

# Interpolation of temperature and non-urban ozone exposure at high spatial resolution over the western United States

E. Henry Lee\*, William E. Hogsett

US Environmental Protection Agency, Western Ecology Division, 200 SW 35th Street, Corvallis, Oregon 97333, USA

**ABSTRACT:** In order to assess the impact of natural and anthropogenic stresses on forest ecosystems, it is necessary to interpolate air temperature and tropospheric ozone (O<sub>3</sub>) exposure values at high spatial resolution over complex terrain. The proposed interpolation approach was selected because of its ability to (1) account for the effect of elevation on temperature and their effects on tropospheric ozone, (2) use auxiliary data at higher spatial resolution than the variables of interest to improve the precision and accuracy of the prediction surfaces, (3) handle large amounts of data, and (4) provide not only a prediction at nonsampled locations but also a prediction standard deviation. The approach used auxiliary digital elevation model (DEM) data at 1 km resolution to improve the precision and resolution of the predictions for temperature and O<sub>3</sub> exposure in the western United States. Initially, the study area was stratified into O<sub>3</sub> regions based on seasonality and variability of monthly SUM06 values at 111 stations for the period 1990–1992 using rotated principal component analysis. Monthly mean daily maximum air temperatures were spatially interpolated using loess nonparametric regression and kriging of the loess residuals and interpolated to 2 km grid points of a DEM and to the ambient air quality monitoring points. Monthly O<sub>3</sub> exposures were spatially interpolated using loess fits to relate O<sub>3</sub> levels to elevation, predicted temperature, and the geographic coordinates and interpolated to 2 km grid points of a DEM. The elevation-based spatial interpolation procedure produced accurate and precise temperature and O<sub>3</sub> exposure surfaces which had desirable statistical properties and were logically consistent with local topographical features and atmospheric conditions known to influence O<sub>3</sub> formation and transport. The leave-one-out cross-validation mean absolute error was 0.93°C for the monthly mean daily maximum temperature and 1.93 ppm-h for the monthly SUM06 index for June 1990 for the western United States, comparable to published results for other regions at smaller spatial scales with less complex terrain.

**KEY WORDS:** Spatial interpolation · Ozone · Temperature · Regionalization · Principal components analysis · Kriging · Loess regression · Western United States

*Resale or republication not permitted without written consent of the publisher*

## 1. INTRODUCTION

Tropospheric ozone (O<sub>3</sub>) is one of the most pervasive and harmful air pollutants known to affect natural and agricultural ecosystems in the United States. The consensus among environmental scientists is that O<sub>3</sub> is a major contributor to the decline in growth of forest species at some locations. Primary evidence comes

from a long-term study in the San Bernardino Mountains of California which concluded that O<sub>3</sub> was the primary cause of foliar injury, premature leaf drop, decreased radial growth and photosynthetic capacity, and death by bark beetles in ponderosa and Jeffrey pine (Miller 1984, Miller & McBride 1999). Chronic exposures to high O<sub>3</sub> concentrations have altered the water, carbon, and nutrient fluxes of the forest ecosystem and predisposed trees to other stresses in the San Bernardino Mountains (Miller 1973). It is clear that even modest concentrations can damage forests. This

\*E-mail: lee.ehenry@epamail.epa.gov

research is concerned with the spatial interpolation of non-urban O<sub>3</sub> exposures and its driving factors on a fine grid for assessing the impact of natural and anthropogenic stresses on forests in the United States.

This study addresses one part of a larger research project to scale our knowledge of vegetative processes, gained from experiments on individuals, to the stand, ecosystem, and higher biological assemblages using a plant growth model with uncertainty. One of the research goals is to examine how uncertainties and variability in parameters and inputs propagate through the model, so it is important to quantify the uncertainties in the spatial predictions of model inputs. Accurate estimates of regionalized variables such as O<sub>3</sub> at unsampled sites and their prediction standard deviations (SDs) at high spatial resolution are critical to characterize the risk of forests to tropospheric O<sub>3</sub> within a GIS-based framework (Hogsett et al. 1997).

The spatial interpolation method used in this study was selected because of its ability to account for relationships between elevation, temperature and O<sub>3</sub> exposure, to regionalize the study area into homogeneous subsets, and to handle large volumes of data. To illustrate the spatial interpolation approach, we have used the western United States (Arizona, California, Colorado, Idaho, Montana, New Mexico, Nevada, North Dakota, Oregon, South Dakota, Utah, Washington, and Wyoming) as the study area. The study area was chosen because of the importance of western forest ecosystems for maintenance of an adequate supply of high quality water and the long-term goal of assessing the risk to important western forest species of exposure to natural and anthropogenic stresses.

We illustrated our spatial interpolation approach using monthly data for June 1990 and compared our results with those in the published literature on the basis of interpolation accuracy and precision. As the true surfaces for monthly temperature and O<sub>3</sub> exposure values were not known, the jackknife procedure of model cross-validation was used to evaluate the statistical properties of our interpolation technique. Data points were deleted one at a time, the remaining points were used to obtain predictions at the deleted point, and the estimated and actual values were compared. The mean error (bias), mean absolute error, standard error for jackknife predictions, mean prediction SD, and correlation between predicted versus observed values were calculated as measures of interpolation accuracy and precision.

## 2. SPATIAL INTERPOLATION

The main difficulties in spatial interpolation of non-urban O<sub>3</sub> exposure over the western United States are:

(1) the spatial dependencies between monitoring sites do not satisfy the assumptions of kriging or cokriging in complex terrain; (2) O<sub>3</sub> monitoring stations in the United States are sparse and are located primarily in urban centers and at lower elevations; (3) spatial variability in O<sub>3</sub> concentrations is high in the western United States, which includes pristine sites (Evans et al. 1983, Lefohn et al. 1990) and the most polluted sites in the country (i.e. San Bernardino County in California, US EPA 1992); and (4) the network of meteorological monitors is lacking at high elevations. Because of complex transport phenomena and chemistry, O<sub>3</sub> concentrations may vary within short distances between areas of different altitudes or topography. For example, local topographical characteristics and their effect on transport of polluted air have a significant influence on the diurnal pattern of O<sub>3</sub> concentrations at non-urban sites on the western slope of the Sierra Nevada in California (Van Ooy & Carroll 1995).

There is no standard procedure available for spatial interpolation of data in complex terrain and nonuniform orography. We propose a semi-parametric approach using ancillary data from a digital elevation model (DEM) at high spatial resolution to model the large-scale variability in the variable of interest and kriging on the regression residuals to model the small-scale variability and spatial dependencies. The elevation function/residual interpolation approach uses the height dependency as the basis for interpolation over complex terrain and compensates for a lack of monitoring of the variable of interest. Previous studies have utilized the more densely sampled DEM data to improve spatial predictions for regionalized variables over smaller regions (<10 million ha) in the United States (Phillips et al. 1992, Ollinger et al. 1995, Dodson & Marks 1997) and Austria (Loibl et al. 1994). However, the elevation-based parametric approaches for predicting precipitation were not applicable for larger regions due to a lack of a consistent precipitation-elevation relationship (Phillips et al. 1992, Ollinger et al. 1995).

Our analysis shows that there is a lack of consistent relationships between temperature, SUM06 and elevation in complex terrain on the spatial scales we are considering. It is necessary to divide the region into a manageable number of homogeneous subregions at spatial scales in which the relationships can be adequately described by the elevation function/residual interpolation approach. The multivariate technique rotated principal component analysis (RPCA) was used to regionalize the western United States into smaller regions whose O<sub>3</sub> exposures exhibit unique homogeneous characteristics in time and space.

The interpolation approach of this paper is a continuation of the work presented in Phillips et al. (1997).

For the geographic range of loblolly pine *Pinus taeda* L. in the southeastern United States, cokriging with auxiliary data on factors known to influence O<sub>3</sub> formation and transport (i.e. NO<sub>x</sub> [nitrogen oxides] emissions and meteorology) was recommended to better predict the monthly SUM06 index. Cokriging allows data from more intensively sampled correlated variables to compensate for a lack of non-urban O<sub>3</sub> monitoring data. Also, cokriging, like kriging, gives optimal estimators in the sense of minimized estimation variance and provides a prediction SD as a measure of precision. Cokriging works well over relatively flat, homogeneous terrain such as in the southeastern United States. However, the assumption of local stationarity of the mean and spatial covariance structure is not valid in the western United States, where the terrain and topography are considerably more complex.

Instead of cokriging, we propose a semi-parametric approach using a loess nonparametric regression model (Cleveland 1979, Cleveland et al. 1992) and kriging of the loess residuals for spatial interpolation of monthly temperature and O<sub>3</sub> exposure in complex terrain. Separate loess-kriging models will be fitted to each spatially cohesive O<sub>3</sub> region in the western United States. The loess regression is computationally efficient, has the ability to adapt to shifting relationships between elevation, climate and O<sub>3</sub> exposure, and optimizes the weighting of nearby sampled points to predict at the cell points on the DEM surface. The loess predictions for temperature and O<sub>3</sub> exposure are based on modeling the mean structure as a function of elevation, location, and local topography. We prefer the direct use of auxiliary data at higher sampling density in modeling the mean structure rather than modeling the spatial covariance structure (i.e. cokriging) to better predict the regionalized variable of interest.

### 3. METHODS

#### 3.1. Digital elevation model data

The study area was defined as the 13 conterminous states west of the 96th meridian which contain the western forest ecosystems. The study area was represented by a 15 arc-second DEM which was projected to an Albers equal-area conic map projection. Elevation data are from the United States Geological Survey (USGS) Conterminous US Advanced Very High Resolution Radiometer (AVHRR) Companion Disc (Love-land et al. 1991) and resampled by the USGS to 1 km resolution. The DEM data will be used to spatially interpolate monthly air temperature at high spatial resolution and, subsequently, with air temperature to spatially interpolate O<sub>3</sub> exposure at the same re-

solution. For illustration of the spatial interpolation method, the time period is June 1990; however, any month or aggregate of several months can be used.

