

Rationale for Spatial Mapping Approach

We selected time-invariant rectilinear grids as our spatial unit rather than ZIP code regions, because the later vary in time and are difficult to use in an areas with rapid population growth, such as Orange, Riverside, and San Bernardino Counties in the 1980-2000 time period. Also, for this study, the geographic unit needed to have large enough populations to generate stable “proportions of asthma-related discharges”. Based on 10x10 km grids, 9 of 7020 values were outliers because of small grid populations. Had we used ZIP code regions as the geographic unit, the outlier problem would have been exacerbated by having more spatial units with small populations.

In this study, monthly average air pollutant concentrations from monitoring stations located typically 20 to 30 km apart are mapped spatially to 10 x 10 km grids. The spatial gradients in monthly average concentrations are small compared to those for daily and hourly concentrations. Under these conditions, the results are not particularly sensitive to the choice of mapping methods. We have found that Kriging (with appropriate parameters) and inverse-distance-squared-weighted interpolation produce very similar results when applied to monthly average concentrations on 10 x 10 km grids. We chose inverse distance square weighted interpolation using data from the 4 closest stations within 50 km of the point of interest when no stations were located within 5 km and assignment of the data from the nearest station when a station was located within 5 km of the point of interest.

As text Figure S1 suggests, the distance from the grid centroids to nearest air quality monitoring stations is less than 50 km in all cases. Table S3 shows the distribution of distances from grid centroids to the nearest air quality stations for O₃, CO, NO₂, and PM₁₀. The distributions vary year-to-year, but for O₃, 13% of the grids, on average, have data measured within the grid, 73% have measured data within 5 to 25 km of the grid, and only 14% have data

measured within 25 to 50 km of the grid. Somewhat larger distances apply to the other pollutants because there are few stations. We believe the spatially-mapped network data are able to represent variations in regional air quality. Small scale variations from traffic or point sources are not resolved. Also, because the spatial coverage of the monitoring network and actual spatial gradients of pollutants vary, variations in the accuracy of the spatially mapped estimates are expected; however, it's unlikely that the variations in accuracy are large enough to significantly bias the results.

Details of Assignment of Hospital Discharges to Spatial Grids:

The finest spatial resolution for which hospital discharge data were available was the 5-digit postal ZIP code of the patient's residence; the patient's street address, 9-digit ZIP code, and census block were not available. Three sets of spatial allocation factors were therefore developed to distribute Zip code-based discharge data to the exposure grids in the 1980-1984, 1995-1994, and 1995-2000 time periods. Each set included separate factors for males and females for ages less than 1 year and 1 to 19 years. A single 5-digit ZIP was allocated spatially, based on population-weighting to multiple exposure grids with ESRI ARcGIS9 software tools (ESRI, Redlands, CA. Most ZIP code regions were allocated to 1 or 2 grids; the largest ZIP code region was allocated to 10 grids. The allocation factors for 1980-1984 were developed using block group centroids and the block group to ZIP code equivalency data provided with the census. For 1985-1994, the 1990 block group boundaries and block group to ZIP equivalency data were used. The allocation factors for 1995-2000 were developed using the 2000 census block centroids and ZIP code boundaries. The exposure grids were overlaid on the population and ZIP data in the GIS to determine the allocation factors for each ZIP code and period.

Spatial allocation of demographic data to exposure grids was based on the smallest geographic unit for which census data were available. The 1980 data were allocated based on census block group data and block group centroids. In 1990, the census block group data and geographic boundaries were available; the data were allocated to grids assuming the population was uniformly distributed within each census block group. The data were available at the census block level in 2000, and were spatially allocated to grids assuming the population was uniformly distributed within each census block. Eight population variables in 1980 and one variable in 1990 and 2000 were renormalized after the spatial allocation to insure consistency across census topics. For example, the population by race was normalized by the total population. The population by sex, age, and race was normalized for consistency with population by race and population by sex. The population by residence was renormalized by the total population age 5 and above, etc.

Population and other demographic parameters were estimated for the intra-census years by linear interpolation of the gridded data for 1980, 1990, and 2000.

Sources:

1. Office of Statewide Health Planning and Development, Healthcare Quality and Analysis Division, *Patient Discharge Data File Documentation*, July 2006.
2. Office of Statewide Health Planning and Development, Healthcare Quality and Analysis Division, data files emailed July 2003, Patient Discharge Data Request #2030417-01.
3. California Department of Health Services Center for Health Statistics
4. U.S. Bureau of Census

Spatial Correlation

A typical assumption that is often not stated explicitly is that the observed data consist of n independent and identically distributed observations from the random variable O with distribution P . In this analysis, we make the assumption that the observed data consist of $n=195$ random variables O_i that describe each spatial/geographical unit i , $i = 1, \dots, n$, each with distribution P_i . Thus, we do not assume that the data from each unit were sampled from one common distribution P but, rather sampled from n distinct distributions, P_i , that may be similar, in particular, for those units that are spatially close. Under this assumption, it follows that mutual independence between the random variables O_i , conditional on the exposure regimen, is a reasonable approximation.

Based on this approximation, we modeled the observed data as n independent and identically distributed observations of the random variable O with distribution P , where P represents average of the distributions P_i across units where each observation receives equal weight, i.e., $P = \frac{1}{n} \sum_{i=1}^n P_i$. The causal parameter of interest is a parameter of the distribution P and is interpreted as an average of the causal effects across the 195 units.

