Developments in Environmental Science, volume 2, pp. 215–247
ISSN: 1474-8177

Chapter 9

Use of geostatistics to estimate surface ozone patterns

Witold Frączek

Environmental Systems Research Institute (ESRI), 380 New York street, Redlands, CA 92373-8100, USA
E-mail: wfraczek@esri.com

Andrzej Bytnerowicz, Michael J. Arbaugh

USDA Forest Service, Pacific Southwest Research Station, 4955 Canyon Crest Dr., Riverside, CA 92507-6099, USA
E-mail: abytnerowicz@fs.fed.us (A. Bytnerowicz), marbaugh@fs.fed.us (M.J. Arbaugh)

Abstract

Models of spatial and temporal distributions of ambient ozone (O_3) in the Sierra Nevada in the spring/summer season of 1999 were developed with the Geostatistical Analyst, an extension to ArcMap^{TM} 8.1.2 (ESRI, Redlands, CA). The models were based on a combination of O_3 concentrations data from passive O_3 samplers and active monitors, digital elevation models, and available meteorological data for the study area. Strong spatial variation of O_3 concentration and weaker temporal variation over the study area were found. Ozone concentrations were the highest at the foothills of the southern Sierra Nevada and the lowest at the high altitudes of the northern part of the range. Summer thunderstorms influenced distribution of O_3 concentrations by setting spatial and temporal pockets that reduced concentrations of pollutants. The number of O_3 monitoring points in the network was not sufficient to ensure a high and relatively uniform level of confidence in the O_3 concentration estimates, especially on the eastern slopes of the southern portion of the Sierra Nevada. High O_3 concentrations in portions of the eastern Sierra Nevada indicate the possibility of a long-range transport of highly polluted air plumes from the California Central Valley and/or the Los Angeles area.

1. Introduction

A quickly growing population and the rapidly increasing motorization in California has resulted in the production of photochemical smog and elevated levels of ozone downwind from the pollution source areas. The prevailing westerly winds move the polluted air masses from the California Central Valley

toward the Sierra Nevada. Sensitive trees, such as ponderosa and Jeffrey pines (*Pinus ponderosa* Dougl. ex Laws. and *P. jeffreyi* Grev. & Balf.), have been severely affected by elevated concentrations of ozone (O_3) in large portions of the western Sierra Nevada (Duriscoe and Stolte, 1989; Peterson and Arbaugh, 1992). If the impact of air pollution on the Californian forests can be determined, this information may provide the basis for rational decisions to protect forest ecosystem of the Sierra Nevada.

Models of air pollution distribution based on emission inventories, weather data and information on conditions requisite to form O_3 have been developed for some areas of Europe (Simpson, 1991) and the United States (Phillips et al., 1997). These models are reasonably accurate for the lowland areas that have uniform forest canopies. However, for mountain terrains, the same models of air pollution distribution and deposition are less reliable and more difficult to quantify. Concentration of O_3, a secondary pollutant formed as a result of complex chemical reactions, is difficult to predict in complex areas and over long time periods. Models of O_3 distribution based on emission inventories are often developed for specific air basins with well-established physical borders (Seinfeld and Pandis, 1998). In the orographically complex mountain terrain, such models are more difficult to apply.

An alternative approach to the traditional emission models and information on ambient O_3 concentrations at receptor sites is to investigate patterns of spatial distribution and estimate O_3 at the areas not directly monitored by the network of established samplers. This capability of spatial analysis and modeling is provided by Geographic Information Systems (GIS). The GIS software and methodologies offer solutions to many specialists including foresters, natural resource managers, and air pollution professionals. By using remote sensing and ground truthing, GIS is able to provide the "big picture" for better analysis and decision-making. Examples of GIS applications include developing long-term supply strategies, forecasting silvicultural stock, determining harvesting system options, evaluation of foliar damage caused by insects or air pollution before it is visible from ground observation, help in protecting forests from fires, control of forest fires, estimating the influence of O_3 and other air pollutants on the health of forest stands, and many more.

The objective of this research was to apply GIS modeling methods by using a combination of passive O_3 samplers and active O_3 monitoring station data to produce mapped distributions of ambient O_3 in the Sierra Nevada, from the Lassen to the Sequoia National Forests over the summer season of 1999. Information from the active monitoring stations alone is not sufficient to develop a meaningful spatial map of ambient O_3 for this region. An extensive network of passive monitors augmented the network of active monitoring stations for one summer to identify the key spatial relationships for locations among active monitoring stations.

2. GIS and geostatistics

GIS is an organized collection of computer hardware, software, geographic data, used by qualified personnel to efficiently capture, store, update, manipulate, analyze, and display all forms of geographically referenced information (Zeigler, 1999). One of the purposes of GIS is to provide a geographic base to support decisions for the intelligent use and management of environmental resources. Foresters, botanists, meteorologists, climatologists, and air pollution scientists are increasingly relying on GIS to help them in making critical decisions. By putting spatial data in an integrated system where it can be organized, analyzed, and mapped, patterns and relationships that were previously unrecognized may emerge.

Geostatistics has become a part of GIS. With Geostatistical Analyst, an extension to the ArcMap™ 8.1.2 software produced by the Environmental Systems Research Institute (ESRI™), Redlands, California, a continuous distribution model or a map of any phenomena can be created from measured sample points. Data collection usually can only be conducted at a limited number of measurement stations due to logistical and financial limitations, but scientists and managers are increasingly interested in continuous surface estimates. In order to generate surface information some type of interpolation technique must be developed to estimate data values for those locations where no samples were taken (Webster and Oliver, 2001). Kriging is a weighted moving average method of interpolation, widely recognized as providing the best estimate of the interpolated value at the unsampled locations. Applying kriging allows to create a continuous layer of information from the set of individual sample points.

By further processing and performing GIS analysis, more output information can be derived to serve as the foundation in an educated decision-making process. For instance, knowing which forested areas are most exposed to high concentrations of O_3 and the distribution of the least resistant tree species to O_3, a new information layer of fire susceptibility might be created.

In order to digitally examine and analyze numerous spatial information layers and to allow the output of such analysis to be useful in another study, all layers have to exist in a common space. This is why one of the requirements of GIS is to use data that are precisely referenced to a specific spot on Earth. Because of the spatially oriented nature of this study, all the data had to be georeferenced with spatial attributes such as the geographic coordinates. All the spatial data were transformed to the Transverse Mercator projection for calculations, geostatistical modeling, and analysis.

3. Geography of the study area

The geographic focus of this study was the Sierra Nevada of California (see Fig. 1 of the Preface, this volume). This area consists primarily of public lands that include four national parks (Lassen Volcanic, Yosemite, Kings Canyon, and Sequoia) and eight national forests (Fig. 1). The four national parks (NP) and eight national forests (NF) form a contiguous area (47,860 km^2), which was designated as the study area for this project. Additional areas outside of these parks and forests were also sampled to improve estimation of areas within the parks and forests.

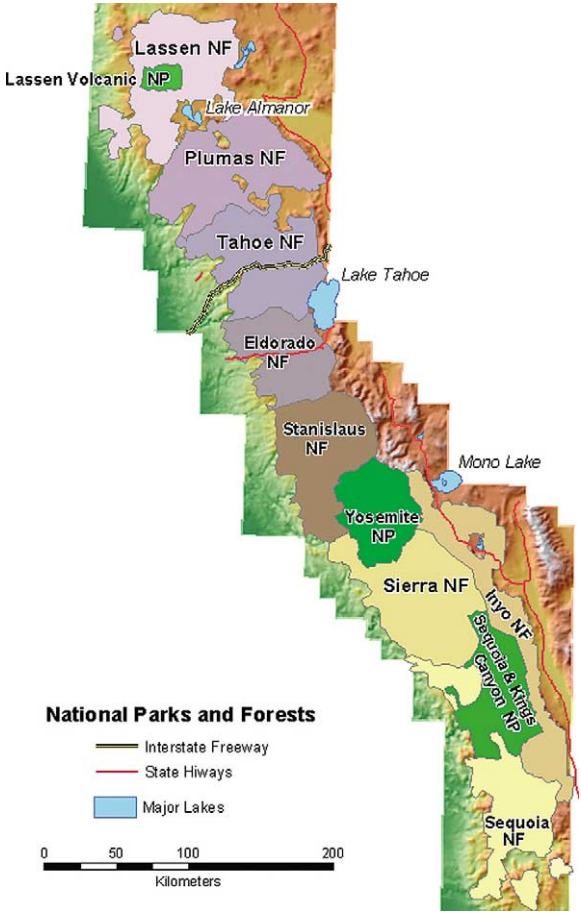


Figure 1. Four national parks and eight national forests constitute the Sierra Nevada Mountains study area.

The predominant type of forest in the Sierra Nevada is the mixed coniferous forest. It covers 78.3% of the combined national parks and national forests territory of the Sierras. Hardwood forests cover 1.5% of the territory; shrub communities and chaparral $\sim$ 10.1%; while barren land, typical for the most elevated areas is present on 8.1% of the study area. Only 0.5% of land is used for residential and agricultural purposes (see Fig. 2 in Preface, this volume).

The Sierra Nevada is the most elevated mountain range in the US (except for Alaska) and its highest summit, Mt. Whitney, reaches 4418 meters a.s.l. The Sierra Nevada range runs roughly from north/northwest to south/southeast. These young mountains (Alpine orogenesis) have in many places remarkable relief with spiky, rocky peaks and deep valleys. Consequently, the diversity of climate, combined with abrupt changes in weather parameters over short distances might result in the significant variability of the concentration of air pollutants.