#### 3.2. Ozone data

The O<sub>3</sub> monitoring data employed in this analysis were obtained from the US Environmental Protection Agency's (EPA) Aerometric Information Retrieval System (AIRS), the US EPA's Clean Air Status and Trends Network (CASTNet), the Sierra Cooperative Ozone Impact Assessment Study (SCOIAS), and the US Department of Agriculture Forest Service, Pacific Southwest Research Station. The EPA's AIRS network is a database of hourly aerometric data collected by state and local agencies at nearly 800 stations nationwide, of which about 27% are from the 13 western states. The EPA's CASTNet database was established in 1986 as a monitoring program to characterize dry deposition patterns and trends at 50 predominantly rural sites in the United States, of which 9 are from the western states (Clarke et al. 1997, Baumgardner & Edgerton 1998).

A major consideration of this study was the spatial interpolation of non-urban O<sub>3</sub> exposures for assessing the potential adverse effects of O<sub>3</sub> on western forest ecosystems. We characterized ambient exposure using the monthly 12 h SUM06 index because the current EPA review of the secondary O<sub>3</sub> standard recommends consideration be given to this exposure index. The SUM06 value was calculated as the sum of all hourly O<sub>3</sub> concentrations  $\geq 0.06$  ppm for each day (08:00 and 20:00 h). The daily values were considered valid if 75% or more of the hourly O<sub>3</sub> concentrations between 08:00 and 20:00 h were available. The daily values were summed over a month, and the monthly SUM06 values were adjusted for missing values. The monthly SUM06 values were considered valid if there were at least 75% valid days in the month.

To minimize the effect of nitrous oxide (NO) scavenging in urban areas, all O<sub>3</sub> monitoring stations described as either rural or suburban and reporting a land use of either forest, agricultural, residential, desert or mobile received highest priority in the analysis. It is well documented that daily maximum O<sub>3</sub> concentrations in non-urban areas can exceed urban concentrations due to a lack of NO scavenging (Logan 1989). Initially, 128 non-urban monitoring stations met our data completeness criterion for June 1990 and were retained for analysis. However, to meet our criterion for spatial completeness and increase the spatial coverage in areas where elevation changes rapidly over short distances, we included an additional 76 monitoring sites that met our data completeness criterion for

June 1990 or June from other years between 1988 and 1994. The 76 supplemental  $O_3$  monitoring sites included 40 urban sites with SUM06 for June 1990 and 27 non-urban and 11 urban sites with SUM06 for June from other years. For the 38 stations with missing monthly SUM06 for June 1990, the June 1990 SUM06 value was imputed as the mean of the monthly SUM06 values for June 1988 to 1994. Our regression analysis indicated that the mean June SUM06 values ( $x$ ) at sampled sites in other years was a good predictor for the June 1990 SUM06 index ( $y$ ) ( $y = 1.03x$ ,  $r^2 = 0.96$ , root mean squared error = 2.16 ppm-h).

These selection criteria resulted in the inclusion of 204 AIRS and other monitoring stations across the western United States, the majority of which are in California along the coast and west of the Sierra Nevada (Fig. 1). The AIRS monitoring sites are located predominantly at lower elevations (<1000 m) and only 8 AIRS stations are located at high elevation (2000 to 2800 m). The 9 CASTNet sites are located at higher elevations (963 to 2926 m) in areas where the AIRS network was lacking in spatial coverage.

### 3.3. Meteorological data

Meteorological data (i.e. maximum daily temperature) came from a combined set of monitoring stations from the National Weather Service (NWS) cooperative network archived at the National Climatic Data Center (NCDC), and the National Resources Conservation Service (NRCS) SNOTEL network. The locations of these stations are shown in Fig. 2. The daily maximum temperature data for 2109 meteorological stations from the NCDC Summary of the Day (SOTD) TD-3200 database were obtained from EarthInfo (EarthInfo 1992a). The daily maximum temperatures were averaged across a month in accordance with the SUM06 calculations. Hourly meteorological data including temperature for 85 weather stations from NCDC Surface Airways (SA) TD-3280 database were also obtained from EarthInfo (Earthinfo 1992b). The hourly temperature data were used to calculate the daily maximum temperature and then averaged across a month at each site. Only 14 of the 85 meteorological stations in the SA data set were not included in the SOTD data set. The NCDC meteorological stations have good spatial coverage at lower elevations

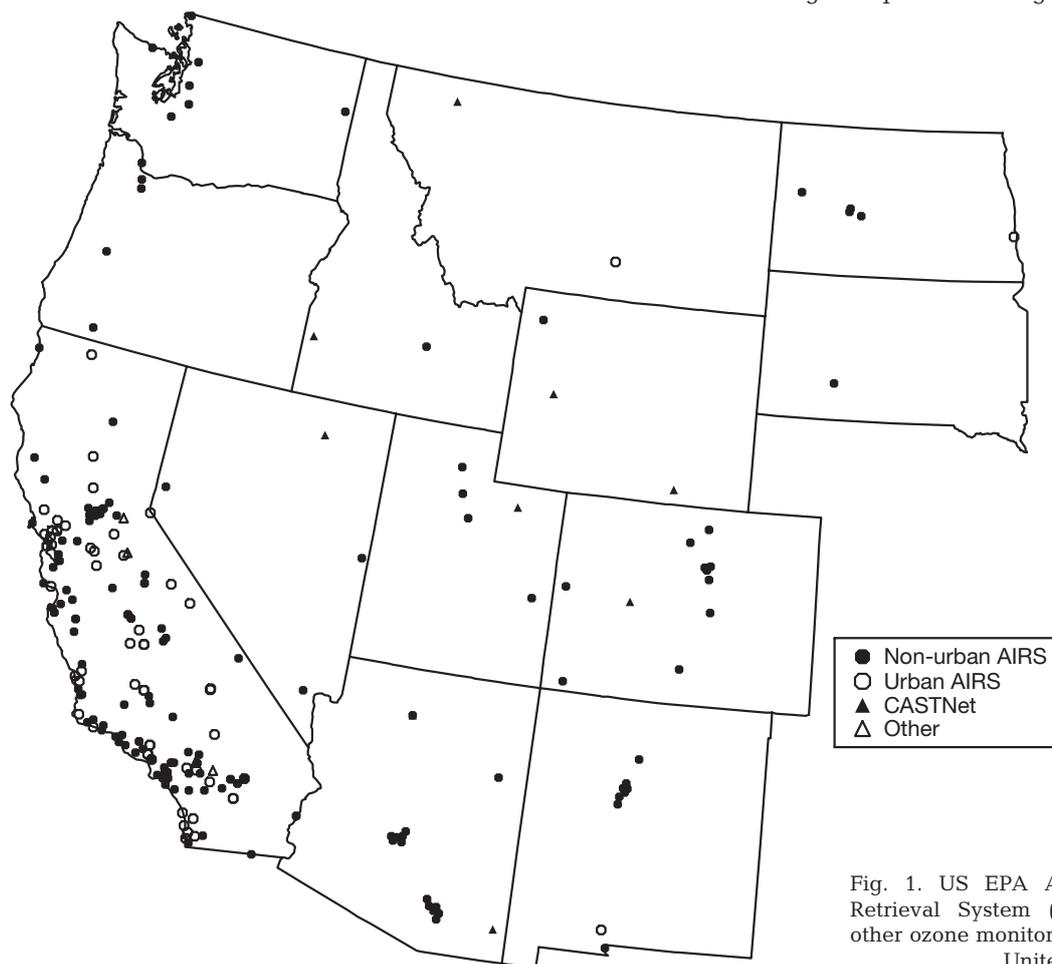


Fig. 1. US EPA Aerometric Information Retrieval System (AIRS), CASTNet and other ozone monitoring sites in the western United States

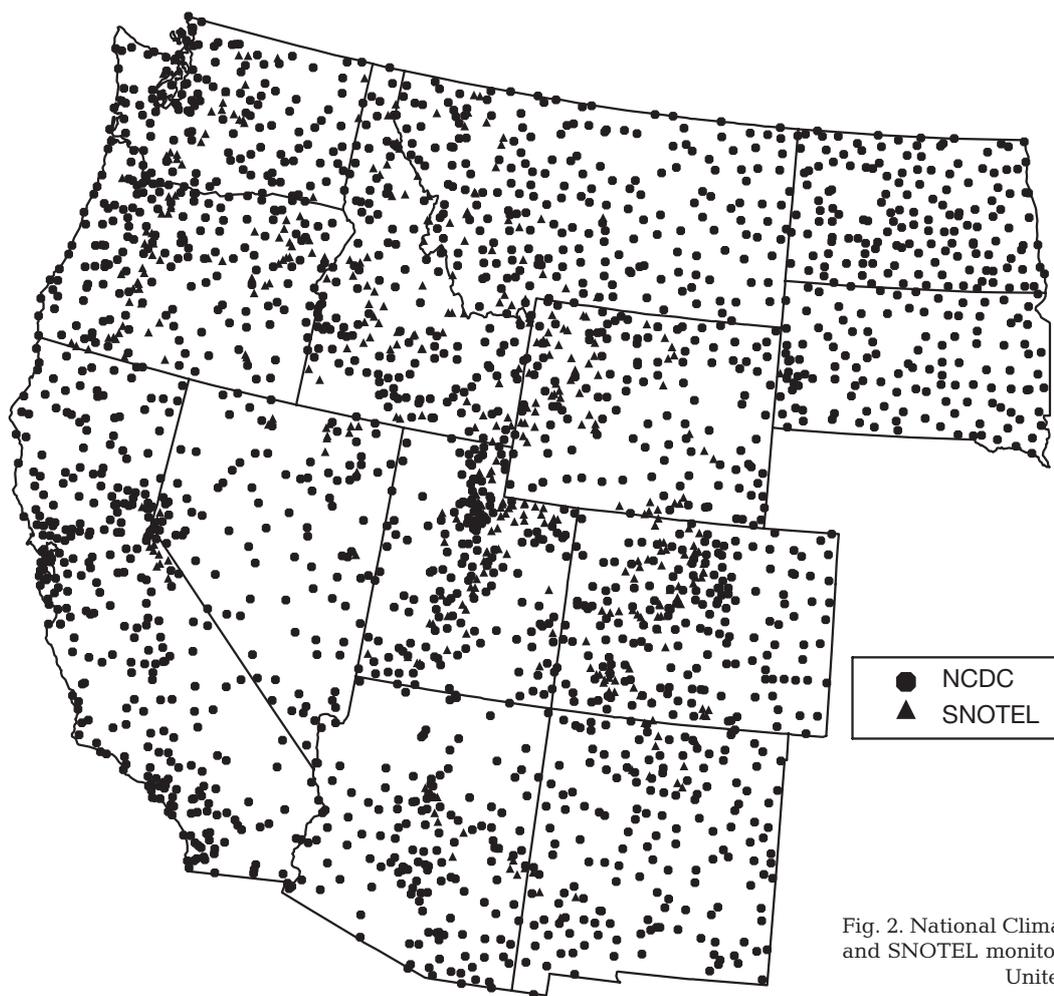


Fig. 2. National Climatic Data Center (NCDC) and SNOTEL monitoring sites in the western United States

