United States
Department
of Agriculture

Forest Service

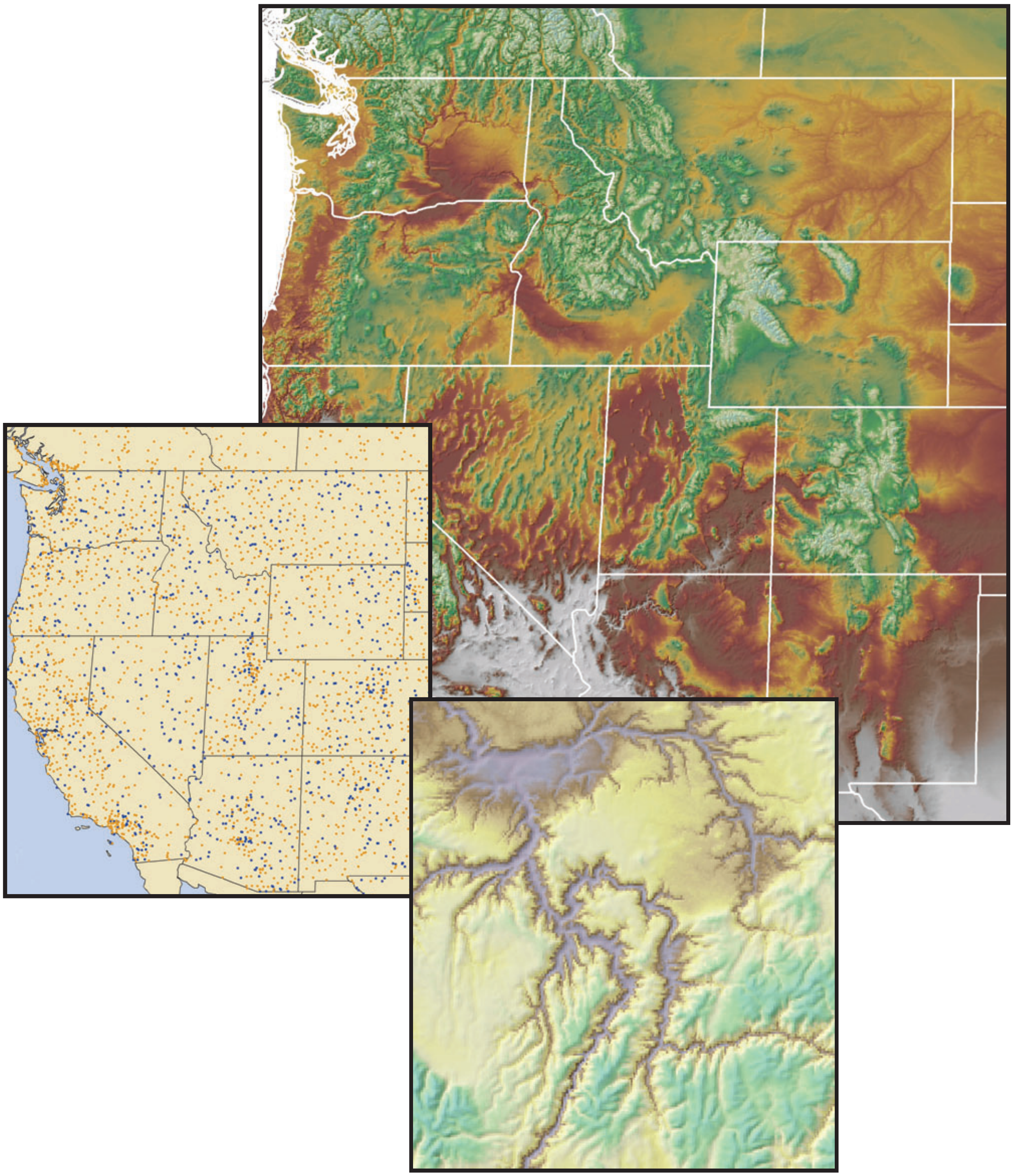
**Rocky Mountain
Research Station**

General Technical
Report RMRS-GTR-165

January 2006

A Spline Model of Climate for the Western United States

Gerald E. Rehfeldt



Rehfeldt, Gerald L. 2006. **A spline model of climate for the Western United States**. Gen. Tech. Rep. RMRS-GTR-165. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 21 p.

Abstract

Monthly climate data of average, minimum, and maximum temperature and precipitation normalized for the period 1961 through 1990 were accumulated from approximately 3,000 weather stations in the Western United States and Southwestern Canada. About two-thirds of these observations were available from the weather services of the two countries while the remaining third were added to the normalized base from daily weather records of stations of short duration. Tests of the procedures used to normalize these supplemental data showed that estimates on average were within 0.2 °C for temperature variables and 2.7 mm for precipitation.

Weather data for the 48 monthlies were fit to geographic surfaces with thin plate splines. Relationships between predicted values and observed monthlies for about 245 records withheld from the modeling process produced values of R^2 that averaged about 0.95 and ranged from 0.87 to 0.99. The slope of the regression line for these relationships was essentially 1.0 for all 48 comparisons. Predictions from the climate model can then be converted to variables of demonstrated importance in plant geography, ecology, or physiology. As an illustration, algorithms are presented and justified for estimating 18 variables derived from predicted values. These derived variables range from the straightforward such as mean annual temperature or mean temperature in the coldest month to those of degree-days >5 °C or freezing dates.

Applications of the model in plant biology are illustrated for (1) generating climate estimates for locations specified by latitude, longitude, and elevation, (2) mapping climate variables, (3) separating species distributions in climatic space, and (4) relating genetic variation among populations to climatic gradients.

Keywords: thin plate splines, climate model, climate normals, climate surfaces, predicting climate, mapping climate

Author

Gerald E. Rehfeldt is a retired Plant Geneticist with the Rocky Mountain Research Station, Forestry Sciences Laboratory, Moscow, ID. He completed M.S. and Ph.D. degrees by 1968 at the University of Wisconsin, and spent the last 39 years studying genetic variation in Rocky Mountain conifers.

Acknowledgment

The development of this climate model and the preparation of this report were dependent on contributions of many people from disparate disciplines. Contributors listed chronologically include Nadejda Tchebakova, Robert Monserud, Run-Peng Wei, William Wykoff, Jeffrey Evans, Andrea Brunelle, Marcus Warwell, Andrew Hudak, Nicholas Crookston, and Andreas Hamann.

Rocky Mountain Research Station
Publishing Services

Telephone	(970) 498-1392
FAX	(970) 498-1396
E-mail	rschneider@fs.fed.us
Web site	http://www.fs.fed.us/rm
Mailing Address	Publications Distribution Rocky Mountain Research Station 240 West Prospect Road Fort Collins, CO 80526

Contents

A Spline Model of Climate for the Western United States

Gerald E. Rehfeldt

Introduction

Humans have been aware for centuries that climate is the primary factor controlling the distribution of plants, but even today, plant-climate relationships are poorly understood (see Woodward 1987). To be sure, assessing the ecological relationships between plants and climate has been hampered by a network of weather stations that largely reflect agronomic interests and population centers. Ecological research, therefore, has been impeded by an inability to predict climate for remote locations. Two recent developments now alleviate this problem. First, concerns about anthropomorphic effects on climate have generated large volumes of climate data. Second, the thin plate splines of Hutchinson (1991, 2000) make possible the fitting of climate data to geographic surfaces. Spline climate models can provide estimates of climate for specific points on the geographic surface, identified by latitude, longitude, and elevation. Such estimates are suited directly for assessing plant responses to climate.

This paper describes a spline climate model largely for the Western United States but also Southwestern Canada, from longitudes between 102°W and 125°W and latitudes between 31°N and 51°N. The model, which yields predictions of monthly temperature and precipitation, is described and verified, algorithms that convert monthly values to variables relevant to plant geography and physiology are presented and justified, and the utility of the model in plant biology is illustrated. The description of the climate model and the format of this paper follow that of McKenney and others (2001) for their climate surfaces of Canada.

Climate Data and Their Normalization

This model is based on the 1961 through 1990 monthly averages of daily minimum, maximum, and average temperature and daily precipitation produced by the weather services of the United States (U.S. Department of Commerce 1994) and Canada (Environment Canada 1994). Because missing observations invariably exist within weather records, adjustments are necessary for standardizing observations between stations. These adjusted records are referred to as the monthly normals by the weather services. For the area of interest herein, the number of stations represented in the normalized data base were approximately 1,900 for the United States and 200 for Southwestern Canada, with the number varying somewhat for each climate variable.

The weather stations contained in these data bases are referred to as standard stations, and in order to qualify as a standard station there must be at least 20 years of record for the 30-year period. To increase the representation of remote locations in this largely agronomic data base, daily data from stations not represented in the data set of monthly normals were normalized from raw daily data provided by EarthInfo, Inc. (1994). Raw data were accepted if the station had at least seven complete years of observations for precipitation and 5 years for temperature, and if the monthlies had less than four missing observations.

The raw data were adjusted to the 1961 through 1990 normals by first selecting the three standard stations that were the closest geographically to the station whose data were being normalized. The three closest standard stations were identified by sequential screenings according to horizontal and vertical distances. The first screening was based on a horizontal distance of 50 km and a vertical distance of 300 m. If this screening failed to yield three stations within the established limits, a second screening was made using a horizontal distance of 50 km and an altitudinal distance of 500 m. Likewise, a third and fourth screenings were implemented when necessary using a horizontal distance of 80 km and, first, an altitudinal distance of 500 m, and, second, an unlimited altitudinal distance. In approximately 80 percent of the cases, the closest three standard stations were identified prior to the third level; less than 1 percent of the cases required level four to locate three suitable standard stations.

