United States Environmental Protection Agency

Office of Research and Development

EPA 600/R-01/028
June 2001
www.epa.gov/ncea

Spatial Variation in Ozone Concentrations in Phoenix, AZ for 1997

Introduction

Statistical analyses of the human health effects of airborne pollutants based on aggregate population time-series data have often relied on ambient concentrations of pollutants measured at one or more central sites in a given metropolitan area. In the particular case of ground-level ozone (O_3) pollution, central-site monitoring has been justified as a regional measure of exposure partly on grounds that correlations between concentrations at neighboring sites measured over time are usually high (U.S. EPA, 1996). In analyses where multiple monitoring sites provide ambient O_3 concentrations, a summary measure such as an average has thus often been regarded as adequately characterizing the exposure distribution. Indeed, a number of studies have referred to multiple-site averaging as the method for measuring O_3 exposure (U.S. EPA, 1996).

This report revisits the practice of multiple-site averaging. The results are drawn from an analysis examining simultaneous mapping of population density and ambient O_3 concentrations on the scale of a single metropolitan area. Because of the ready availability of data associated with a related research effort, the city selected for this analysis was Phoenix, AZ. The Phoenix metro area is situated within Maricopa County, the most populous county in Arizona.

Description of Data

Two types of data were retrieved from Internet sites maintained by U.S. government agencies. The first type of data consists of O_3 concentrations reported at seventeen monitoring sites operating in the greater Phoenix area during calendar year 1997, extracted from the U.S. EPA Aerometric Information Retrieval System (AIRS) data base. The latitude and longitude of each monitor were also retrieved from the AIRS data base. Although far from uniformly distributed, the seventeen monitor sites overlay roughly a 35×30-mile rectangle covering the Phoenix area. The AIRS data base provides concentrations recorded at each monitoring site by date and hour. For each monitor reporting data on a given date, the maximum eight-hour moving average was calculated between the hours of 8 am and 8 pm. Any 8 am - 8 pm data block for a monitor was excluded from consideration if it did not contain all twelve hourly observations. To identify potential reporting problems, the daily eight-hour maximum for each monitor was compared to the eight-hour maxima at nearby monitors and extreme differences were investigated through inspection of the hourly records. The most common pattern observed was a precipitous drop of O_3 levels to near zero reported hourly concentrations, persisting for a day or more. These cases (a total of twenty-three monitor-days distributed across six different monitors) were attributed to instrument malfunction and were excluded from the analysis.

In order to characterize the geographic distribution of the local population on a sufficiently detailed scale, total (all-ages) population and population-by-age data were obtained for each census tract in Maricopa County in 1990, using the Decennial Census lookup facility provided on the U.S. Census Bureau website. Although more current population data would have been preferred, 1990 is the most recent year for which complete census tract data (including distribution by age) were publicly available at the time of this analysis. Census tracts can vary considerably in geographic size and some have irregular shapes. To facilitate map placement of the population for each census tract, the geographic centroid (latitude and longitude) of each tract was also obtained using query features provided as part of the Decennial Census lookup. The population numbers and centroid for each census tract were then combined according to the FIPS state-county-tract identifier codes established by the Census Bureau.

Description of Maps and Spatial Smoothing Approach

The initial objective of the project was to develop a visual characterization of pollution monitor placement with respect to local population distribution. Applying the graphics capabilities of SPLUS 2000 (MathSoft, Inc.) to the data, detailed population density maps were produced with O_3 monitor locations superimposed. The two-dimensional LOESS smoothing routine available in SPLUS was then applied to the 8-hour O_3 maxima across monitoring sites (span = .75) to generate concentration isopleths covering much of the densely populated part of Maricopa County. The isopleths are intended to provide a convenient visual summary of the data across the network of monitoring sites. By superimposing these isopleths on the population/monitor maps, composite maps were obtained showing the variation in O_3 concentrations across the populated area. (The spatial smoothing routine was applied to the O_3 concentration data on a scale treating one degree latitude as representing the same distance as one degree longitude. In fact, the distance represented by one degree latitude is roughly 20 percent greater than the distance represented by one degree longitude at this location on the globe, which introduced minor distortions in the smoothing process. However, all maps have been sized so that one inch on the vertical scale represents the same distance in miles as one inch on the horizontal scale, resulting in proper depiction of greater vertical distance per degree than horizontal distance per degree.) Notably, Phoenix has several large population clusters of persons $\geq$ 65 years of age concentrated on the fringes of the metro area. Therefore, maps were also produced reflecting the geographic distribution of this special population.