(<2000 m) in the western United States but are lacking at high elevations (>3000 m). About 90% of the NCDC stations are located at elevations <2000 m and less than 1% are located at elevations >3000 m. Daily maximum temperature data for 411 NRCS SNOTEL monitoring stations were obtained from the Western Regional Climate Center (WRCC) Web page (<http://www.wrcc.dri.edu/index.html>). The SNOTEL stations are located at higher elevations within the 4 major mountain ranges in the western United States. About 67% of the SNOTEL stations are at elevations >2000 m.

Because  $O_3$  formation and transport are dependent upon radiant energy and atmospheric conditions,  $O_3$  concentrations correlate well with atmospheric variables. Many studies of  $O_3$  concentrations and meteorology in rural and urban areas in the United States have found that ambient air temperature was the most important variable for modeling trends in  $O_3$  concentrations or identifying high ozone days (Wolff & Liroy 1978, Chock et al. 1982, Kuntasal & Chang 1987, National Research Council 1992, Eder et al. 1994, Van Ooy &

Carroll 1995, Bloomfield et al. 1996). The monthly mean daily maximum temperature met our objective of using a highly correlated auxiliary variable for the interpolation of the monthly SUM06 exposure index in non-urban areas in the western United States. This atmospheric variable was chosen because: (1) the correlation between  $O_3$  concentrations and daily maximum temperature was generally the strongest and was stable over time and space; (2) methodology exists for spatial interpolation of temperature over complex terrain based on elevation data; and (3) temperature data were available at sufficient spatial coverage in urban and non-urban areas in the United States.

### 3.4. Regionalization of $O_3$ regimes

For spatial interpolation using loess regression and kriging, it was necessary to divide the western United States into a manageable number of homogeneous regions in which spatial dependencies can be modeled

as a function of relative spatial distance. Initially, RPCA with orthogonal rotation was used to identify pollutant exposure regions based on seasonality and variability of monthly SUM06 values at 111 monitoring stations for the period 1990–1992. In this study, RPCA in conjunction with topography was used to identify spatially cohesive  $O_3$  regions because mountains act as physical barriers to  $O_3$  transport and, to that extent, topography influences the spatial dependencies between monitoring stations.

We applied S-mode PCA (multiple stations over time) using a correlation matrix to monthly SUM06 values. Use of the correlation matrix, as opposed to the covariance matrix, allows monitoring stations with high and low  $O_3$  concentrations to be directly compared for seasonal patterns. Stations with the same seasonal timing of  $O_3$  concentrations will be grouped together, presumably because of the same atmospheric conditions. The scree test and eigenvalues  $>1$  were used to determine the number of PCs to retain (Cattell 1966). In our case, we retained 4 PCs (i.e. regions) explaining 88% of the total variance in the data. To enhance interpretation, we considered both oblique and orthogonal rotations of the PCs and found better results using orthogonal rotations. We chose the varimax orthogonal rotation and assigned each station to the component having the highest score.

Fig. 3 displays the 4 contiguous regions identified by the RPCA with modifications so that the boundaries occurred along natural formations (i.e., the Cascade

Range in the north and the Sierra Nevada Range in central California). When drawing the regionalization boundaries, it was important to bisect the sampled stations along the periphery so that there were stations on either side to avoid edge effects in spatial interpolation. Sampled stations on the periphery were assigned to 1 or both regions depending upon the number and proximity of nearby stations and local topographical features. RPCA and natural geologic formations identified 4 distinct regions which in order are: (1) Central California region, which includes the San Joaquin Valley and southern part of the Mountain Counties; (2) South California region, which includes the South Central and South Coast areas; (3) NW Coast region, which lies to the east of the Cascade Range in Oregon and Washington and extends south into central California; and (4) the Intermountain West region, which is bounded on the west by the Cascade Range in the north and the Sierra Nevada Range in the south and includes the 10 inland states. Local topographical features were given greater consideration in defining the extent of the Intermountain West region, because sites in the northern states were sparse and had uniformly low monthly SUM06 values with no discernable seasonal pattern. California was stratified into 4 regions which are physically reasonable and similar to the California Air Resources Board's (CARB) geographic division of their state into air basins for managing state air resources.

### 3.5. Spatial interpolation of temperature

For purposes of spatially interpolating monthly mean daily maximum temperature data in mountainous terrain, an elevation function/residual interpolation method was used to account for the orographic effects on temperature. Initially, loess regression was used to explore the dependence of temperature on elevation and geographic location for each region. Because sites near a large water body may experience climatic conditions that are significantly different than those just a short distance inland, a coastal proximity variable, defined as the shortest distance to water, was included in the loess regression to better predict temperature for the S California and NW Coast regions. When the sampled points displayed small-scale spatial dependencies, ordinary kriging of the loess residuals was used to model the variogram. The loess and kriged predictions were summed to produce a surface of predicted temperature at the spatial resolution of the DEM surface. The loess regression circumvents the specification of a constant linear lapse rate based on hydrostatic and sea-level equivalent temperature equations which is likely to fluctuate at the regional scale.



Fig. 3. RPCA regionalization based on monthly SUM06 values for 1990–1992. The 4 regions were delineated based on the maximum principal component score for the varimax rotation

Loess regression and kriging were used to model the large-scale and small-scale variabilities, respectively, of the variable of interest. Both methods interpolated values for points or areas which have not been sampled, using data from surrounding sampled points. The 2 methods differed in their assumptions and in their assignment of weights to the neighboring points to be used in interpolating at a non-sampled point. In both methods, the weights assigned to the neighboring points decreased as the relative spatial distance from the point of interest increased. Loess regression assumes that the surface is locally well approximated by a low-order polynomial and the smoothness of the loess fit depends on specification of the neighborhood parameter,  $\alpha > 0$ . A locally quadratic loess model was adequate for describing the relationship between temperature and elevation and the Albers projection coordinates for each region. Criteria for selecting a loess fit included maximizing  $r^2$ , minimizing the residual standard error, and parsimony.

Kriging assumes that the spatial data represent a realization of random variables at particular points whose covariance structure (i.e. variogram) can be modeled as a function of relative spatial location. Points close together are more highly correlated than points further apart. The spatial dependencies of the data can be quantified by specifying a theoretical model for the variogram for purposes of spatial interpolation. For each region, the weighted nonlinear least-squares approach was used to fit a theoretical variogram model of the 'spherical' type to the empirical one (Cressie 1985). Kriging assigns weights to the neighboring points to have a minimum variance of prediction based on the variogram model. All calculations were performed in MathSoft Splus V5.1 and its accompanying module S+Spatialstats V1 (MathSoft 1996, 1998).

An important property of loess regression and kriging is the spatial prediction at a nonsampled point is linear in the variable of interest ( $y_j$ ). This linearity results in distribution properties of the predictor that are analogous to classical parametric fitting and facilitates calculation of the variance (or SD) of prediction. Under the assumption of Gaussian error, the prediction SD or prediction interval at nonsampled points may be used to evaluate the precision of the spatial interpolation using either loess and/or kriging. When kriging was used to model small-scale dependencies in the loess residuals, the spatial interpolation for temperature was calculated as the sum of the loess and kriged predictions. The SD for loess prediction is calculated as  $[(\text{SD for loess surface estimation})^2 + (\text{loess residual SE})^2]^{1/2}$ . The SD for the sum of the loess and kriged predictions is calculated as  $[(\text{SD for loess surface estimation})^2 + (\text{SD for kriged prediction})^2]^{1/2}$ .

### 3.6. Spatial interpolation of $O_3$

A similar loess/kriging method was used to spatially interpolate the monthly SUM06 values at the sampled sites to non-monitored sites for each region. Initially, June 1990 mean daily maximum temperatures were spatially interpolated at the AIRS and CASTNet stations using the procedure described above. While hourly ambient temperature data were available for the CASTNet stations, the monthly mean daily maximum temperature values were consistently lower than those at nearby meteorological stations and so were not used in the analysis. Loess regression was used to relate June 1990 SUM06 with the predicted June 1990 mean daily maximum temperature, elevation, and the Albers projection coordinates. Our analysis indicated that the coastal proximity variable was not required for predicting the monthly SUM06 index. When spatial dependencies existed in the loess residuals, ordinary kriging was used to spatially interpolate the SUM06 residuals to a grid of high spatial resolution. The predicted SUM06 at the grid points was calculated as the sum of the loess prediction and the kriged prediction when appropriate. Negative predicted SUM06 values were set to 0 ppm-h. High predictions for SUM06 were truncated to 10% above the maximum of the sampled SUM06 values for that region.