Temperature normals were calculated by averaging deviations between monthly means from data being normalized and those of the three geographically proximal standard stations: Let N_1, N_2, and N_3 be monthly normals for the three standard stations closest geographically to station i. Then a normalized value of raw temperature data from station i in month j becomes:

$$N_{ij} = \frac{\left[N_{1j} - \frac{\sum\left(X_{ijk} - X_{1jk}\right)}{n_{jk}}\right] + \left[N_{2j} - \frac{\sum\left(X_{ijk} - X_{2jk}\right)}{n_{jk}}\right] + \left[N_{3j} - \frac{\sum\left(X_{ijk} - X_{3jk}\right)}{n_{jk}}\right]}{3}$$

where N_{ij} is an estimate of the 1961 through 1990 normal for station i in month j; X_{ijk} is raw data for station i in month j for year k common between station i and the standard station; X_{1jk}, X_{2jk}, and X_{3jk} are raw data for the standard stations; and n is the number of years k in common between the two stations in month j.

Precipitation normals were calculated similarly by using ratios of the raw data to those of a standard station:

$$N_{ij} = \frac{\left[N_{1j} \times \frac{X_{ij\bullet}}{X_{1j\bullet}}\right] + \left[N_{2j} \times \frac{X_{ij\bullet}}{X_{2j\bullet}}\right] + \left[N_{3j} \times \frac{X_{ij\bullet}}{X_{3j\bullet}}\right]}{3}$$

where $X_{ij\bullet}$, is raw data for station i in month j summed for all years k in common between station i and the standard station; and $X_{1j\bullet}$, $X_{2j\bullet}$, and $X_{3j\bullet}$ are raw data for the three standard stations summed for the years in common between the two stations. Other variables are defined above.

To test the adequacy of the normalization process, six standard stations of disparate climate were removed from the normalized data base of eight States and were normalized anew according to the procedures detailed above. In total, 48 observations were available for comparing monthly normals of standard stations against those produced by the normalizing procedures. On average, differences in monthly precipitation were within 2.7 mm, the standard deviation was 3.6 mm, and the largest difference was 21 mm, the last of which was for Santiam Pass, Oregon, where precipitation averages 1,963 mm and the error, therefore, is about 1 percent. For average, minimum, and maximum temperatures, the mean difference was 0.2 °C, the standard deviation was 0.14 °C, and the largest difference was 1.1 °C. These comparisons thus provide a strong validation of the procedures.

The procedures yielded an additional 1,200 weather stations with normalized monthlies. Data from stations that were either duplicates of or in proximity to standard stations were discarded, the latter group of which was removed in order to break up clusters of data points. This left approximately 1,000 additional stations available for the spline models. These additional data were appended to the normals of the standard stations to obtain a total of about 3,000 geographically diverse stations for the modeling of precipitation monthlies (table 1, fig. 1) and 2,700 for temperature monthlies.

Table 1—Altitudinal range and number of weather stations within the normalized precipitation database listed according to administrative units

Administrative unit	Number of weather stations	Altitude (m) Minimum	Altitude (m) Maximum
Alberta	27	717	1,439
Arizona	241	37	2,588
British Columbia	122	1	1,520
California	497	–59	3,801
Colorado	235	1,034	3,460
Idaho	167	303	2,226
Montana	297	585	2,274
Nebraska	26	1,010	1,479
Nevada	141	174	2,279
New Mexico	225	993	3,257
North Dakota	40	552	909
Oklahoma	3	1,226	1,326
Oregon	234	2	2,141
Saskatchewan	52	497	1,070
South Dakota	54	680	1,951
Texas	51	732	1,695
Utah	213	495	3,260
Washington	20	33	1,796
Wyoming	170	1,075	2,535

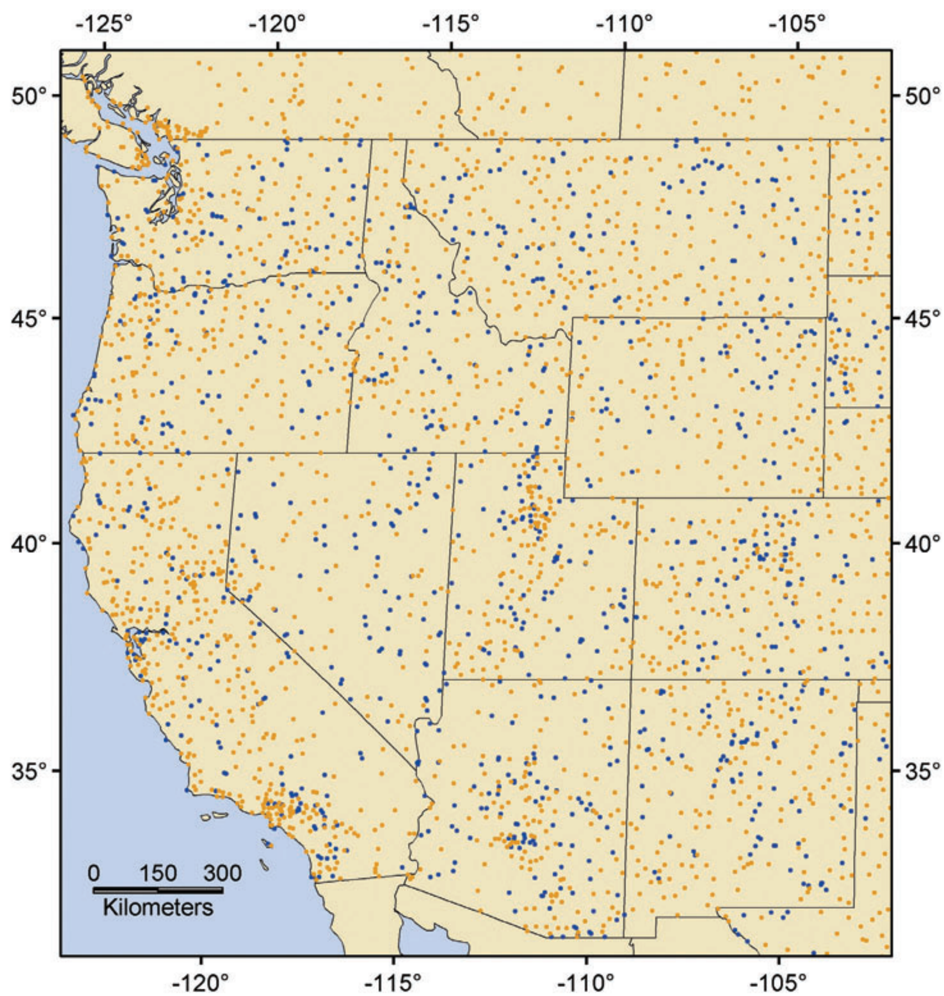


Figure 1—Location of weather stations represented in the data base of normalized monthly precipitation available from national weather services of the United States and Canada (orange) and those appended (blue) to this data base by procedures described herein.

Fitting Thin Plate Splines to Monthly Climate Normals

Thin plate splines of Hutchinson (2000) were fit to the twelve monthly values of four climate variables in two stages. For both stages, the square root transformation was used on precipitation data, but output is presented in the original units of measure. To ease computation in large data sets, Hutchinson recommends the use of knots, which, in this case, are stations located at regular geographic intervals. For the present analyses, 1,300 knots were used for the temperature splines and 1,600 for precipitation. In the first stage, a data set was created for judging the effectiveness of the spline model. This data set contained roughly equal contributions of data from stations (a) removed from the normalized data in order to break up geographic clusters, (b) deleted because of duplication, and (c) included to obtain a geographically representative sample. The total number of stations represented in this data set was 242 for maximum, minimum, and average temperature and 254 for precipitation. Hutchinson emphasizes the necessity of using withheld data to judge the effectiveness of these nonparametric procedures. In the second stage, the data withheld at level 'c' were returned to the normalized data base before refitting the splines.

Relationships between the observed monthly normals of withheld data and those predicted from the first stage of modeling are strong, with values of R^2 ranging from 0.87 to 0.98 (table 2). Examples of the strongest and weakest relationships are illustrated in figure 2. Table 2 also shows that the regression coefficient is near to 1.0 for all months but that precipitation may be underestimated slightly when it is high. Table 2 and figure 2 thus demonstrate that the fit of the splines is quite good and is probably as good as can be expected given the original distribution of the weather stations and the geographic diversity of Western United States.

For the second and final stage of the model development, approximately one-third of the withheld data were then returned to the data set of monthly normals, and the splines

Table 2—Statistics derived from the linear regression of observed monthly temperature and precipitation of data withheld from the modeling on values predicted from the spline model. Regressions were of the form: $Y = a + bX$ where Y is the observed value from either 241 or 254 weather stations, and X is predicted from the spline model.