The Sierra Nevada acts as a giant water trap, capturing the moisture from the Pacific storms as they move east. This is a bonus to California agriculture, since the water from the Sierra flows into the Central Valley. It also creates a vast rain shadow desert that stretches for five hundred miles to the east.

Much of the Sierra Nevada has a Mediterranean climate of warm, dry summers and cool, wet winters. Most precipitation falls between October and April, though the Sierra crest is subject to summer thunderstorms. Because prevailing westerly winds bring moisture from the Pacific Ocean, western slopes of the range receive relatively intense precipitation, characterized by considerable rainfall during summer storms (a single event might bring more than 10% of the annual total precipitation for the area) and plenty of snowfall in winter. The eastern, much steeper slopes are located in the rain shadow area of the most elevated part of the mountains, which are especially dry in the southern part of the range (Inyo National Forest). Meteorologically, the eastern and western sides of the main range are very different, especially in the south where the highest section of the mountains with the most dramatic relief creates an extensive physical barrier.

According to weather data published by National Oceanic and Atmospheric Administration (NOAA, 1999), there is more rain in the northern Sierra than in the southern end of the range. Precipitation ranges from less than 250 mm per year on the Kern River in the southern Sequoia National Forest to more than 2000 mm per year in the mountains surrounding the Feather River in the northern Plumas National Forest (see Fig. 1). Even though the Sierra crest is more than twice as high in the south, maximum precipitation is only half of that at the northern end of the range.

The Sierras are located close to areas of highly polluted air, such as the California Central Valley stretching along the Sierras and the Los Angeles Basin, which is situated directly south (see Fig. 1 of the Preface, this volume). The

local centers of air pollution within the Central Valley are Fresno, Sacramento, and Bakersfield urban-rural areas. Because of the proximity to these areas and the prevailing westerly winds, the Sierra Nevada are continuously exposed to the O_3-polluted air masses.

4. Available data

Monitoring of O_3 was performed on 89 sites throughout the entire range of the Sierra Nevada (see Fig. 2, Arbaugh and Bytnerowicz, Chapter 6, this volume). Passive samplers were changed every 2 weeks between May 12 and October 13, 1999 and were calibrated against active, real-time O_3 instruments in nine collocated Sierra Nevada sites. Based on those calibrations an equation was developed that transformed nitrate formation rates ($\mu g\, NO_3^-/h$) obtained from passive samplers into O_3 concentrations (ppb) (Lee, Chapter 7; Preisler and Schilling, Chapter 8; Arbaugh and Bytnerowicz, Chapter 10, this volume).

Because it is known from other research that O_3 ambient concentration is usually correlated to air temperature (Finlayson-Pitts and Pitts, 2000), frequency and amount of precipitation (Seinfeld and Pandis, 1998), and elevation above sea level (Brace and Peterson, 1998; Smidt and Gabler, 1995), data supplemental to O_3 concentration were acquired for this study. These three collateral sets of spatial data were included in the analysis to enhance quality of the geostatistical modeling of O_3 concentration based on the network of air pollution measurement stations. Thus, the data used to model the distribution of O_3 in this study included:

- Ozone concentration: The monitoring sites (see Fig. 2, Arbaugh and Bytnerowicz, Chapter 6, this volume) ranged in elevation from 223 to 2796 meters above sea level. There were 10 biweekly measurement periods during the season. The number of monitoring sites varied from 60 in early May to 89 in September. Four of the measurement stations located in Yosemite National Park provided weekly, instead of biweekly, measurements of O_3 concentration. The data from these stations were averaged to calculate the biweekly values to be consistent with the data received from all the remaining O_3 measurement stations. The highest measured average value of O_3 during the biweekly measurement unit of time was 142 ppb and occurred in the first half of September at station number 67, located at the western foothills of the Sierra Nevada in the Sierra National Forest, followed by 136 ppb recorded in the first half of June at station 42, also positioned at the foothills in the Sequoia National Forest. The lowest measured biweekly average O_3 concentration was 18 ppb and occurred in mid-May in the upper section of San Joaquin River Valley (Table 1).

Table 1. Summary of ozone concentrations (2-week averages in ppb) during the measurement season of 1999

Period number	Starting day	Minimum value	Maximum value	Mean value	Standard deviation	Number of samples	Number of sites
1	May 12	18.6	102.4	54.1	12.2	60	60
2	May 26	23.7	87.5	49.5	11.0	74	74
3	June 9	25.1	135.6	61.7	18.3	91	87
4	June 23	23.4	97.3	58.8	16.0	91	87
5	July 7	28.5	107.1	61.8	14.7	92	88
6	July 21	27.1	85.8	58.2	12.5	92	88
7	August 4	23.7	77.6	49.8	10.4	92	88
8	August 18	23.7	101.0	58.8	13.9	92	88
9	September 1	31.5	141.7	63.4	17.8	93	89
10	September 16	23.4	83.7	55.8	12.0	93	89

- Maximum temperature: Meteorological data were obtained from the network of 61 weather stations (see Fig. 2, Arbaugh and Bytnerowicz, Chapter 6, this volume), which provided the maximum temperature recorded for each period of time that was critical for this study. The meteorological monitoring stations were spread all over the Sierra Nevada and located across a wide variety of elevations (52 to 2551 meters above sea level). The highest recorded biweekly average maximum temperature was 39 °C (102 °F) and occurred during the first half of July at the foothills of Sequoia National Park.
- Precipitation: Quantitative rainfall information was acquired from 87 weather stations. The precipitation data were available from the same source as the maximum temperature and from additional 26 stations, all synchronized with the time frame of O_3 measurements. The rainfall data could not be directly applied for the geostatistical calculations because most of the recorded values at all time period were zeros, and relatively few had meaningful records. Thus, these data did not fulfill the criteria required by cokriging so that the regionalized variables could be correlated. Still, these data were used for the analysis of the obtained geostatistical results. The stations recording precipitation data were located at elevations from 52 to 2940 meters above sea level (Table 2).

Digital Elevation Model (DEM): The DEM had elevations ranging from 19 to 4418 meters and was used as collateral data to enhance the quality of the geostatistical estimation of O_3, the primary variable. The relevant, 1-arc second (30-meter) resolution elevation data for almost 200 topographic quadrangles was downloaded from the US Geological Survey Web site http://edcwww.cr.usgs.gov/webglis/, resampled to a coarser resolution and merged into a single, continuous map. An effort was made to determine the optimal resolution of the DEM for this study. Consequently, a DEM of 1 km

Table 2. Summary of precipitation (in mm) during the measurement season of 1999

Period number	Starting day	Maximum value at any station	Mean value of all stations	Sum value from all stations
1	May 12	112	7.8	678
2	May 26	170	46.6	4053
3	June 9	1	0.0	1
4	June 23	9	0.1	10
5	July 7	100	10.3	896
6	July 21	10	0.2	22
7	August 4	107	14.9	1299
8	August 18	57	10.7	933
9	September 1	37	2.4	213
10	September 16	176	12.1	1057

resolution (see Fig. 2, Arbaugh and Bytnerowicz, Chapter 6, this volume) was utilized for most of the geostatistical modeling of this study, as it was the best fit with the density of the other datasets.

The two sampling networks, O_3 concentration and maximum temperature, were spatially independent from each other. However, the measurement periods for the meteorological data were summarized to the same time intervals as the passive sampler data. The O_3 concentration distribution models of enhanced accuracy were developed based on both O_3 concentration measurements as the primary and the maximum temperatures as the secondary variables.

5. The geostatistical approach

5.1. Defining the geostatistical problem

The goal of this study was to provide spatial and temporal information on O_3 concentration over the forested areas of the Sierra Nevada. The time frame of the study was focused on the 20 weeks of potentially highest concentrations of O_3 in the study area, from the middle of May to the end of September 1999.

Ozone concentration (the average values during all of the 10 biweekly periods) was monitored in the selected sites, ranging over time in number from 60 to 89 (Table 1). Considering the spread of the study area (over 650 km) and its significant geographic and climatologic diversity, the number of the O_3 monitoring sites was not quite sufficient to estimate the concentration of O_3 with equally high reliability over the entire study area. The total number of the O_3 monitoring sites was not adequate to apply a standard geostatistical technique such as kriging. Moreover, the kriging method could not be successfully utilized because the spatial distribution of the monitoring sites was far from optimal with some significant parts of the study area not being monitored. The use of kriging as well as any of the deterministic methods (e.g., inverse distance

weighting, IDW) would not be able to compensate for the lack of sufficient sampling in some regions (Johnston et al., 2001).

5.2. Introduction to kriging and variograms

Among numerous geostatistical methods, kriging is widely recognized as the one providing the most accurate results and allowing for the most flexibility. Kriging is done in two parts:

1. The sample semivariance is used to estimate the shape of the variogram (a curve that represents the semivariance as a function of distance). The variogram describes the spatial relationship between the data points.
2. The estimated semivariance function is used to determine the weights needed to define the contribution of each sampled point to the interpolation. Sample points close to the location for which an estimated value is to be generated contribute the most to the interpolation (Babish, 2000).

A variogram is a graphical display of a variance of measurements over the distance between the measurement sites. If there are spatial dependencies, the variance between the observations on two points normally increases with increasing distance until a maximum value, called "sill", is reached. From this point, the semivariance no longer increases, causing a flat region to occur on the variograms. A distance from the zero value on the x-axis to the beginning of the sill is called a "range" of the regionalized variable. Within this range, locations are related to each other, and all known samples contained in this region (the neighborhood) must be considered when estimating unknown points (Babish, 2000).

