapse rate required only a single regression: elevation versus mean annual temperature. The NSA method, on the other hand, depends on a cross-validation analysis to derive the bias correction factor and then requires the additional processing of the bias correction procedure. The LLRA4.8 method requires 2 cross-validation analyses in order to derive the 'optimal' lapse rate where the bias correction factor is exactly zero. Note, however, that the NSA method, run at a monthly timestep, produced essentially the same bias correction factor as the daily timestep (Table 2) while reducing the computation time by a factor of 30. This time-reduction may also hold true for deriving an 'optimal' LLRA lapse rate.

Our results suggest that the best of the methods investigated is LLRA, using a lapse rate determined by the relationship between mean annual temperature and elevation (LLRA3.9 in our case). LLRA3.9 was by far the simplest method to implement, and differs only slightly from the 'optimal lapse rate' method of LLRA4.8 (e.g. the maximum MAD of the 2 methods was only 0.31°C; Table 4). If time and resources permit a number of cross-validation analyses, then it might be worthwhile to compute an optimal LLRA lapse rate (LLRA4.8 in our case).

8. CONCLUSIONS

The temperature interpolation methods considered in this study were selected specifically for the task of processing very large amounts of data. As such, they make some unrealistic simplifying assumptions. The NSA effectively imposes a lapse rate close to the dry adiabatic lapse rate. Since there are numerous non-adiabatic processes at work in the environment, such as radiational heating/cooling at the surface and horizontal advection, the lapse rates imposed by NSA tend to be too steep, causing a bias with respect to vertical extrapolation (the ΔZ bias). The ΔZ bias is later corrected, however the effect on areas outside the elevational range of the station data is unknown. The LLRA uses a better approximation of the overall environmen-

---

[2]Exact times varied depending on system and network load. The cross-validation jobs were not run at maximum priority. The cross-validation procedure was implemented as a set of linked PERL scripts which use IPW routines (Frew 1990, Longley et al. 1992) to perform spatial processing tasks

tal lapse rate, however the notion of a lapse rate which is constant over space and time is flawed. Lapse rates fluctuate at many scales: seasonally, diurnally, and regionally, and in the case of temperature inversions may even change sign (Barry & Chorley 1987, p. 56). Other sources of error which have undoubtedly introduced noise into all of the analyses include recording errors, instrument bias, uncertainty in the measured air temperatures, incorrectly reported locations and elevations of the measurement stations, DEM uncertainty, and the presence of missing values in the temperature data set. Also note that daily T_{min} and T_{max} measurements tend to be much more noisy, and therefore more difficult to interpolate, than monthly or annually averaged temperature data.

More sophisticated procedures exist for modeling a changing relationship between 2 variables over space, such as the moving-window regression logic of the PRISM model (Daly et al. 1994), where smoothed elevation is regressed on precipitation over a set of topographically similar slope facets. However, such models are difficult to implement, usually require considerable parameterization, and are often too slow to be used over large areas at high resolutions.

Air temperature surfaces are rarely developed as an end in themselves. Often they are used as one of a set of inputs to a spatially distributed model. Unfortunately, there are few instances in the literature where a spatially-distributed ecological or hydrologic modeling project included a rigorous accuracy assessment of all input data. Indeed, statements of accuracy in digital spatial databases are still somewhat rare (Goodchild 1993). An accuracy statement for the best temperature surfaces created in this study would read something like the following:

At elevations below about 2500 m, errors in interpolated temperatures have an expected value near zero and an average magnitude (MAE) of about 1.3°C. Individual errors with magnitudes up to 7 or 8°C are likely. At elevations above 2500 m, accuracy is unknown, but is probably less than the accuracy at lower elevations.

More sophisticated methods will undoubtedly improve on accuracy as measured by cross-validation, but error will always be present, especially as temporal resolution, spatial resolution, and ruggedness of the terrain increase. In particular, the bias toward lower elevations in meteorological station networks will always make interpolated values in high-elevation mountainous regions suspect. Cross-validation is at best a rough indicator of the accuracy of the interpolated surface. Given the reasonable levels of cross-validation accuracy obtained by the LLRA and NSA methods, it is debatable whether the extra effort of a highly complex interpolation method is warranted in order to shave a few tenths of a degree off of the cross-validated MAE. The methods investigated in this study require very little parameterization, are easy to implement and fast to execute (LLRA in particular), and predict temperature at a reasonable level of accuracy. They should provide adequate inputs for use in spatially distributed modeling in mountainous regions. Further work on temperature interpolation methods is certainly needed, but perhaps more important are methods for modeling the errors themselves. Meaningful and realistic error surfaces would enable sensitivity analyses to be performed on the spatial models which rely on air temperature surfaces as input.