### 3.7. Cross-validation procedures

The jackknife procedure of model cross-validation was used for both loess and kriging portions of the analyses for monthly temperature and SUM06 values. Sampled points were removed one at a time and predicted from its neighbors using loess regression and/or kriging. Spatial accuracy and precision can be inferred from the actual-minus-predicted values and the prediction standard deviations. When kriging was used for prediction, the nugget, sill and range coefficients were not changed at each iteration because our analysis found that the variogram was not influenced by any single point.

## 4. RESULTS AND DISCUSSION

### 4.1. Temperature prediction

Spatial predictions for the June 1990 mean daily maximum air temperature over the western United States had desirable statistical properties and were logically consistent with local topographical features and orographic effects. The locally quadratic loess fits, and kriging of the loess residuals when appropriate,

Table 1. Locally quadratic loess fits for 1990 June mean daily maximum air temperature as a function of elevation, geographic location and coastal proximity

Stratum	Region	No. of observations	Residual SE (°C)	r <sup>2</sup>	Span $\alpha$	Coefficient of variation (%)
1	Central California	63	1.25	0.97	0.50	4.1
2	S California	81	1.90	0.92	1.40	6.4
3	NW Coast	265	1.52	0.92	0.49	6.7
4	Intermountain West	2031	1.21	0.95	0.35	4.6

produced accurate and precise predictions for monthly mean temperature values over complex terrain. The quadratic loess fit for each region included first- and second-order terms for elevation, the Albers projection coordinates, and coastal proximity (for S California and NW Coast regions). The loess fits explained much of the total variation in temperature ( $r^2 \geq 0.92$ ) and had residual standard errors between 1.21 and 1.90°C for the 4 regions (Table 1). Because the range for elevation at the sampled sites was least for the S. California region, a larger neighborhood size (i.e.  $\alpha = 1.4$ ) was used to predict monthly temperatures at unsampled sites in the S California region, resulting in a smoother surface. While the inclusion of a proximity-to-water variable improved the loess fits for the coastal regions, the predictions for monthly temperatures were less precise for these regions, in particular the S California region. The poorer fit for S California was attributed to complex terrain and different climatic controls affecting this region, which resulted in large gradients in monthly temperatures over short distances (Comrie & Glenn 1998).

Examination of the residuals from the loess regressions indicated the presence of small-scale spatial dependencies for all regions, except the Central California region, and a tendency for loess predictions to understate the extreme temperature values for all 4 regions. The directional sample variograms for the residuals from the loess fits did not show any signs of anisotropy. Kriging was used to fit separate spherical variogram models for the 3 regions displaying an omnidirectional spatial correlation structure to improve the loess predictions for temperature (Table 2). For the Intermountain West region, kriged predictions for elevation-adjusted temperature assigned nonzero weights to the support points within a 299 km radius from the nonsampled point of interest. For the S and Central California regions, where elevation and temperature changed rapidly over short distances, kriged predictions were heavily influenced by neighboring sampled points within a 48 to 72 km radius.

The spatial pattern for the predicted June 1990 mean daily maximum temperatures at 2 km resolution (4 km for Intermountain West region) on the DEM surface had the expected elevational gradients and was consistent with local topographical features (Fig. 4). The transition in the temperature surface along the boundaries of the 4 regions was smooth due to the high spatial density of the meteorological monitoring stations. Mean daily maximum temperatures displayed a north-south gradient as well as an east-west gradient along the coast due to the cooling effect of the dominant Pacific anticyclone in the summer. The lowest temperature values were predicted at higher elevations in the 4 major mountain ranges. In the intermountain areas in the northern states, temperatures were generally between 21 and 27°C. In the western Great Basin, the spatial pattern for temperature was consistent with a mid-latitude desert terrain interspersed with many north-south oriented fault-block mountains. Inland valleys including the Willamette Valley in Oregon and the Central Valley in California were clearly delineated as having warmer temperatures than the Coast Range to the west and the Cascade Range to the east.

The elevation-based spatial interpolation approach produced precise predictions for the monthly mean daily maximum temperature values with a prediction SD of 1.5°C or less for over 90% of the western United States (Fig. 5). Precision was greatest in the Intermountain West region, where temperature values were predicted to within 1.25°C at all DEM grid points, except along its eastern boundary in Central Califor-

Table 2. Kriging models for loess residuals for 1990 June monthly mean daily maximum air temperature as a function of spatial distance. (Kriging was not required for the Central California region because there were no small-scale spatial dependencies in the loess residuals for temperature)

Stratum	Region	No. of observations	Range (km)	Sill (°C <sup>2</sup> )	Nugget (°C <sup>2</sup> )
1	Central California	63	NA	NA	NA
2	S California	81	72	1.26	1.93
3	NW Coast	265	48	1.13	0.50
4	Intermountain West	2031	299	0.30	1.06

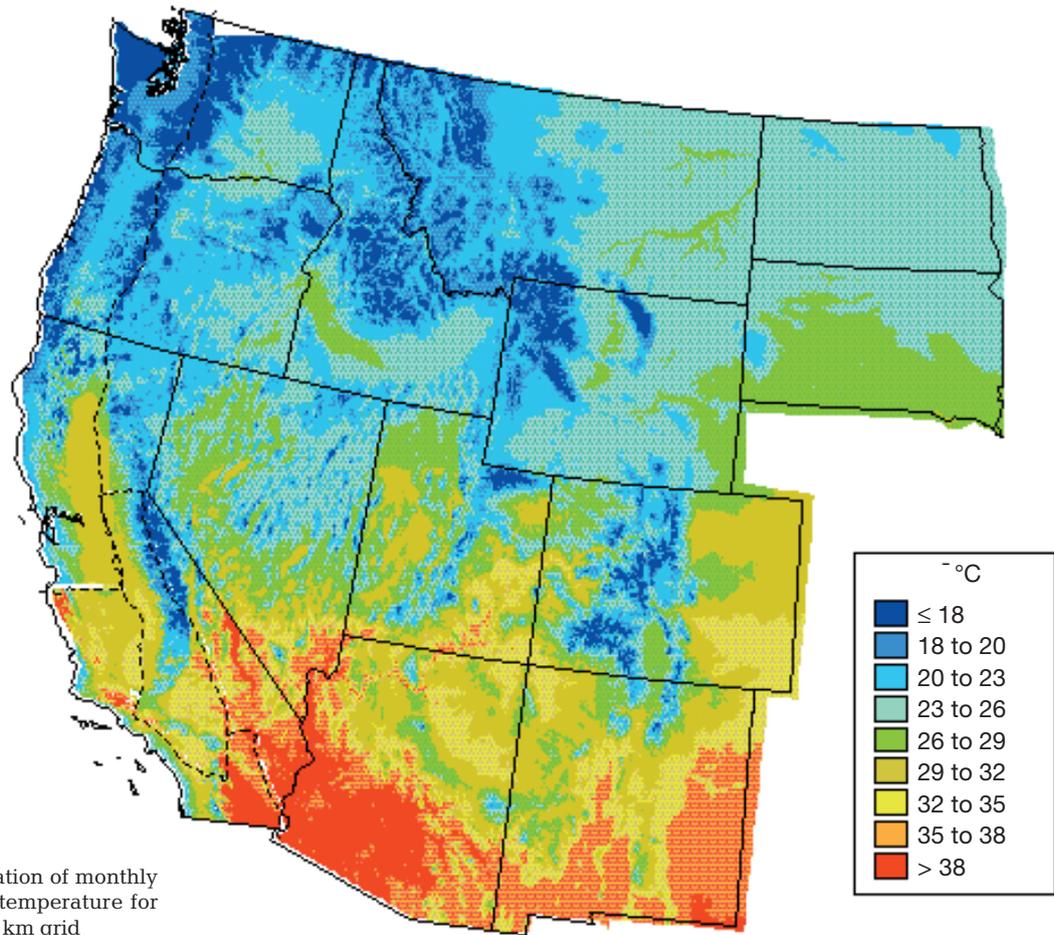


Fig. 4. Spatial interpolation of monthly mean daily maximum temperature for June 1990 on 2 km grid

nia. In the NW Coast and Central California regions, temperature values were less precisely predicted with prediction SDs ranging from 0.5 to 6.2°C. The spatial precision in temperature predictions was least in the S California region—SD ranged from 0.9 to 16.6°C—due to several factors. The northern and southeastern parts of the S California region experienced a mixture of different climate controls, which resulted in large temperature changes over short distances. Also, there were gaps in the spatial coverage in areas where spatial variability was high (Fig. 2). Although most meteorological monitoring was at an elevation less than 3000 m, the spatial coverage was adequate for producing a high-resolution surface for monthly temperature values with reasonable accuracy and precision.

#### 4.2. Cross-validation of elevation-based predictions for monthly temperature

The leave-one-out cross-validation results indicated that the elevation-based predictions for the June 1990 mean daily maximum temperatures at the 2438 meteo-

rological monitoring stations were accurate and precise in the 4 regions, but were the least precise in the S California region (Table 3). When the sampled point was excluded, the loess predictions were nearly unbiased (i.e. mean error ranged between -0.16 and 0.18°C) for each region and the level of precision for the loess predictions as measured by the mean absolute error (MAE) ranged from 0.93°C for the Intermountain West region to 1.75°C for the S California region. Kriging in conjunction with loess regression improved the level of accuracy and precision in the spatial predictions for temperature for the NW Coast and Intermountain West regions, but less so for the S California region. Our spatial predictions for temperature for the western United States were comparable in accuracy and precision to published elevation-based predictions for temperature at smaller spatial scales and less complex terrain (Dodson & Marks 1997, Kurtzman & Kadmon 1999). The loess/kriged prediction SDs for the S California and NW Coast regions may have understated the actual precision by a small amount because 92% of the prediction errors were less than 2 prediction SD in absolute value.