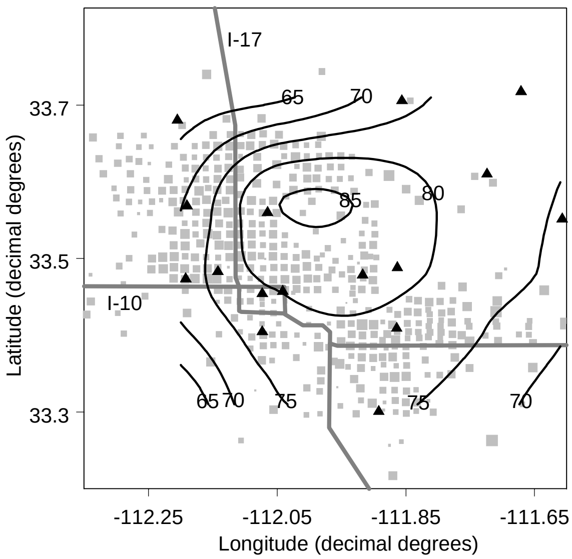
Among the numerous maps generated, significant spatial variation is often apparent in the concentrations reported across monitoring sites. Figures 1 a - d show a sampling of the types of spatial patterns observed. In each map the triangle symbols denote the locations of the monitors reporting on the indicated date. The range of concentrations indicated by the contour lines for each plot is generally conservative due to the wide-span smoothing, i.e., the contour lines generally span a narrower numerical range than do the individual concentrations across monitoring sites. Each square on a map represents a single census tract. The size of each square is proportional to the 1990 all-ages population residing within the corresponding census tract, and the placement of each square is based on the census tract centroid. The map area represents roughly a 43.1-mile square. Selected segments of the Interstate and U.S. highway system are also included for reference.

It might be noted from the maps that several monitors, most notably the four monitors in the extreme east and northeast of the metro area, are at sites with very low nearby population. These are nevertheless useful in the spatial analysis since they act as "high-leverage" data points in the smoothing process. Because the LOESS smoothing procedure available in SPLUS does not extrapolate outside the range of the coordinates, the coverage area for the concentration isopleths is restricted to a rectangle roughly 35 miles from east to west and 30 miles from north to south.

The maps for 6/8/1997 and 7/5/1997 (Figures 1 a and d) indicate a "bullseye" pattern with the highest concentrations observed near the center of the metro area. The map for 6/19/1997 (Figure 1 b) indicates a gradient in O_3 concentrations increasing in a general west-to-east direction, with notable westward intrusions of the concentration zones. The map for 6/28/97 (Figure 1 c) indicates relative spatial uniformity of concentrations.

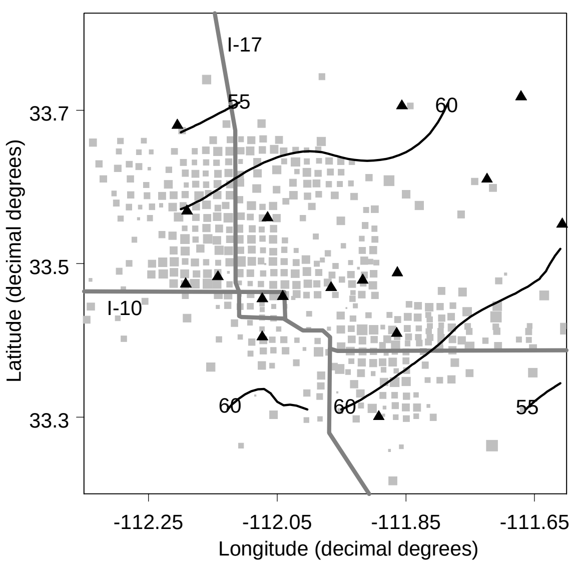
Based on Figures 1 a - d, the monitoring network appears to provide reasonably effective coverage of the all-ages population. A second set of maps for the same four days (Figures 2 a - d), showing the population distribution of individuals $\geq$ 65 years of age, provides something of a contrast to the first set of maps. Given the observed spatial variation in O_3 concentrations on selected days, the monitors do not appear to be ideally situated to allow the assessment of exposure for this at-risk population. Although there may be some discrepancies created by superimposing 1997 monitor sites on 1990 population maps, there is reason to expect that the maps are representative of the situation in 1997. Based on more recent U.S. Census Bureau aggregate-level estimates, population growth in Phoenix proper between 1990 and 1997 was about 20% while in the remainder of Maricopa County population growth was about 34%. Indications are that the areas to the northwest (e.g., Peoria and Surprise City) and to the southeast (e.g., Chandler, Gilbert, and Mesa), which do not fall directly under the monitoring network, experienced this same disproportionate growth. Based on these growth patterns, it is unclear whether the discrepancy between the geographic distribution of persons $\geq$ 65 years of age and that of the general population persists, although it seems unlikely to be substantially changed from 1990 given that suburban Phoenix is well-known for its retirement communities.