Rationale for use of Linear Models

For both the traditional and HRMSM approaches, working models considered were semi-parametric linear models. An alternative to the linear fit is a binomial logistic fit for the number of asthma-related hospital discharges, conditional on the number of individuals in the corresponding grid at the corresponding point in time. The binomial logistic fit is equivalent to a

weighted logistic fit with the outcome being the proportion of asthma-related hospital discharges and the weights being the total number of individuals in a given grid at the given time point. Our motivation to use a linear model was twofold: **1)** The proportion of asthma discharges had a very small range of values (0 to 4.98×10^{-3}) that was not close to spanning the 0 to 1 interval. A main factor in selection of the binomial logistic model is to constrain the probabilities between 0 and 1; however, since the range of the proportions was so small, this was not an important consideration. The small range of values also has the effect that we only would focus on a small portion of the logistic curve. Any such small part of the curve can be approximated adequately with a line; and **2)** The binomial model fit is driven by observations with large weights, that is, large populations in a given grid; thus, a small number of observations could influence the fit significantly but may not be representative of the whole population (Neugebauer and van der Laan 2007).

History-Restricted Marginal Structural Model Details:

Marginal Structural Models (MSM) represent the effect of an entire exposure history (since the start of the study) that precedes a time-specific outcome. In this analysis, the entire history of O_3 is not relevant in estimating the effect of O_3 on the proportion of asthma-related hospital discharges. Instead we investigate the effect of the level of O_3 at the current quarter only on the outcome during that same quarter. History-restricted MSM (HRMSM) were proposed to allow for such a flexible analysis considering only part of the exposure history (Feldman et al. 2004; Joffe et al. 2001; Neugebauer, 2007). In the standard MSM framework, the effect of interest would be investigated with a model for the distribution of the counterfactual outcomes, $Y_{\bar{a}(t-1)}(t)$ for all possible treatment histories $\bar{a}(t-1)$ and time-specific outcomes of interest, $Y(t)$. In the HRMSM framework adopted, the effect of interest is instead investigated

with a model for the distribution of the counterfactual outcomes $Y_{a(t-1)}(t)$. We used the following HRMSM for the expectation of the counterfactual outcomes,

$$E(Y_{a(t-1)}(t)) = \beta_0 + \beta_1 a(t-1).$$

HRMSMs are interpretable causally, directly at the population level, that is, β_1 can be interpreted as the population-level effect of the same quarter level of O₃ exposure on the proportion of asthma-related hospital discharges during any given quarter.

We implemented two estimators of HRMSM causal parameters (Neugebauer, 2007): the Inverse Probability of Treatment Weighted (IPTW) and G-computation (G-comp). Both estimators rely on the Sequential Randomization Assumption (no unmeasured confounders).

To obtain the G-computation estimate $\beta = (\beta_0, \beta_1)$, we used the DSA selected model for $E(Y(t) | A(t-1), W^*(t-1))$ that we obtained in the traditional approach. We note that in linear models for $E(Y(t) | A(t-1), W^*(t-1))$ when there is no interaction between $A(t-1)$ and $W^*(t-1)$, the G-computation estimate of the coefficient estimate for $a(t-1)$ in the MSM is equivalent to the coefficient estimate for $A(t-1)$ in the traditional model, $E(Y(t) | A(t-1), W^*(t-1))$. We also note that consistent G-computation estimation relies on consistent estimation of $E(Y(t) | A(t-1), W^*(t-1))$.

The IPTW estimate for β is obtained by regressing $Y(t)$ on $A(t-1)$ using the HRMSM with weights defined by $1/g(A(t-1) | W^*(t-1))$. Consistent IPTW estimation relies on consistent estimation of the density $g(A(t-1) | W^*(t-1))$ known as the treatment mechanism distribution, and the Experimental Treatment Assignment (ETA) assumption which is discussed in more details below. We used the DSA to select a Gaussian model for $g(A(t-1) | W^*(t-1))$ with

constant standard deviation. The weights were estimated as $1/g_n(A(t-1)|W^*(t-1))$, where g_n is the estimated model for g , and were subsequently truncated between 1/0.1 and 1/0.9 to mitigate the poor finite sample performance of IPTW estimation due to the practical violation of the ETA assumption, i.e. improve efficiency.

The ETA assumption, on which the IPTW estimate relies, states that there are no values of $W^*(t-1)$ for which treatment is assigned deterministically. This assumption requires that there is experimentation among all possible levels of ozone at time $t-1$ denoted by $\mathcal{A}(t-1)$, or that ozone is not assigned deterministically based on $W^*(t-1)$. The IPTW estimates also suffer from finite sample bias and poor efficiency when the ETA assumption is violated practically (Wang et al. 2006). Practical violations of the ETA assumption are not uncommon for a continuous treatment variable since it is almost inevitable that for given strata defined by $W^*(t-1)$, some levels of treatment are rare or even non-existent. A plot of the observed treatment against the predicted treatment from the treatment mechanism model can be used to visually check this assumption. When the points are clustered around the $y=x$ line then there is evidence of an ETA violation as this indicates that treatment was assigned in a deterministic fashion. In addition to this visual check, we also applied a diagnostic tool that has been developed based on parametric bootstrap sampling from an estimated data-generating distribution to assess the bias in IPTW estimation due to ETA violation (Wang et al. 2006) (the software is available at: <http://www.stat.berkeley.edu/~laan/Software/>).

We assumed that the missingness occurred at random; thus, the observations with missing values for covariates in the DSA-selected model were dropped from the data prior to fitting the model. Dropping the observations for which covariate values $W^*(t-1)$ are missing can be done without loss of consistency if the missingness mechanism is believed to be

uninformative. Due to the small number of missing values and the fact that the data was not self-reported, this is a reasonable assumption.

Confidence intervals and p -values for both HRMSM estimates were obtained using 10,000 bootstrap iterations where re-sampling was based on the 195 independent grids. Due to the computation time required by the DSA procedure, we were not able to repeat the data-adaptive model selection procedure for each bootstrap sample. Thus, we treated the selected models for the treatment mechanism and $E(Y(t) | A(t-1), W^*(t-1))$ as given for the purposes of the bootstrap by simply refitting them for each bootstrap sample. This approach ignored the extra variability of our estimators that is introduced by the model selection procedure and thus possibly underestimated the variance (Shen and Huang 2004).