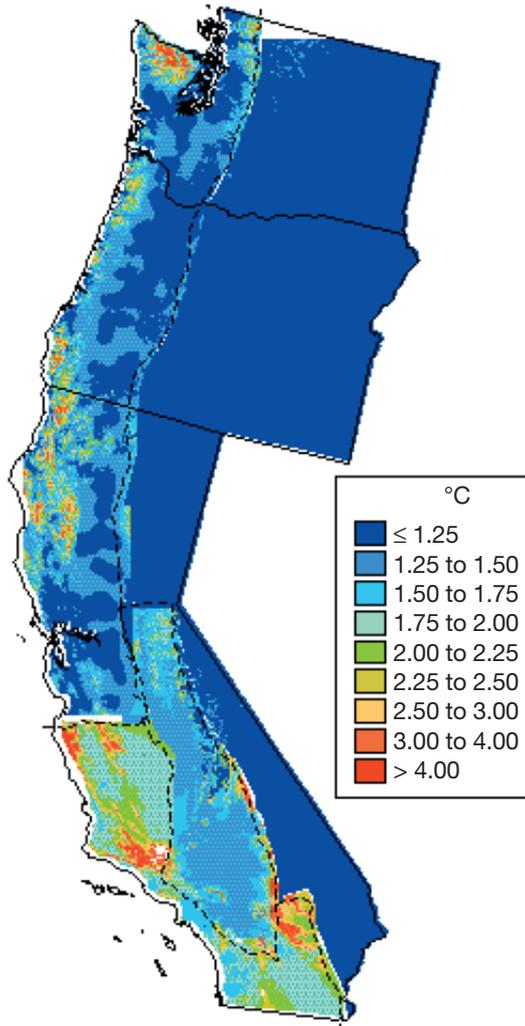


Fig. 5. Prediction standard deviations for mean daily maximum temperature for June 1990 interpolated on a 2 km grid. Standard deviations for Intermountain West region were uniformly less than 1.25°C

No meaningful relationships were found between the prediction errors, geographic location and elevation, indicating that the loess/kriging models were appropriate. There was a tendency to predict warmer temperatures than expected at high-elevation sites in the Intermountain West region and in the Coast Range in Oregon and California, where temperatures were less than 23°C. Predicted temperatures were more than 4°C warmer than expected for 8 sampled sites located within 50 km of the coast in Oregon and California (denoted B, C, E–H, J–L in Fig. 6). Warmer temperatures than expected were also predicted for a number of higher elevation sites in the Intermountain West regions (M, N, P had elevation between 1213 and 3048 m), because these sites had few nearby neighbors which were generally at lower elevation. The largest

Table 3. Summary of cross-validation temperature predictions for the NCDC and SNOTEL stations within the 13 western states based on loess, alone and in conjunction with kriging

Stratum	Minimum (°C)		Maximum (°C)		Mean error (°C)		Mean error/loess+krige (°C)		Jackknife SE (°C)		Mean pred. SD (°C)		% of errors within 95% pred. interval		Correlation obs. vs predicted	
	observed	loess+krige	observed	loess+krige	loess only	loess+krige	loess only	loess+krige	loess only	loess+krige	loess only <sup>a</sup>	loess+krige <sup>b</sup>	loess only	loess+krige	loess only	loess+krige
Central California <sup>c</sup>	17.4, 19.3, na	na	40.9, 39.1, na	na	-0.16, na	na	1.15, na	na	1.39, na	1.70, na	na	100, na	0.96, na			
S California	19.0, 21.1, 20.5	30.1, 30.1	42.8, 41.8, 41.3	30.1, 30.1	0.04, 0.01	29.9, 30.1, 30.1	1.75, 1.71	2.12, 2.14	2.12, 2.14	2.12, 1.92	2.12, 1.92	94.8, 92.2	0.93, 0.93			
NW Coast	11.8, 10.0, 10.5	22.8, 22.8	33.5, 32.8, 32.8	22.8, 22.8	0.18, 0.02	22.8, 22.8	1.27, 1.14	1.76, 1.59	1.76, 1.59	1.73, 1.31	1.73, 1.31	95.4, 91.5	0.93, 0.94			
Intermountain West	11.8, 15.0, 15.2	26.5, 26.5	43.7, 42.3, 42.3	26.5, 26.5	-0.00, 0.01	26.5, 26.5	0.93, 0.86	1.20, 1.12	1.20, 1.12	1.22, 1.09	1.22, 1.09	94.7, 94.4	0.98, 0.98			
Overall	11.8, 10.0, 10.5	26.3, 26.3	43.7, 42.3, 42.3	26.3, 26.3	0.02, 0.00	26.3, 26.3	1.00, 0.93	1.31, 1.23	1.31, 1.23	1.31, 1.14	1.31, 1.14	94.9, 94.1	0.97, 0.98			

<sup>a</sup>The SD for loess prediction is calculated as [(SD for loess surface estimation)<sup>2</sup> + (loess residual SE)<sup>2</sup>]<sup>1/2</sup>

<sup>b</sup>The SD for loess and kriged prediction is calculated as [(SD for loess surface estimation)<sup>2</sup> + (SD for kriged prediction)<sup>2</sup>]<sup>1/2</sup>

<sup>c</sup>na = not applicable; kriging was not used due to lack of local spatial pattern in the loess residuals

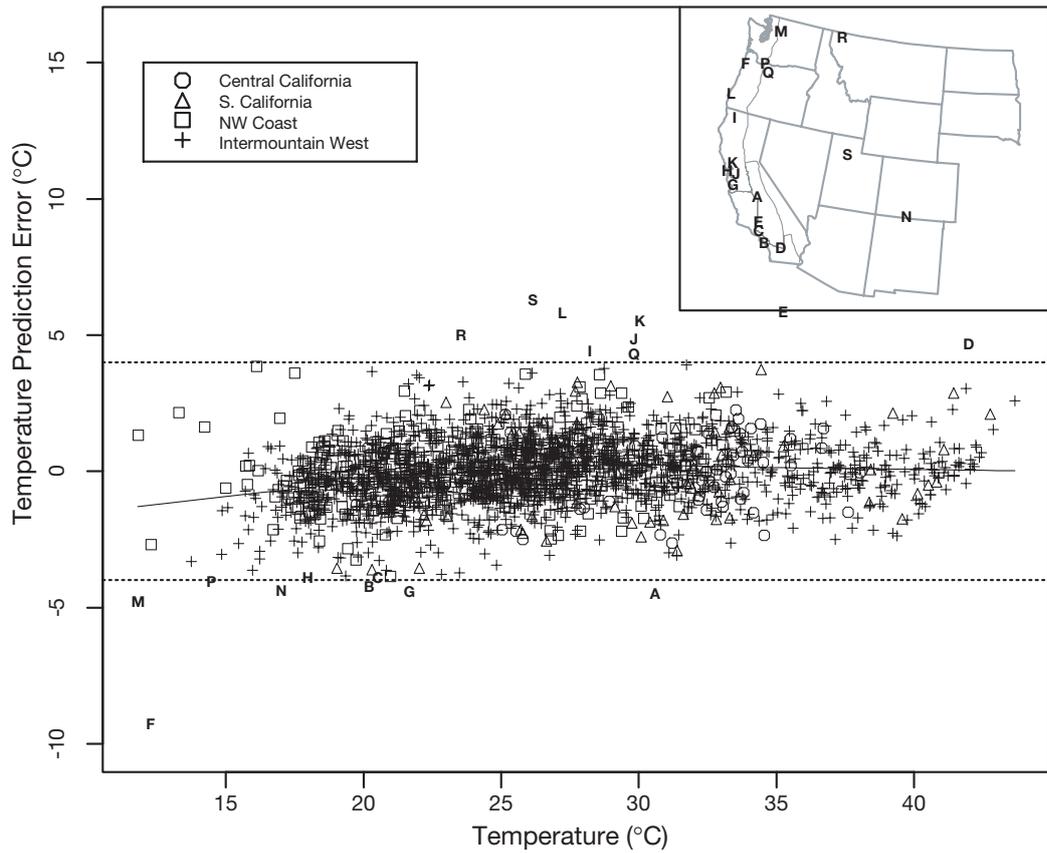


Fig. 6. Observed temperature vs prediction errors for June 1990 at 2438 monitoring stations. Prediction errors >4°C in absolute value are denoted by letters A–E in S California region, F–L in NW Coast region, and M–N and P–S in the Intermountain West region

overprediction occurred at a SOTD site (F, 1094 m elevation) in the Coastal Range in northern Oregon, which had much lower temperature than its nearest neighbor, at 128 m elevation. Predicted temperatures were at least 4°C cooler than expected for 2 coastal sites in the San Francisco Bay area in California (J, K), 1 near-coastal site in southern Oregon (L), and 6 inland sites in California (D, E, I), Oregon (Q), northern Montana (R) and Utah (S), which were at higher (or lower) elevations than their nearest neighbors.

### 4.3. SUM06 prediction

Despite the poor spatial coverage in pollution monitoring stations in the western United States, the locally quadratic loess fits were able to resolve O<sub>3</sub> features in complex terrain. The optimum loess fits had r<sup>2</sup> ranging from 0.57 in the Intermountain West region to about 0.9 in the other 3 regions (Table 4). Based on the residual standard error (RSE), the loess fit was best in the NW Coast region, where SUM06 values and their spatial variability were

low. The RSE was highest in the Intermountain West region, where local variability in SUM06 values was high, the terrain was complex, and spatial coverage was poor. Spatial coverage was lacking particularly in the northern latitudes and at higher elevations, which made spatial interpolation in these areas difficult (Fig. 1). While the SUM06 values were considerably higher in the Central and S California regions, the loess fits had RSEs comparable to that for the Intermountain West region. The loess fits for the SUM06 index required a larger span parameter than those for temperature, because there were fewer O<sub>3</sub> monitoring stations.

