Journal of Exposure Analysis and Environmental Epidemiology (2004) 14, 404–415

www.nature.com/jea

Comparison of spatial interpolation methods for the estimation of air quality data

DAVID W. WONG,[a] LESTER YUAN[b] AND SUSAN A. PERLIN[b]

[a]*School of Computational Sciences, George Mason University, Fairfax, VA, USA*
[b]*National Center for Environmental Assessment, Washington, US Environmental Protection Agency, USA*

We recognized that many health outcomes are associated with air pollution, but in this project launched by the US EPA, the intent was to assess the role of exposure to ambient air pollutants as risk factors only for respiratory effects in children. The NHANES-III database is a valuable resource for assessing children's respiratory health and certain risk factors, but lacks monitoring data to estimate subjects' exposures to ambient air pollutants. Since the 1970s, EPA has regularly monitored levels of several ambient air pollutants across the country and these data may be used to estimate NHANES subject's exposure to ambient air pollutants. The first stage of the project eventually evolved into assessing different estimation methods before adopting the estimates to evaluate respiratory health. Specifically, this paper describes an effort using EPA's AIRS monitoring data to estimate ozone and PM10 levels at census block groups. We limited those block groups to counties visited by NHANES-III to make the project more manageable and apply four different interpolation methods to the monitoring data to derive air concentration levels. Then we examine method-specific differences in concentration levels and determine conditions under which different methods produce significantly different concentration values. We find that different interpolation methods do not produce dramatically different estimations in most parts of the US where monitor density was relatively low. However, in areas where monitor density was relatively high (i.e., California), we find substantial differences in exposure estimates across the interpolation methods. Our results offer some insights into terms of using the EPA monitoring data for the chosen spatial interpolation methods.
Journal of Exposure Analysis and Environmental Epidemiology (2004) **14,** 404–415. doi:10.1038/sj.jea.7500338

Keywords: *AIRS monitors, ozone, particulate matter, respiratory health, NHANES, spatial interpolation, kriging.*

Introduction

Many aspects of the outdoor environment, including air quality, can adversely affect health in general. But outdoor air quality may be the most influential factor, among others, in affecting respiratory health. A number of studies have examined the relationships between ambient air quality and respiratory health effects (Detels et al., 1987; Abbey et al., 1991; Schwartz et al., 1994; Brunekreef et al., 1995; Pope et al., 1995a, b; Kunzli et al., 1997; McConnell et al., 1999; Norris et al., 1999). One approach to investigating these relationships is to collect environmental data coincidentally with health data so that the environmental data can directly support analysis of health outcomes (Stern et al., 1989; Thurston et al., 1992; Schwartz et al., 1994; Linn et al., 1996; Vedal et al., 1998; Peters et al., 1999). As studies of this type are costly, they are typically conducted over relatively confined areas with small sample sizes. An alternate approach is to use existing health and environmental databases.

Unfortunately, most databases for general health surveys do not have detailed environmental and/or air quality information. Commonly used health data sets include the series of National Health and Nutrition Examination Surveys (NHANES). The National Center for Health Statistics (NCHS) completed the third NHANES (NHANES-III, 1988–1994). This NHANES data set provides very detailed health information on a statistically representative sample of the US population. NHANES-III

1. Abbreviations: AIRS, Aerometric Information Retrieval System; BG, block group; CO, carbon monoxide; EPA, Environmental Protection Agency; FIPS, Federal Information Processing System; IDW, inversed distance weighting; NAAQS, National Ambient Air Quality Standards; NAMS, National Air Monitoring Stations; NCHS, National Center for Health Statistics; NHANES, National Health and Nutrition Examination Survey; NO2, nitrogen dioxide; O3, ozone; OAQPS, Office of Air Quality Planning and Standards; PM10, particulate matter below 10im mean diameter; SLAMS, State and Local Air Monitoring Stations; SO2, sulfur dioxide; TSP, total suspended particulates.

2. Address all correspondence to: David W. Wong, School of Computational Sciences, George Mason University, Fairfax, VA 22030, USA. Tel.: +703-993-1212; Fax: +703-993-1980.
E-mail: dwong2@gmu.edu

3. The views expressed in this paper are those of the authors and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency.

Received 22 May 2002; accepted 11 November 2003

includes a series of survey questions on various aspects of health history of the subjects, including detailed respiratory health data and quantitative spirometry tests of respiratory function. These data provide a rich source of health outcome data that potentially can be linked to exposure estimates of ambient air pollution, but they do not provide any air quality information. The lack of air quality information is not unique in HNAHES, but is quite common among general health survey data.

Many researchers have used EPA air-monitoring data to estimate ambient air exposure as risk factors for human health effects. Using NHANES as an example, we are aware of only several studies that have already linked this data set to environmental exposure estimates derived from EPA air quality data to assess potential associations between environmental exposures and health effects. Chestnut et al. (1991) used EPA air-monitoring data to examine the association between total suspended particulates and lung function in adults from an earlier NHANES-I survey. Schwartz (1989) estimated ambient air concentrations of pollutants based on EPA monitoring data and then linked these estimates to NHANES-II children's respiratory function data. A recent study by Schwartz (2001) used NHANES-III data to examine air pollutants as cardiovascular risk factors.

In these studies, air pollution and health data were collected in independent surveys, so that human subjects' residences and air monitors were not colocated. As a result, air pollution concentrations needed to be interpolated between monitors in order to estimate exposures to study subjects. At this time, no consensus exists as to the appropriate method for interpolating between irregularly spaced monitors. Indeed, the NHANES-based studies cited above all used different interpolation methods. Chestnut et al. (1991) identified a central urban area for each NHANES-I location and averaged all monitoring data located in that area. Schwartz (1989) assigned subjects the average of pollutant levels from all monitors located within 10 miles of the population centroid of the census tract in which they resided. Schwartz (2001) used the weighted average of all monitors in each subject's county of residence plus the adjoining counties. These weights were proportional to the inverse of the square of the distance between the centroid of the residence's block group and the monitor.

Given that different interpolation methods are found in the literature for estimating ambient air quality to assess health outcomes, our objective is to evaluate spatial interpolation methods methodically to assess the degree to which different methods influence the estimated air pollutant levels. Many papers have attempted to interpolate levels of air pollutants and some efforts intended to incorporate more than one measurement (e.g., Brown et al., 1994) and explore their relationships to health outcomes (e.g., Duddek et al., 1995). The current paper addresses a much simpler type of situation with only one measurement in each estimation process. Our ultimate objective is to evaluate how different interpolation methods will affect the assessment of relationships between air pollutant levels and respiratory health captured by NHANES data, but our results reported in this paper will be applicable to the analysis of other health outcomes related to air pollutants.