As a cautionary note, although the cases presented here need not be regarded as particularly anomalous, neither should they be regarded as representative or typical; they were selected simply as interesting examples of the types of spatial patterns observed.

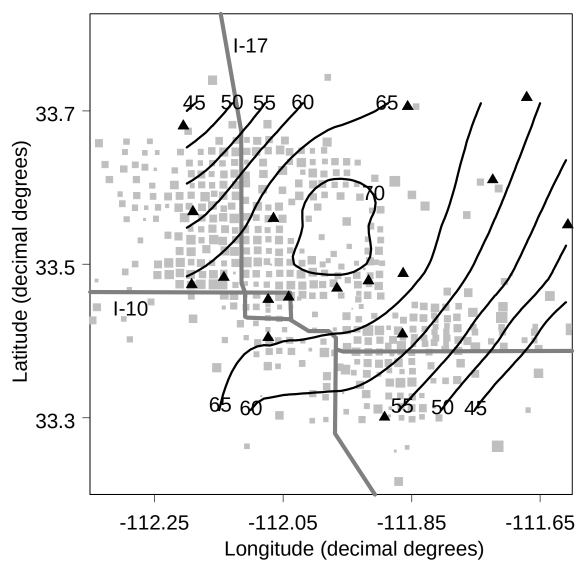


(a) 6/8/97

(b) 6/19/97

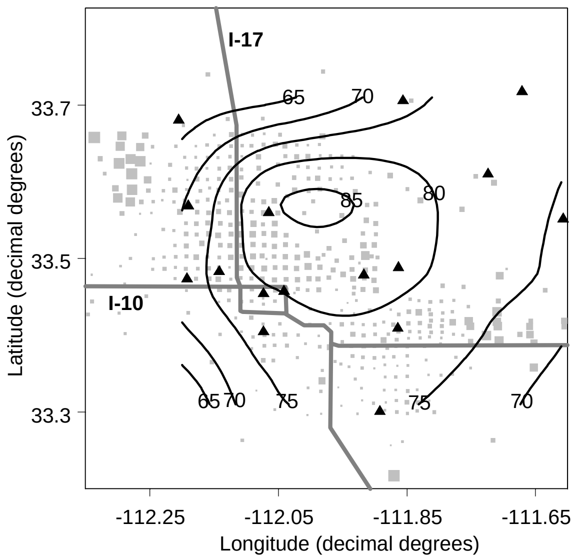


(c) 6/28/97

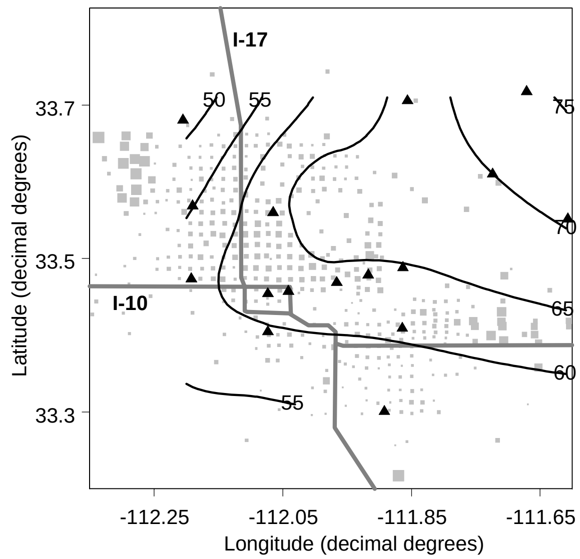


(d) 7/5/97

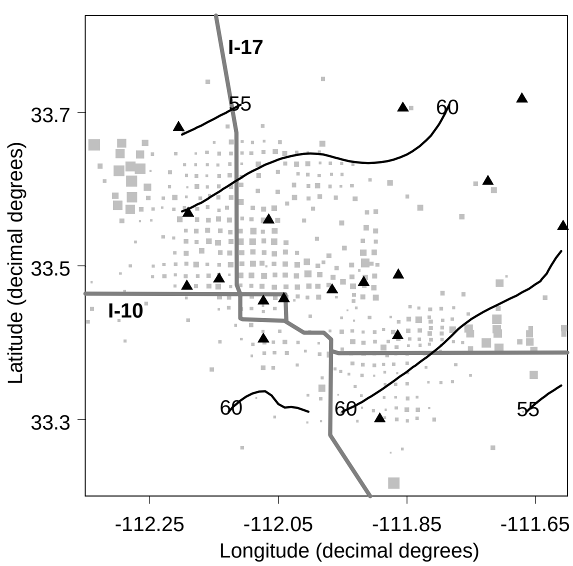
Figures 1 (a) - (d). Eight-Hour Maximum O_3 Concentrations (ppb) over Greater Phoenix, AZ for selected days in 1997. Squares denote census tracts and are proportional in size to 1990 all-ages population. Triangles show locations of O_3 monitors reporting for each day. The maps represent an area approximately 43 miles square.



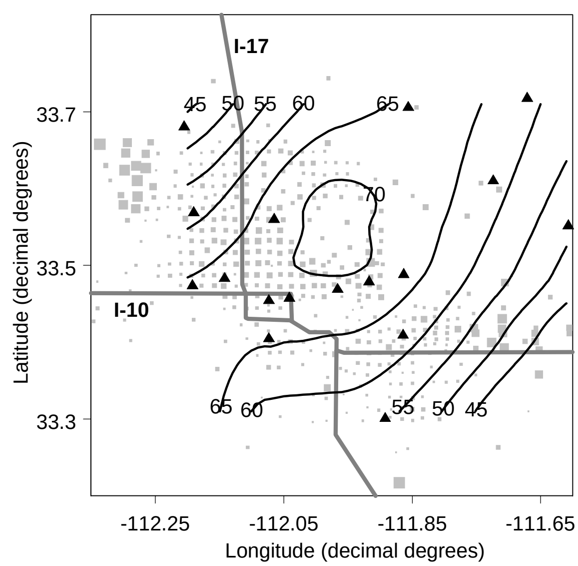
(a) 6/8/97



(b) 6/19/97



(c) 6/28/97



(d) 7/5/97

Figures 2 (a) - (d). Eight-Hour Maximum O_3 Concentrations (ppb) over Greater Phoenix, AZ for selected days in 1997. Squares denote census tracts and are proportional in size to the 1990 population ≥ 65 years of age. Triangles show locations of O_3 monitors reporting for each day. The maps represent an area approximately 43 miles square.

Loss of Information on Spatial Variability Due to Averaging

Despite the high correlations often observed between O_3 concentrations measured over time at neighboring sites, the correlations are not perfect and the maps presented in the previous section suggest that the shapes of the concentration isopleths do not always persist from day to day. Consequently, cross-site averages of O_3 concentrations might not fully describe changes in exposure patterns day to day. This hypothesis can be studied somewhat more systematically by calculating the cross-site average and standard deviation of O_3 concentrations for each day, and then plotting daily averages against daily standard deviations. If the daily standard deviations exhibit variability which is unrelated to the daily average concentrations, a loss of information about exposure patterns will be realized if the average alone is used to characterize O_3 exposure. For this phase of the analysis, the period from January through mid-July 1997 was considered. This period is of particular interest because it was during this time only that an O_3 monitor was in operation in the northwest part of the metro area. This monitor (visible in all of the maps) often indicated concentrations much lower than other monitors in the network, and moreover it is the monitor nearest to much of the population cluster $\geq$ 65 years of age in the northwest corner of the metro area. (Maricopa County government staff confirmed that quality control and quality assurance procedures had been followed regarding data from this monitor.) In order to make the analysis precise, only those dates with complete hourly readings for all monitors in a fixed subset were admitted into the analysis. Fourteen monitors operated more or less continuously over the January to mid-July period, and on 47 days all fourteen monitors reported hourly readings for all twelve hours between 8 am and 8 pm. The locations of the subset of fourteen monitors used in this analysis are shown in Figure 3.