Kriging can provide a measure of an error or uncertainty of the estimated surface. Since the estimation variances can be mapped, the confidence placed in the estimates can be calculated.

When studying two or more regionalized variables, which are correlated with each other as in this case, a technique called cokriging can be used. Cokriging uses one set of data to help explain and improve the description of variability of the main variable. Cokriging can be used to reduce the estimation variance when the variable of interest has been undersampled. The existence of the collateral data was essential to allow for cokriging, as the geostatistical technique can estimate the distribution of one phenomenon based on the distribution of the other one, presuming that there is a strong correlation between the two (Cressie, 1993).

5.3. Applied solution—geostatistical methodology

The Geostatistical Analyst extension to ArcMap^{TM} 8.1.2 of ESRI was applied to generate models of O_3 concentration for the study area. Because of

the limitations to the reliability of the prediction as mentioned above, an attempt was undertaken to estimate O_3 concentration even in the areas where the sampling network was not sufficiently dense and where the topographic obstacles create obvious interpolation problems for the standard geostatistical methods.

The data of maximum biweekly temperatures obtained from 61 weather stations spread all over the entire study area, including the most sparsely sampled areas for O_3 like the Inyo National Forest and the fine DEM of the Sierra Nevada, enabled the prediction of O_3 distribution. These two collateral data sets were used as the secondary and tertiary variables in the applied ordinary cokriging geostatistical method.

There is a significant correlation between O_3 concentration and maximum temperature (Seinfeld and Pandis, 1998). Similarly, it is a well-known fact that the temperature decreases with increased altitude (environmental lapse rate of 6.5 °C for every 1000 meters). A workaround was necessary because there was not enough evidence to establish a strong correlation between O_3 concentration and elevation. This correlation would have to be "customized" for the local climatic and topographic conditions, possibly stratifying the data. Unfortunately, not enough measurements of O_3 and temperature were available for this study at the higher elevations in order to determine the correlation for the upper elevations of the Sierra Nevada. Thus, the final output surfaces for all 10 biweekly periods of O_3 measurements in 1999 as well as the model of O_3 concentration for the average values of O_3 for the entire 20-week sampling season were generated applying a sequence of several steps to ensure the most accurate results possible.

For every time period, the appropriate set of maximum temperature measurements was cokriged with the elevation data to generate a precise spatial model of maximum temperatures. To eliminate the influence of the existing general trend in temperature distribution and to derive the most accurate output surface, the first order detrend surface was removed for both variables. Because kriging is an exact interpolator and the two applied data sets were very different in terms of the number of samples, the output surface of temperatures would demonstrate unnecessarily detailed local variation. To avoid the unreliable local variation and to smooth the output, a nugget effect was applied. The resulting continuous surface of predicted maximum temperatures was converted to a dense (about 3600 meters between neighboring points) regular network of points with the total number of estimated values close to 20,000. The generated much denser network of maximum temperatures was used as the cokriging secondary variable to predict O_3 concentrations throughout the study area. When creating the O_3 surface, detrending was not used because a spatial trend was not detected among the O_3 concentration measure-

ments. As above, to account for the error of a local variation, a small value for the nugget parameter was entered to the model.

The output surfaces of O_3 concentration created using the described procedure had remarkably low standard prediction errors. This is why the above method was selected over direct cokriging of O_3 in which maximum temperature and DEM were used as the secondary and tertiary variable, respectively. Consequently, this was the method applied to create models of O_3 concentration for all ten measurement periods and the season's average (see Appendix A for the details).

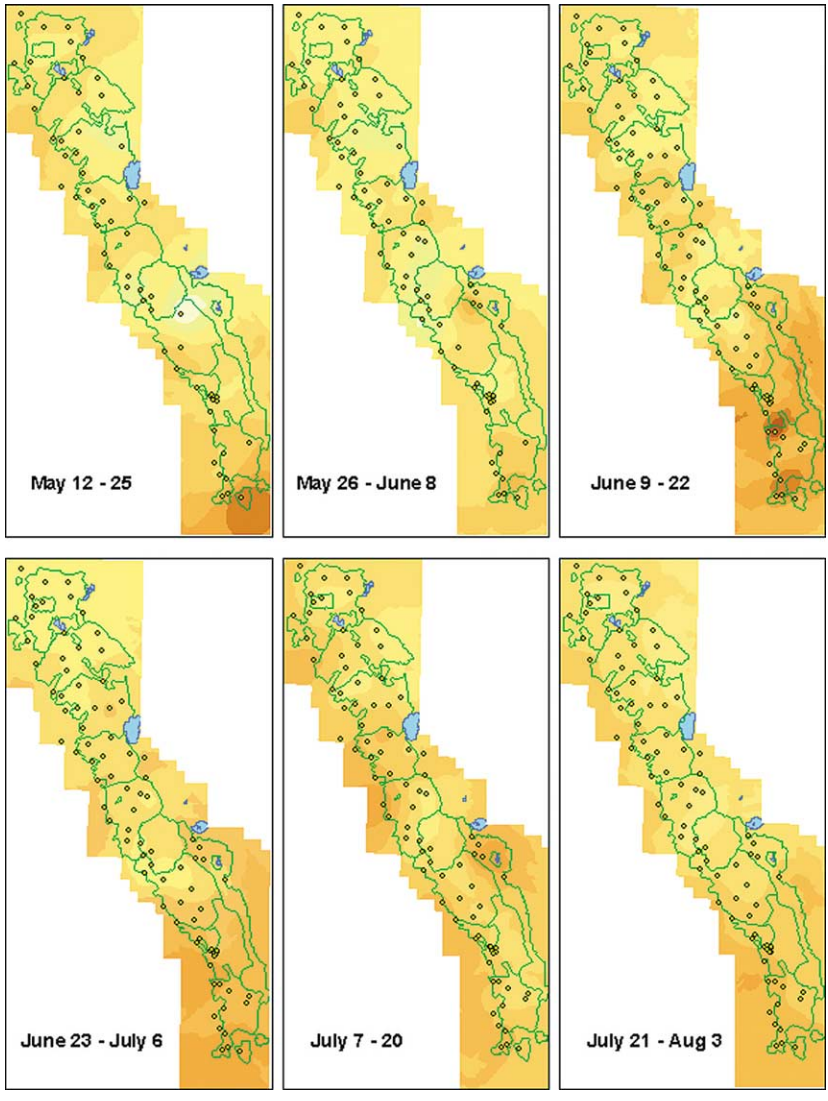
6. Results and discussion

6.1. Spatial and temporal changes of O_3 concentrations

A set of ten models of O_3 concentration for the biweekly measurement periods and one supplemental synthesis model for the entire season were prepared. These models enabled the analysis of the spatial distribution of O_3 concentration. The comparison of the models and organizing them into a time-series model helped to analyze the temporal changes of O_3 concentration (Fig. 2(a), (b)). The time-series model based on ten snapshots taken in every 14 days has identical intervals for the categories of O_3 concentration and the standardized color symbols. Therefore, it represents the concentration of O_3 from its lowest occurrences recorded (18 ppb) to the highest (142 ppb). The time-series is showing that, in general, higher O_3 concentrations were found in the south than in the north, and at low elevation sites. Very low O_3 concentrations were almost always present at the most elevated, eastern part of the Stanislaus National Forest and in the neighborhood of Lassen Volcanic National Park. On the contrary to the minimums, the time-series showed that the local maximums of O_3 concentration significantly moved spatially within time. The sequence of the ten models demonstrated continuous changes of both spatial and temporal O_3 distribution. The overall intensity of O_3 concentration varied notably in between the temporarily sequential models.

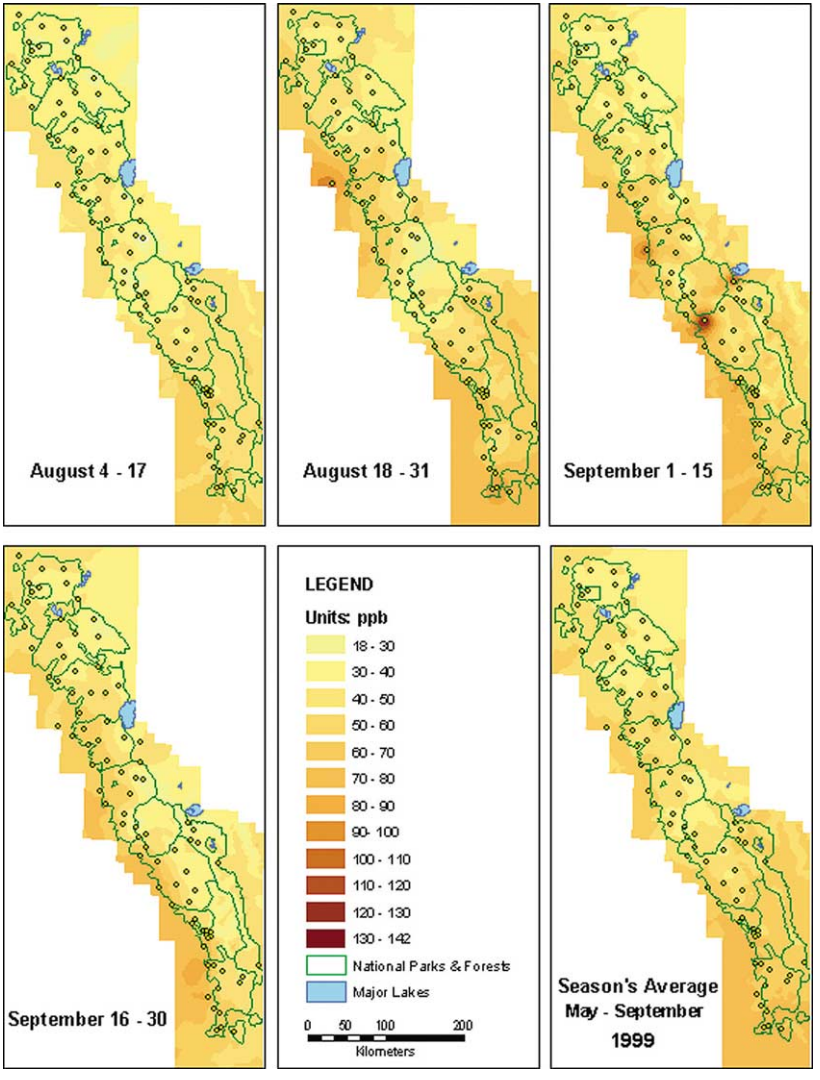
Generally, the O_3 concentration varied significantly over time (Fig. 2(a), (b)). Consequently, the range of O_3 concentrations on some maps was so small that it was graphically presented by only four to five concentration categories. In order to present more legible maps emphasizing the spatial variation of O_3 distribution, another full set of maps was generated, with each particular biweekly O_3 concentration range shown in 10 color categories. Selected maps of this cartographic version of the models clearly illustrating certain features of O_3 distribution are presented on Figs. 3–6.