Methods and materials

Data

In NHANES-III, data were gathered from 82 unique counties across the country from 1988 to 1994, including almost 14,000 children up to 16 years of age. NCHS collected address information to geocode each subject to the nearest census block group (BG), which, in general, contains between 600 and 3000 people with an optimal size of 1500 people (http://www.census.gov). But NCHS does not release this information to the public in order to protect subjects' confidentiality. Only the identification of the county in which the subject resided is available to the public, provided the county had a population of greater than 500,000. Approximately half of the children in NHANES-III resided in these large counties. The remaining children resided in small counties that are not identified by NCHS in the public database.

In order to meet Clean Air Act requirements, every state has established a network of air-monitoring stations for regulated pollutants (CO, NO_2, O_3, lead, PM_{10}, and SO_2). Data from these monitors are submitted to and stored in EPA's AIRS database. The vast majority of these monitors sample air quality in the country's densely populated urban areas. Based on the current scientific evidence of associations between exposure to various air pollutants and adverse respiratory health effects (Detels et al., 1987; Abbey et al., 1991; Schwartz et al., 1994; Brunekreef et al., 1995; Pope et al., 1995a, b; Kunzli et al., 1997; McConnell et al., 1999; Norris et al., 1999), we limited our study to estimating exposures to O_3, PM_{10}, NO_2, and SO_2. This paper presents our methods and results for estimating exposures to outdoor ambient levels of O_3 and PM_{10} for 1990 for illustrative purposes only. Owing to confidentiality constraints, we needed to estimate ambient air concentrations in every BG in every county in which NHANES-III subjects resided. These estimates were then linked to subjects by NCHS personnel using the confidential location information for subsequent analyses. For this preliminary work, we linked all the NHANES-III children to the 1990 concentration estimates, but eventually, we will link children by the year when they were examined to the year-specific monitored data.

Exposure Metrics

The AIRS database reports O_3 as an hourly average concentration. O_3 concentrations usually exhibit a daily diurnal pattern and a seasonal pattern. Owing to the strong seasonality of O_3 concentrations, many states limit their O_3 monitoring to a certain portion of the year, termed the ozone season, the length of which varies from one area of the country to another (US EPA, 1996a). As we were interested in chronic health effects, we sought to identify one statistic that would characterize long-term exposure to O_3 at each location. Little consensus exists as to which statistic is the most appropriate. The US EPA (1996a) recommended the maximum 8-h moving average as an appropriate statistic. Based upon this recommendation and our analysis of O_3 diurnal and seasonal patterns in the NHANES-III counties, we decided, as an initial effort, to compute a "summer daytime ozone" metric defined as the hourly average from May to September, between 10:00 am and 6:00 pm.

The AIRS database reports PM_{10} concentrations every sixth day for a 24-h sampling period (US EPA, 1996b). We characterized exposure to PM_{10} in 1990 using two metrics. The first one was the annual average provided by AIRS. As there is a strong seasonal and geographical pattern in daily PM_{10} levels, we also wanted to capture potential exposure during peak PM_{10} seasons. To do this, we selected the calendar quarter with the highest average for each monitor as the maximum quarterly average, our second PM_{10} metric. Both averages are weighted, with the weights adjusting for changes to the scheduled sampling frequency of the monitors.

Quality Assurance of AIRS Data

To develop estimates of O_3 and PM_{10} concentrations for each BG with an NHANES-III child, we tested four spatial interpolation methods. Before conducting the interpolation procedures, we performed validation procedures to ensure that (1) the monitored data were valid and/or sufficiently complete to be used in the study, and (2) the monitor location information was correct.

(1) Evaluation of monitoring data completeness/validity: Based on the current literature (Schwartz, 1989; Abbey et al., 1991; Kinney et al., 1998; Nikiforov et al., 1998), we established the following rules for determining if monitored data were valid and appropriate for inclusion in our study. For daytime summer O_3, we required: (1) at least eight valid hourly readings out of nine during the specified time interval (10:00 am to 6:00 pm, at the hour) for that day's reading to be included into the daily average; and (2) at least 75%, or 115 days (out of 153 days), of valid day readings from May to September for that monitor to be included into the summer daytime average.

For each PM_{10} monitor, AIRS records the percent of actual data values that were reported as compared to the number of data values that should have been reported for the monitor/year or quarter (PCT_OBS). We computed the minimum value of PCT_OBS for each quarterly interval for each monitor, and required that at least 75% of the possible samples be taken in every quarter of the year for that monitor to be included. Using this value ensured that all four quarters in the year had been sufficiently sampled.

(2) Evaluation of monitor location information: Primary location information presented in the AIRS database consists of (1) the latitude and longitude of each monitor, and (2) a monitor identification number whose first five digits are the FIPS code of the state and county in which the monitor is supposed to be located. We mapped monitor locations based on their reported latitude and longitude readings. We compared counties as identified by the first five digits of the monitor identification number and the counties where the monitors were located according to their latitude and longitude readings, assuming that the first five digits of the monitor identification numbers represented the monitor's correct state and county. Thus, any mistakes in AIRS on the location of a monitor were assumed to be the result of incorrect latitude/longitude entries. As a result, we made minor latitude or longitude adjustments on two of 739 O_3 and two of 732 PM_{10} monitors.

Interpolation Methods

We selected four different interpolation methods to estimate O_3 and PM_{10} air concentrations at the census BG level. These methods have been used either in NHANES-related studies (Schwartz, 1989, 2001; Chestnut et al., 1991) or other studies (Abbey et al., 1991; Kinney et al., 1998; Mulholland et al., 1998; Nikiforov et al., 1998). The four methods were (1) spatial averaging, (2) nearest neighbor, (3) inverse distance weighting, and (4) kriging.

All four methods are *weighted average* methods, and they all have the same basic mathematical formulation. That is, we want to compute the air pollution concentration, z at an unsampled point, x_0, given a set of neighboring sampled values z_i, sampled at locations denoted by x_i. The interpolating relationship is

$$z(x_0) = \sum_{i=1}^{n} \lambda_i \cdot z(x_i)0 \text{ and } \sum_{i=1}^{n} \lambda_i = 1 \tag{1}$$

where λ_i represents the weights assigned to each of the neighboring values, and the sum of the weights is one. For these approaches, interpolation involves: (a) defining the search area or neighborhood around the point to be predicted; (b) locating the observed data points within this neighborhood; and (c) assigning appropriate weights to each of the observed data points.

The four interpolation methods differ only in their choice of sample weights. In brief, with spatial averaging, we selected all sampled values within a fixed distance from the point of interest and assigned the same fractional weight, based on the number of monitors, to each of them. With the nearest-neighbor method, we chose only the single sampled value that was closest to the point of interest and assigned it a weight of 1. With inverse distance weighting, we assigned samples that were closer to the point of interest of correspondingly larger weights; and with kriging, we assigned weights based on the spatial autocorrelation statistics of the sampled data set.