Since the primary interest is in population exposure, the daily cross-site averages and standard deviations were population-weighted. Specifically, the all-ages population for each census tract was assigned to the nearest monitor. Each monitor was then given a weight proportional to its assigned population. Figure 4 shows the resulting daily cross-site averages plotted against the daily cross-site standard deviations for the 47 dates with complete hourly data for the subset of fourteen monitors. The population-weighted cross-site daily averages range between 20.8 and 71.0 ppb, while the population-weighted standard deviations range between 3.0 and 8.8 ppb. It is apparent from the scatterplot that for this particular data set, the cross-site averages and standard deviations are weakly related (sample correlation coefficient $r = -.06$). At all observed average concentrations the pattern of ambient concentrations may reflect relatively low or relatively high site-to-site variability, depending on the particular day. As a specific example, the population-weighted averages corresponding to Figures 1 c and d are 62.3 ppb and 64.8 ppb, respectively, while the respective population-weighted standard deviations are 4.2 ppb and 8.6 ppb, indicating a substantial difference in site-to-site variability for two days with comparable cross-site averages. This difference is readily apparent in the two figures. Repeating this analysis under a weighting scheme based on the population $\geq$ 65 years of age produced qualitatively comparable results, although the daily standard deviations were somewhat greater than those based on all-ages population weighting.

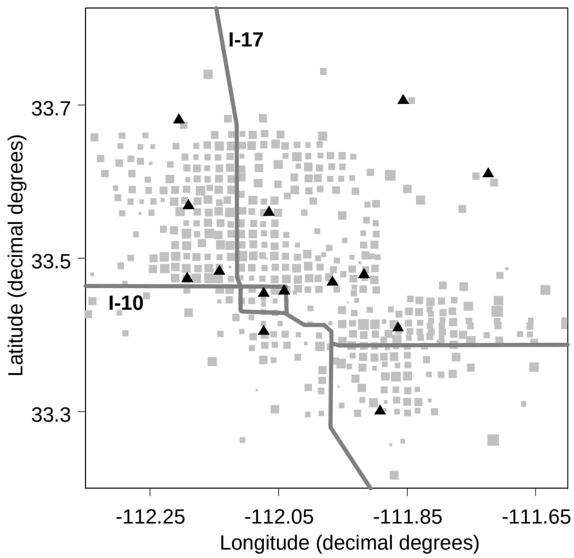


Figure 3. Locations of O_3 monitors used in January to mid-July 1997 analysis. Squares denote census tracts and are proportional in size to 1990 all-ages population.

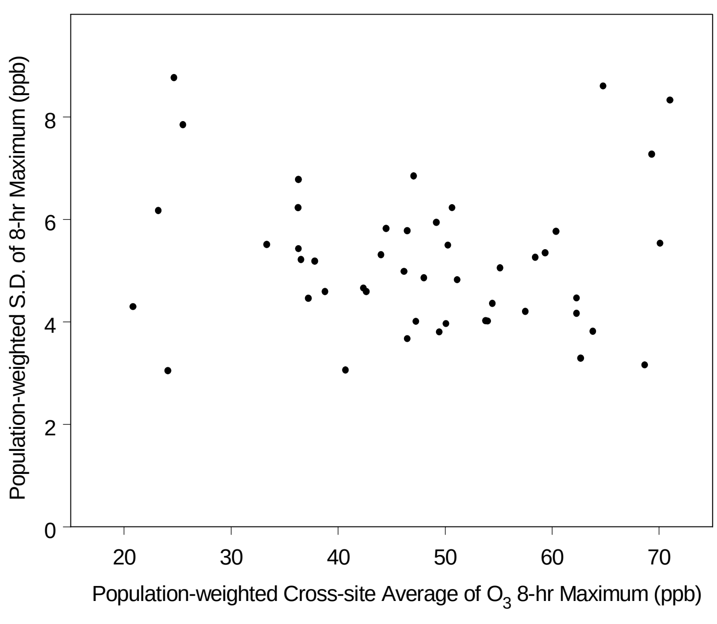


Figure 4. Comparison of population-weighted cross-site averages and population-weighted site-to-site standard deviations for O_3 8-hour maximum concentrations.