Table 4. Locally quadratic loess fits for 1990 June SUM06 as a function of elevation, 1990 June mean daily maximum air temperature and the Albers projection coordinates

Stratum	Region	No. of observations	Residual SE (ppm-h)	r <sup>2</sup>	Span $\alpha$	Coefficient of variation (%)
1	Central California	46	2.31	0.91	2.0	12.0
2	S California	58	2.31	0.93	1.5	22.4
3	NW Coast	51	1.52	0.87	2.2	60.8
4	Intermountain West	70	2.57	0.57	2.5	55.4

Examination of the loess SUM06 residuals indicated no signs of spatial dependency, but there was a tendency to understate the extreme SUM06 values. The directional sample variograms for the loess residuals did not show any signs of anisotropy. The omnidirectional variogram indicated that the loess residuals did not display spatial autocorrelation, so kriging was not required for any region. The exclusion of superfluous urban O<sub>3</sub> monitoring stations may have reduced the small-scale variability and spatial dependencies in the data. In the Intermountain West and NW Coast regions, the spatial coverage for the AIRS network was poor, and distances between O<sub>3</sub> monitoring stations were typically greater than 500 km, which was beyond the range for the variogram model for SUM06 reported by Phillips et al. (1997).

The spatial pattern for the predicted June 1990 SUM06 index at the 2 km resolution (4 km for Intermountain West region) on the DEM surface had the expected elevational gradients and was consistent with local topographical features, temperature and other atmospheric properties governing O<sub>3</sub> transport and formation (Fig. 7). The lower spatial coverage in pollutant monitoring resulted in a smoother SUM06

surface than the temperature surface, a greater total area for which SUM06 predictions were not possible, and more discontinuities along the boundaries between contiguous regions.

The discontinuities in the SUM06 surface at the region boundaries formed along natural formations (i.e., the Cascade Range in the north and the Sierra Nevada Range in Central California) were due to a combination of poor spatial coverage along the periphery and at higher elevations and mountains acting as barriers to pollutant transport (Fig. 7). While the SUM06 surface was less well defined at the boundaries, the SUM06 gradients across regions were logically consistent with local topographical features and atmospheric conditions. Ozone plumes cannot be transported over mountains extending above the boundary layer, which typically is between 1 and 2 km during the day and may shrink to less than 100 m at night (Oke 1990). Elevation effects on O<sub>3</sub> transport and formation were prominent in the SUM06 surface as evidenced by local minima found in the highest elevations within the 4 major mountain ranges.

The spatial pattern for the June 1990 SUM06 index was particularly complex in the Central and S California regions, where SUM06 values changed rapidly over short distances between locations at differing elevations. The highest SUM06 levels in the western

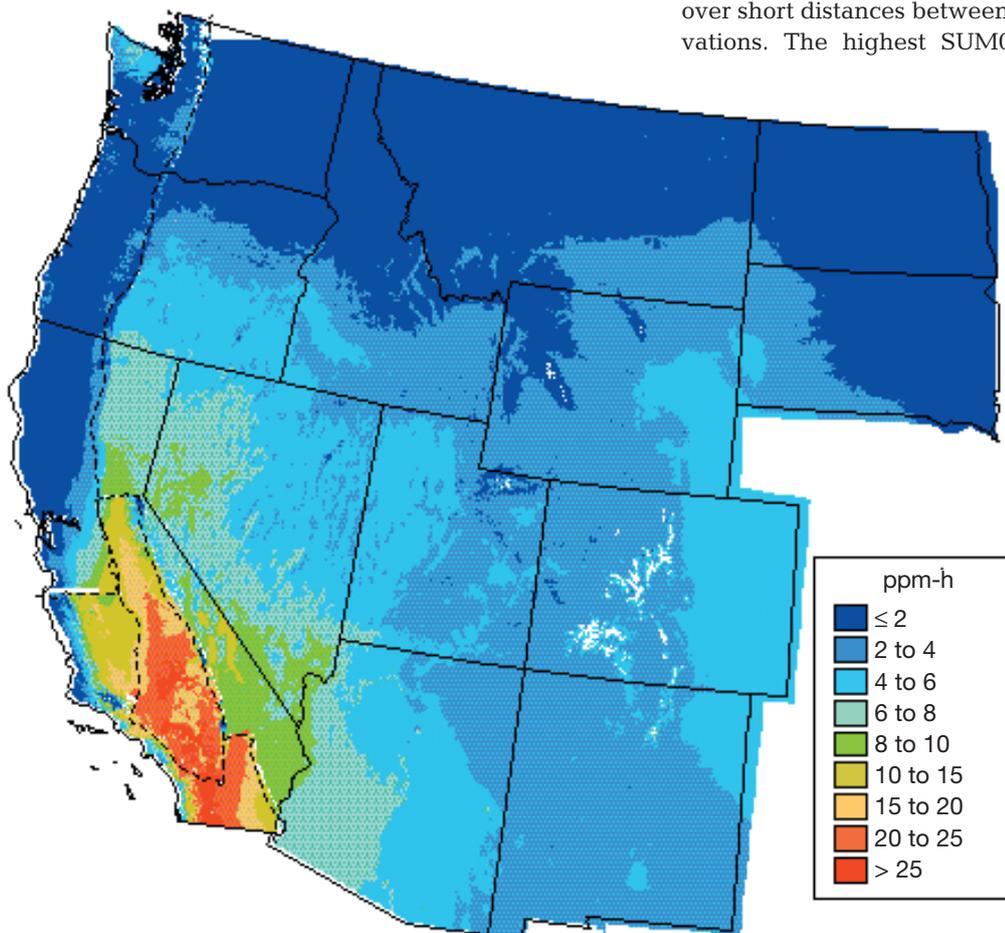


Fig. 7. Spatial interpolation of the 12 h SUM06 index for June 1990 on a 2 km grid

United States were predicted in the inland portions of Kern, San Bernardino, Riverside, and San Diego Counties of California. The lowest SUM06 levels in the S California and NW Coast regions were predicted along the coast, where temperatures were lower and atmospheric conditions were not conducive to O<sub>3</sub> formation. An east-west gradient with increasing SUM06 values inland was attributed to the westerly winds from the dominant North Pacific anticyclone and associated west coast ridge. In California, these patterns are consistent with observational support that urban O<sub>3</sub> concentrations were trapped within the coast subsidence inversion and transported to the heated inland mountain slopes (Edinger 1973).

Prediction SDs for monthly SUM06 values at grid points on the DEM surface were less than 3.5 ppm-h for over 90% of the western United States (Fig. 8). Precision was greatest in the Puget-Willamette Lowland in Washington and Oregon and the northern part of the Central Valley of California, where SUM06 values were low, terrain was uniform, and spatial coverage was high (Fig. 1). In this area, SUM06 values were predicted to within 2 ppm-h. The prediction SDs west of the Cascade Range in Washington and Oregon ranged

between 2 and 5 ppm-h due to lower spatial resolution in the pollutant monitoring network. Precision was least in the Klamath-Siskiyou Mountains in Oregon and northern California and another section of the Coast Range in central California (SD > 7 ppm-h), where the terrain was complex, spatial coverage was poor, and, in the southern latitudes, local variability in SUM06 values was high. The SUM06 predictions were less precise around the periphery of the 4 regions, at higher elevations and remote locations where observations were sparse and not representative of the physical geography. Discontinuities in the surface for the prediction SDs were pronounced and reflected the differences in spatial coverage, distribution, and variability among the 4 regions along their boundaries formed by the Cascade and Sierra Nevada Ranges.

#### 4.4. Cross-validation of loess predictions for monthly SUM06

The cross-validation results indicated that accurate and precise predictions for monthly SUM06 values were possible in complex terrain but were dependent upon local topography and the number and spatial distribution of the sampled points (Table 5). Our spatial interpolation procedure produced more precise and

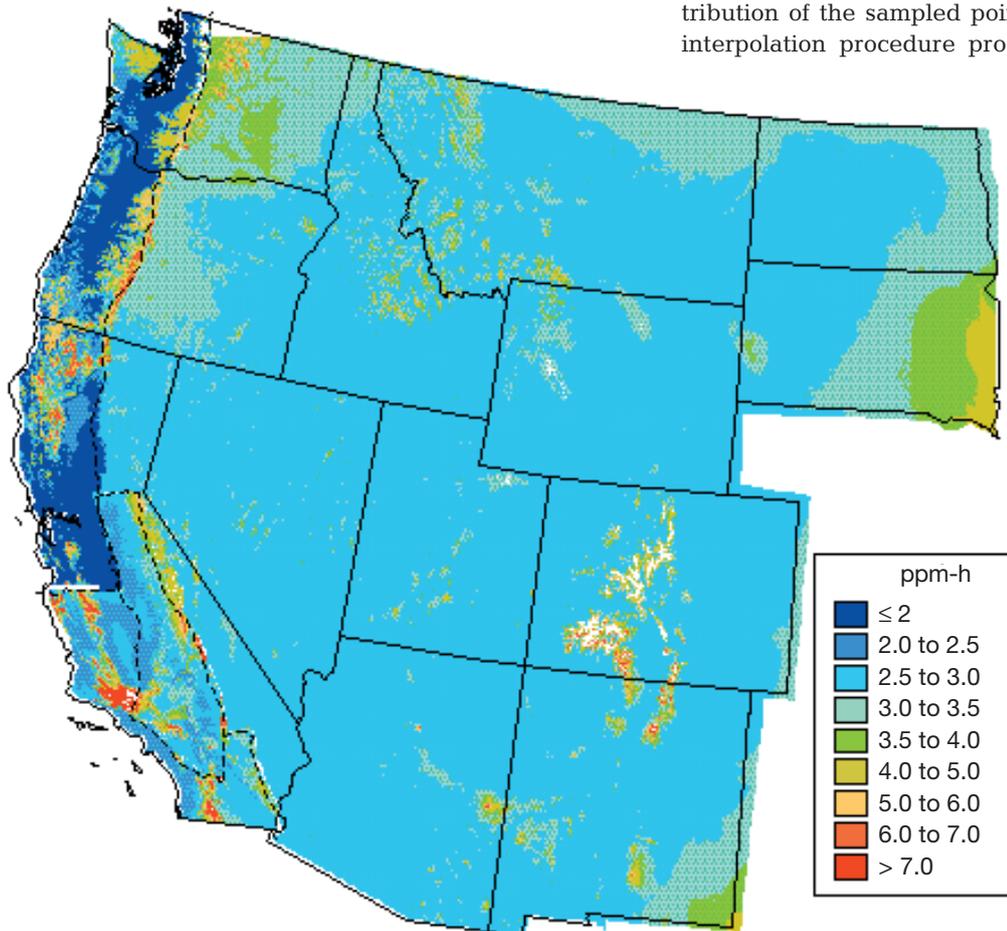


Fig. 8. Standard deviations of prediction for the 12 h SUM06 index for June 1990 on a 2 km grid