	Month											
Statistic	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Mean average temperature												
b	1.00	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.00
R^2	0.97	0.98	0.98	0.98	0.97	0.96	0.95	0.95	0.96	0.96	0.97	0.97
s	1.05	0.91	0.78	0.75	0.80	0.94	1.02	1.04	1.00	0.94	0.86	0.96
Mean minimum temperature												
b	1.01	1.02	1.03	1.03	1.02	1.02	1.00	1.00	1.01	1.03	1.02	1.01
R^2	0.94	0.94	0.95	0.92	0.90	0.87	0.87	0.87	0.88	0.88	0.92	0.93
s	0.91	0.79	0.71	0.76	0.84	0.92	0.98	1.00	0.86	0.76	0.69	0.82
Mean maximum temperature												
b	0.99	0.99	0.99	1.00	1.00	1.00	1.01	1.01	1.00	0.99	1.00	0.99
R^2	0.98	0.98	0.99	0.98	0.97	0.97	0.96	0.95	0.97	0.98	0.99	0.98
s	1.66	1.52	1.28	1.28	1.40	1.63	1.74	1.79	1.82	1.76	1.56	1.62
Mean precipitation												
b	1.04	1.04	1.03	1.02	1.03	1.02	0.97	1.03	1.07	1.10	1.04	1.03
R^2	0.93	0.93	0.91	0.89	0.91	0.95	0.92	0.92	0.88	0.92	0.95	0.94
s	18.05	14.82	14.54	9.75	8.17	5.86	6.59	5.94	6.83	11.03	15.85	17.78

Note: s = standard deviation of the residuals

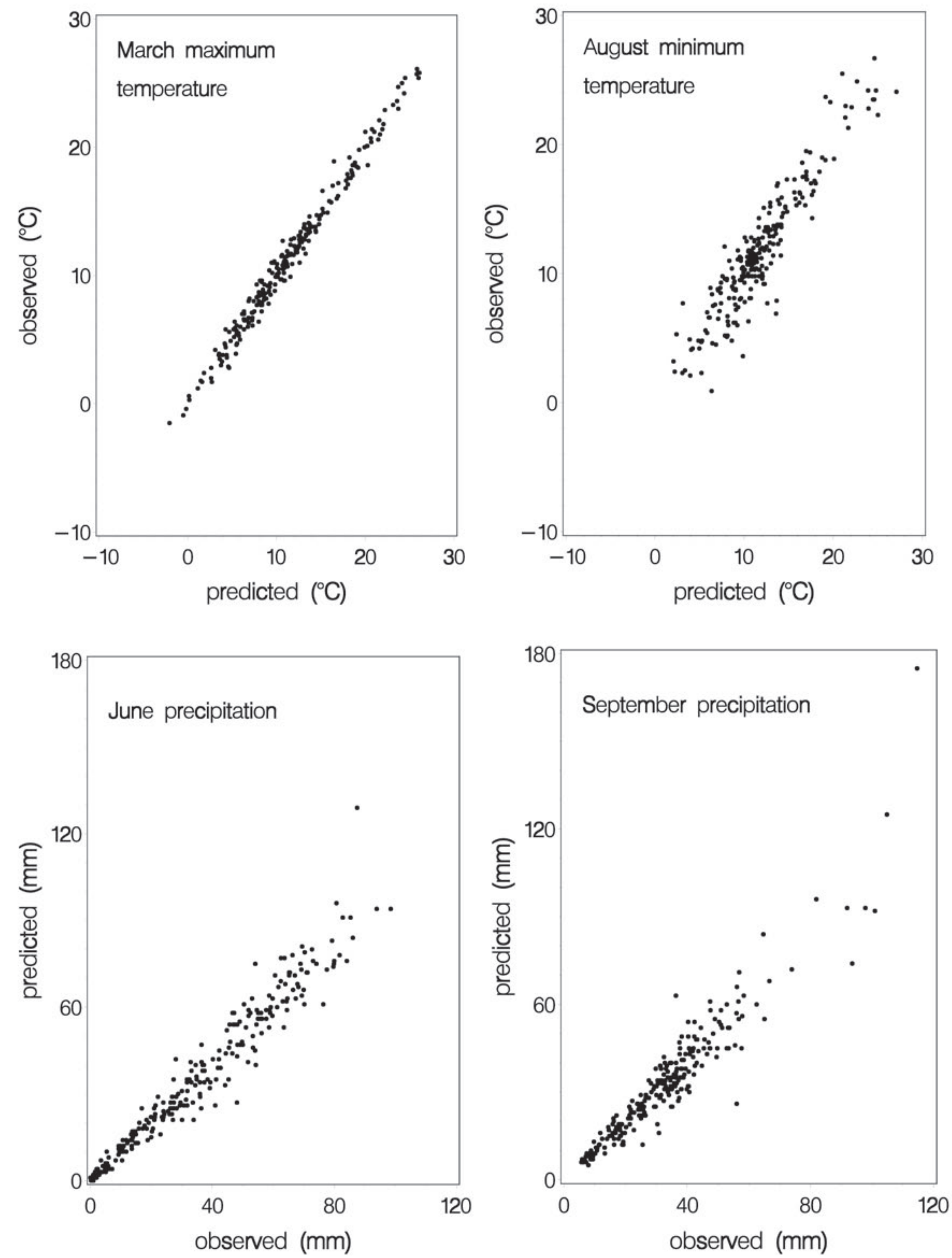


Figure 2—Predicted values of monthly means plotted against the actual illustrating one of the best fits (left) and poorest fits (right) for 241 stations withheld from the normalized data base for temperature (upper) variables and 251 stations for precipitation (lower). Statistics describing these relationships are in table 2.

were fit anew. Hutchinson (2000) notes that among the diagnostic statistics, three in particular are useful in judging the fit of the splines: the signal, the root mean square error (RTMSE), and the root of the generalized cross validation statistic (RTGCV). The signal is indicative of the degrees of freedom associated with the surface, which in a well-fitting model should be approximately one-half of the number of observations (or knots); RTGCV is a spatially averaged standard error, plus and minus two times of which is a rough indication of the 95 percent confidence limits about a predicted point; and the RTMSE is a measure of the standard error after the data error has been removed; it is, therefore, an optimistic estimate of the surface error.

Table 3 presents diagnostic statistics for the second stage of the modeling. These statistics show that the ratio of the signal to the number of knots is about 0.5 for the temperature surfaces, but averages 0.64 for the precipitation surfaces. Hutchinson notes that values larger than 0.5 would indicate either an inherently noisy system (large variability) or a lack of adequate data. McKenney and others (2001) present ratios for precipitation in excess of 0.7 and conclude, therefore, that the surfaces would be improved if more data were available.

Table 3 also shows that values of the RTMSE are generally low. For precipitation surfaces, this statistic, expressed as a percentage of the surface mean, ranged from 7 percent in the summer months to about 11 percent in the winter. McKenney and others (2001) presented values as high as 28 percent of the mean and noted that winter precipitation is a more spatially complex phenomenon. For temperature surfaces, the standard errors are approximately 0.4 °C for the average and maximum temperature but about 0.7 °C for minimum temperature. McKenney and others again note that larger errors for minimum temperature are to be expected in mountainous terrain where cold air drainages are prominent.

Errors of prediction expressed in the RTGCV suggest that predictions would be surrounded by a confidence interval ($\alpha \approx 0.05$) of about ±2 °C for average and maximum temperature; ±3 °C for predictions of the minimum temperature; and about ±20 mm for precipitation. Table 3 also shows that the ratio of the RTMSE to the RTGCV ranged from 0.42 to 0.47, indicating that the spline model overcame a substantial amount of noise in the monthly normals (McKenney and others 2001). It is likely that some of this noise was introduced by accepting data from stations with limited records (5 years for temperature

Table 3—Ratio of signal to the total number of knots, the root generalized cross validation (RTGCV), and the root mean square error (RTMSE) for spline monthly temperature and precipitation surfaces. Number of knots was 1,400 for the temperature variables and 1600 for precipitation.

	Variable											
	Average temperature			Minimum temperature			Maximum temperature			Precipitation		
Month	ratio	RTGCV	RTMSE[a]	ratio	RTGCV	RTMSE[a]	ratio	RTGCV	RTMSE[a]	ratio	RTGCV	RTMSE[a]
1	0.58	1.00	0.46	0.56	1.59	0.72	0.60	0.84	0.39	0.64	15.2	7.16
2	0.55	0.90	0.41	0.54	1.47	0.66	0.55	0.78	0.35	0.64	12.7	6.00
3	0.49	0.74	0.33	0.50	1.25	0.54	0.49	0.75	0.32	0.65	12.5	5.95
4	0.48	0.73	0.32	0.48	1.24	0.54	0.48	0.82	0.36	0.68	8.4	4.03
5	0.50	0.79	0.35	0.49	1.32	0.58	0.49	0.91	0.40	0.63	6.8	3.22
6	0.52	0.91	0.41	0.51	1.49	0.66	0.51	1.03	0.46	0.60	5.4	2.50
7	0.52	1.00	0.45	0.54	1.58	0.71	0.50	1.19	0.52	0.56	5.3	2.39
8	0.51	1.00	0.41	0.53	1.63	0.73	0.50	1.15	0.50	0.62	5.2	2.42
9	0.48	0.94	0.39	0.50	1.68	0.74	0.48	0.99	0.43	0.64	5.4	2.57
10	0.47	0.89	0.37	0.49	1.68	0.74	0.51	0.81	0.36	0.62	8.3	3.87
11	0.51	0.83	0.38	0.50	1.49	0.66	0.53	0.69	0.31	0.64	13.1	6.20
12	0.57	0.97	0.45	0.55	1.58	0.72	0.61	0.80	0.37	0.64	14.5	6.87

[a]Approximate untransformed value

and 7 years for precipitation) but most undoubtedly stems from the variable nature of the climate, particularly in mountainous terrain.

Table 4 lists the eight stations with the largest RTMSE, calculated as a composite for the 12 splines for each of the four variables. This table shows that for predictions of average temperature, the stations with the largest composite residual were randomly distributed geographically and, except for a coastal station from California, tended to be from high elevation. Those with the largest residuals for the minimum temperature were from high elevations where temperatures tend to be the coolest. The stations with the worst fit for maximum temperature were from warm, maritime climates largely in California, while those for monthly precipitation were all from northern coastal regions where precipitation is high. Notice also that the source of the monthly normals had no apparent effect on the size of the largest residuals. This suggests, therefore, that the normalization procedures described herein did not unduly increase error variances.

Table 4—Eight stations and their locations with the highest aggregate root mean square residual (RTMSR) for the 12 monthly surfaces of four climate variables.