Site-to-site Correlation Before and After Seasonal Filtering

Site-to-site correlation of measured pollutant concentrations as a function of distance has been previously investigated for various airborne pollutants, including O_3. Based on data collected from a network of sixteen monitors across the greater Detroit area in the summer of 1981, the correlations between O_3 concentrations measured at a central city monitor and the other fifteen monitors were generally high (median r = .78) but also were observed to decrease with distance between monitors (Kelly et al., 1986). Site-to-site correlations of O_3 concentrations measured across a network of four monitors in this same geographic area over a multi-year period were similarly high (median r = .83) and decreased gradually as the distance between monitors increased (Lippman et al., 2000). Similarly, based on O_3 concentrations measured across a network of ten monitors in the greater Chicago metro area from 1985 - 1994, site-to-site correlations were generally high (median r = .75) but clearly decreased with distance (Ito and Thurston, 1999).

Many contemporary environmental epidemiology time-series analyses employ some form of filtering to remove seasonality and other types of trends from the data prior to the correlational modeling phase. In some instances, filtering "on both sides" is the preferred approach (Kinney and Özkaynak, 1991; Çakmak, et al. 1998). Under this approach the analyst would take steps to detrend not only the health outcome time-series data but also the pollutant time-series data. In the case of O_3, it stands to reason that site-to-site correlation is driven in part by its strong seasonal profile (Ito, et al. 1998). Hence, we may anticipate that seasonal detrending is likely to reduce the relatively high site-to-site correlation commonly observed between monitors across an urban airshed.

An analysis of the site-to-site correlation as a function of distance for the 1997 Phoenix data suggests the usual high site-to-site correlation even at distances greater than 30 miles. For consistency, this site-to-site analysis was based on data for the same dates and monitors analyzed in the previous section. The selected monitors provided enough geographic spread to allow an evaluation of the distance effect. Figure 5a shows the 91 site-to-site correlations plotted against distance in miles. The trend appears approximately linear, and an OLS regression line was fit through the scatter of points (estimated intercept = .98, estimated slope = -.0046).

To examine the effect of seasonal detrending on the site-to-site correlations, the time series of eight-hour maxima for each monitor was filtered using a wide-span LOESS smoother. The within-site standard deviations of the filtered concentrations ranged from 6.1 ppb to 9.5 ppb for the fourteen monitors, indicating that substantial day-to-day variation remained after detrending. The site-to-site correlations of the filtered concentrations were then plotted against distance and a regression line was again fit to the scatter of points (estimated intercept = .92, estimated slope = -.0079). The results are shown in Figure 5b. Not surprisingly, the site-to-site correlations for the filtered concentrations are weaker and decline more rapidly with distance than do those based on unfiltered concentrations. The smaller intercept suggests an overall loss of correlation even between neighboring monitors, and the decline in correlation with distance is $-.0079 / -.0046 \cong 1.7$ times as fast as a result of seasonal filtering.

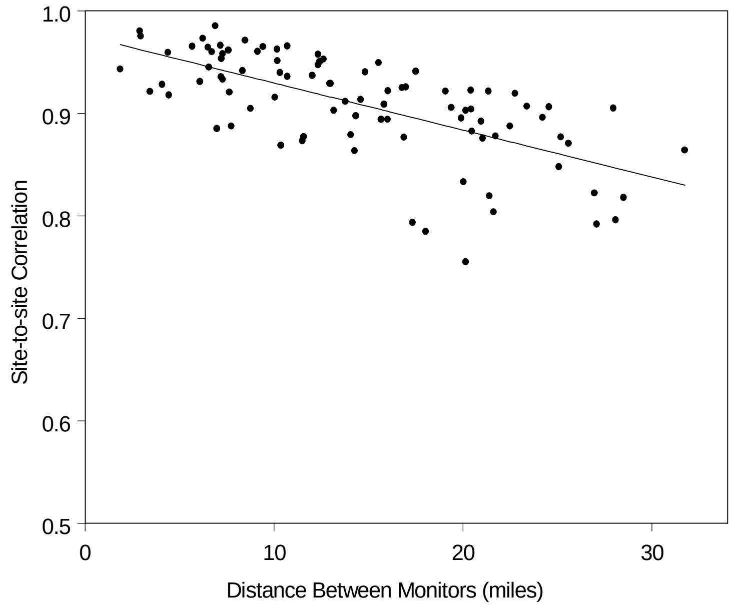


Figure 5a. Site-to-site correlation of eight-hour maximum O_3 concentrations as a function of distance between O_3 monitors. The O_3 maximum concentrations have not been seasonally detrended.