(1) Spatial averaging: Several studies (Chestnut et al., 1991; Kinney et al., 1998) have used what we refer to as the spatial averaging method. Schwartz (1989) also used spatial averaging to study air pollution as a risk factor for decrements in children's lung function. Even though we believe that using 10 miles as the neighborhood definition in Schwartz's study was subjective and not well justified, we replicated Schwartz's (1989) approach so that we could directly compare our results to his work. Therefore, for each county with an NHANES-III child, we selected all O_3 and PM_{10} monitors within 10 miles of each BG centroid, calculated the average of the values from the selected monitors and assigned that value to the appropriate BG. If no monitor was found within 10 miles of a particular BG, then that BG was assigned a missing value for the specific pollutant.

(2) Nearest neighbor: Each BG was assigned the air concentration level of the monitor nearest to its centroid regardless of how far away the monitor was located. This eliminated the search radius as a parameter in the interpolation (Detels et al., 1987; Schwartz and Zeger, 1990; Stern et al., 1994; Kunzli et al., 1997; Vedal et al., 1998).

(3) Inverse distance weighting (IDW): Interpolation weights in IDW are computed as a function of the distance between observed sample sites and the site at which the prediction has to be made (Gunnink and Burrough, 1996). As observed values that were closer to the point of interest were more heavily weighted, a larger search window could be used that still preserved some of the local variations in pollutant levels. Data from all monitors within a 250 km search window were included in the IDW interpolation for both PM_{10} and O_3. The 250 km radius was chosen to ensure that all block groups in the counties of interest had an interpolated value for both O_3 and PM_{10}. We have tested other smaller windows (60 and 100 km), but through cross-validation, we found that the accuracy of interpolation was not too sensitive to the window size. However, choosing smaller windows will leave some of the block groups of interest without an estimated value. Similar to Abbey et al. (1991), Kinney et al. (1998), and Kunzli et al. (1997), we used $\lambda_i = 1/d_i$ as the weight, where d_i is the distance between sets of data points and the point to be predicted (Burrough and McDonnell, 1998).

(4) Kriging: The spatial variation of air pollution measurements is complex but is not generally unstructured. It is almost always spatially dependent on some scale, with this dependence referred to as spatial autocorrelation (Griffith, 1988; Bailey and Gatrell, 1995). This structure may then be overlain by more or less random local variation. The whole can be described by a variogram that summarizes the variation.

A variogram expresses the degree of similarity between two observations separated by a given distance (the lag). An empirical variogram can be computed from sampled data using the following expression:

$$\gamma(h) = \frac{1}{2M(h)} \sum_{i=1}^{M(h)} \{z(x_i) - z(x_i + h)\}^2 \qquad (2)$$

where $\gamma(h)$ is the estimated semivariance at a separation distance, or lag h, $z(x_i)$ and $z(x_i + h)$ are the observed values at x_i and $x_i + h$ separated by h, of which there are $M(h)$ pairs (Oliver, 1996). In most cases, the semivariance increases as the distance separating pairs of points (i.e., the lag) increases, indicating that points close together tend to have more similar values than those far apart.

The basic idea behind a kriged interpolation is to use the variogram to compute weights λ_i, which minimize the variance in the estimated value. The first step in this process is to fit a function to the empirical variogram such that semivariances can be computed at all separation distances, or lags. For this study, we used spherical variograms for all of the models (Burrough and McDonnell, 1998). We chose the spherical specification because it is one of the most commonly chosen forms, and the purpose of our study is to compare commonly used methods. Once the model variogram is fit to the empirical data, it is used to compute the weights λ_i such that the estimation variance is less than the variance for any other linear combination of the observed values (Isaaks and Srinivasta, 1989).

Computing the empirical variogram requires that one first selects regions over which the sampled processes can be reasonably assumed to be homogeneous, with a constant mean and variance. For PM_{10}, we followed the precedent of EPA (1996b), in which the country was divided into seven aerosol regions, each with similar PM_{10} characteristics. Owing to the location of the monitors relative to our counties of interest and the boundaries of the seven regions, we made minor changes to the boundaries of two regions and ended up with five PM_{10} regions (Northwest, Southern California, Southeast, Industrial Midwest, and Northeast) covering most of the NHANES counties (see Figure 1). We computed separate variograms for annual average and quarterly maximum PM_{10} for each of our five regions. As indicated in Figure 1, we did not specify a Southwest region. Although this area had several NHANES-III counties of interest, there were too few PM_{10} monitors, as indicated in Figure 2, to support kriging.

Our development of variograms for O_3 kriging differed from the approach we used for PM_{10}. We examined the spatial statistics for each of the counties of interest and looked for evidence of spatial continuity, or the presence of spatial autocorrelation patterns. Unfortunately, all regions of

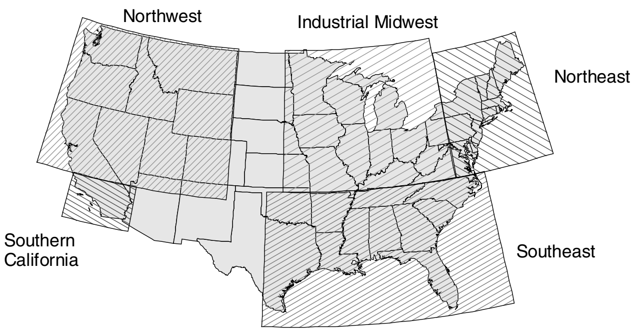

Figure 1. The modified five aerosol regions.

interest, except Southern California, did not exhibit a spatial autocorrelation pattern of O_3. Therefore, we could not justify dividing the country into O_3 regions analogous to the PM_{10} regions.

Several different forms of kriging are commonly used. With simple kriging one assumes that the mean value is known, while with ordinary kriging the mean value is determined during the interpolation. For nonstationary variation, where there is drift (or trend) in the data, universal kriging or kriging with intrinsic random functions should be used (Oliver, 1996; Mulholland et al., 1998). We adopted ordinary kriging for all the interpolations for O_3 and PM_{10}.

Figure 2. Distributions of ozone and PM_{10} monitors.

In the current study, we observed situations in which a statistically significant drift could be detected in the sampled data; however, the sources of the drift were not evident. As such, we believed that it was better to apply the more conservative ordinary kriging approach, rather than to risk extrapolating the data drift to areas in which no drift exists.