Administrative unit	Station	Latitude (°N)	Longitude (°W)	Altitude (m)	RTMSR
Mean average temperature					
Colorado	3113	39.95	105.83	2610	2.84
Arizona	4453*	34.75	112.12	1599	2.54
Montana	7750	38.03	114.18	1814	2.44
South Dakota	7227	44.13	103.73	1635	2.40
Montana	9072	41.66	115.79	1898	2.33
California	3191*	38.52	123.25	34	2.27
California	5356	38.70	119.78	1688	2.26
Utah	2057*	40.40	111.53	1607	2.18
Mean minimum temperature					
Arizona	4453*	34.75	112.12	1599	5.26
Montana	7750	38.03	114.18	1814	4.70
Colorado	3113	39.95	105.83	2610	4.47
Utah	5607	41.22	112.64	1286	4.46
South Dakota	7227	44.13	103.73	1635	4.14
Nevada	8761*	39.30	119.63	1933	4.04
Utah	2696*	40.27	112.08	1488	3.99
New Mexico	3505	36.43	106.97	2263	3.98
Mean maximum temperature					
California	7953*	34.00	118.50	4	4.50
California	9792	37.43	122.25	116	4.36
California	5866*	35.37	120.85	35	3.91
California	3191*	38.52	123.25	34	3.79
Oregon	4133	42.64	124.05	107	3.56
California	7767*	37.77	122.50	10	3.34
California	1484*	34.18	118.57	241	3.29
California	7714*	36.02	121.25	23	3.08
Mean precipitation					
British Columbia	6330*	49.52	123.48	8	4.45
Washington	1760*	46.07	122.20	201	2.94
British Columbia	0590*	49.35	124.17	15	2.85
British Columbia	7200*	49.43	122.97	244	2.77
Washington	3333*	46.38	123.57	30	2.70
Washington	7538	48.08	123.10	57	2.54
California	0673	37.08	122.08	135	2.39
California	6154	34.95	119.68	658	2.29

*Data normalized by procedures described herein

Deriving Climate Variables Relevant to Plant Biology

The spline climate surfaces predict 48 monthly values of temperature and precipitation. These values can be used to derive variables of demonstrable effectiveness in plant geography, physiology, and ecology (see Tuhkanen 1980). The following 18 are a sample:

1. mean annual temperature
2. average temperature in the coldest month
3. minimum temperature in the coldest month
4. average temperature in the warmest month
5. maximum temperature in the warmest month
6. annual precipitation
7. growing season precipitation, April through September
8. summer-winter temperature differential, variable #4 minus #2
9. degree-days >5 °C
10. degree-days <0 °C
11. minimum degree-days <0 °C
12. Julian date of the last freezing date of spring
13. Julian date of the first freezing date of autumn
14. length of the frost-free period
15. degree-days >5 °C accumulating within the frost-free period
16. Julian date when the sum of degree-days >5 °C reaches 100
17. annual moisture index, the ratio of variable #9 to #6
18. summer moisture index, the ratio of variable #15 to #7

Note that the first eight of these variables come directly from the monthlies, and four others involve simple ratios or differences, but estimation of numbers 9 to 16 require elaboration.

Estimation of Degree-Days

Degree-days are temperature sums above or below a threshold value. Three annual degree-day sums are used as derived variables in the climate model: degree-days >5 °C (DD5), degree-days <0 °C (DD0), and minimum degree-days <0 °C (MINDD0). The first is widely accepted among plant geographers as a general indication of the warmth of the growing season; the second is viewed as an indicator of the coldness of the winter; and the third is included because many of the weather stations used herein from mild climates record freezing temperatures but nonetheless have a DD0 equal to zero.

Degree-days are ordinarily calculated in three steps from daily temperatures:

1. Calculate for each day, *i*, the difference (T_{di}) between the mean temperature of the day (T_i) and the threshold temperature (T_t):

$$T_{di} = T_i - T_t$$

where *i* is the Julian date.

2. Apply a condition that is dependent on whether DD5 or DD0 is being summed:
(a) if DD5 is the statistic of interest, all values of $T_{di} < 5$ are equated to zero.
(b) if DD0 is the statistic of interest, all values of $T_{di} > 0$ are equated to zero.

3. Summing T_{di} across the period of interest. For an annual sum of DD5:

$$DD5 = \sum_{i=1}^{365} T_{di}$$

Notice, however, that these steps use daily temperatures while the spline climate model predicts monthly temperatures. Obviously, T_{di} could be calculated as above using monthly average temperatures and multiplying by the number of days in the month to estimate the degree-days that accumulated for that month. The problem, however, is that in doing so, one ignores daily variations about the monthly mean. This would mean that for the calculation of DD5, days with an average temperature > 5 °C would be ignored if the monthly mean was < 5 °C. The resulting sum, therefore, would underestimate the actual.

To provide an unbiased estimate of DD5 and DD0, regressions using monthly normals of the National Climate Data Center (U.S. Department of Commerce 1994) and Environment Canada (1994) were developed for predicting monthly degree-day sums from monthly average temperatures. Because months have a different number of days, the latter statistic was expressed as an average daily value by dividing the monthly sum by the number of days in the month.

Degree-days > 5 °C—A complete set of monthly normals for average temperature and DD5 were available for 2,143 weather stations for Western North America 1,868 stations from the United States and 275 stations from Canada. The relationship between average monthly temperature and degree-days >5 °C, expressed as the average daily accumulation, is shown in figure 3, left. The figure illustrates that at temperatures >9.5 °C, the daily

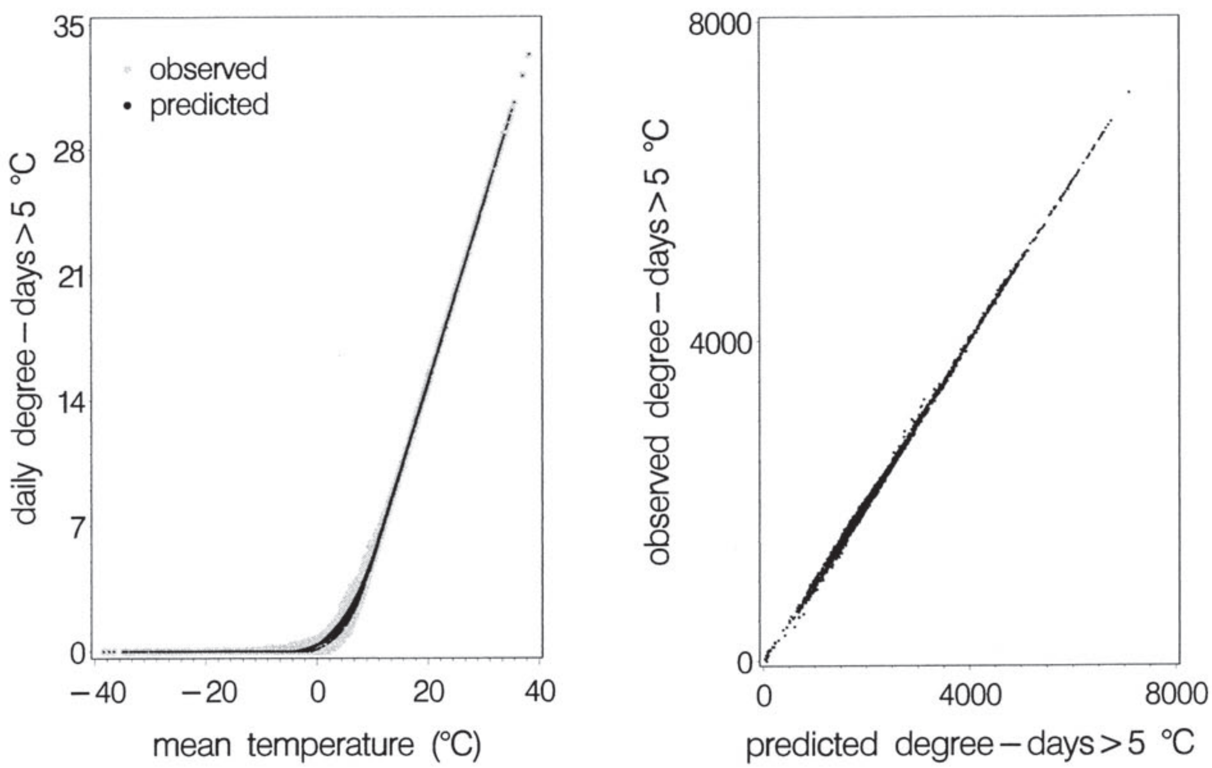


Figure 3—Plots of average daily temperature monthly means against the number of degree-days > 5 °C that accumulate per day during the month overlain with values predicted by the three-stage regression model (left), and relationship between observed annual sums of DD5 with the predicted sum (right).

accumulation is essentially equal to the amount by which the average temperature exceeds 5 °C; at mean daily temperatures below –13 °C, the daily contribution equals zero; and for the interval of –13 °C to 9.5 °C, the daily contribution is exponential. Modeling this response, therefore, could be done in three stages.

Stage 1: define γ as the daily contribution to the monthly sum of DD5, and equate γ to zero for all temperatures < –13 °C.

Stage 2: fit the following power function to γ for temperatures between –13 °C and +9.5 °C for each month:

$$\gamma_{ij} = a + be^{(T'_{ij})^c}$$

where γ is daily degree-days >5 °C for station i in month j; a, b, and c are regression coefficients; e is the base of the Napierian logarithms; and T' is the average temperature for station i in month j transformed to a value between 0 and 1:

$$T'_{avg} = \left(T_{avg} + 13\right) / 24$$

To provide a suitable number of degrees of freedom for the regression, observations for June, July, and August were combined. As a result, the regressions were based on as many as 1,891 observations in April and as few as 93 for the summer months.

All regressions were statistically significant ($p < 0.01$), accounting for an average of 95 percent of the variance in the dependent variable, a minimum of 92 percent (April), and a maximum of 98 percent (January and September).

Regression coefficients for each month are:

	a	*b*	*c*
January	–4.3302	4.2501	3.8847
February	–4.3866	4.2652	3.8238
March	–4.2910	4.1651	3.5863
April	–4.9051	4.0704	2.2671
May	–4.2351	4.0838	3.8330
June	–4.5055	4.2100	3.8902
July	–4.5055	4.2100	3.8902
August	–4.5055	4.2100	3.8902
September	–4.2853	4.0123	3.3756
October	–4.2737	4.0056	3.1198
November	–4.2447	4.0844	3.5669
December	–4.4445	4.3440	3.7888

Stage 3: fit the following linear regression for temperatures >9.5 °C for each month:

$$\gamma_i = b_0 + b_1 T_i + b_2 M_i$$

where γ is daily degree-days >5 °C for station i in month j; the b's are regression coefficients; and M is the mean annual temperature of station i. A minimum of 127 observations was available for these regressions (January), while the maximum was 2,118 (July and August). The smallest R^2 was 0.996.