The map of the season's average O_3 concentration (Fig. 2(b)) is "flattened" as a result of "averaging" temporally changing O_3 distribution patterns. Still



(a)

Figure 2. Temporal and spatial variation in ozone concentration for the entire 1999 spring–summer monitoring period. The spatial differences resulted from altitude, latitude, air temperature, proximity to the sources of air pollution, and precipitation. The temporal variation is controlled mostly by changes of O_3 concentrations in the pollution-source areas, temperature, winds, and precipitation. (a) Models for May, June, and July; (b) Models for August, September, and the average of the entire monitoring period.



(b)

Figure 2. (*Continued*)

some features were persistent enough to be considered typical for O_3 distribution in the study area and observed on the great majority of generated models:

- Overall tendency of higher O_3 concentration at the low altitudes than at the highly elevated areas.

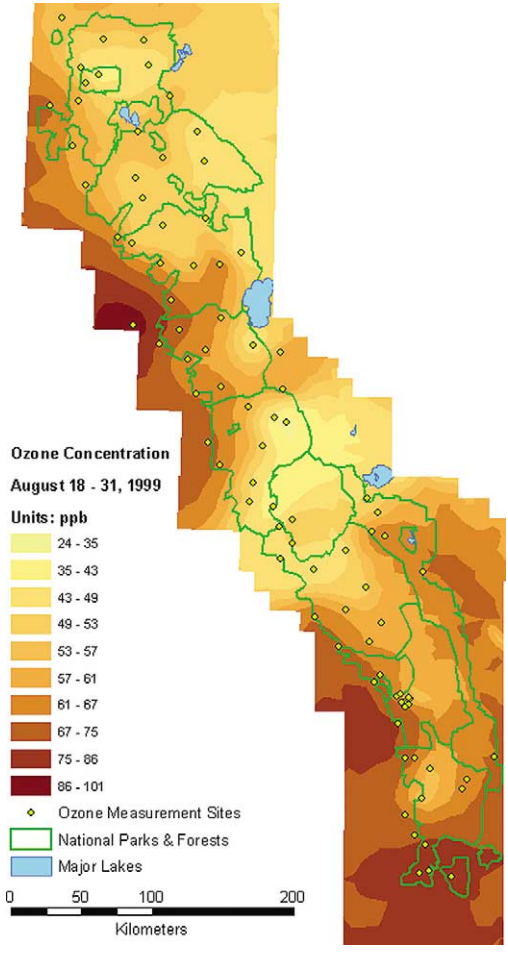


Figure 3. Distribution of O_3 concentration during the second part of August, with the intrusion of O_3 into the Sierra Nevada from the Central Valley, as well as high concentrations of O_3 in the southern part of the range and in the Owens Valley.

- Clear pattern present during the entire season showing the highest O_3 concentrations at the western foothills of the mountains in the proximity of the highly polluted California Central Valley.
- A tendency for higher O_3 concentrations in the south than in the north, related to the variation in temperatures and precipitation.

The last tendency was heavily interrupted during the prevailing part of the season by the existence of a few persistent local maximums of O_3 concentra-

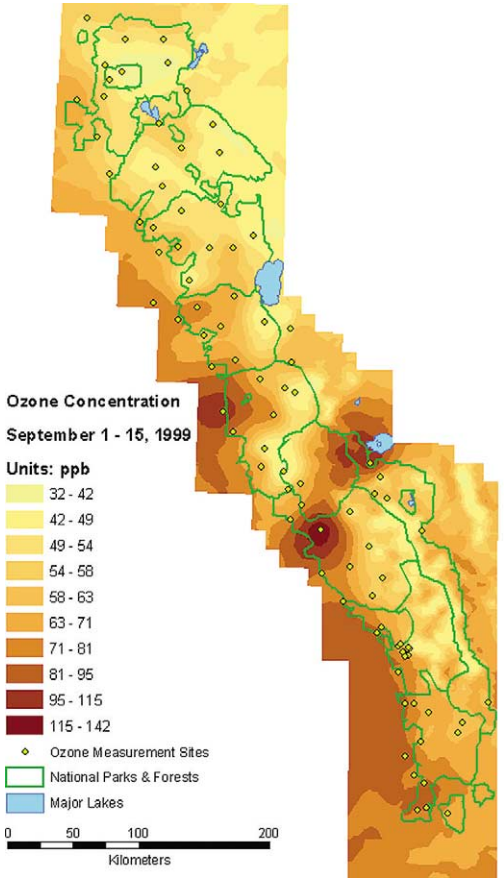


Figure 4. Distribution of O_3 concentration during the first part of September, with the presence of high O_3 concentration in the Mono Lake area.

tion. The biggest one stretched across the southern part of Eldorado National Forest. It seemed to be related to the transport of highly polluted air of the Central Valley into the Lake Tahoe area along the Mokelumne River Valley (particularly its South Fork) corridor. The corridor is especially well visible for the second part of May (Fig. 2(a)) and the second part of August (Fig. 3).

Another significant exception to the above trend, though intermittent, was the local anomaly stretching along the San Joaquin River Valley to Mammoth Lakes. It is not certain whether the elevated concentrations of O_3 at Mammoth Lakes—observed especially during the first half of June, the first half of July, and to even stronger degree, during the first half of September (Fig. 4)—are solely due to a transport of polluted air masses from the Central Valley. The

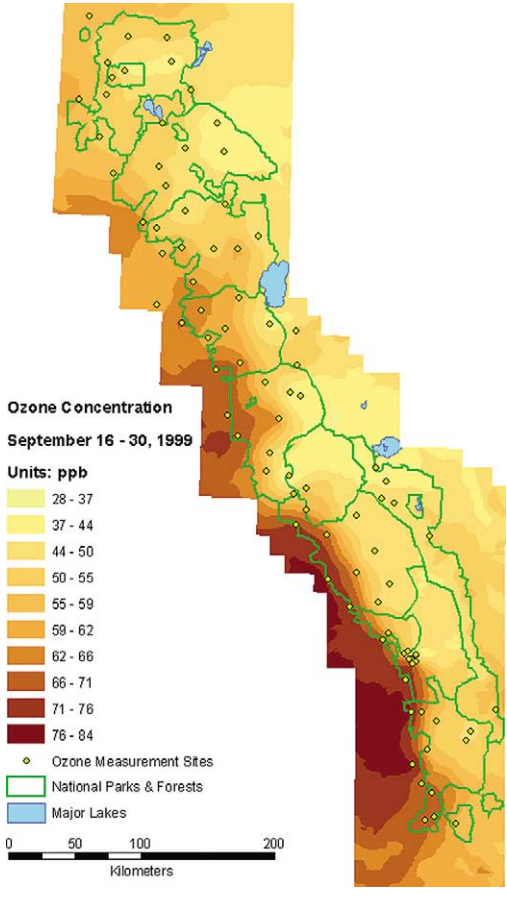


Figure 5. Distribution of O_3 concentration during the second part of September, including the significant difference between the lowlands of the Central Valley and the elevated areas of the Sierra Nevada.

Mammoth Pass between the San Joaquin River Valley and the Mammoth Lakes area is at 2800 meters, which seems to be low enough to allow transport of the O_3 contaminated air masses into the eastern side of the mountain range by southwesterly winds.