Results

Of the 830 ozone monitors in the US and its territories in 1990, 739 met our data validity criteria. Of the 1473 PM_{10} monitors in the US and its territories in 1990, 768 met our validity criteria. Several PM_{10} monitors were located at the same site. We computed the mean values of the colocated monitors, resulting in 732 distinct monitoring locations. Two O_3 monitors and two PM_{10} monitors were in the correct state, but actually several counties away from the correct county indicated by the monitor's ID. By adjusting either the reported latitude or longitude, but not both for each monitor, assuming that errors were incurred on entering the coordinate information, we were able to place all four monitors into the correct counties, but not necessarily the exact locations, based on their identification numbers. The locations of the 739 O_3 monitors and 732 PM_{10} monitors are shown in Figure 2. Although the O_3 and PM_{10} monitors were separate and distinct, both types of monitors appear to be colocated in some instances.

Table 1 presents summary descriptive statistics for the 1990 data of the 739 O_3 and 732 PM_{10} monitors. Most of the extreme O_3 values (highest and lowest) were recorded in California. Maximum PM_{10} levels were observed in the eastern part of the country primarily in the summer, and in the western part of the country primarily in the fall and winter. This is similar to the patterns noted in EPA (1996b).

Summer Daytime Ozone: Interpolated Values
Only the variograms computed for California showed characteristics of spatial continuity for summer daytime ozone. Variograms computed for the other counties of interest did not display consistent trends with separation distances, and model variograms could not be fitted to the data. These findings imply that simple kriging methods may not be appropriate for those places. Therefore, kriging was conducted for O_3 in California only. Table 1 also presents summary statistics based on the results of the four interpolation methods for all the counties of interest. Of the 56,721 BGs within the counties of interest, we could not conduct spatial averaging in 15% of these BGs because they did not have at least one O_3 monitor within 10 miles of their centroid. Also, the summary statistics for kriging were based only on the counties in California where variograms could be constructed.

The four methods produced fairly different estimates of air pollution concentrations between sample locations. Mapped examples of the interpolations of O_3 for Los Angeles county are shown in Figure 3. IDW predicted a fairly uniform concentration across the entire county, with a slight decrease toward the coast. Kriged values increased much more strongly from the coast to the inland. Nearest-neighbor interpolations were a patchwork of areas of constant values, and spatial averaging was notable in the number of locations in the county for which interpolated values were not computed. Note that air pollution concentrations are displayed in these maps only at the resolution of BGs, so isopleths reflect BG boundaries and are not smooth.

Table 1. Summary statistics comparing the monitor values and the results of the four different interpolation methods for summer daytime ozone, annual average PM_{10} (Ann. avg.), and maximum quarterly average PM_{10} (Max. Q.) at the block group (BG) level.

	Statistics	25th Percentile	Median	Mean	75th Percentile	No. BG not applicable
Summer daytime ozone (in ppb)	Monitor ($n=739$)	40	45.2	45.9	51.2	—
	Kriging	36.9	47.5	49.4	61.2	42,298
	IDW	39.9	43	45	49.8	0
	Nearest neighbor	35.9	40.9	42.8	47.5	0
	Spatial averaging	36.7	41.4	43.4	48	10,667
Ann. avg PM10 (μg/m^3)	Monitor ($n=732$)	23	27.7	29	33.6	—
	Kriging	27.1	29.9	33.2	37	2,597
	IDW	27.8	30.5	33.5	37.5	0
	Nearest neighbor	26	30.4	33.7	39	0
	Spatial averaging	27.7	32.4	34.8	39.4	17,861
Max. Q. avg PM 10 (μg/m^3)	Monitor ($n=732$)	29	35.6	37.8	43.7	—
	Kriging	34.8	38.2	42.6	46.9	2,584
	IDW	34.1	38.9	42.9	51.4	0
	Nearest neighbor	33.5	38.1	43.3	52	0
	Spatial averaging	34.5	39.8	44.3	52.5	17,861

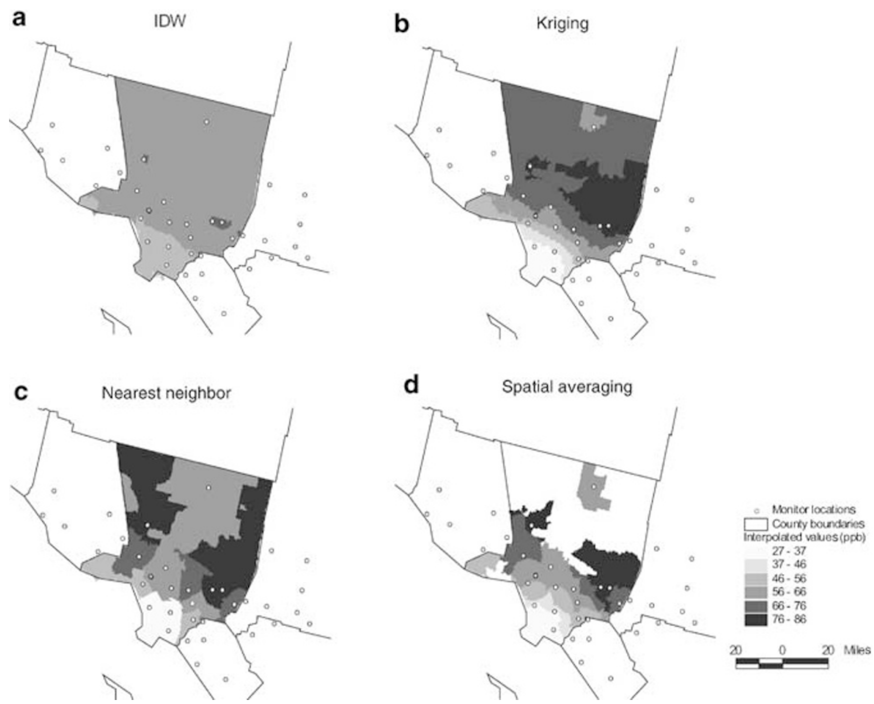

Figure 3. Estimated of four methods for summer daytime O_3 for Los Angeles.

As kriging was limited to California, we could only compare all four methods for counties in that state. The method-specific mean summer daytime O_3 values for all 13,326 BGs in California were: 53.3 ppb (IDW), 49.5 ppb (spatial averaging), 49.4 ppb (kriging), and 49.2 ppb (nearest neighbor). The mean value for IDW was significantly larger than the other three means (based upon Duncan and Tukey significance of means tests, $P < 0.05$), while the mean values for the other three methods were not significantly different from each other. These comparison tests were not strictly valid statistically because significant spatial autocorrelation existed in each sample, but they provided a useful means of interpreting the trends in the data.