Details of Model Selection for Table 3

Figure S7_a shows the cross-validated (CV) risk for the fit of the model reported in Table 3 in the main text. The x-axis indicates the size of the model (maximum of 11 allowed for this analysis=10 variables + intercept); the y-axis indicates the CV risk for the best model of size N. The optimal model is that with the lowest CV risk (size 9). The CV risk curve for a model with no interaction (order=1) and two-way interactions (order=2) are shown. In general, for any model size, the differences in CV risk between interaction orders are very small.

Figure S7_b presents the details of the models selected at each step of the DSA process. All variables listed in Table S2 below, plus quarterly average 1-hour daily maximum O₃ (O3Dmax) could have entered the modeling at any step. The only pollutant selected is O₃, and it is first selected at a model of size 6 (broken arrow). Based on Figure S7_a, model 9 is the optimal model (dotted arrow in Figure S7_b). Note that the two additional models, sizes 10, 11 do not include any other pollutant or meteorological variables. The CV risk declines from

1.005×10^{-7} for model size 2 to 9.51×10^{-8} for model size 9 and increases to 9.64×10^{-8} for model size 11. O3Dmax produces a small change in the CV risk when selected as part of the best 6-variable model (CV risk model size 5: 9.62×10^{-8} ; CV risk model size 6: 9.59×10^{-8}) and remains in all other models thereafter.

References

Feldman, H, Joffe, M, Robinson, B, Knauss, J, Cizman, B, Guo, W, Franklin-Becker, E, and Faich, G. 2004. Administration of parenteral iron and mortality among hemodialysis patients. *J Am Soc Nephrol* 15: 1623–1632.

Joffe, M., Santanna, J., and Feldman, H. (2001). Partially marginal structural models for causal inference [abstract]. *Amer. J. Epidemiology* 153, S261.

Neugebauer, R, Joffe, MM, Tager, IB, van der Laan, MJ. 2007. Causal Inference in Longitudinal Studies with History-Restricted Marginal Structural Models. *Electronic Journal of Statistics* 1: 119-54.

Neugebauer, R, van der Laan, M. 2007. Non-parametric causal effects based on marginal structural models. *J Stat Plan Inference* 137:419-34.

Shen, X, Huang, H-C, Ye, J. 2004. Inference after model selection. *JASA* 99:751-762.

Wang, Y, Petersen ML, Bangsberg, D, van der Laan, MJ. 2006. Diagnosing Bias in the Inverse Probability of Treatment Weighted Estimator Resulting from Violation of Experimental

Treatment Assignment. Available: <http://www.bepress.com/ucbbiostat/paper211/> [accessed 1 March 2007].

Table S1: Variables with Univariate Association with Asthma and Ozone Exposure^a

Demographic Variables

% completed 1-3 years of high school	% foreign borne
% completed high school	% Hispanic
% completed 1-3 years college	% White
% complete college or greater	% native born but not in California
% females not in the labor force	% residence in another state for 5-years prior to given quarter
% household income < \$19,900	% residence in same house for 5-years prior to given quarter
% household income \$20,000-\$39,999	Median household income
% household income > \$100,000	% female

Pollutant and Weather Variables

Quarterly average 24-hour CO ₂ (ppm)	Quarterly average 24-hour NO ₂
Previous year quarterly average 24-hour CO ₂ (ppm)	Previous year quarterly average 24-hour NO ₂
Quarterly average 24-hour average PM ₁₀	
Previous year quarterly average 24-hour average PM ₁₀	
Quarterly average 24-hour average relative humidity (%)	
Previous year quarterly average 24-hour average relative humidity (%)	
Quarterly average 24-hour average temperature (F)	
Previous year Quarterly average 24-hour average temperature (F)	

Time Variables	Quarter number (0-35)	Quarter indicator
	Year number	

Table S2: Variables Selected for Possible Inclusion in Statistical Models

Demographic Variables

Percent Asian
Percent African American
Previous year CO 24-hr average (ppm)
Percent Completed From Preschool to 8 Grades of Elementary School
Percent completed 1 to 3 Years of High School
Percent Graduated High School
Percent Completed 1 to 3 Years Undergraduate Work in College
Percent Graduated College or More
Percent Female, Not In Labor Force
Percent Male, Not In Labor Force
Percent Female, In Labor Force
Percent Male, In Labor Force
Percent Female
Percent Foreign Born (“Foreign”)
Percent Hispanic
Percent household Income \$0-\$19,999
Percent household Income \$20,000-\$39,999 (“income2”)
Percent household Income \$40,000-\$74,999
Percent household Income \$60,000-\$99,999
Percent household Income > \$100,000
Median Household Income (“median_income”)
Percent Native, Born In Different State or Abroad
Percent Born in California
Percent Other Race (2variables: “person 1”, “other”)
Percent 1-2 persons in household
Percent > 5 persons in household
Percent Above poverty level
Previous Year Proportion asthma-related hospital discharges
Percent Residence in Different House, Same County 5 years prior
Percent Residence in state other than California 5 years prior
Percent Residence in same house 5 years prior
Percent Residence in Different House and County, Same State 5 years prior
Percent White (“white”)

Pollutant and Weather Variables

CO 24-hr average (ppm)
Previous Year CO 24-hr average (ppm)
NO₂ 24-hr average (ppb)
Previous Year NO₂ 24-hr average (ppb)
PM₁₀ 24-hr average (ug/m³)
Previous Year PM₁₀ 24-hr average (ug/m³)
SO₂ 24-hr average (ppb)
Previous Year SO₂ 24-hr average (ppb)
Relative Humidity 24-hr average (%)
Previous Year Relative Humidity 24-hr average (%) (“Rhavg14”)
Temperature 24-hr average (deg F) (“Tavg”)

Previous Year Temperature 24-hr average (deg F)

Time Variables

Quarter Number (0-35)

Quarter Indicator

Year Number (0-17)