Owens Valley had high O_3 concentrations throughout the entire summer season. Origin of elevated levels of O_3 in the Owens Valley may reflect both high background level of the pollutant on the North American continent (Lefohn et al., 2001) and transport of polluted air masses from the Central Valley or the Los Angeles Basin. The geographic conditions in the area, such as low elevations, high temperatures, negligible precipitation, and the lack of clearing

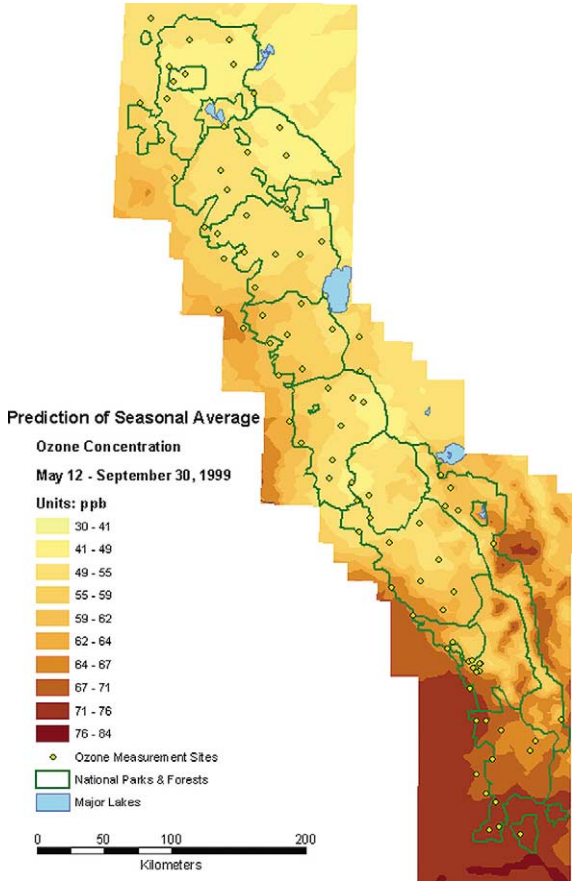


Figure 6. The "summary" of O_3 distribution for the 1999 monitoring period, including the pattern of increasing concentration of O_3 toward the south and high O_3 concentration in the Owens Valley.

winds in this deep rift valley were all favorable for high O_3 concentration to develop.

A clear decline in O_3 concentration was observed at the western foothills of the Sierra Nevada in the second half of September. The uninterrupted pattern of high concentrations of O_3 at lowlands of the westerly foothills transformed rapidly along the transition borderline parallel to the westerly slopes of the mountains. It may indicate that with the beginning of cooler temperatures, the higher O_3 concentrations are not reaching mountainous areas as they do earlier in the season (Fig. 5).

Most of the above spatial features of O_3 distribution are also visible in the synthesis model (Fig. 6). To generate the synthesis map, all ten records of O_3 concentration for the measurement season were averaged, and an output surface produced. On the season's average model, the range of values was shrunken, as the maximum value was about the same as the smallest of all the biweekly maximums while the minimum one was similar to the largest out of the recorded values of all biweekly minimums. Despite this necessary generalization, the O_3 concentration model shows that the highest concentration of O_3 was occurring at the southern foothills of the range, with the lowest ones being at the northern highlands. The low altitudes of the mountain range adjacent to the Central Valley and Owens Valley were characterized by high concentration of O_3.

6.2. Possible effects of precipitation

The general influence of rains and storms on O_3 concentration in the summer season is remarkable. Because the nature of the typical storm rainfall is sporadic and limited to a relatively small area on the one hand and short lasting and intense on the other, no direct and measurable correlation with O_3 concentration was determined. However, a strong overall pattern was observed. During the periods when the O_3 concentrations were the highest (the first half of June, the first half of September), the overall precipitation over the study area was the lowest. Similarly, the periods of the lowest concentration of O_3 were characterized by the highest amounts of precipitation (the second half of May, the first half of August, and the second half of September. The large reduction in O_3 concentration was a result of rainfall (compare Tables 1 and 2 and the maps in Figs. 7 and 8). For instance, during the second half of May, the sum of all measured records of precipitation received from all 87 stations was the highest (4053 millimeters). At the same time, the mean value of O_3 was 49.5 ppb, which was the lowest recorded value for all time periods. Interestingly, during the first half of August the area of Lake Tahoe received intense (60–90 millimeters) precipitation and the O_3 concentration was very low in the area. During the next time frame (the second half of August) with no rain there, the concentrations of O_3 significantly increased. Unfortunately, it was not feasible to mathematically quantify the described associations.

In addition to temperature, the distribution of the storms over the mountains affects the spatial distribution of O_3. As a result, the spatial variation of O_3 concentration for a single measurement period was significant, as the ratio between the maximum and minimum values of O_3 concentration was typically 3 : 1 to 4 : 1, or even more than 5 : 1 in the second half of May and the first half of June. These variations were caused by the combination of temperatures, elevation, precipitation, and latitudinal location (lower latitudes have more direct, high angled sun radiation).

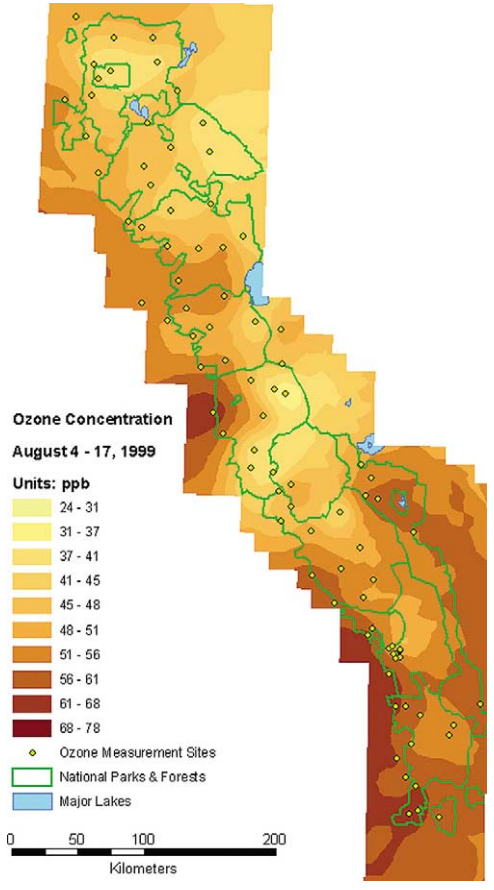


Figure 7. Distribution of O_3 concentrations during the first part of August. Patterns of local high and low concentrations can be compared with amounts of received precipitation during the same period (see Fig. 8).

The temporal disparity during the season was significantly smaller than the spatial variation of O_3 concentration. The ratios between the highest values of O_3 concentration and the lowest ones of the season for the maximum, minimum, and mean values of O_3 were 1.8 : 1, 1.7 : 1, and 1.3 : 1, respectively.

6.3. Possibility of a trans-Sierran transport of polluted air masses

Large variations in O_3 concentrations within the Sierra Nevada and especially occasional high levels of the pollutant on the eastern side of the range (where no significant sources of photochemical smog exist) indicate a possibility of

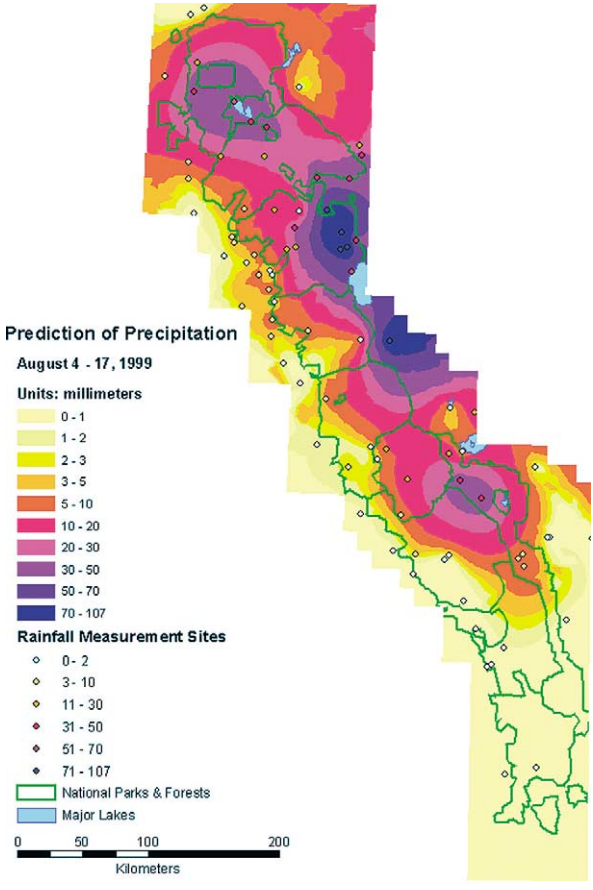


Figure 8. Distribution of precipitation over the Sierra Nevada during the first part of August. Areas that received the largest and the smallest amount of rain can be compared with the distribution of O_3 concentration during the same period (see Fig. 7).

a trans-Sierra transport of polluted air masses. This event could occur during conditions of favorable winds moving contaminated air masses over specific mountain passes. Most likely the heavily polluted air of the Central Valley could migrate along the river valley corridors most suitable for such a transfer when the prevailing winds are from west/southwest or southwest. A significant influx of O_3 could have happened at least occasionally when the winds were favorable to move the masses of polluted air through the valleys and the passes.

The highly elevated topography of the southern Sierra Nevada creates a considerable problem for the interpolation of O_3 concentration. It is known from

the other studies that urban pollution plumes can reach only certain altitudes in the surrounding mountains (Turco, 1997). What these exact altitudes might be is disputable, as it depends on the climate (especially received rain, thickness of inversion layer in pollution sources areas, speed and direction of prevailing winds) and the topography itself. Certainly, the mountain range rising above 3000 and partially even more than 4000 m a.s.l. creates a major physical barrier for transport of the polluted air masses (Fig. 9). A separate study would need to be conducted to specifically determine local conditions for O_3 migration in the study area. Here, we assume that O_3 does not typically cross over 3000 m elevational barriers. Up to this elevation, the O_3 concentrations could increase in the complex terrain of the Austrian Alps (Puxbaum et al., 1991). On the western Sierra Nevada slopes, the highest concentrations of O_3 originating from the Central Valley were found at about 1300 m a.s.l. (Bytnerowicz et al., 2002), and in the San Bernardino Mountains of southern California, the highest concentrations of O_3 generated in the Los Angeles Basin occurred at about 1500 m a.s.l. (Bytnerowicz et al., 1999). However, a hypothesis to be tested in future studies is that during episodes of high photochemical smog in the Central Valley and favorable weather and topographic conditions, urban air pollution plume may pass over 3000 m. Obviously, O_3 concentrations will decrease as the distance between the pollution source and receptor sites increases (dilution and deposition effects).