slightly less accurate predictions for monthly SUM06 values across the western United States than previous methods using kriging, inverse distance weighting and cokriging with correlated auxiliary variables for the southeastern United States (Phillips et al. 1997). For example, our overall mean error was  $-0.11$  ppm-h, versus  $0.19$  ppm-h for July 1988 SUM06 predictions using cokriging for the southeastern United States, and our overall MAE was  $1.93$  ppm-h, versus  $2.04$  ppm-h for the cokriged predictions. The loess predictions for June 1990 SUM06 were fairly accurate for each region, although there was a tendency to underpredict in all regions except the NW Coast region, more so for the Intermountain West and Central California regions. The level of precision for the loess predictions for SUM06 based on auxiliary variables for elevation and temperature was highest in the NW Coast region, where air quality was good to moderate (Table 5). Less-accurate and less-precise predictions were found in areas of relatively low monitoring density and poor air quality or in areas where elevation changes rapidly over short distances. The prediction errors were generally less than  $5$  ppm-h in absolute value, and 95% of the errors fell within the 95% prediction interval. Overall, the correlation between observed and predicted SUM06 values was  $0.95$  for the western United States.

No meaningful relationships were found between the prediction errors, geographic location, elevation, and land use, indicating that the loess models were appropriate. There was a tendency to predict worse air quality than expected at the most pristine sites and, for the Intermountain West and NW Coast regions, to predict better air quality than expected at sites with moderate air quality (Fig. 9). When the sampled point was excluded from the analysis, prediction errors  $>5$  ppm-h in absolute value were found at 12 sites, mostly in California (denoted by A–I and L in Fig. 9). Large prediction errors and/or large prediction SDs were due to poor spatial coverage, large spatial variability, and/or large elevational gradients over short distances. This was most evident for an isolated high-elevation site (H, elevation =  $1788$  m) in the northwest corner of the Sierra Nevada Range in California, whose nearest neighbors were  $184$  km distant and at a much lower elevation ( $<400$  m) on the windward side (Fig. 9). Accuracy and precision were low at 3 sites (C, D, E) located within a  $30$  km radius area on the western boundary of the Mojave Desert in southern California, where elevation changed rapidly over short distances.

## 5. CONCLUSION

The availability of high-resolution digital elevation data has promoted the development of surfaces with

Table 5. Summary of cross-validation SUM06 predictions for the AIRS and CASTNet stations within the 12 western states (plus North Dakota) based on loess regression. Kriging was not required as the loess residuals did not display any spatial dependencies for any region. Results for Intermountain West region reported including and excluding one site: AIRS ID = 60893003, Manzanita Lake, Shasta County, CA, which was located along the western periphery, had no nearby neighbors, and had the minimum monthly mean daily maximum temperature among ozone monitoring sites

Stratum	No. of missing jackknife pred.	Minimum observed, loess (ppm-h)	Maximum observed, loess (ppm-h)	Mean observed, loess (ppm-h)	Mean error (ppm-h)	Mean error! (ppm-h)	Jackknife prediction SE (ppm-h)	Mean loess pred. SD (ppm-h)	% of errors within 95% prediction interval	Correlation obs. vs predicted
Central California	1	3.3, 4.4	31.8, 29.7	19.2, 19.0	0.14	2.25	2.77	2.93	95.6	0.90
S California	3	0.2, 0.0	25.4, 25.1	10.3, 10.3	0.02	2.10	2.65	2.73	98.2	0.93
NW Coast	1	0.0, 0.0	15.2, 19.2	2.5, 2.6	-0.12	1.11	1.67	1.89	96.0	0.89
Intermountain West	0	0.0, 0.0 0.0	14.4, 15.1 10.5	4.6, 4.4 4.6	-0.10 0.09	2.36 2.21	3.17 2.78	2.95 2.94	94.3 95.6	0.49 0.62
Overall	2	0.0, 0.0	31.8, 29.7	8.0, 8.1	-0.11	1.93	2.61	2.65	95.5	0.95

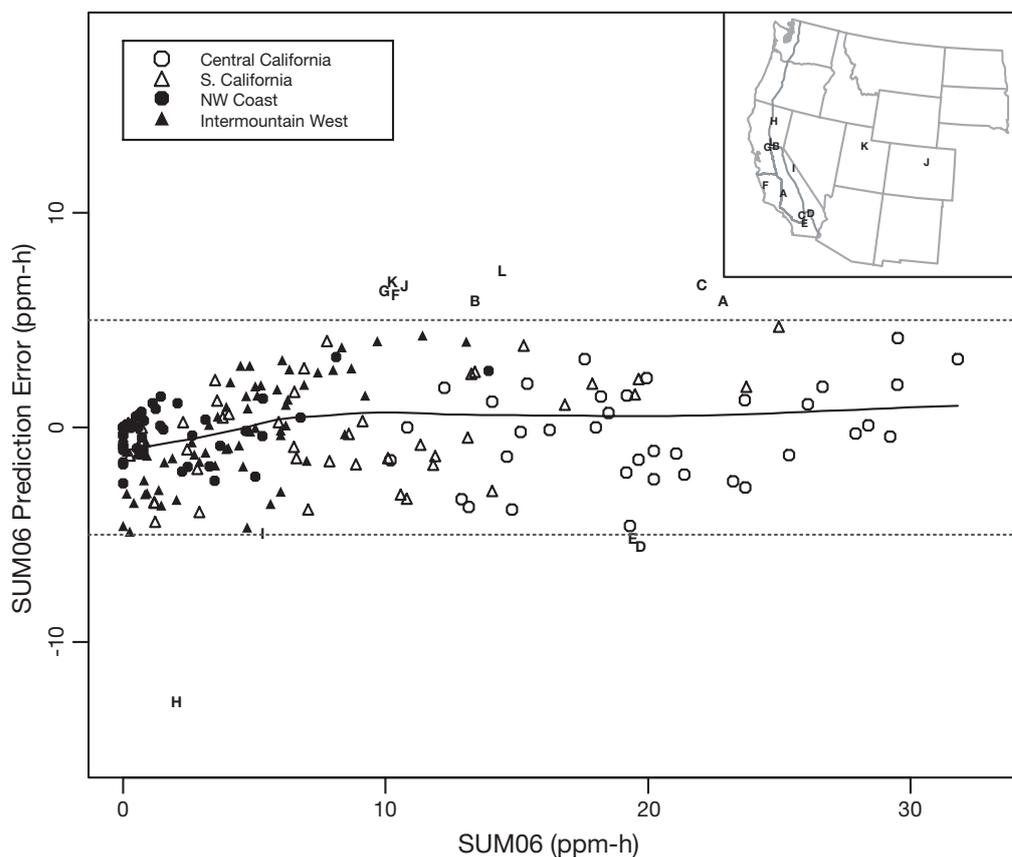


Fig. 9. Observed SUM06 for June 1990 vs cross-validation prediction errors at 204 monitoring stations. Prediction errors  $>5$  ppm-h in absolute value are denoted by letters: A–C in Central California region; D–F in S California region; G in NW Coast region; and H–L in Intermountain West region

high spatial resolution for regionalized variables including temperature, solar radiation, leaf area index, timber volume and ambient air quality. Because temperature data are widely available and have been shown to correlate well with  $O_3$  exposures, it is possible to utilize this additional information to aid in producing a surface for non-urban  $O_3$  exposures over complex terrain at high spatial resolution. The spatial interpolation methods considered in this study were selected specifically for the task of processing very large amounts of data and to predict air quality including SD for non-urban locations. The inclusion of elevation and geographic position substantially improved the ability to predict temperature and  $O_3$  exposure on a fine grid at reasonable cost and overcomes the limitations imposed by the scarcity and poor spatial distribution of non-urban  $O_3$  monitoring stations. This was accomplished primarily through the use of loess regression to model the variable of interest as a non-parametric function of elevation, geographic coordinates, and other auxiliary variables sampled at high spatial resolution.

The elevation function/residual interpolation procedure offers advantages over cokriging and ad hoc procedures such as inverse distance weighting. The loess/kriging approach provides not only a prediction

of ambient air quality with desirable statistical properties but a prediction SD as well. Because data generated in this study were to be used in a plant growth model, the prediction SD was needed in order to examine how uncertainty and variability in inputs propagate through the model. The elevation-based interpolation procedure is applicable to any measure of ambient air quality or temperature defined over any time period, and the surface output generated from this interpolation approach can be easily implemented in a GIS-based risk assessment model at any spatial and temporal resolution. In the western United States, interpolation methods such as kriging have limited application for predicting  $O_3$  exposure because the proximity of the sampled air-quality monitoring data exceeds the range for which small-scale spatial dependencies can be detected and modeled. Further, an interpolation which ignores elevation would seriously under- or overestimate temperature and  $O_3$  concentrations in complex terrain.