Table S3: Distribution of Distances from Grids to Nearest Air Quality Stations

Pollutant	Percent of Grids with the Closest Air Quality Station		
	Within the Grid	Within 5 to 25 km of Grid Centroid	Within 25 to 50 km of Grid Centroid
O ₃	13 ± 2	73 ± 3	14 ± 5
CO	9 ± 1	58 ± 5	32 ± 4
NO ₂	10 ± 2	58 ± 9	32 ± 10
PM ₁₀	7 ± 3	69 ± 16	23 ± 14

Table S4: Selected Demographic and Pollutant Variables for Persons Birth to 19-Years

Used in Analysis: Quarters Two (April-June) and Three (July-September) of 1983-2000

VARIABLE	Spatial Grid Values (median, IQR, range)
Population	
Males	51.4 (50.5 - 52.5; 35.4 - 69.6)
Females	48.6 (47.5 - 49.5; 30.4 - 64.6)
Asthma Discharges – all grids	
Number over time	1,719 (1,497 - 2,013; 1,097-2,882)
Number 1 st listed discharge	1,540 (1,328 - 1,690; 1026-2,302)
Proportion of total age-specific population	4.3*10 ⁻⁴ (4.0*10 ⁻⁴ - 4.6*10 ⁻⁴ ; 3.1*10 ⁻⁴ – 6.2*10 ⁻⁴)
Residence (%)	
Same house entire period	43.8 (37.3 - 49.8; 12.5,9.2-71.6)

Different House, same county	30.4 (25.1 - 35; 9.9,8.6-61.2)
Different CA county	14.7 (6.5 - 24.8; 18.3,1.3-59.2)
Different State	9.1 (6.7 - 11.8; 5.1,1.0-58.5)
Race/Ethnicity (%)	
Percent Hispanic	15.3% (9.6 - 24.7; 0 - 78.9)
Percent White	74.7% (60.4 - 83.3; 1.9 - 100)
Percent African American	2.2% (1.0 - 5.1; 0 - 49.8)
Percent Asian	3.1 (1.1 - 6.6; 0 - 31.6)
Percent Other	1.6 (0.9 - 2.8; 0 - 33.4)
Percent Below poverty level (%)	9.5 (6-13.2; 0 - 38.1)
Place of Birth (%)	
Born in California	51.5 (45.6 - 57.0; 22.4 – 76.0)
Native, Born In Different State or Abroad	34.3 (28.9 - 40.1; 6.5 - 60.8)
Foreign Born	11.2 (7.4 - 18.9; 0 - 55.6)
Temperature 24-hr average (F)	68.4 (65.2 - 73.5; 49.5 - 84.3)
Relative Humidity 24-hr average (%)	58.5 (52.5 - 64.8; 18.3 - 80.7)
Pollutants	
O3 daily max (ppb)	87.8 (69.9 - 110.6; 28.6 - 199.9)
NO ₂ 24-hr average (ppb)	28.4 (20.1 - 36.7; 4.8 - 61.8)
CO 24-hr average (ppm)	0.9 (0.7 - 1.2; 0 - 2.6)
SO ₂ 24-hr average (ppb)	1.2 (0.6 - 1.9; 0 - 10)
PM ₁₀ 24-hr average (µg/m ³)	45.7 (36.5 - 57.6; 16.3 - 119.7)

Table S5: Distribution of Pollutants for Selected Years

	1980 Qtr2,3 ^a Qtr4,1	Qtr2,3 Qtr4,1	Qtr2,3 Qtr4,1	Qtr2,3 Qtr4,1	Qtr2,3 Qtr4,1
Pollutants (ppb)					
O ₃ 1-hr max	108 (86-139; 42-208) ^b 53 (46-74; 30-102)	113 (87-128; 42-174) 52 (48-56; 32-73)	90 (72-106; 42-147) 51 (47-54; 26-65)	76 (65-97; 34-136) 44 (40-53; 21-67)	73 (6-81; 37-91) 38 (34-43; 17-54)
O ₃ 10 am – 6 pm	80 (64-97; 30-148) 38 (33-50; 18-69)	82 (66-93; 31-121) 38 (34-41; 21-49)	67 (55-79; 28-107) 38 (33-42; 16-51)	60 (52-73; 23-99) 34 (29-38; 11-52)	59 (49-66; 26-72) 30 (25-35; 10-43)
Pollutants (ppb) ^c	1980				
NO ₂ 24-hr	42 (32-51; 6-97)	36 (29-44; 7-69)	32 (25-39; 5-63)	27 (20-35; 6-63)	24 (20-31; 6-50)
CO 24-hr	2 (2-3; 0-7)	2 (1-2; 0-5)	1 (1-2; 0-4)	1 (1-1; 0-4)	1 (1-1; 0-3)
PM ₁₀ 24-hr	55 (42-69; 22-173)	67 (56-76; 37-115)	48 (40-58; 19-116)	37 (27-50; 7-104)	36 (29-43; 14-69)
SO ₂ 24-hr	1985 5 (3-7; 1-15)	1990 2 (1-3; 0-11)	1995 1 (0-2; 0-9)	2000 2 (1-2; 0-5)	1 (1-2; 0-3)

^aQuarters of year: 1=January-March.; 2=April.-June; 3=July-September; 4=October-December