Regression coefficients for each month are:

	b_0	b_1	b_2
January	–4.65118	0.96051	0.00807
February	–4.67189	0.95861	0.01169
March	–4.63736	0.90456	0.05482
April	–4.55469	0.93601	0.03633
May	–4.95205	1.01226	–0.01927
June	–4.98636	0.99862	0.00046
July	–4.99574	0.99926	0.00038

August	–4.99671	0.99959	–0.00023
September	–4.87564	1.00664	–0.01928
October	–4.68604	1.02324	–0.04397
November	–4.76599	0.96969	0.01103
December	–4.31612	0.94154	0.00331

Note that these regressions predict daily degree-days from temperature by month. To obtain the total for the month, the prediction must be multiplied by the number of days in the month, n_j. The annual accumulation of DD5 is then the sum of the monthly contributions:

$$DD5 = \sum_{j=1}^{12} \left(\gamma_{ij} n_j \right)$$

When this three-stage model is applied to the annual accumulation of DD5 for the 2,143 weather stations, a linear regression of observed values on the predicted produced an R^2 of 0.9994, an intercept of –15.5 DD5, and a regression coefficient of 1.0068 (fig. 3, right). The standard deviation of the residuals was 28.9 degree-days while the residuals themselves ranged from –132 to +142.

Degree-days <0 °C—Negative degree-day monthly normals were available for 1,868 stations for the United States and 275 for Canada west of 100°. The relationship between average monthly temperature and degree-days <0 °C expected to accumulate each day is shown in figure 4, left. The figure shows that at average daily temperatures >10 °C, the daily accumulation of negative degree-days is essentially zero, and that at temperatures < –10 °C, the daily accumulation of degree-days is essentially equal to the

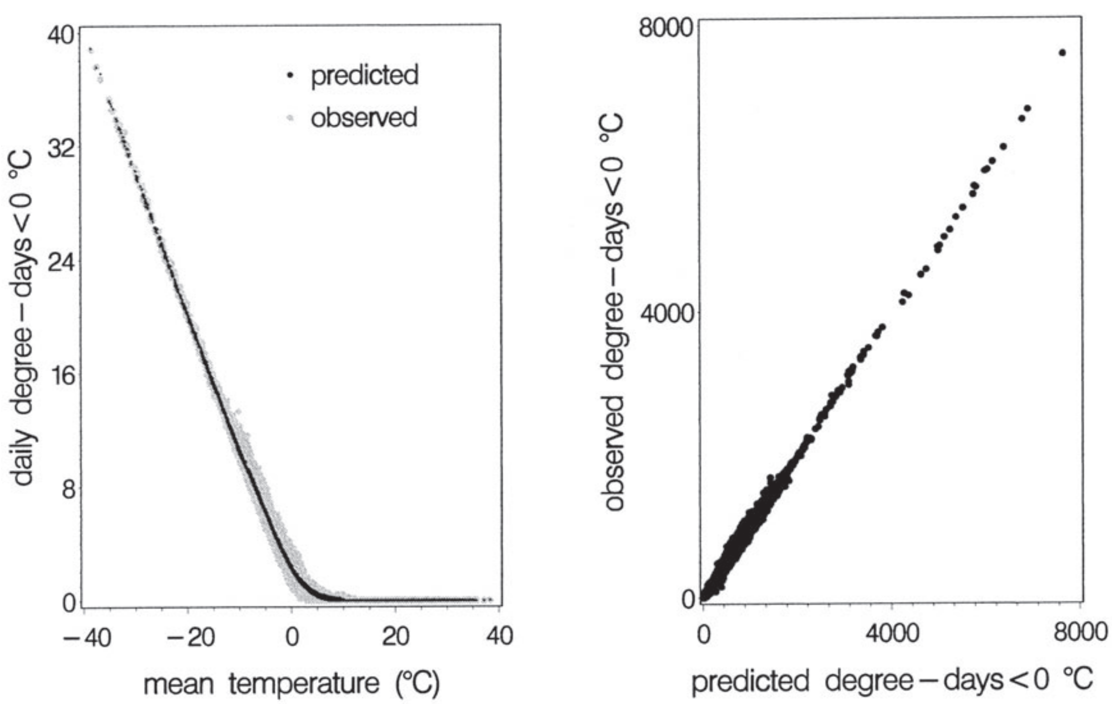


Figure 4—Plots of average daily temperature monthly means against the number of degree-days < 0 °C that accumulate per day during the month overlain with values predicted by the three-stage regression model (left), and relationship between observed annual sums of DD0 with the predicted sum (right).

average temperature. This means, first, that bias in estimating DD0 would accrue when 10 °C > T_{avg} > –10 °C and, second, that a three-stage model would also be appropriate for estimating DD0. As with DD5, therefore, regression models were used to fit mean monthly temperatures to the average daily contribution to the monthly total of DD0.

Stage 1: define φ as the daily contribution to the monthly sum of DD0, and equate φ to zero for all temperatures >10 °C.

Stage 2: fit the following function to daily for temperatures between –10 °C and +10 °C for each month:

$$\varphi'_{ij} = e^{a+b(T'_{ij})^{C}}$$

where φ'_{ij} is daily degree-days <0 °C for station *i*; *a, b,* and *c* are regression coefficients; *e* is the base of the Napierian logarithms; and *T'* is the average temperature for station *i* transformed as:

$$T'_{avg} = \left(T_{avg} + 75\right) / 20$$

This transformation produced values between 3.25 and 4.25 for T' which provided smoother transitions between stages than values between zero and 1.

The regression was based on slightly more than 21,000 observations, was statistically significant ($p < 0.01$), and accounted for 97 percent of the variance in the dependent variable. Regression coefficients for the best fitting model were:

a = 2.7862, b = -0.00000222, and c = 10.3641.

Despite the high degree of fit for this model, residuals were reduced further with the following multiple regression:

$$\varphi_{ij} = b_0 + b_1\hat{\varphi}'_{ij} + b_2 T_{i(j+1)} + b_3 C_{ij}$$

where $\hat{\varphi}'_{ij}$ is the predictedvalue of φ_{ij}; and *T* is the average temperature for station *i* in the month following $(j+1)$ month *j*, and *C* is the average temperature in the coldest of the two months adjacent to *j*.

Regression coefficients were: b_0 = –0.09082, b_1 = 1.00547, b_2 = 0.01629, and b_3 = –0.01658. This model accounted for 98 percent of the variance in the dependent variable, an increase of only 1 percent over the previous model. However, with more than 20,000 degrees of freedom, a seemingly small increase in R^2, produces a substantial decrease in the residual mean square. Also, because the model predicts daily contributions to the degree-day total, omitting a small correction for bias produces errors that are magnified when daily contributions are summed to provide annual estimates.

Stage 3: fit the following linear regression for temperatures < –10 °C:

$$\varphi_{ij} = b_0 + b_1 T_{ij} + b_2 T_{ij}^2 + b_3 W_{ij}$$

where *W* is the average temperature in the warmest of the 2 months adjacent to *j*.

The model accounted for 0.9969 of the variance in the dependent variable, and the regression coefficients were: b_0 = 1.35535, b_1 = –0.88994, b_2 = 0.00143, and b_3 = –0.03380.

Note, again, that these calculations predict daily degree-days from temperature by month. To obtain the total for the month, the prediction must be multiplied by the number of days in the month, n_j. The annual accumulation of DD0 is then the sum of the monthly contributions:

$$DD0 \sum_{j=1}^{12} \left(\varphi_{ij} n_j\right)$$

When this three-staged model is applied to the annual accumulation of DD0 for the 2,143 weather stations, a linear regression of observed values on the predicted produced an R^2 of 0.9993, an intercept of 3.45 DD0, and a regression coefficient of 0.991 (fig. 4, right). The standard deviation of the residuals was 53.5 degree-days, while the residuals themselves ranged from –234 to +257.

Minimum degree-days < 0 °C—Predicting MINDD0 can be done with the algorithms presented above for DD0 but by using monthly mean minimum temperatures instead of monthly average temperatures.

Estimation of Freezing Dates

Stepwise regression was used to develop algorithms for predicting the Julian date of the last freeze of spring (SDAY), the first freeze of fall (FDAY), and, from them, the length of the frost-free period (FFP) from monthly means of minimum temperatures (M_i) from 1,376 weather stations in Western North America (Koss and others 1988; Environment Canada 1994). Dependent variables included the first and second powers of the 12 monthly minimums plus interpolated estimates of the Julian dates when temperatures reached one of three threshold values: –2, 5, and 11°C. By assuming that the monthly mean temperature occurred on the median date, Julian dates of each threshold temperature could be interpolated for the spring (S_{-2}, S_5, and S_{11}) and autumn (F_{-2}, F_5, and F_{11}). Also included as independent variables in the stepwise model were the 12 interactions between the monthly mean temperatures and either S or F.

Weather station data showed that if the January mean minimum temperature was >6.8 °C, then SDAY=0; and if the December minimum temperature was greater than 7.5 °C, then FDAY=365. The 35 observations for which these conditions were true were removed from the data set, but as shown below, the conditions themselves were used as conditional expressions in the estimation of the freezing dates.

The remaining observations were assorted into three groups for the regression analyses. Group 1 comprised the 1,291 stations for which the minimum temperature in either July or August was greater than 5.0 °C; group 2 included the 52 observations for which the July or August temperature was less than 5.5 °C; and group 3 included the 151 observations for which the January or December temperature was greater than 2 °C. Notice that in providing a sufficient number of observations for regression analyses within these groups, some stations were represented in more than one group. Grouping was necessary to accommodate the disparity in climate among the weather stations. The interpolated dates appropriate to the first group were S_5 and F_5; $S_{-2,}$ and F_{-2} for the second group; and $S_{11,}$ and F_{11} for the third group. Regressions were fit for SDAY and FDAY for each group. The best fitting regression models were judged according to the significance of the coefficients, reduction in the residual mean square, and the Mallows statistic.