PM_{10} Annual Average and Maximum Quarterly Average: Interpolated Values

Model variograms were fit to empirical data in each of the five PM_{10} aerosol regions (Figure 4). The strongest signals of spatial continuity, as shown in Figure 4, were found in the Southeast, Northwest, and Southern California regions. Evidence of spatial continuity in the Northeast and Industrial Midwest was much less definitive. In particular, semivariances computed for the Industrial Midwest remained less than the total sample variance until separation distances of approximately 700 km. This separation distance indicates a lack of spatial continuity in that region.

Table 1 also provides summary statistics for annual average PM_{10} based on the results of the four different interpolation methods for all the counties of interest. We could not conduct spatial averaging in 26% of the 56,721 BGs, because they did not have at least one PM_{10} monitor within 10 miles of their centroid. Kriging was not performed in 4.6% of the total BGs where they were in the four counties in southwestern US not included in the five PM_{10} regions (see the kriging discussion in the last section).

The four interpolation methods were also performed, where appropriate, for the maximum quarterly PM_{10} averages. The last part in Table 1 presents summary statistics based on the results of the four different interpolation methods for all the counties of interest. Even though the absolute levels of the estimates based upon the maximum quarterly averages were higher than those based upon the annual averages, they correlated very well spatially so that the relative levels for PM_{10} maximum quarterly averages were similar to those using annual average PM_{10} metric.

Interpolated Method Comparisons

To compare interpolation results, we computed correlation coefficients for pairs of methods across counties. The coefficients between predictions from pairs of interpolation methods were generally high (from 0.80 to 0.97). The lowest correlation (0.80) was observed in comparing O_3 estimates between IDW and kriging and between IDW and the nearest neighbor. Correlation coefficients for most pairs of methods exceeded 0.9.

We also examined the range of estimates across all of the interpolations within each BG by computing the difference between the highest and lowest interpolated values. We were

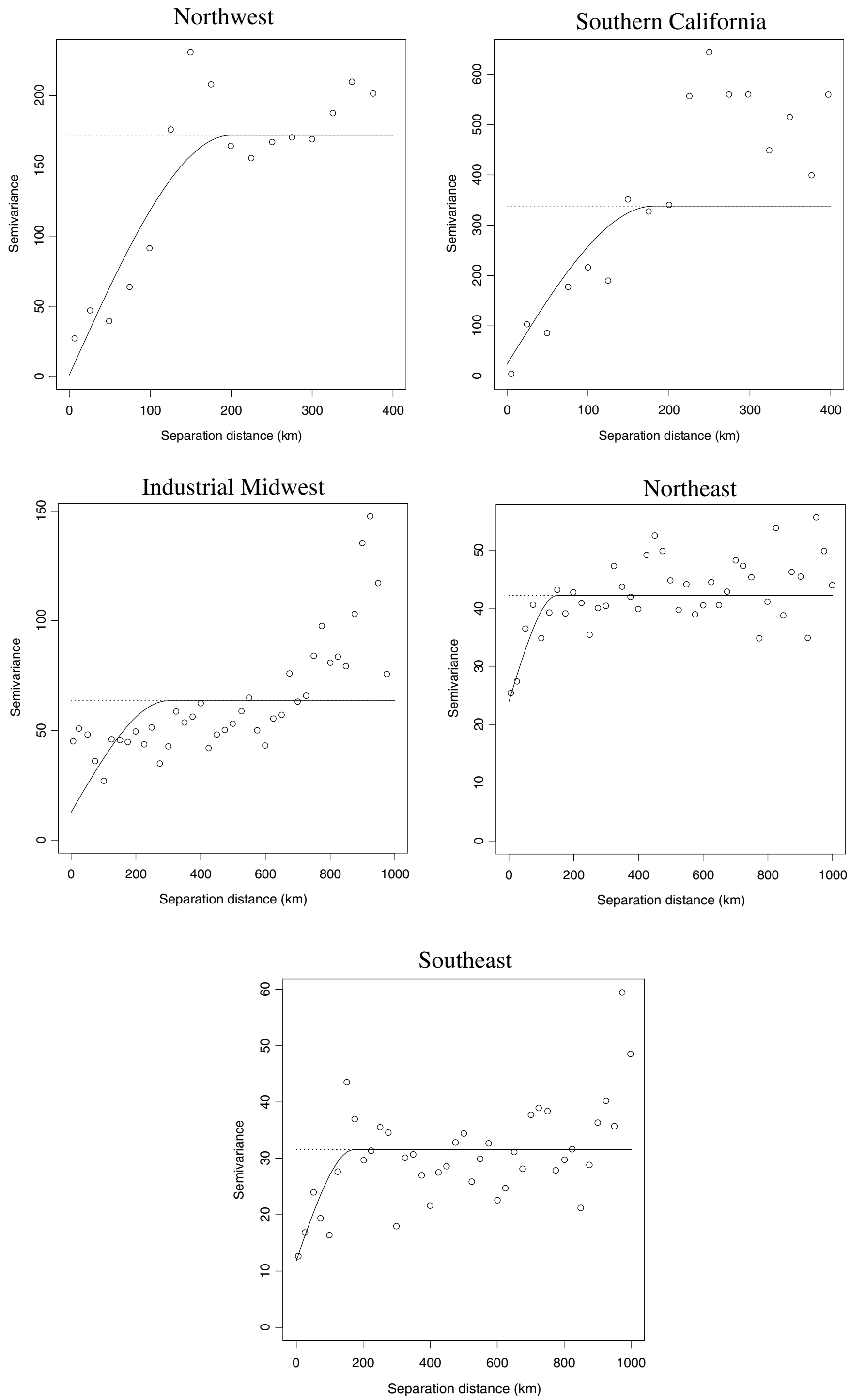

Figure 4. Empirical variograms for the five modified aerosol regions.

able to perform all four methods only in California (11,934 BGs). The range values for O_3 were distributed approximately normally, implying that differences in interpolation estimates were moderately high for a substantial percentage of BGs. The distributions of the ranges for the annual and maximum quarterly averages of PM_{10} were similar to each other in that most BGs have ranges below 15 $\mu g/m^3$, with the majority of values less than 7 $\mu g/m^3$. As less than 25% of all the BGs of interest had interpolated O_3 values from all four methods, we relaxed the comparison criterion for O_3 slightly by including BGs, provided two or more interpolation methods were used. This resulted in a positively skewed distribution of the range similar to that observed for PM_{10} annual and quarterly maximum averages. In this case, more than 50% of BGs had a difference of less than 4 ppb O_3 between the highest and lowest interpolation estimates.