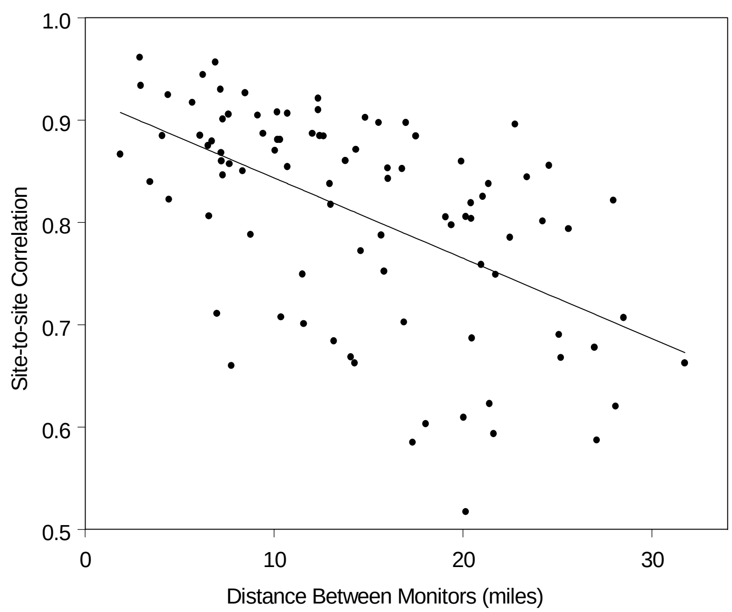


Figure 5b. Site-to-site correlation of eight-hour maximum O_3 concentrations as a function of distance between O_3 monitors after seasonal detrending.

Summary

The results of the data analyses presented in this report indicate that in the Phoenix metro area: (1) significant spatial variation in daily maximum eight-hour O_3 concentrations is occasionally observed and patterns of spatial variation do not always persist from day to day; (2) monitor placement may not be ideal for measuring population exposure, particularly for the population of individuals $\geq$ 65 years of age; (3) multiple-site averaging does not provide full information on the distribution of ambient concentrations; and (4) high site-to-site correlations are due in part to common seasonal variation at all monitoring sites and do not remain nearly as high after seasonal detrending. Findings (1) and (2) are presented visually and might be considered anecdotal, but nevertheless may be potentially useful in suggesting general hypotheses. Findings (3) and (4) result from systematic analyses and may have implications for the statistical evaluation of health effects.

References

Çakmak S., Burnett R., and Krewski D. (1998). Adjusting for temporal variation in the analysis of parallel time series of health and environmental variables. *Journal of Exposure Analysis and Environmental Epidemiology*, **8(2)**: 129-144.

Ito K., Thurston G.D., Nadas, A., and Lippman, M. (1998). Multiple pollutants' spatio-temporal exposure characterization errors. In *Proceedings of a Specialty Conference Cosponsored by the Air & Waste Management Association and the U.S. Environmental Protection Agency's National Exposure Research Laboratory*, Cary, NC.

Ito K. and Thurston G.D. (1999). A pilot study of the effects of Berkson-type exposure measurement errors on regression models of mortality and morbidity: final report. Unpublished manuscript.

Kelly N.A., Ferman M.A., and Wolff G.T. (1986). The chemical and meteorological conditions associated with high and low ozone concentrations in southeastern Michigan and nearby areas of Ontario. *Journal of the Air Pollution Control Association*, **36**: 150-158.

Kinney P.L. and Özkaynak H. (1991). Associations of daily mortality and air pollution in Los Angeles County. *Environmental Research*, **54**: 99-120.

Lippman, M., Ito K., Nadas, A., Burnett, R.T. (2000). Association of particulate matter components with daily mortality and morbidity in urban populations. HEI Research Report No. 95. Boston: Health Effects Institute.

U.S. EPA (1996). *Air quality criteria for ozone and related photochemical oxidants*. Office of Research and Development, EPA/600/P-93/004cF, Washington, DC.

Disclaimer

This document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

For additional information contact:

National Center for Environmental Assessment
U.S. Environmental Protection Agency
MD-52
Research Triangle Park, NC 27711