As an example, the stepwise model for predicting SDAY for the first group of stations was of the general form:

$$SDAY = b_0 + b_1 S_5 + b_2 M_1 ..... + b_{13} M_{12} + b_{14} M_1^2 ..... + b_{25} M_{12}^2 + b_{26} S_5 M_1 ..... + b_{37} S_5 M_1$$

where M's are monthly means of the month identified by the subscript for minimum temperatue.

The best fitting regression models for this group produced values of R^2 of 0.93 for SDAY and 0.97 for FDAY; the corresponding values for analyses of the second group were 0.73 and 0.78, and those for the third group were 0.78 and 0.79. The three stages together produced a model that accounted for about 97 percent of the variance in SDAY and FDAY. FFP was then calculated by subtracting SDAY from FDAY, and the relationship between this estimate and the actual value accounted for 97 percent of the variance in FFP of the 1376 stations. The relationships between the observed and predicted values are shown in Figure 5 for the three variables as are the dispersion of residuals against predicted values of FDAY.

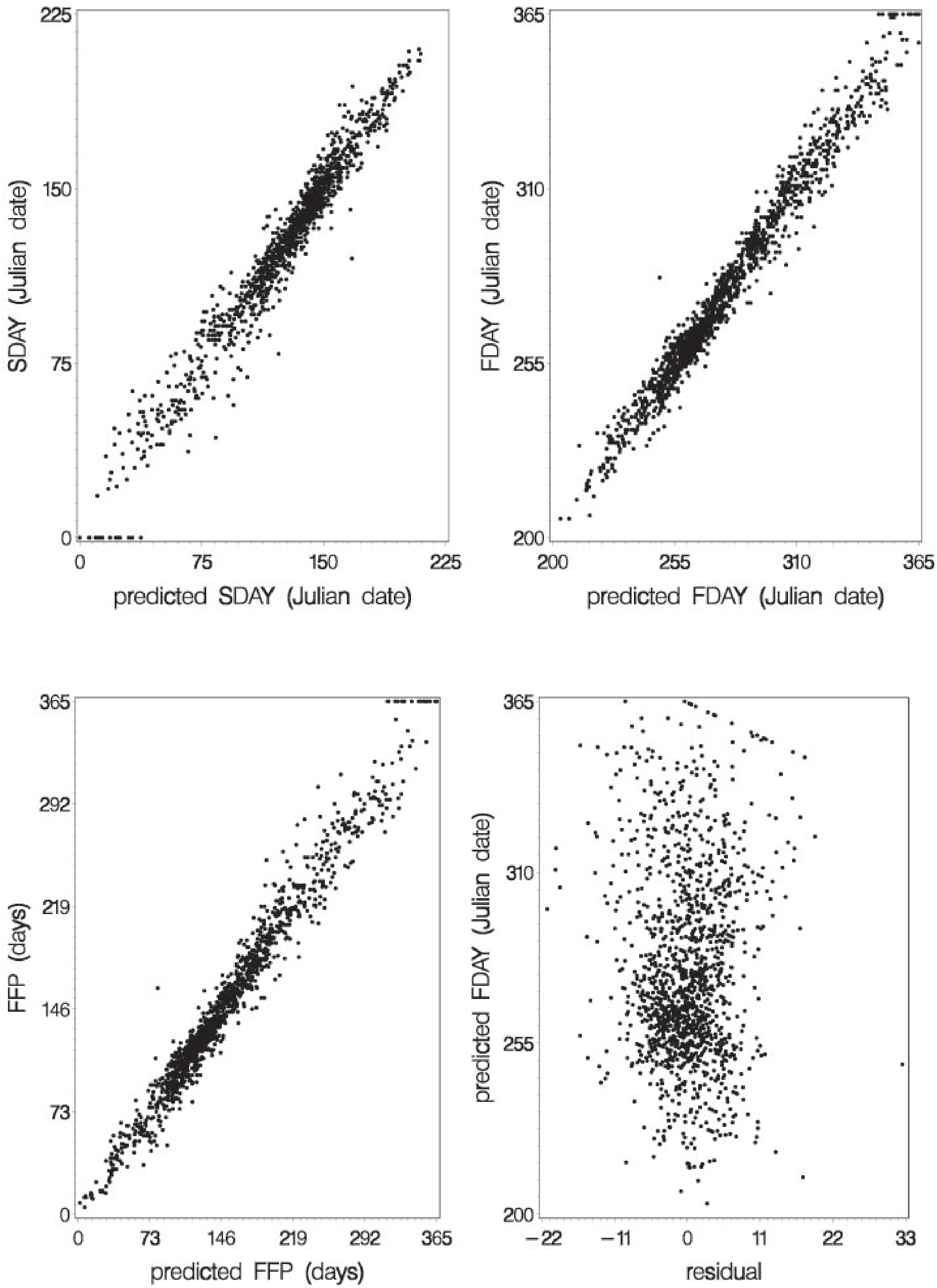


Figure 5—Plots of observed freezing data against those predicted by the three-stage regression models for SDAY, the last frost of spring (upper left), FDAY, the first frost of autumn (upper right), and FFP, the length of the frost-free period (lower left). At lower right is the relationship between predicted values of FDAY and the residual from the regression model.

Models of the three groups of data together with the conditional expressions produced the following general model for the freezing variables:

If M_7 or $M_8 \geq 5.5$ then:

$$SDDAY = -1.08 + 93S_5 + 2.08M_2 + 1.9M_{11} - 3.85M_{12}$$
$$FDAY = 30.28 + 0.92F_5 - 1.80M_6 + 1.84M_9$$

If M_7 or $M_8 < 5.5 then$:

$$SDAY = 213.11 - 0.08M_{10}^2 - 2.65M_9 - 0.04S_{-2}M_7$$
$$FDAY = 211.97 + 5.75M_8 - 9.23M_6 + 0.05F_{-2}M_6$$

If M_1 or $M_{12} > 2.0$ and $M_7 > 11$ then:

$$SDAY = 147.13 - 12.24M_2 - 0.07S_{11}M_4 + 0.03S_{11}M_6$$
$$FDAY = 282.28 + 8.95M_2 - 0.46M_2^2 + 6.35M_{12}$$

IF JAN > 6.8 THEN SDAY = 0;
IF DEC >7.5 THEN FDAY = 365;
IF SDAY < 0 THEN SDAY = 0;
IF FDAY > 365 THEN FDAY = 365;
FFP = FDAY – SDAY
IF SDAY > FDAY THEN FFP = 0

Estimating Temperature Sums for Specified Periods

The procedures described above for estimating degree-days use the daily contribution to the total for each month. These contributions can be summed for each day to produce an annual accumulation (fig. 6) from which estimates of degree-days within a specified period can be made. Two examples follow.

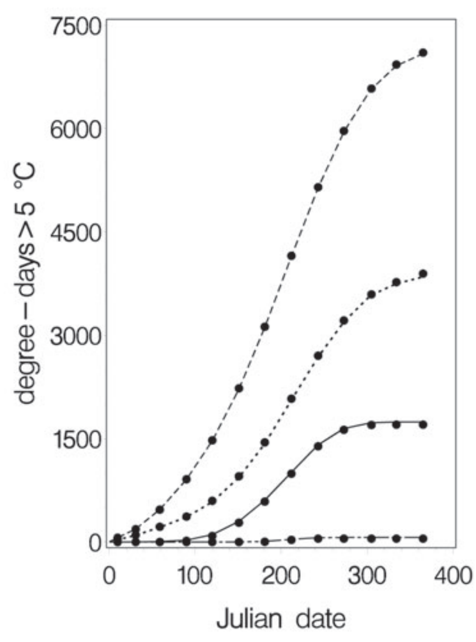


Figure 6—Plots of observed (dots) and predicted (lines) sums of degree-day > 5 °C according to Julian date for four stations with different summer temperature regimes. From upper to lower: Death Valley, CA, Placerville, CA, Priest River Experimental Forest, ID, Barrow, AK.

At low elevations (for example, 950 m) at the Priest River Experimental Forest in northern Idaho, winter dormancy in most organisms (from trees to mosquitoes) has been arrested by May 1. Climate normals from the Experimental Forest show that DD5 on May 1 averages 109. One can define a climate variable, therefore, as $DD5_{100}$, the Julian date on which degree-days > 5 °C reach 100. This value can be obtained directly by from the dataset of daily degree-day sums.

Another example may involve the amount of summer warmth available for plant growth and development. Such a variable can be defined as the number of degree-days > 5 °C that accumulate during the frost-free period ($DD5_{ffp}$). This value can be estimated from the difference in degree-days between FDAY and SDAY.

Applications

The spline surfaces have many uses in biology. McKenney and others (2001), for instance, have already used such surfaces to revise and remap plant hardiness zones for Canada. The following examples provide a sample of additional applications.

Point Estimates

In ecological genetics, researchers deal with the ecological bases for genetic differentiation among populations. Because climate is the primary factor controlling differentiation, ecological genetic concepts must be couched in terms of climate. This would require, therefore, estimates of climate for individual populations and test sites. For a sample of 10 populations of *Picea engelmannii*, table 5 lists six climate variables that described the climatic distribution of the species south of latitude 51 °N. The table shows that across a latitudinal spread of about 17 ° and an altitudinal distribution of nearly 3,000 m, populations of this species occupy climates differing by as much as 10 °C in mean annual temperature, 1,500 degree-days > 5 °C, 1700 degree-days < 0 °C, 700 mm in annual precipitation, 11 °C in mean temperature in the coldest month, and 120 days in the length of the frost-free season. Nonetheless, populations as geographically disparate as north-central Washington (altitude = 762 m) and southern New Mexico (altitude = 2,622 m) can occupy remarkably similar climates.