Given these phenomena, O_3 concentration could not be accurately predicted for some forested areas in the eastern sections of the Sierra Nevada range. Predictions of O_3 concentration for those particular areas were problematic for the following reasons:

- Lack of an adequate number of measurement sites in the Inyo National Forest, the eastern part of Yosemite National Park, and to a lesser degree in the eastern Plumas National Forest.
- Inability to properly interpolate and extrapolate the O_3 values from the sites located in the southwestern part of the Sierras because of (1) the distance exceeding the reasonable radius of prediction from the nearest existing sites and (2) the high physical elevational barrier between the existing O_3 measurement sites on the western slopes and the eastern slopes of the southern Sierras.

The problem is particularly critical in the southeastern parts of the study area. Because of the existing topographical barrier (Fig. 9) of the highly elevated mountain range (3000–4000 m a.s.l.) stretching in the direction perpendicular to the prevailing westerly winds, O_3 concentrations in these areas were probably estimated with low accuracy. The correlations between O_3 concentration and temperature as well as between temperature and elevation, which were established for the western parts of the mountains, may not be accurate in the

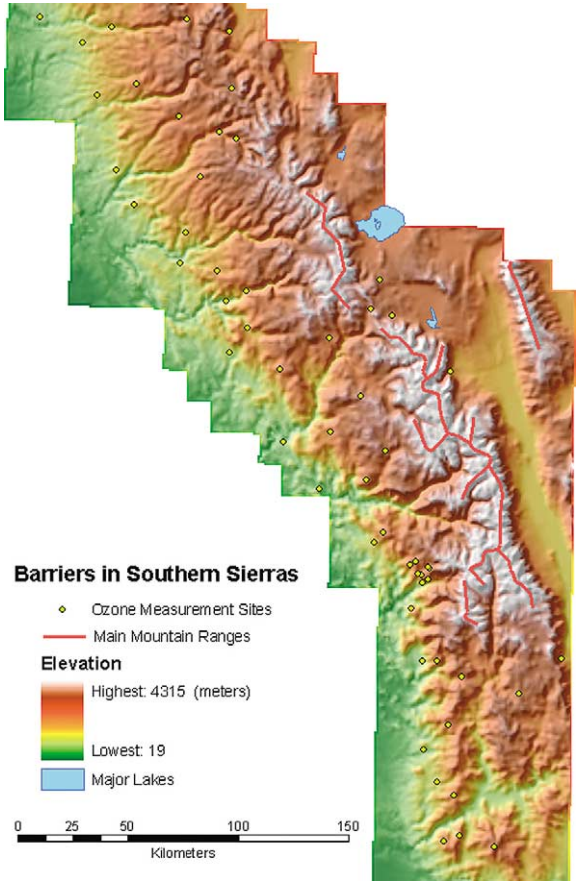


Figure 9. Highly elevated mountain ranges create significant physical barriers for O_3 transport. Because of the presence of these barriers (3000 m) interpolation and prediction of O_3 concentration in the area are problematic.

easterly sections of the mountain range. Therefore, in future studies it would be advantageous to add supplemental monitoring sites there. This is mainly because it is likely that polluted air masses from the Los Angeles Basin migrate north through Owens Valley along the eastern slopes of the Sierra Nevada. Moreover, specific topography, almost complete lack of precipitation during the long, extremely hot summer season, and the lack of clearing winds generate favorable conditions for the occurrence of elevated O_3 concentrations. Once generated or brought into the Owens Valley, high levels of O_3 may be maintained for prolonged periods of time.

7. Reliability of the model

The density of the O_3 monitoring network was not extensive enough to ensure uniformly reliable predictions of O_3 for the entire study area, especially in the southern and eastern portions of the Sierras (Fig. 10). Difficult access to some areas characterized by abrupt changes in topography may be partially responsible. In addition, some of the measurement sites did not provide the complete records of O_3 concentration for the entire season. Thus, several of

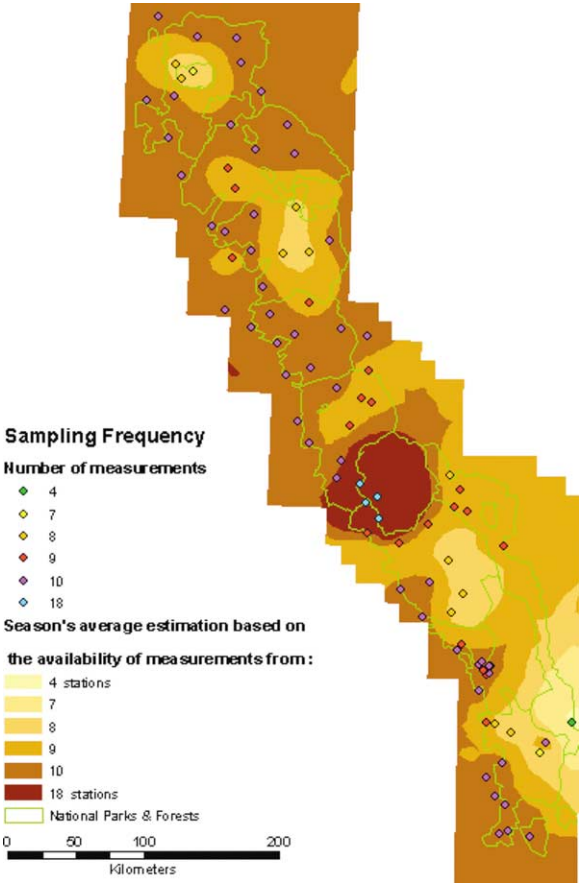


Figure 10. The uneven numbers of ozone concentration measurement samples contribute to the uncertainty of prediction at the areas not adequately monitored for O_3 concentration, especially in the southeast Sierra Nevada. Since some of the monitoring stations did not provide results from all biweekly sampling periods, the season's average O_3 concentrations were based on various number of measurements.

the generated models are based on networks of measurements with missing sites, especially in the vicinity of the Inyo National Forest.

To indicate the specific areas where more sampling is necessary and to provide explicit numbers characterizing the reliability of prediction, the Geostatistical Analyst's calculating values function for a map of prediction errors was used. The cokriging interpolation technique was applied to generate the map of standard prediction errors (Fig. 11). The utilized algorithm was iden-

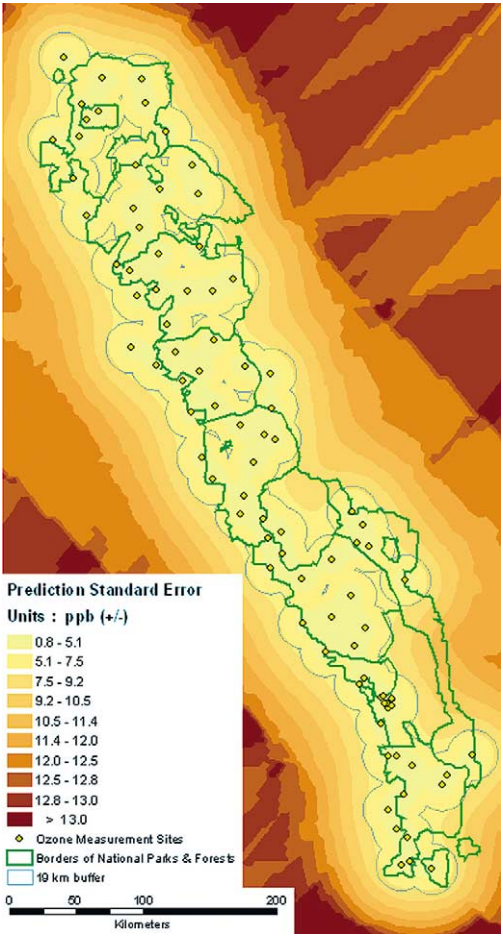


Figure 11. The certainty of prediction is based on the proximity to the sample points and the density of the measurement network. Beyond certain distance, the accuracy of estimation starts dropping rapidly. The buffer of 19 km from the O_3 monitoring sites was used to approximate the threshold value of the acceptable prediction error.

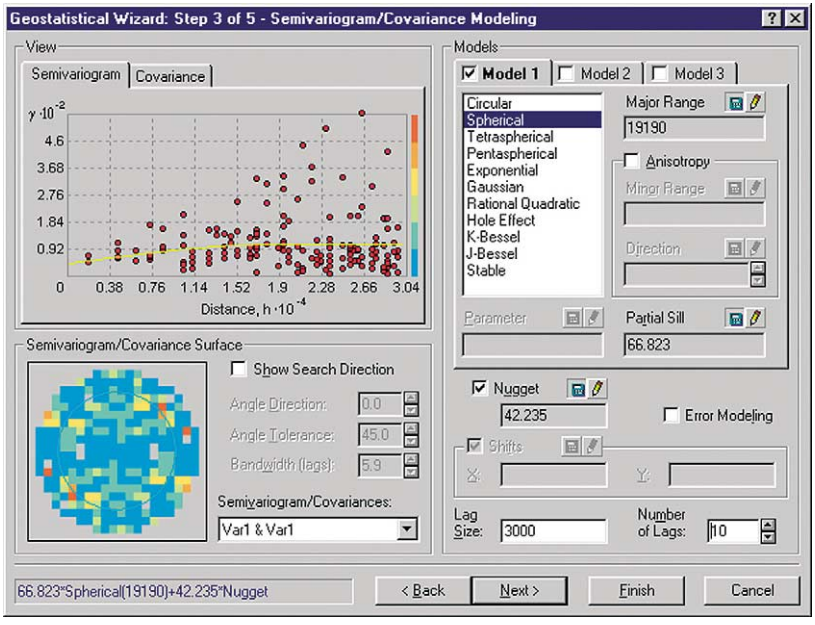


Figure 12. The semivariogram indicates that the concentration of O_3 might be reasonably estimated up to about 19 km from a sample point. This screen-saved illustration of the Geostatistical Analyst wizard shows that the correlation between measured values exists within the range of approximately 19,190 meters. Similar values might be determined from the shape of the semivariogram curve.

tical to the one used to generate the set of prediction models of O_3 concentration.