To further evaluate the differences in results from different interpolation methods, we selected a small fraction of the BGs that actually contained an NHANES subject. We repeated our analysis of range using these BGs. We created frequency distributions of block groups by the range distributions of O_3, and annual and maximum quarterly averages of PM_{10}. For all three metrics, the distributions were positively skewed, indicating that differences in the estimates derived from different interpolation methods were relatively small in most of the BGs covered by the NHANES. χ^2 statistics indicate that these new distributions were not significantly different ($P < 0.05$) from the range distributions for all 56,721 BGs. We also computed new range distributions for California, as it was the only state having estimates of O_3 from all four methods. The distribution of the range of O_3 for the selected BGs in California was also not significantly different from the range distributions of all counties in California having all four methods applied (χ^2, $P < 0.05$). To assess the geographical distribution of differences between methods, we computed the maximum observed range for different annual PM_{10} estimates within each of the 82 counties. More than half of the 82 counties had a maximum range for PM_{10} estimates less than 10 μg/m^3, a relatively small difference.

So far, our methods adopted to evaluate results from different interpolation methods are quite lenient. Therefore, "leave-one-out" cross-validations were also performed to provide a more quantitative comparison of the interpolation methods. By this method, samples were dropped sequentially and the remaining samples were used to interpolate the value at the location of the dropped sample. The interpolated values were plotted and regressed against the actual measured values, and the R^2 value and coefficients from the resulting regression were examined. Cross-validations for different interpolation methods (kriging, nearest neighbor, and spatial average) were performed only in Southern California where results from all three methods were available. Table 2 reports some of the results. All methods were minimally biased, with regression coefficients close to 1. The R^2 value for kriging was the lowest among the three methods for this limited subset of the data. All three coefficients are significant statistically with $P<0.05$. The coefficients for the nearest and spatial averaging methods are not significantly different from each other using a 95% confidence interval, but they are significantly different both from one and the coefficient for the kriging method (although not significantly different using 90% confidence intervals). However, the regression coefficients indicate that kriging produces slightly less biased results than the other two methods, despite the fact that overall the three methods perform reasonably well in Southern California.

Table 2. Results from cross-validation based upon ozone data in Southern California.

	Kriging	Nearest neighbor	Spatial averaging
R^2	0.72	0.80	0.88
Regression coefficient	1.058	0.751	0.828
n	55	55	33

Discussion

In this initial phase of this project, we have compiled ambient air pollution data sets and implemented interpolation methods to estimate air pollution concentrations across the country at NHANES-III locations. We have also performed some preliminary comparisons between methods. Based on this work, we have identified several issues regarding the implementation of the different interpolation methods, and we have also drawn some conclusions regarding the differences between the estimates produced by the different interpolation methods.

Implementation of Interpolation Methods

Two points regarding the implementation of these different interpolation methods merit further discussion. First, the choice of search radius can be important for interpolation by IDW and by spatial averaging. A large search radius incorporates monitor values that are very distant from the point of interest and can lead to smoother interpolated surfaces. However, a large search radius may also use a sample that is so distant that its relationship to the air pollution levels at the location of interest is tenuous. Conversely, a small search radius can fail to generate interpolated values for a large subset of locations, but generally provides more conservative estimates of concentration values. Little guidance exists for establishing an optimal value for this search radius, so our methods for choosing search radii were somewhat ad hoc. For IDW, we selected a search radius that ensured that the interpolations produced results for all of the counties of interest in the study. For spatial averaging, we followed the precedent of Schwartz (1989) and selected a 10-mile search radius. These decisions were arbitrary, and exploring the effects of different search radii was beyond the scope of the present study. Furthermore, in our comparisons between interpolation methods, we found only small differences for a majority of the study areas, and for these locations, the search radii probably did not exert a strong influence on the results.

Second, we achieved limited success in fitting model variograms to air-monitoring data. Other researchers have reported similar difficulties in computing variograms from air pollution concentrations (Lefohn et al., 1987, 1988; Casado et al., 1994; Liu and Rossini, 1996). For O_3, we were able to fit variograms for counties in California, but in general, it appeared that the density of O_3 air monitors outside of California was too sparse to accurately estimate variograms for other counties of interest. The success of developing the variograms in California can be explained by two factors. First, the density of O_3 monitors in California was much greater than anywhere else in the country (Figure 2). Second, the topography of California is unique, as the urban areas at the coast are typically bordered by mountains slightly inland. The mountains present a barrier to the movement of O_3, so high O_3 concentrations persist inland at the base of the mountain chains, and low O_3 concentrations are found near the coast. Thus, the spatial patterns in the distribution of O_3 in California tend to be relatively static. Outside of California, we found no compelling reason to force variogram models where the empirical data did not show characteristics of spatial continuity and did not appear to support interpolation by kriging. In other words, applications of kriging techniques to interpolate air pollutant levels have to be justified by the presence of spatial relationship captured by the empirical data. Forcing the data into a kriging framework without empirical evidence may commit model specification errors.

For PM_{10}, we observed reasonable evidence of spatial autocorrelation for three of our five regions (Northwest, LA Basin, and Southeast), but did not find strong signals in the Northeast and Industrial Midwest. Again, the different observed behaviors probably are caused by different monitor distributions and densities. A visual inspection of monitor locations suggested that the monitor network was more highly clustered in the Northeast and Industrial Midwest, and thus, poorly suited for assessing the degree of spatial variability in the monitored data. As variograms computed for PM_{10} in some of the regions appeared robust, we felt that it was likely that some spatial continuity existed in the Industrial Midwest and Northeast. As such, we specified model variograms for these two regions that best represented the regional trends suggested by the semivariance values and that agreed with the patterns established in the other three regions. This approach may not be the best to deal with ill-behaved variograms, but for demonstration and illustrative purposes, we believe the method still produced reasonable results.

Overall, we believe that the air-monitoring network is poorly suited for estimating spatial autocorrelation of aggregated air pollution measurements. The monitors have historically been placed in locations where exceedences of air quality standards are expected, and therefore tend to be clustered in urban areas. Rural areas have relatively few monitors, and so the distribution of monitors is biased toward areas of higher pollution concentrations. To compute accurately the spatial statistics of air pollution distributions, more evenly distributed monitors are required. This will be an important issue to consider when we evaluate the linked NHANES-III and interpolated air concentration data, since almost half of the children live in more urbanized areas and half in more rural areas.

Comparisons of Interpolation Methods

In general, correlation coefficients and our analysis of the range of interpolation estimates in each BG suggest that different interpolation methods did not generate dramatically different results in most of the BGs and for most of the counties visited by NHANES-III. BGs with larger ranges of interpolated values were mostly found in California within several major metropolitan areas, and in several cities along the Northeast Corridor. Note that California is one of the largest states in terms of population and 25% of the NHANES-III children live in California.