Mapping Climate Surfaces

Hutchinson's ANUSPLIN software (Hutchinson 2000) can output a grid file in that point estimates are made for grids which are suitable for mapping. Figure 7 illustrates a map of degree-days > 5 °C for western North America south of latitude 51 ° N that uses a digital elevation model on a 1 km grid (0.0083333 degrees) (GLOBE 1999). There are nearly 7 million pixels in this map, and for each, an estimate of degree-days > 5 °C has been made with Hutchinson's software. Predicted values range from –2,331 to 6,700. Values less than zero arise from extrapolation, are absurd, and can be equated to zero. Negative values as large as –500 occurred, for instance, in the ice fields of the Wind River Range in Wyoming, but the largest negative values occurred on the volcanoes in the Cascade Range.

Figure 8 shows a portion of the map of figure 7 at a scale exposing the underlying pixels. The map can easily be confused for the elevation model that was used in its construction, but nonetheless represents degree-days. The map includes the communities of Lewiston, ID, and Clarkston, WA, in the upper left plus the Grande Ronde, Snake, Salmon, and Clearwater Drainages. The regions with the coolest summers are in the Eagle Cap Wilderness toward the southwest. Values of RTGCV

Table 5—Point estimates of six climate variables for 10 climatically disparate populations of *P. engelmannii.*

	Location			Climate variable					
Administrative unit	**Latitude (°N)**	**Longitude (°W)**	**Altitude (m)**	**Mean annual temperature (°C)**	**Degree-days >5°C**	**Degree-days <0°C**	**Mean annual precipitation (mm)**	**Coldest month minimum temperature (°C)**	**Length of frost-free period (days)**
Colorado	39.67	105.88	3598	–2.2	340	1932	789	–17.2	10
Utah	40.72	110.85	3178	–0.4	574	1582	741	–14.8	36
British Columbia	49.37	117.45	1997	0.3	599	1405	1157	–12.5	62
Montana	44.82	113.30	2226	1.8	934	1268	346	–16.3	33
Arizona	33.93	109.63	2927	4.0	997	668	839	–12.0	60
Idaho	44.67	116.13	1616	4.4	1288	829	686	–12.3	59
Montana	47.42	115.63	1159	5.7	1452	593	955	–8.4	93
Idaho	46.58	114.62	1204	5.8	1507	610	713	–9.6	77
Washington	48.93	118.88	762	6.7	1781	577	395	–9.2	115
New Mexico	32.80	105.70	2622	8.1	1783	209	694	–6.3	131

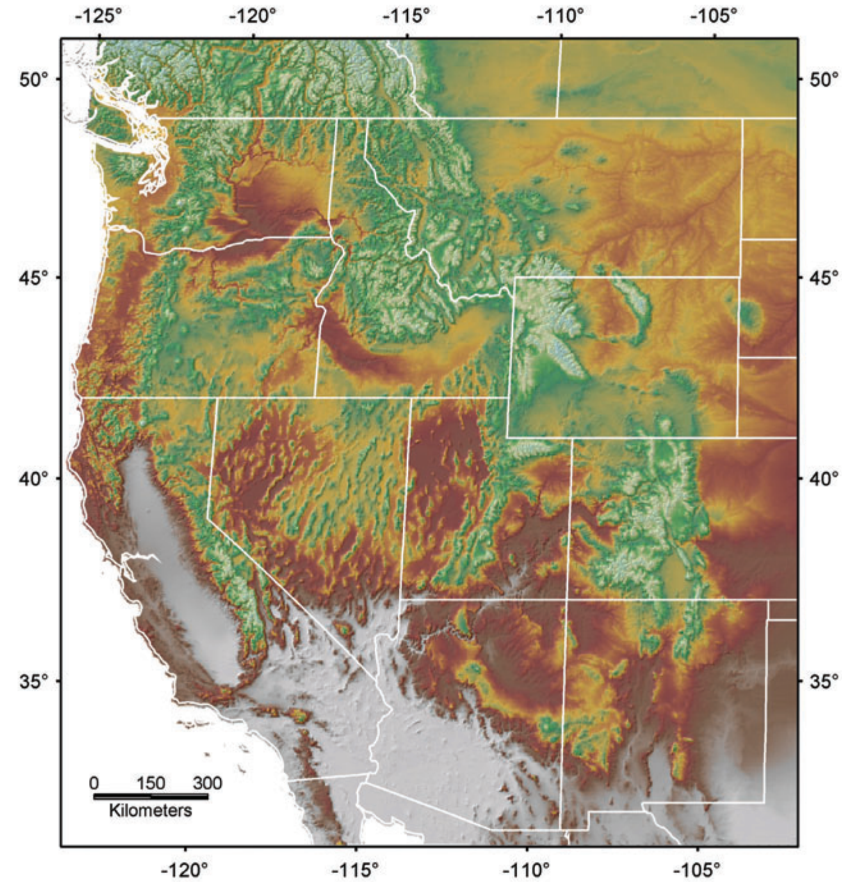


Figure 7—Map of degree-days >5 °C ranging from zero (light blue) to 6,700 (silver).

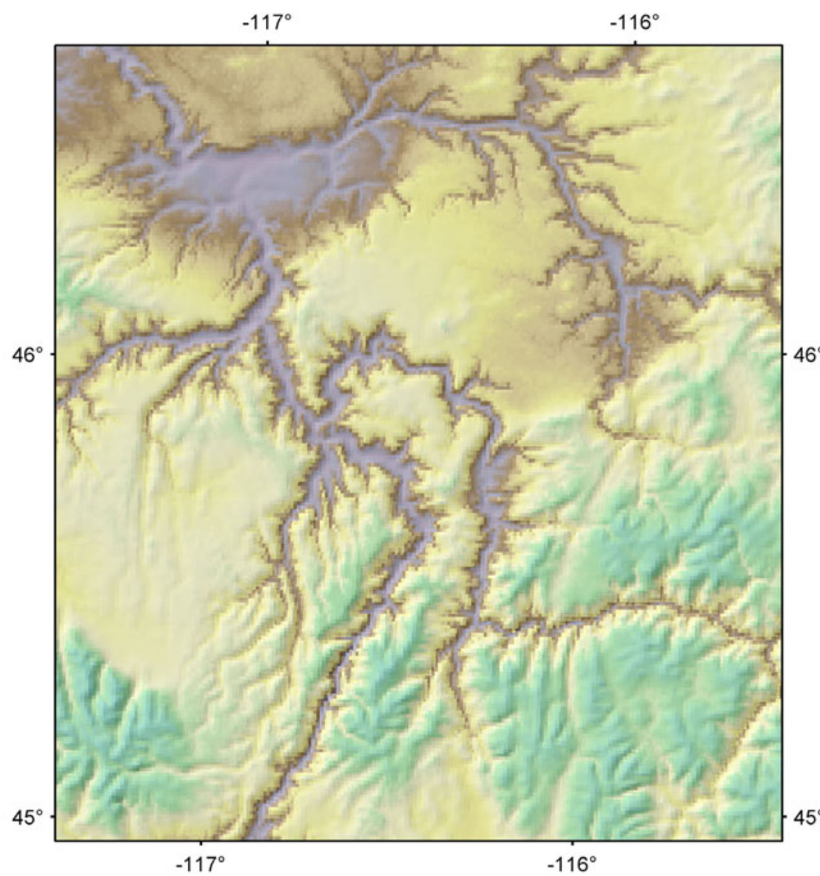


Figure 8—Map of degree-days > 5 °C using shades of grey to code 19 classes ranging from zero (light blue) to 3,700 (gray) for an area surrounding Lewiston, ID, and Clarkston, WA, (upper left). Major drainages counterclockwise are the Grande Ronde, Snake, Salmon, and Clearwater.

for the degree-day spline suggest that 95 percent confidence intervals about predicted points are about ±45 degree-days.

Separating species in climatic space—Paleoecologists commonly use the fossil pollen records of vegetation to infer climate. This can become problematic because in many instances, distributions in climate space are unknown for taxa of interest. For example, reconstructing the composition of plant communities at high altitudinal sites in the Northern Rocky Mountains of the United States requires knowledge of the conditions under which either *P. albicaulis* or *P. contorta* can be dominate (Brunelle and Whitlock 2003; Brunelle and others 2005). To address this problem, the spline climate model was used to make point predictions for 277 locations known to be inhabited by *P. contorta* and 74 locations inhabited by *P. albicaulis*. Although these two species can be sympatric, it is not known at which of these locations they co-occur. Eighteen derived variables were calculated for each population, and a principal component analysis was made for the total assemblage. The species separated nicely according to the first and second principal components (fig. 9, left). The first component was strongly influenced by the summer temperature and the second by the annual moisture index. Consequently, the species could be separated nearly as well by these variables as by the principal components (fig. 9, right). The conclusion would be that *P. contorta* generally occurs at locations that are warmer and drier than those inhabited by *P. albicaulis*. This suggests for paleoecological reconstructions that forests dominated by *P. albicaulis* are cooler and wetter than those dominated by *P. contorta*. While this is an intuitive conclusion, this spline model provides the first definite evidence of this relationship.