The range of spatial autocorrelation was determined to be about 19 km (Fig. 12). Beyond this distance, the estimated values rapidly loose accuracy. Based on the entire applied dataset, a geostatistical ratio of the prediction error of O_3 concentration versus the distance from monitoring sites was established. The values of that surface of prediction errors at the distance of 19 km from sample points, at the direction where extrapolation had to be applied, were about ± 8.2 ppb. For cartographic purposes, this continuous surface of O_3 concentration reliability was transformed into a map showing four categories of confidence (Fig. 13) in the previously described O_3 concentration models.

The map on Fig. 14 reveals that even the nearest proximity to the sampling points carry the inherited error of prediction larger than ± 0.79 ppb. It can be interpreted that even at its best fitting, the cokriging output surfaces smooth the distribution of O_3 concentration from the directly recorded values (Babish,

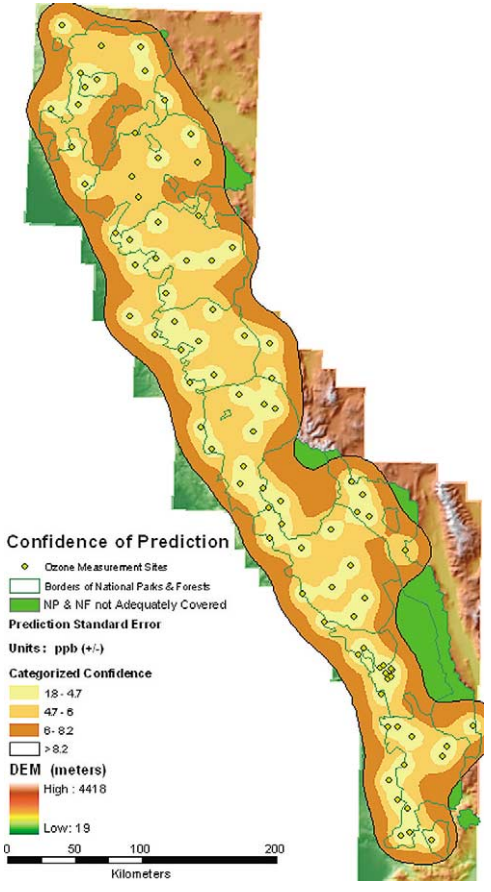


Figure 13. The areas outside the solid black line representing roughly the limit of the acceptable accuracy of prediction, should be considered as those not adequately sampled for O_3 concentration. For that reason, the estimated values of O_3 concentration over these areas on all presented models should be treated as uncertain.

2000). In the case of this dataset, the magnitude of surface generalization was smaller than 1 ppb. This may indicate the unavoidable presence of the measurement error and/or small-scale irregularities of O_3 concentration.

8. Simulated ozone monitoring network

An enhanced version of the monitoring network for possible future implementation was developed using a simulation exercise. To generate this new optimized network, it was assumed that all of the sampling sites of the 1999 net-

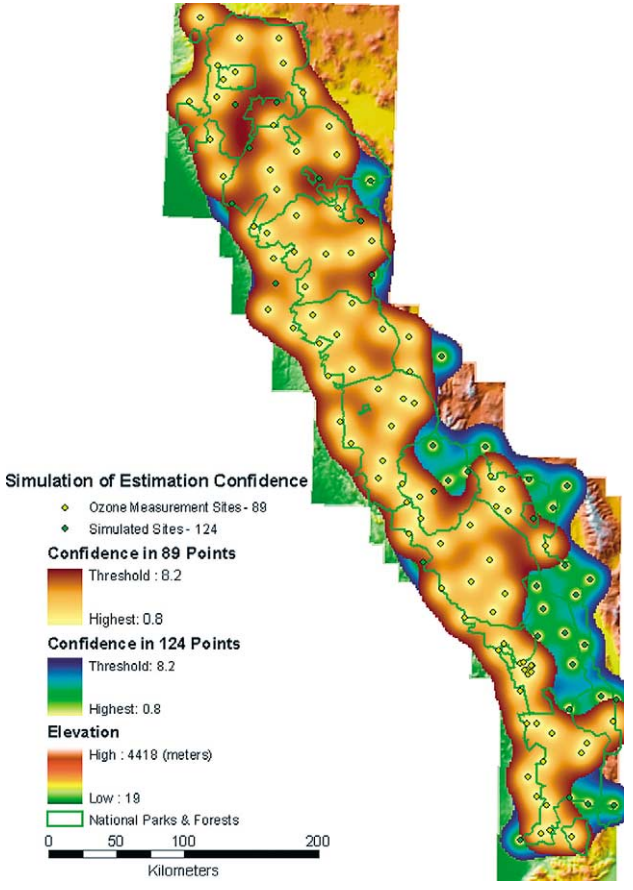


Figure 14. Distribution of points of the simulated network. With the additional 35 points, O_3 concentration could be adequately estimated on the entire study area.

work should be included in the proposed new network. Maintaining the same sites would ensure consistency needed for comparison of the results in time. Proposed supplemental O_3 monitoring sites should be located at:

- Areas not covered by the sampling network of 1999 made of the 89 points to be able to reliably estimate O_3 concentration for the entire Sierra Nevada range.
- Areas of the highest occurrences of O_3 concentration (during any of the measurement periods) to verify reliability of the 1999 measurements and possibly to detect even higher levels of O_3 concentration.

- Areas of low elevations located below the sites of the highest detected concentration of O_3 to determine sources of air pollution and directions of pollution plume migration.
- Areas along the major river valleys perpendicular to the main range to check whether the migration of O_3 from the heavily polluted Central Valley to the eastern side of the mountains truly occurs.
- Areas located at elevations above 2600 meters (9000 ft) to determine the local elevational gradient/environmental lapse rate of O_3 concentration.
- Along roads to ensure an easy access to the proposed new monitoring sites.

By following the above criteria, locations of the proposed supplemental sites were digitally placed on a map. A new model of the prediction error was generated using the Geostatistical Analyst. This was achieved by adapting the same algorithm as was done for the 1999 network of 89 points. The simulation indicated that an additional 35 new points were necessary to create the new network of the total of 124 points.

The densified network of 124 points will allow adequate monitoring of O_3 concentration in the combined area of the 12 national parks and forests of the Sierra Nevada. An additional advantage of the densified network would be the increased average reliability of the prediction of O_3 concentration for the study area, which would be improved from ± 7.8 to ± 6.5 ppb. This would result not only from sufficiently covering the areas where the trustworthiness was not sufficient, but also by improving the reliability of prediction where it was not adequate, e.g., central part of Plumas National Forest. Supplemental sites were added mainly in the Inyo National Forest, eastern foothills of the range adjacent to Owens Valley, the wilderness area (the eastern part) of Kings Canyon National Park, the area around Mono Lake, and the elevated eastern section of Yosemite National Park. Individual additional sites were also suggested in the northern part of the study area where the 1999 network was sparse and in several instances along the foothills of the western Sierras (Fig. 14).

9. Conclusions

Strong spatial variation of O_3 concentration over the study area was observed and was caused by a combination of several reasons. The primary factor was obviously proximity of the pollution source areas. Abrupt changes in the geographic conditions of the mountainous topography played a major role in distribution of O_3 concentrations. Significant longitudinal extent of the study area and the large range of elevations above sea level affected ambient air temperatures. As a result, the O_3 concentrations were highest at the foothills of the southern Sierras and lowest at the high altitudes of the northern part of the mountain range. This pattern is a function of distance (both vertical and horizontal) from the O_3 sources.

Summer thunderstorms typically bring intense rains for spatially limited areas. Almost random spatial and temporal distribution of the storms over the mountains heavily influenced the temporal and spatial variation of O_3 concentrations by setting spatial and temporal pockets of reduced pollutant concentrations.

The temporal variation during the 20-week measurement season was much smaller than the geography-driven spatial variation. The fact that the spatial variation was more (about threefold) diversified than the temporal one indicates that most of the time the Central Valley O_3 source area controlled the distribution of O_3 concentrations. However, elevated O_3 levels in the eastern Sierra during some periods may indicate that pollution from the Los Angeles Basin could also get into that area. The observed less significant temporal variations were the function of the climatic and weather patterns—for the most part random distribution of summer storms, which were reducing the O_3 concentration in their spatial and temporal vicinity. Hot temperatures observed mostly at the foothills around the southern part of the Sierra Nevada locally magnified the already high concentrations of O_3.

It has been determined that the total number of sampling points in the network of air pollution monitoring sites was not sufficient to ensure high and relatively uniform level of confidence in the O_3 concentration estimates. Consequently, in order to ensure high quality and trustworthiness of model predictions, it is necessary to establish additional monitoring sites, especially on the eastern slopes of the southern portion of Sierra Nevada (Inyo National Forest, the eastern Yosemite National Park, and others).