To better understand the relationships between different interpolation methods, we compared pairs of methods in scatter plots. Examples of these plots are shown in Figure 5, where kriged estimates are plotted against IDW estimates of annual average PM_{10}. In Figure 5a, interpolated values for all of the BGs in the study are shown. The two estimates are strongly correlated, but internal structures are discernible within the cloud of points. The origin of these structures can be better understood when individual scatter plots are drawn for each county. In Figures 5b and c, we present two examples of these county-specific plots (Figure 5b: Los Angeles County, CA, USA and Figure 5c: Duval County, FL, USA, respectively). These two counties were selected to

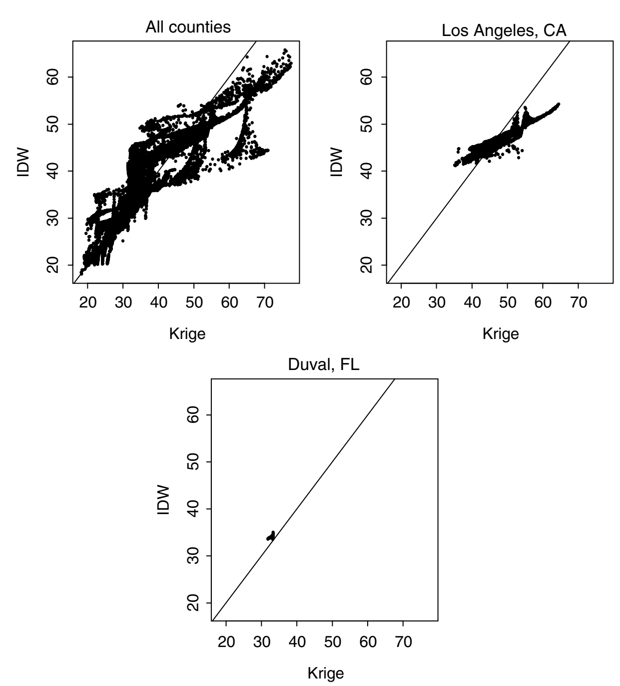

Figure 5. Scatter plot comparison between kriged and IDW estimates of PM_{10} for (a) all counties, (b) Los Angeles, CA, USA, and (c) Duval, FL, USA.

show examples of the two distinct types of relationships that we observed between kriged and IDW estimates.

In Figure 5b, we observed a point pattern that occurred frequently when comparing IDW and kriging. The plotted points fell along a shallowly sloped line with several sharp deviations. These two interpolation methods presented very different pictures of the distribution of air pollution concentrations across the county. IDW severely damped variations in value, so that most estimated values tended toward a narrow range of values, while kriging estimates spanned a wider range. Each deviation from the diagonal line was a manifestation of the IDW-generated egg-shaped region of sharply higher or lower concentrations that surrounded monitors within the county (see Figure 3).

More importantly, in the plot for the second county (Figure 5c: Duval, FL, USA), both IDW and kriged estimate values fell within a narrow range relative to the range of possible values observed across the entire country. In counties with very few or no closely located monitors, we would expect a narrow range of estimates across the county, as the variability of air pollution at that location is poorly quantified, and interpolation methods cannot impute variability without more monitored values. We observed a similar narrow distribution of interpolation estimates for more than half of the NHANES-III counties.

The contrast between Figures 5b and c suggests that the importance of the interpolation method depends strongly upon the nature of the local monitor network. In regions of the country in which monitors are sparse, all interpolation methods converge to a similar, narrow range of predictions. When the nearest monitor for a given BG is relatively far away, different interpolation methods will use the same monitor for estimation, and therefore the results will be similar or identical. In areas of the country where monitors are dense, different interpolation methods can generate vastly different predictions of the spatial distribution of pollutants because monitors being used will be weighted differently by different methods. This conclusion is counterintuitive, but is well-supported by the results of our study.

Owing to the exploratory nature of this project, many of our choices of parameters and interpolation methods seemed arbitrarily constrained by the types of data that were available and by the nationwide scope of the study area. Nevertheless, our findings suggest that for a majority of NHANES study areas where monitors are sparse, the simplest interpolation method, nearest neighbor, may perform as well as any of the other methods. In California and other locations, where monitor density is relatively high, the selection of interpolation method must be approached more carefully. The fundamental issue is whether the spatial sampling intensity is adequate to capture the spatial structure, if any, possessed in the data. If the sampling is too sparse, we may not have sufficient information to describe the spatial structure if it does exist, and therefore different methods may not perform dramatically different. But if we have relatively intense samples and the spatial structure may be captured, then the selection of the method should be made with care so that the interpolation method can fully utilize the spatial information captured by the samples. Surely, how dense the sample should be is phenomenon- and geography-dependent.

Kriging is generally thought to produce the most realistic estimates of values between monitors, and we have successfully applied the method for O_3 in California and for three out of the five regions defined for PM_{10}. Further study will be required to refine our estimates of the air pollution concentration and the variances in the estimates. California, in particular, merits special attention because many (approximately 25%) NHANES-III subjects resided there and because gradients in air pollution are relatively strong between the coastal and inland areas. Therefore, different interpolation methods will likely change pollutant concentration estimates significantly for a large number of subjects.

In the very near future, we will attempt to find associations between air pollution estimates and respiratory health outcomes using the nationwide NHANES-III sample. However, the wide range of monitor densities across the country suggests that the accuracy of the interpolation estimates will vary greatly. This difference in accuracy may influence our abilities to draw strong conclusions regarding the effects of air pollution on respiratory health.

Acknowledgments

We acknowledge the assistance of Milton Siegel of U.S. EPA National Center for Environmental Assessment — Washington Division, in analyzing the data, and Stephen Andrews of Indus Corporation in retrieving the AIRS data.

References

Abbey D., Mills P., Petersen F., and Beeson W.I. Long-term ambient concentrations of total suspended particulates and oxidants as related to incidence of chronic disease in California seventh-day adventists. *Env Health Perspect* 1991: 94: 43–50.

Bailey T.C., and Gatrell A.C. *Interactive Spatial Data Analysis*. Longman: London, 1995.

Brown P.J., Le N.D., and Zidek J.V. Multivariate spatial interoplation and exposure to air pollutants. *Can J Stat* 1994: 22: 489–509.

Brunekreef B., Dockery D., and Krzyzanowski M. Epidemiologic studies on short-term effects of low levels of major ambient air pollution components. *Env Health Perspect* 1995: 103 (Suppl. 2): 3–13.

Burrough P.A., and McDonnell R.A. *Principles of Geographical Information Systems*. Oxford University Press, Oxford, 1998.

Casado L.S., Rouhani S., Cardelino C.A., and Ferrier A.J. Geostatistical analysis and visualization of hourly ozone data. *Atmos Environ* 1994: 28: 2105–2118.