^bmedian, interquartile range, range

^c For all 4 quarter

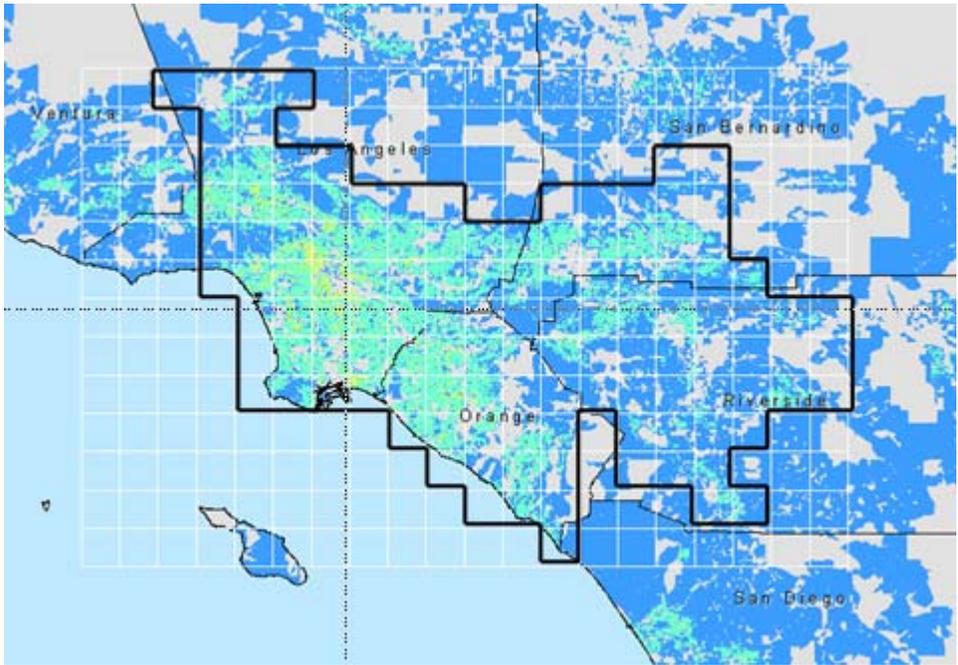


Figure S1: Example of grid structure for South Coast Air Basin

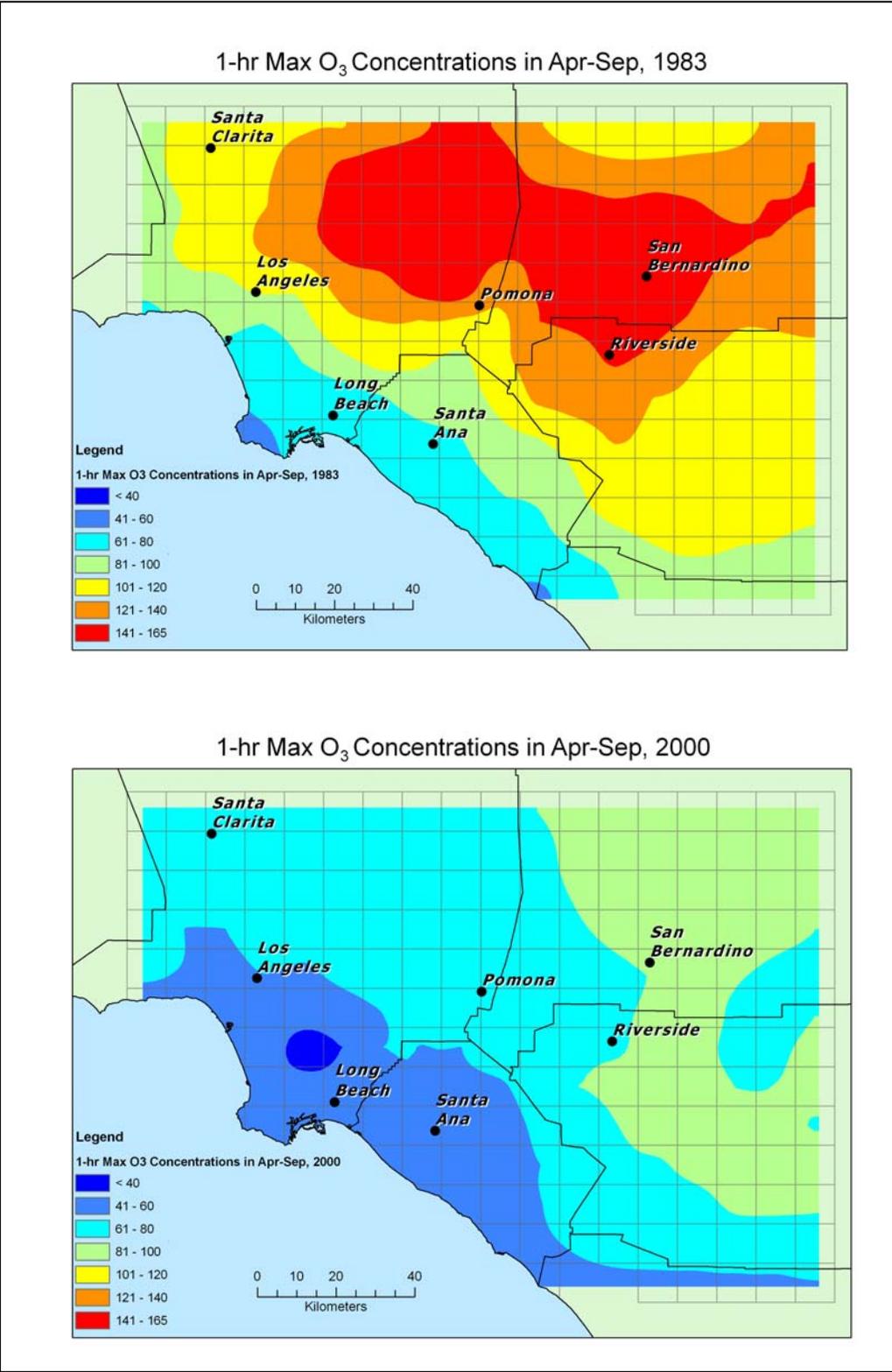
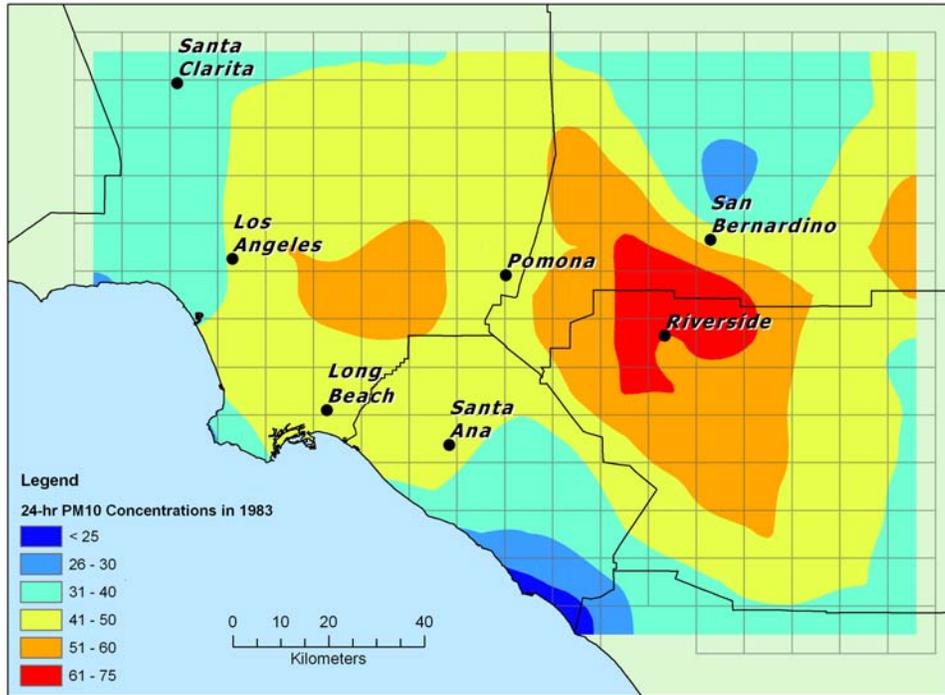