High O_3 concentrations in portions of the eastern Sierra Nevada indicate possibilities of the trans-Sierra transport of polluted air masses from the California Central Valley and long-range transport of the Los Angeles pollution plume north along the Owens Valley. It is possible that there are three or more corridors along river valleys crossing the Sierra Nevada where with favorable winds O_3 passed from California's Central Valley to the valleys of the eastern site. To verify this hypothesis, additional O_3 monitoring sites should be established along valleys crossing the Sierra Nevada (e.g., the San Joaquin River Valley) and along eastern slopes of the range in the proximity of Owens Valley. More monitoring of O_3 in these corridors would also help to indicate what meteorological conditions are necessary for the influx O_3 to take place. Information about wind direction and speed would greatly help in understanding of O_3 distribution in the Sierra Nevada. Similarly, it would be important to determine what are the sources of air pollution affecting the study area. These probably would be the California Central Valley, but the Los Angeles Basin area or other smaller source areas or point sources in Owens Valley or the Mojave Desert have to be considered.

These general concerns regarding the missing pieces of information should be investigated and addressed so that a clear understanding of O_3 distribution is reached. This in turn should help in reducing air pollution to the levels safe for the sensitive ecosystems of the Sierra Nevada. The evidence of forests affected by elevated levels of air pollutants can be used to increase environmental awareness of the general public and agencies responsible for implementation of national and state air pollution standards. Reliable information on O_3 distribution is also needed for proper planning of prescribed fires in national forests. Emissions from forest fires contain significant amounts of the O_3 precursors (mainly NO_x and hydrocarbons) that could deteriorate air quality of the surrounding areas above the accepted levels. In such situations, permits for conducting prescribed fires should not be issued.

In geostatistics, every spatial dataset might be considered unique. For each one of them, even a change of one value from a single sampling point, or one parameter during processing of data will affect the final output. Similarly, a change of one of the numerous variables (geostatistical parameters) offered by the Geostatistical Analyst will influence the model's final product. Therefore, geostatistical modeling is a process where the real optimization of the final geostatistical surface requires a combination of expertise in geostatistics and knowledge of a particular branch of science (air chemistry, meteorology, and geography). However, because even experts opinions are prone to some error, it may not always be possible to determine beyond a reasonable doubt that the end result of hard, prolonged labor fully reflects the reality. However, to the best of the authors' knowledge the presented models are reasonably assessing O_3 distribution in the Sierra Nevada.

Acknowledgements

We would like to thank numerous individuals who helped in the success of this study. Bill Lyons and Trent Procter of the Sequoia National Forest helped in the management of field monitoring efforts. Dozens of USDA Forest Service and USDI National Park Service employees changed passive samplers. Diane Alexander, Enrique Jimenez, and Tony Davila of the Forest Service extracted and analyzed passive samplers. Rocío Alonso helped with O_3 concentration calculations. We also thank the Environmental Systems Research Institute for its continuous support and for making the Geostatistical Analyst software available for the project.

Appendix A

We used an algorithm to derive the relevant final geostatistical surfaces of O_3 distribution in the Sierra Nevada study area in 1999. Naturally, due to

the changing number of O_3 input samples as well as their variant values, the variograms, and consequently the parameters of prediction errors, differed for every data collection period. The applied methodology was consistent for all the presented final geostatistical models of O_3 distribution. The provided example depicts the case of the first half of July 1999.

1. From 61 weather stations with available measurements of maximum biweekly temperature located in the study area, a surface of maximum temperatures was created by cokriging with the elevation as a secondary variable. The geostatistical prediction surface was based on *Ordinary Cokriging* with the following parameters and options:

   - Detrending was applied to both datasets: the *Constant* Order of Trend Removal with 65% global for maximum temperature and the *First* for the DEM with 80% global;
   - Lag size of 18 km;
   - Nugget of 8 (which for interpretation purposes could be roughly translated to allowing the maximum difference from the forced matching of $\pm 1.4\,^{\circ}$F) for temperature and of 5000 for elevation (loosely translated to allow difference from the ideal matching of 35 meters—smoothing effect);
   - Default value of 32.8 km for the major range;
   - Model applied: *Spherical*;
   - A minimum of two neighbors taken into account in each of four sections of the *Searching Neighborhood* circle;

2. The prediction errors of the resulting surface were the following:

   - Mean: 0.749;
   - Root-Mean-Square: 5.30;
   - Average Standard: 5.28;
   - Mean Standardized: 0.139;
   - Root-Mean-Square Standardized: 0.9984;
   - Difference between Root-Mean-Square and Average Standard: 0.014;

3. The output surface has the maximum value of $97\,^{\circ}$F and the minimum of $70\,^{\circ}$F.
4. The output surface was rasterized into ESRI's grid format. For practical reasons, it was decided that the output resolution is 3594 meters (a consequence of the software default number of rows and columns for the size and the geographic extent of the input data).
5. The grid representing maximum temperatures was then converted to a point coverage and subsequently into a shapefile. The number of points depicting the maximum temperature in the new information layer was 19,796.

6. The output shapefile with the densified regularized network of maximum temperature values was applied as a secondary variable in the generation of a new surface of O_3 concentration.
7. The parameters and options used to generate the final surface of O_3 concentration:

   - No trend removal was applied to generate this surface;
   - Lag size of 10.7 km;
   - Nugget of 66 for O_3 concentration (± 4 ppb) and 3.14 for the maximum temperature (less than 1°F);
   - Default major range of 77.66 km;
   - Model applied: *Spherical*;
   - A minimum of two neighbors taken into account in each of four sections of the *Searching Neighborhood* circle;

8. The prediction errors of the resulting surface were the following:

   - Mean: -0.231;
   - Root-Mean-Square: 14.3;
   - Average Standard: 12.4;
   - Mean Standardized: -0.0125;
   - Root-Mean-Square Standardized: 1.171;
   - Difference between Root-Mean-Square and Average Standard: 1.92;

9. The output surface has a maximum value of 107.1 ppb and a minimum one of 28.5 ppb.

References

Babish, G., 2000. Geostatistics without tears. A practical guide to geostatistics, variograms and kriging. Environment Canada.

Brace, S., Peterson, D.L., 1998. Spatial patterns of tropospheric ozone in the Mount Rainier of the Cascade Mountains, USA. Atmos. Environ. 32, 3629–3637.

Bytnerowicz, A., Fenn, M.E., Miller, P.R., Arbaugh, M.J., 1999. Wet and dry pollutant deposition to the mixed conifer forests. In: Miller, P.R., McBride, J.R. (Eds.), Oxidant Air Pollution Impacts in the Montane Forests of Southern California. In: Ecological Series, Vol. 134. Springer, New York, pp. 235–269.

Bytnerowicz, A., Tausz, M., Alonso, R., Jones, D., Johnson, R., Grulke, N., 2002. Summer-time distribution of air pollutants in Sequoia National Park, California. Environ. Pollut. 118, 187–203.

Cressie, N., 1993. Statistics for spatial data. John Wiley & Sons, New York.

Duriscoe, D.M., Stolte, K.W., 1989. Photochemical oxidant injury to ponderosa (Pinus ponderosa Laws.) and Jeffrey pine (Pinus jeffreyi Grev. and Balf.) in the national parks of the Sierra Nevada of California. In: Olson, R.K., Lefohn, A.S. (Eds.), Effects of air pollution on western forests. Air & Waste Management Association, Pittsburgh, PA, pp. 261–278.

Finlayson-Pitts, B.J., Pitts Jr., J.N., 2000. Chemistry of the upper and lower atmosphere. Academic Press, San Diego, CA.

Johnston, K., Ver Hoef, J., Krivoruchko, K., Lucas, N., 2001. Using ArcGIS Geostatistical Analyst. Environmental Systems Research Institute.

Lefohn, A.S., Oltmans, S.J., Dann, T., Singh, H.B., 2001. Present-day variability of background ozone in lower troposphere. J. Geophys. Res. 106, 9945–9958.

NOAA, 1999. National Climatic Data Center, Daily Surface Data, October 8, 1999. Ashville, North Carolina.

Peterson, D.L., Arbaugh, M.J., 1992. Mixed conifer forests of the Sierra Nevada. In: Olson, R.K., Binkley, D., Bohm, M. (Eds.), The Response of Western Forests to Air Pollution. In: Ecological Studies, Vol. 97. Springer-Verlag, New York, pp. 433–459.

Phillips, D.L., Lee, E.H., Herstrom, A.A., 1997. Use of auxiliary data for spatial interpolation of ozone exposure in southeastern forests. Environmetrics 8, 43–61.

Puxbaum, H., Gabler, K., Smidt, S., Glattes, F., 1991. A one-year record of ozone profiles in an Alpine valley. Atmos. Environ. 25A, 1759–1765.

Seinfeld, J.H., Pandis, S.N., 1998. Atmospheric chemistry and physics. John Wiley & Sons, New York. 1326 p.

Simpson, D., 1991. Long period modeling of photochemical oxidants in Europe. EMEP, Cooperative Programme for Monitoring and Evaluation of the Long Range Transmission of Air Pollutants in Europe, Meteorological Synthetizing Centre—West, The Norwegian Meteorological Institute, Oslo, Norway.

Smidt, S., Gabler, K., 1995. SO_2, NO_x and ozone records along "Achenkirch altitude profiles". Phyton (Horn, Austria) 34, 33–44.

Turco, R.P., 1997. Earth Under Siege. Oxford Univ. Press, Oxford.

Webster, R., Oliver, M., 2001. Geostatistics for Environmental Scientists. John Wiley & Sons, New York.

Zeigler, M., 1999. Modeling Our World. The ESRI guide to geodatabase design. ESRI Press.