Chestnut L., Schwartz J., Savitz D., and Burchfiel C. Pulmonary function and ambient particulate matter: epidemiological evidence from NHANES-I. *Arch Environ Health* 1991: 46(3): 135–144.

Detels R., Tashkin D., Sayre J., Rokaw S., Coulson A., Massey F., and Wegman D. The UCLA population studies of chronic obstructive respiratory disease, 9: lung function changes associated with chronic exposure to photochemical oxidants, a cohort study among never-smokers. *CHEST* 1987: 92(4): 594–603.

Duddek C., Le N.D., Zidek J.V., and Burnett R.T. Multivariate imputation in cross-sectional analysis of health effects associated with air pollution. *J Environ Ecol Stat* 1995: 2: 191–212.

Griffith D.A. *Advanced Spatial Statistics*. Kluwer Academic Publishers: Dordrecht, 1988.

Gunnink J.L., and Burrough P.A. Interactive spatial analysis of soil attribute patterns using exploratory data analysis (EDA) and GIS. In: Masser I., Salge F. (Eds). *Spatial Analytical Perspectives on GIS*. Taylor & Francis, New York, 1996, pp. 87–99.

Isaaks E.H., and Srinivasta R.M. *Applied Geostatistics*. Oxford University Press: Oxford, 1989.

Kinney P.L., Aggarwal M., Nikiforov S.V., and Nadas A. Methods development for epidemiologic investigations of the health effects of prolonged ozone exposure. Part III: an approach to retrospective estimation of lifetime ozone exposure using a questionnaire and ambient monitoring data (U.S. sites). *Health Effects Inst Res Rep* 1998: 81: 79–107.

Kunzli N., Lurman F., Segal M., Ngo L., Balmes J., and Tager I. Association between lifetime ambient ozone exposure and pulmonary function in college freshman—results of a pilot study. *Environ Res* 1997: 72: 8–23.

Lefohn A.S., Knudsen H.P., Logan J.A., Simpson J., and Bhumralkar C. An evaluation of the kriging method to predict 7-h seasonal mean ozone concentrations for estimating crop losses. *JAPCA* 1987: 37: 595–602.

Lefohn A.S., Knudsen H.P., and McEvoy L.R. The use of kriging to estimate monthly ozone exposure parameters for the Southeastern United States. *Environ Pollut* 1988: 53: 27–42.

Linn W., Shamoo D., Anderson K., Peng R., Avol E., Hackney J., and Gong H. Short-term air pollution exposures and responses in los angeles area school children. *J Exp Anal Environ Epi* 1996: 6(4): 449–472.

Liu L.-J.S., and Rossini A.J. Use of kriging models to predict 12-hour mean ozone concentrations in metropolitan Toronto — a pilot study. *Environ Int* 1996: 22: 677–692.

McConnell R., Berhane K., Gilliland F., London S., Vora H., Avol E., Gauderman J., Margolis H., Lurmann F., Thomas D., and Peters J. Air Pollution and Bronchitis Symptoms in Southern California Children with Asthma. *Environ Health Perspect* 1999: 107(9): 757–760.

Mulholland J.A., Butler A.J., Wilkinson J.G., Russell A.G., and Tolbert P.E. Temporal and spatial distributions of ozone in Atlanta: regulatory and epidemiologic implications. *J Air Waste Mngt Assoc* 1998: 48: 418–426.

Nikiforov S.V., Aggarwal M., Nadas A., and Kinney P.L. Methods for spatial interpolation of long-term ozone concentrations. *J Exposure Anal Environ Epidemiol* 1998: 8: 465–481.

Norris G., Youngpong S., Koenig J., Larson T., Sheppard L., and Stout J. An association between fine particles and asthma emergency department visits for children in Seattle. *Environ Health Perspect* 1999: 107(6): 489–493.

Oliver M.A. Geostatistics, rare disease and the environment. In: Masser I., Salge F. (Eds). *Spatial Analytical Perspectives on GIS*. Taylor & Francis, New York, 1996, pp. 67–85.

Peters J., Avol E., Navidi W., London S., Gauderman W., Lurmann F., Linn W., Margolis H., Rappaport E., Gong H., and Thomas D. A study of twelve southern california communities with differing levels and types of air pollution. *Am J Respir Crit Care Med* 1999: 159: 760–767.

Pope C.A., Bates D., and Raizenne M. Health effects of particulate air pollution: time for reassessment? *Environ Health Perspect* 1995a: 103(5): 472–480.

Pope C.A., Dockery D., and Schwartz J. Review of epidemiological evidence of health effects of particulate air pollution. *Inhal Toxicol* 1995b: 7: 1–18.

Schwartz J. Lung function and chronic exposure to air pollution: a cross-sectional analysis of NHANES II. *Environ Res* 1989: 50: 309–321.

Schwartz J. Air pollution and blood markers of cardiovascular risk. *Environ Health Perspect* 2001: 109(Suppl. 3): 405–409.

Schwartz J., Dockery D., Neas L., Wypij D., Ware J., Spengler J., Koutrakis P., Speizer F., and Ferris B. Acute effects of summer air pollution on respiratory symptom reporting in children. *Am J Respir Crit Care Med* 1994: 150: 1234–1242.

Schwartz J., and Zeger S. Passive smoking, air pollution, and acute respiratory symptoms in a diary study of student nurses. *Am Rev Respir Dis* 1990: 141: 62–67.

Stern B., Jones L., Raizenne M., Burnett R., Meager J.C., and Franklin C.A. Respiratory health effects associated with ambient sulfates and ozone in two rural canadian communities. *Environ Res* 1989: 49: 20–39.

Stern B., Raizenne M., Burnett R., Jones L., Kearney J., and Franklin C.A. Air pollution and childhood respiratory health: exposure to sulfate and ozone in 10 Canadian rural communities. *Environ Res* 1994: 66: 125–142.

Thurston G., Ito K., Kinney P., and Lippmann M. A multi-year study of air pollution and respiratory hospital admissions in three new state metropolitan areas: results for 1988 and 1989 summers. *J Exp Anal Environ Epi* 1992: 2(4): 429–450.

US Environmental Protection Agency. *Air Quality Criteria for Ozone and Related Photochemical Oxidants*. Vol. I of III Office of Research and Development, Washington, DC, July, 1996a, EPA/600/P-93/004aF.

US Environmental Protection Agency. *Air Quality Criteria for Particulate Matter*. Vol. I of III Office of Research and Development, Washington, DC, April, 1996b, EPA/600/P-95/001aF.

Vedal S., Petkau J., White R., and Blair J. Acute effects of ambient inhalable particles in asthmatic and nonasthmatic children. *Am J Respir Crit Care Med* 1998: 157: 1034–1043.