Figure S2: Spatial distribution of ozone concentration in the SoCAB for selected years.

24-hr PM₁₀ Concentrations in 1983



24-hr PM₁₀ Concentrations in 2000

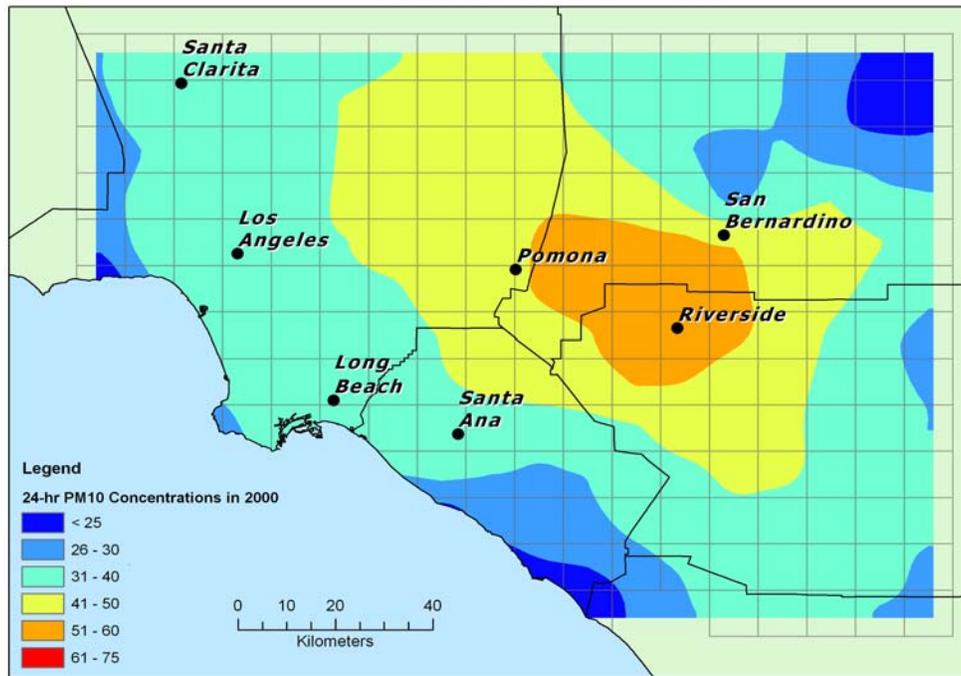
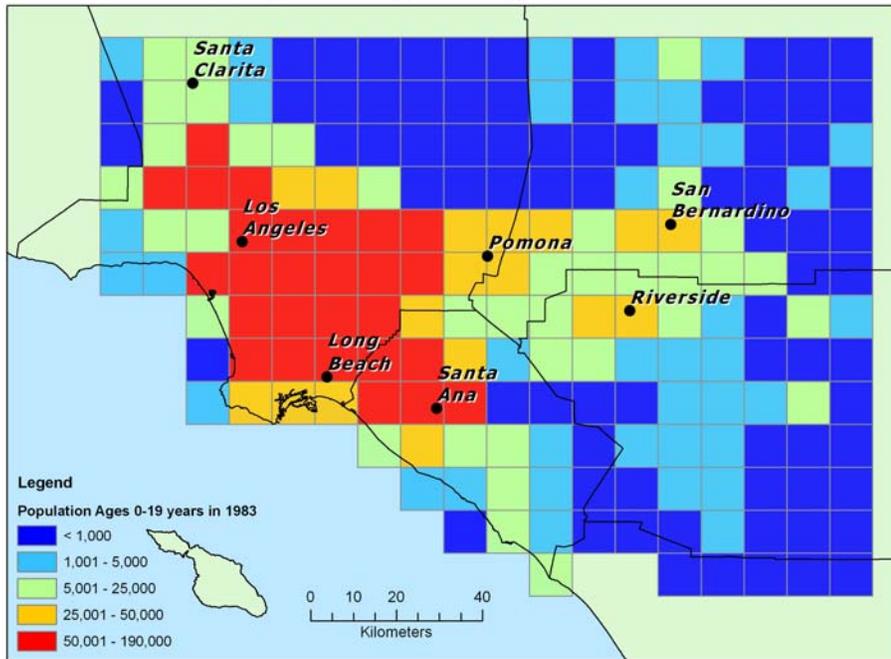


Figure S3: Spatial distribution of PM₁₀ concentration in the SoCAB for selected years.