LITERATURE CITED

Barry RG, Chorley RJ (1987) Atmosphere, weather and climate, 5th edn. Routledge, London

Byers HR (1974) General meteorology, 4th edn. McGraw-Hill, New York

Daly C, Neilson RP, Phillips DL (1994) A statistical-topographic model for mapping climatological precipitation over mountainous terrain. J Appl Meteorol 33:140–158

Davidson F (1996) Principles of statistical data handling. Sage Publications, Thousand Oaks, CA

Dolph J, Marks D, King GA (1992) Sensitivity of the regional water balance in the Columbia River Basin to climate variability: application of a spatially distributed water balance model. In: Naiman RJ, Sedell JR (eds) New perspectives for watershed management: balancing long-term sustainability with cumulative environmental change. Springer-Verlag, New York, p 233–265

EarthInfo (1990) Programmer's toolkit user's manual for EarthInfo's climate-data CD-ROM summary of the day, Version CD3. EarthInfo, Inc, Boulder, CO

Environment Canada (1989) Climatological station catalogue. Atmospheric Environment Service, Downsview, Ontario

EROS Data Center (1991) Conterminous U.S. AVHRR 1990 biweekly composites. U.S. Geological Survey, EROS Data Center, Sioux Falls, SD

Frew JE (1990) The image processing workbench. PhD dissertation, Department of Geography, University of California, Santa Barbara

Goodchild MF (1993) Data models and data quality: problems and prospects. In: Goodchild MF, Parks BO, Steyaert LT (eds) Environmental modeling with GIS. Oxford University Press, New York, p 94–103

Greenland D (1994) The Pacific Northwest regional context of the climate of the H. J. Andrews experimental forest. Northwest Sci 69(2):81–96

Hutchinson MF (1989) A new objective method for spatial interpolation of meteorological variables from irregular networks applied to the estimation of monthly mean solar

radiation, temperature, precipitation. and windrun. Tech Memo 89(5):95–104, CSIRO, Australia, Division of Water Resources, Victoria

Isaaks EH, Srivastava RM (1989) An introduction to applied geostatistics. Oxford University Press, New York

Leemans R Cramer WP (eds) (1991) The IIASA database for mean monthly values of temperature, precipitation, and cloudiness on a global terrestrial grid. International Institute for Applied Systems Analysis, Laxenburg

Legates DR, Willmott CJ (1990) Mean seasonal and spatial variability in global surface air temperature. Theor Appl Climatol 41:11–21

Longley KD, Jacobsen D, Marks D (1992) Supplement to the Image Processing Workbench (IPW): modifications, procedures, and software additions, Revision 2.0. Tech Rep, EPA-COR EPA/600/9-92/217, U.S. Environmental Protection Agency, Environmental Research Laboratory, Corvallis

Marks D (1990) A continental-scale simulation of potential evapotranspiration for historical and projected doubled-CO_2 climate conditions. In: Guchinski H, Marks D, Turner D (eds) Biospheric feedbacks to climate change: the sensitivity of regional trace gas emissions, evapotranspiration, and energy balance to vegetation redistribution. EPA/600/3-90/78. U.S. Environmental Protection Agency, Corvallis

Marks D, Dozier J, Davis RE (1992) Climate and energy exchange at the snow surface in the alpine region of the Sierra Nevada: 1. Meteorological measurements and mon-

itoring. Water Resour Res 28(11):3029–3042

Marks D, King GA, Dolph J (1993) Implications of climate change for the water balance of the Columbia River Basin, USA. Clim Res 2:203–213

Myers DE (1994) Spatial interpolation: an overview. Geoderma 62(1):17–28

Phillips DL, Marks D (1996) Spatial uncertainty analysis: propagation of interpolation errors in spatially distributed models. Ecol Model 91:213–229

Robeson SM (1993) Spatial interpolation, network bias, and terrestrial air temperature variability. Publ Climatol 46(1): 1–51

Robeson SM (1994) Influence of spatial interpolation and sampling on estimates of terrestrial air temperature change. Clim Res 4(2):119–126

Robeson SM (1995) Resampling of network-induced variability in estimates of terrestrial air temperature change. Clim Change 29:213–229

Tukey JW (1977) Exploratory data analysis. Allison-Wesley, Reading, MA

USDA-SCS (1988) Snow survey and water supply products reference. U.S. Department of Agriculture-Soil Conservation Service, Western National Technical Center, Portland

Willmott CJ, Matsuura K (1995) Smart interpolation of annually averaged air temperature in the United States. J Appl Meteorol 34:2577–2586

Willmott CJ, Robeson SM (1995) Climatologically aided interpolation of terrestrial air temperatures. Int J Climatol 15(2):221–229