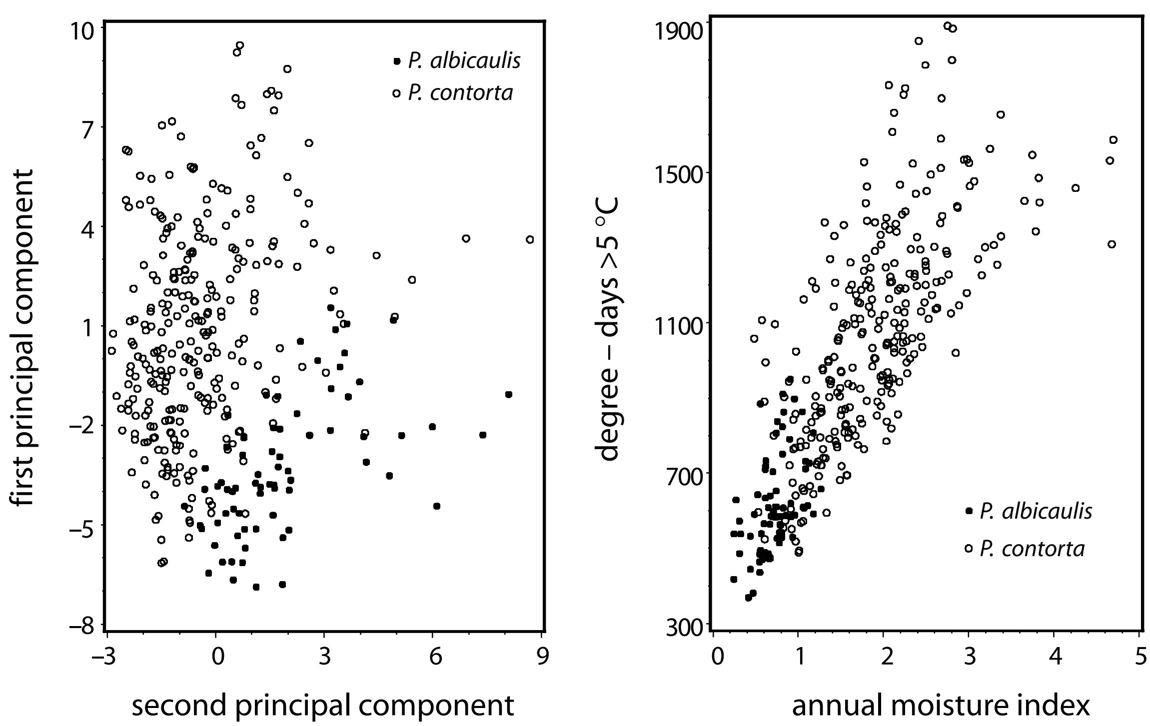


Figure 9—Scatter plots of 277 populations of *P. contorta* and 74 populations of *P. albicaulis* according to the first two principal components of 18 climate variables describing their provenance (left), and to degree-days > 5 °C and annual moisture index of the provenance (right) (from Rehfeldt, unpublished).

Evaluating genetic variation along climatic gradients—For years, forest geneticists have been using geographic variables to describe genetic differences among populations as they exist across landscapes. The underlying assumption in these analyses has been that the geographic variables were acting as surrogates for climate variables that were difficult to observe. The spline climate model, however, offers an opportunity to replace the surrogates with operative climate variables. Figure 10 illustrates genetic variation among 215 populations of *Picea engelmannii* from the Western United States in relation to the best fitting geographic variable (left) and the best fitting climate variable (right). The geographic surrogate, elevation, is a weak (R^2 = 0.07) but nonetheless a statistically significant ($p < 0.05$) predictor of variation among populations. By contrast, the winter temperature of the provenance, degree-days < 0 ºC, is a strong predictor (R^2 = 0.62) of genetic differentiation. In this example, geographic variables were poor surrogates for climate and led, therefore, to a fallacious view of the species' genetic structure.

Conclusions

Climate models developed from thin plate splines provide geographic surfaces that can be addressed to obtain point estimates of climate. Errors of prediction of these estimates are available, and the estimates can be converted to variables known to be relevant in biology. Spline climate models have unlimited applications in assessing plant responses to climate.

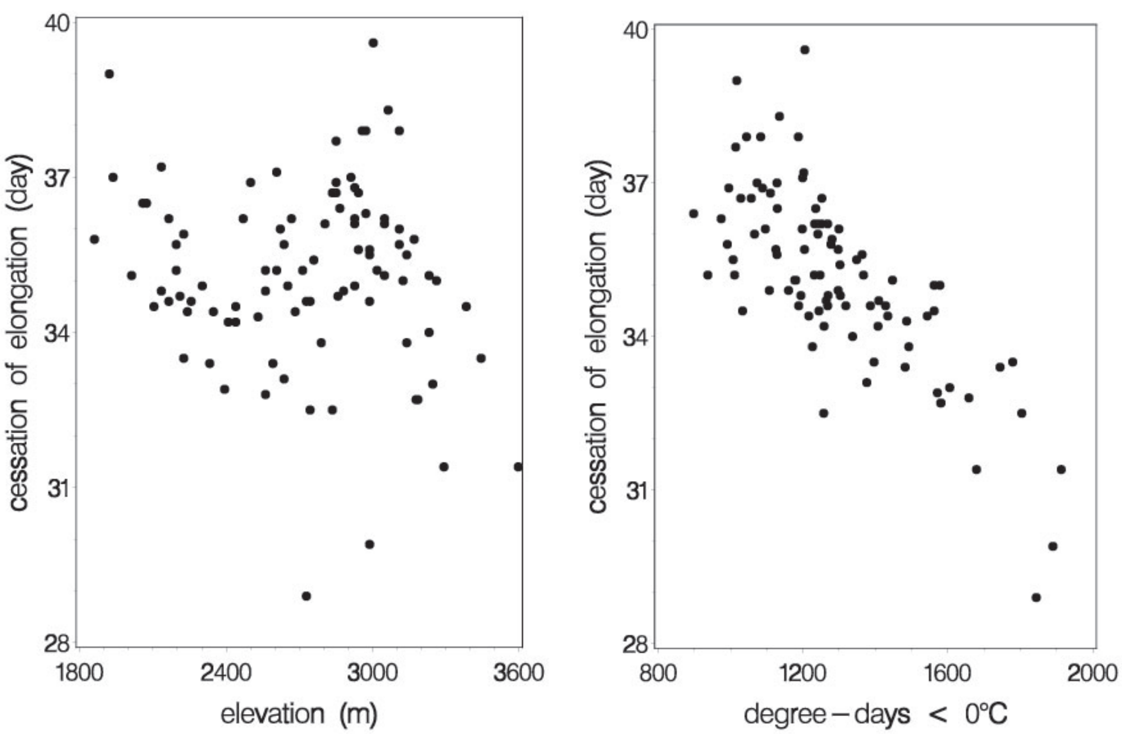


Figure 10—Mean date of the cessation of shoot elongation of 2-year *Picea engelmannii* populations measured in greenhouse-shadehouse studies plotted against the provenance elevation (left) and provenance degree-days < 0° (right) (from Rehfeldt, et al. 2004).

References

Brunelle, A.; Whitlock, C. 2003. Postglacial fire, vegetation, and climate history in the Clearwater Range, Northern Idaho, USA. Quaternary Research. 60: 307-318.

Brunelle, A.; Whitlock, C.; Bartlein, P.J.; Kipfmueller, K. 2005. Holocene fire and vegetation along environmental gradients in the Northern Rocky Mountains. Quaternary Science Reviews. 24: 2281-2300.

EarthInfo, Inc. 1994. Database Guide. Earthinfo, Inc. Boulder, CO.

Environment Canada. 1994. Canadian Monthly Climate Data and 1961-1990. Normals. Prairie and Northern Region Commercial Weather Services Division, Edmonton, Alberta.

GLOBE Task Team. 1999. The Global Land One-kilometer Base Elevation (GLOBE) Digital Elevation Model, Version 1.0. National Oceanic and Atmospheric Administration, National Geophysical Data Center, 325 Broadway, Boulder, Colorado 80303, U.S.A.

Hutchinson, M. F. 1991. Continent wide data assimilation using thin plate smoothing splines. Pages 104-113, In: J.D. Jasper, ed. Data assimilation systems. BMRC Research Report 27, Bureau of Meteorology, Melbourne, Australia.

Hutchinson, M. F. 2000. ANUSPLIN Version 4.1 User's Guide. Australian National University, Centre for Resource and Environmental Studies, Canberra.

Koss, W.J.; Owenby, J.R.; Steurer, P.M. Devoyd, S.E. 1988. Freeze/frost data. Climatography of the U. S. No. 20. Supplement No. 1. National Oceanic and Atmospheric Administration, National Climatic Data Center, Ashville, North Carolina.

McKinney, D.W.; Hutchinson, M.F.; Kesteven, J.L.; and Venier, L.A. 2001. Canada's plant hardiness zones revisited using modern climate interpolation techniques. Canadian Journal of Plant Science. 81: 129-143.

Rehfeldt, G.E.; Tchebakova, N.M.; Parfenova, E. 2004. Genetic responses to climate and climate change in conifers of the temperate and boreal forests. Recent Advances in Genetics and Breeding. 1:113-130.

Tuhkanen, S. 1980. Climatic parameters and indices in plant geography. Acta Phytogeographica Suecica. 67:1-105.

United States Department of Commerce. 1994. U.S. Divisional and Station Climatic Data and Normals. Volume 1. National Oceanic and Atmospheric Administration, National Climatic Data Center, Ashville, North Carolina.

Woodward, F.I. 1987. Climate and plant distribution. Cambridge University Press, London.

RMRS

ROCKY MOUNTAIN RESEARCH STATION

The Rocky Mountain Research Station develops scientific information and technology to improve management, protection, and use of the forests and rangelands. Research is designed to meet the needs of National Forest managers, Federal and State agencies, public and private organizations, academic institutions, industry, and individuals.

Studies accelerate solutions to problems involving ecosystems, range, forests, water, recreation, fire, resource inventory, land reclamation, community sustainability, forest engineering technology, multiple use economics, wildlife and fish habitat, and forest insects and diseases. Studies are conducted cooperatively, and applications may be found worldwide.

Research Locations

Flagstaff, Arizona
Fort Collins, Colorado*
Boise, Idaho
Moscow, Idaho
Bozeman, Montana
Missoula, Montana
Reno, Nevada
Albuquerque, New Mexico
Rapid City, South Dakota
Logan, Utah
Ogden, Utah
Provo, Utah

*Station Headquarters, Natural Resources Research Center, 2150 Centre Avenue, Building A, Fort Collins, CO 80526