Population Ages 0-19 years in 1983



Population Ages 0-19 years in 2000

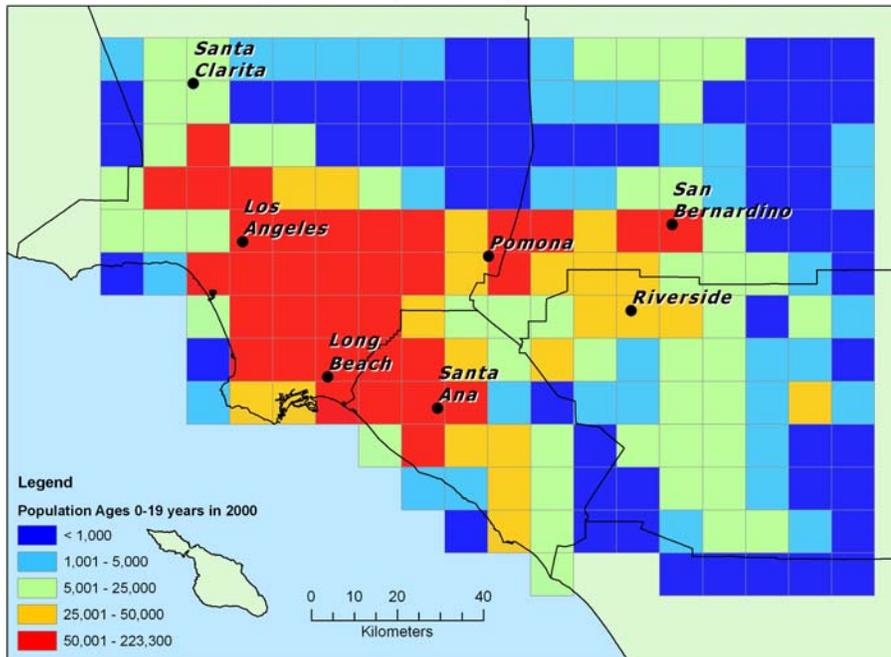


Figure S4: Spatial Distribution of Population of SoCAB for selected years.

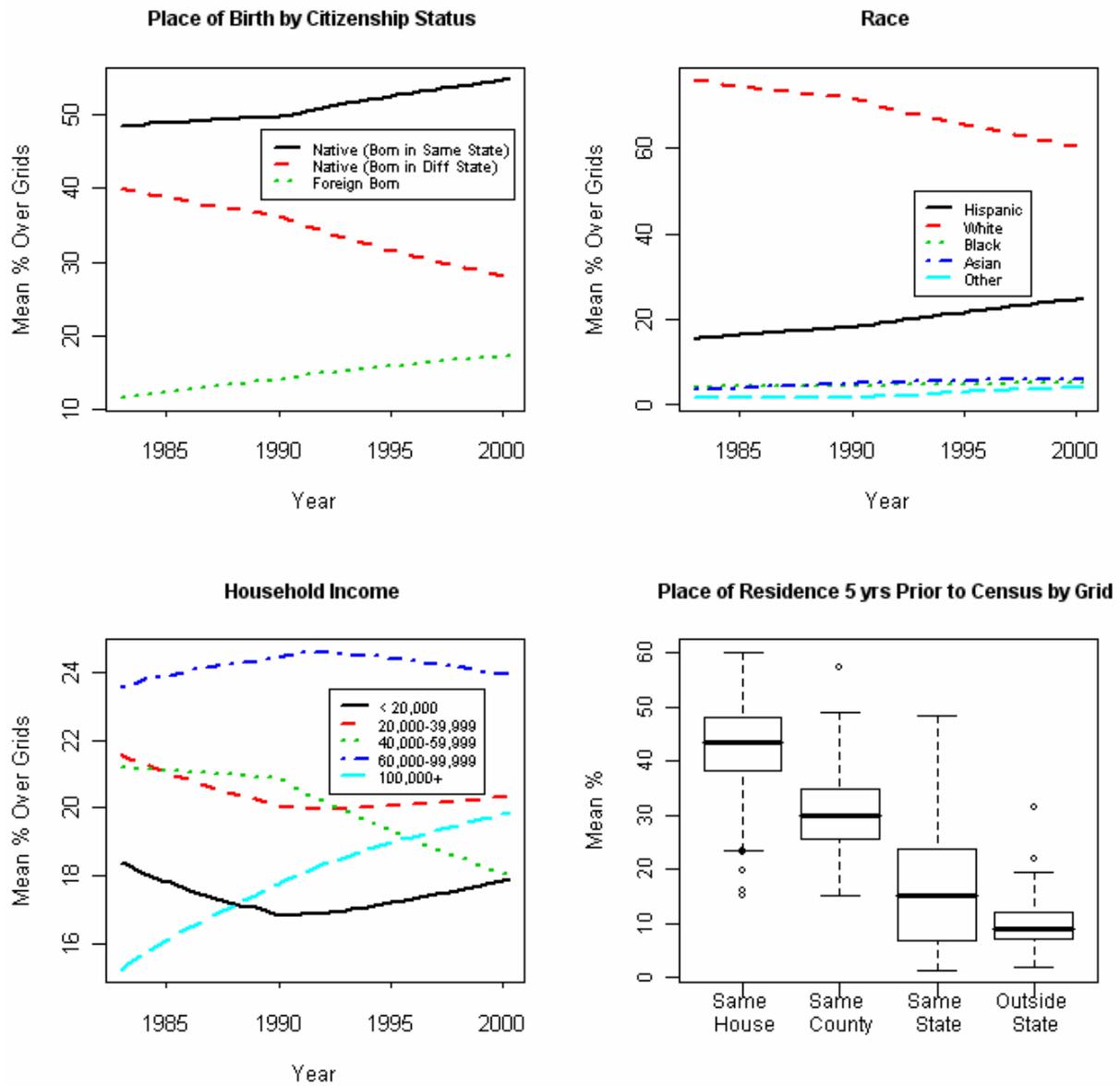


Figure S5: Temporal changes and spatial distribution for selected variables in SCAQMD. Years 1980, 1990, 2000 are based on actual census data. Estimates for inter-census years based on linear interpolation.