While the elevation-based spatial interpolation procedure has desirable statistical properties and good predictive ability in complex terrain, it is important to recognize the limitations of the approach. The loess and kriging models used to spatially interpolate an  $O_3$  exposure index are not rigorous, mechanistic models of

atmospheric chemical and transport processes. Rather, the models were designed to generate climate and pollutant inputs at the temporal and spatial levels required for a growth model simulation and a GIS-based risk assessment of forest ecosystems in the western United States. The loess and kriging models were empirical, and, consequently, the predictions for non-urban O<sub>3</sub> exposures were sensitive to the regionalization of O<sub>3</sub> regimes and the number and spatial arrangement of monitoring stations in the periphery of the regions. The calculation for the SUM06 prediction SDs does not allow for the uncertainty in the predicted temperature values at sampled and nonsampled points for O<sub>3</sub>, the derived weights assigned to the neighboring support points, and the supplemental monthly SUM06 values for urban sites and mean monthly SUM06 values for sites from other years. The first 2 sources of uncertainty are perhaps second-order issues and are not major factors, as evidenced by the fact that 95% of the cross-validation errors fell within the 95% prediction intervals.

We believe that interpolation approaches such as the one we have described can improve assessments by reducing the uncertainties in the temperature and pollutant data surfaces. For GIS-based risk assessment using ozone exposure-response functions for western forest species as seedlings, the uncertainty in the prediction surface for monthly SUM06 values is considerably less than the inherent variability in the growth and biomass responses from field exposure studies conducted in open-top chambers. Except possibly for Central and Southern California, where air quality is poor, the predicted seedling biomass responses for western forest species under ambient conditions would be largely unaffected when the prediction SDs for SUM06 values are included in the calculation. For some western forest species, there were localized areas in its range that would be predicted to have greater losses and risk under ambient conditions with uncertainty.

Our work is also applicable to the construction of optimal spatial sampling designs for monitoring networks to examine the spatial trends of environmental processes at various scales. Our spatial interpolation approach can be used to redesign an existing network, by deleting redundant sites and adding new sites, to meet the needs of environmental scientists and regulators. The elevation-based loess/kriging approach provides the spatial predictions and their SDs and the spatial covariance structure at sampled and non-sampled locations which are required by most methods for redesigning an existing network. A promising new development in pollution research is the use of inexpensive passive ozone samplers in mountainous regions to potentially fill in data gaps in air-quality

monitoring (Brace & Peterson 1998, Peterson et al. 1999). Our interpolation approach provides important inputs to make informative decisions on the placement of these passive samplers in conjunction with the existing ozone monitors to quantify the spatial variability of tropospheric ozone exposure. Future research will consider the construction of optimal network designs using a combination of continuous ozone monitors and passive ozone samplers for quantifying ambient air quality in complex terrain.

*Acknowledgements.* The authors acknowledge contribution of data from the US Environmental Protection Agency's Clean Air Status and Trends Network, the National Resources Conservation Service SNOTEL network, SCOIAS, and Dr Nancy Grulke, US Department of Agriculture Forest Service, Pacific Southwest Research Station. We thank Jeffrey Kern for providing GIS support and the digital elevation model data from the US Geological Survey. We appreciate the technical and scientific comments provided by Dr Donald Phillips, Dr David Peterson, Dr Danny Marks, and Dr John Ray.

*Disclaimer.* The information in this document has been funded wholly by the US Environmental Protection Agency. It has been subjected to the Agency's peer review and administrative review, and it has been approved for publication as an EPA document. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

#### LITERATURE CITED

- Baumgardner RE, Edgerton ES (1998) Rural ozone across the eastern United States: analysis of CASTNet data, 1988–1995. *J Air Waste Manage Assoc* 48:674–688
- Bloomfield P, Royle JA, Steinberg LJ, Yang Q (1996) Accounting for meteorological effects in measuring urban ozone levels and trends. *Atmos Environ* 30:3067–3077
- Brace S, Peterson DL (1998) Spatial patterns of tropospheric ozone in the Mount Ranier region of the Cascade Mountains, U.S.A. *Atmos Environ* 32:3629–3627
- Cattell RB (1966) The scree test for the number of factors. *Multivar Behav Res* 1:245–276
- Chock DP, Kumar S, Herrmann RW (1982) An analysis of trends in oxidant air quality in the south coast air basin of California. *Atmos Environ* 16:2615–2624
- Clarke JF, Edgerton ES, Martin BE (1997) Dry deposition calculations for the clean air status and trends network. *Atmos Environ* 31:3667–3678
- Cleveland WS (1979) Robust locally-weighted regression and smoothing scatterplots. *J Am Stat Assoc* 83:596–610
- Cleveland WS, Grosse E, Shyu WM (1992) Local regression models. In: Chambers JM, Hastie TJ (eds) *Statistical models in S*. Wadsworth & Brooks/Cole Computer Science Series, Pacific Grove, CA, p 309–376
- Comrie AC, Glenn EC (1998) Principal components-based regionalization of precipitation regimes across the southwest United States and northern Mexico, with an application to monsoon precipitation variability. *Clim Res* 10:201–215
- Cressie N (1985) Fitting variogram models by weighted least squares. *Math Geol* 17:563–586
- Dodson R, Marks D (1997) Daily air temperature interpolated at high spatial resolution over a large mountainous region. *Clim Res* 8:1–20

- Earthinfo (1992a) Earthinfo's NCDC summary of the day user's manual. Earthinfo, Inc, Boulder
- Earthinfo (1992b) Earthinfo's NCDC surface airways user's manual. Earthinfo, Inc, Boulder
- Eder BK, Davis JM, Bloomfield P (1994) An automated classification scheme designed to better elucidate the dependence of ozone on meteorology. *J Appl Meteorol* 33: 1182–1199
- Edinger JG (1973) Vertical distribution of photochemical smog in the Los Angeles basin. *Environ Sci Technol* 7: 247–252
- Evans G, Finkelstein P, Martin B, Possiel N, Graves M (1983) Ozone measurements from a network of remote sites. *J Air Pollut Control Assoc* 33:291–296
- Hogsett WE, Weber JE, Tingey DT, Herstrom AA, Lee EH, Laurence JA (1997) An approach for characterizing tropospheric ozone risk to forests. *Environ Manage* 21: 105–120
- Kuntasal G, Chang TY (1987) Trends and relationships of O<sub>3</sub>, NO<sub>x</sub>, and HC in the South Coast Air Basin of California. *J Air Pollut Control Assoc* 37:1158–1163
- Kurtzman D, Kadmon R (1999) Mapping of temperature variables in Israel: a comparison of different interpolation methods. *Clim Res* 13:33–43
- Lefohn AS, Krupa SV, Winstanley D (1990) Surface ozone exposures measured at clean locations around the world. *Environ Pollut* 63:189–224
- Logan JA (1989) Ozone in rural areas of the United States. *J Geophys Res* 94:8511–8532
- Loibl W, Winiwarter W, Baumann R (1994) Estimating the spatial distribution of ozone concentrations in complex terrain. *Atmos Environ* 28:2557–2566
- Loveland TR, Merchant JW, Ohlen DO, Brown JF (1991) Development of a land-cover characteristics database for the conterminous United States. *Photogramm Eng Remote Sens* 57:1453–1463
- MathSoft (1996) S+SPATIALSTATS user's manual, version 1.0. MathSoft, Inc, Seattle
- MathSoft (1998) S-PLUS 5 for UNIX guide to statistics. Data Analysis Products Division, MathSoft, Seattle
- Miller PR (1973) Oxidant-induced community change in a mixed conifer forest. *Adv Chem Ser* 122:101–117
- Miller PR (1984) Ozone effects in the San Bernardino National Forest. In: Davis DD, Millen AA, Dochinger L (eds) *Air pollution and the productivity of the forest: proceedings of the symposium, October 1983, Washington, DC*. Izaak Walton League of America, Arlington, p 161–197
- Miller PR, McBride JR (1999) *Oxidant air pollution impacts in the montane forests of southern California*. Springer-Verlag, New York
- National Research Council (1992) *Rethinking the ozone problem in urban and regional air pollution*. National Academy Press, Washington, DC. <http://www.nap.edu/index.html>
- Oke TR (1990) *Boundary layer climates*. Routledge, London
- Ollinger SV, Aber JD, Federer CA, Lovett GM, Ellis JM (1995) *Modeling physical and chemical climate of the northeastern United States for a Geographic Information System*. General Technical Report NE-191, United States Department of Agriculture, Forest Service, Northeastern Forest Experimental Station, Radnor, PA
- Peterson DL, Bowers D, Brace S (1999) Tropospheric ozone in the Nisqually River drainage, Mount Ranier National Park. *Northwest Sci* 73:241–254
- Phillips DL, Dolph J, Marks D (1992) A comparison of geostatistical procedures for spatial analysis of precipitation in mountainous terrain. *Agric For Meteorol* 58:119–141
- Phillips DL, Lee EH, Herstrom AA, Hogsett WE, Tingey DT (1997) Use of auxiliary data for spatial interpolation of ozone exposure in southeastern forests. *Environmetrics* 8:43–61
- US Environmental Protection Agency (1992) *National air quality and emissions trends report, 1991*. EPA/450-R-92-001. US Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park
- Van Ooy DJ, Carroll JJ (1995) The spatial variation of ozone climatology on the western slope of the Sierra Nevada. *Atmos Environ* 29:1319–1330
- Wolff GT, Liroy PJ (1978) An empirical model for forecasting maximum daily ozone levels in the northeastern U.S. *J Air Pollut Control Assoc* 28:1034–1038

*Editorial responsibility: Brent Yarnal,  
University Park, Pennsylvania, USA*

*Submitted: April 10, 2000; Accepted: January 28, 2001  
Proofs received from author(s): July 13, 2001*