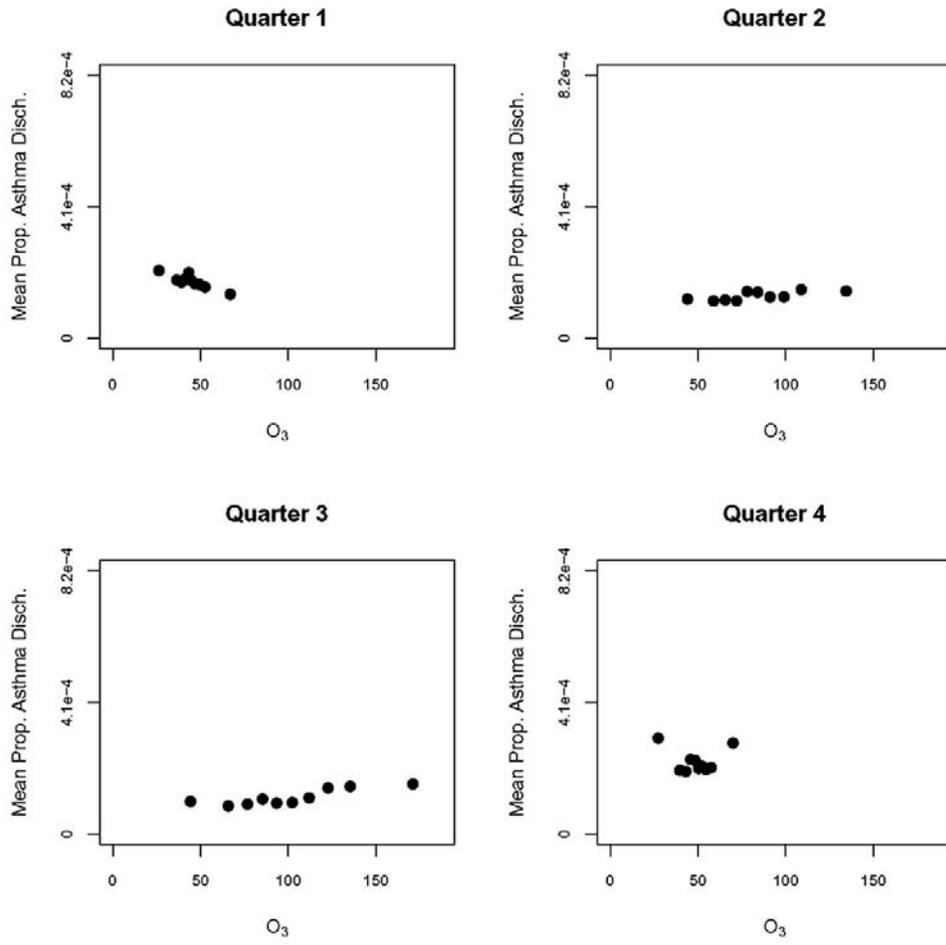


Figure S6: Pooled ozone distributions for all quarters 1980-2000

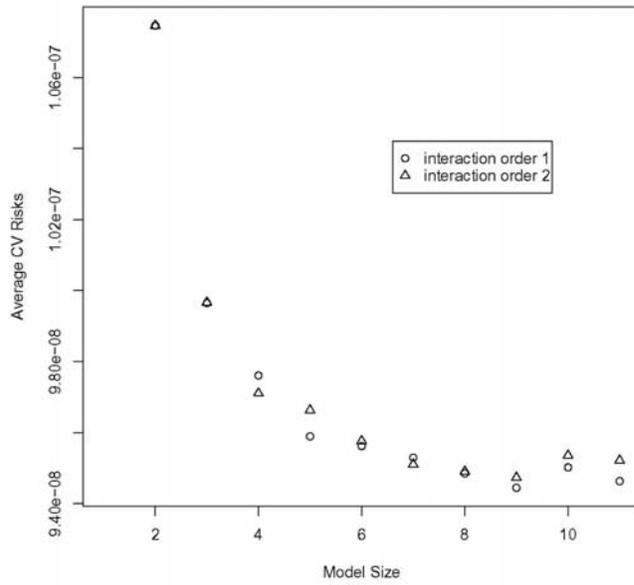


Figure S7_a: Cross-validated risk by model size

Model Size	Model Selected
2	I(white)
3	I(white) + I(income2)
4	I(white) + I(income2) + I(Tavg)
5	I(white) + I(income2) + I(Tavg) + I(rhavgl4 ³)
6	I(white) + I(income2) + I(Tavg) + I(rhavgl4 ³) + I(O3Dmax) ← - - - -
7	I(white) + I(income2) + I(Tavg) + I(rhavgl4 ²) + I(O3Dmax) + I(med_income)
8	I(white) + I(income2) + I(Tavg) + I(rhavgl4 ³) + I(O3Dmax) + I(med_income) + I(foreign ²)
9	← ····· I(white) + I(income2) + I(Tavg) + I(rhavgl4 ³) + I(O3Dmax) + I(med_income) + I(foreign ³) + I(white ³)
10	I(white) + I(income2) + I(Tavg) + I(rhavgl4 ³) + I(O3Dmax) + I(med_income) + I(foreign ³) + I(income1 ²) + I(med_income ²)
11	I(white) + I(income2) + I(Tavg) + I(rhavgl4 ³) + I(O3Dmax) + I(med_income) + I(foreign ³) + I(income1 ²) + I(med_income ²) + I(white ³)

Figure S7_b: Models Selected for each model size considered with no interaction terms based on the empirical risk.
(I=indicator, O3Dmax=quarterly average 1-hour daily maximum O₃; see Table A2 remainder of variable names (appear in bold type in parentheses). The intercept-only model constitutes model size 1.

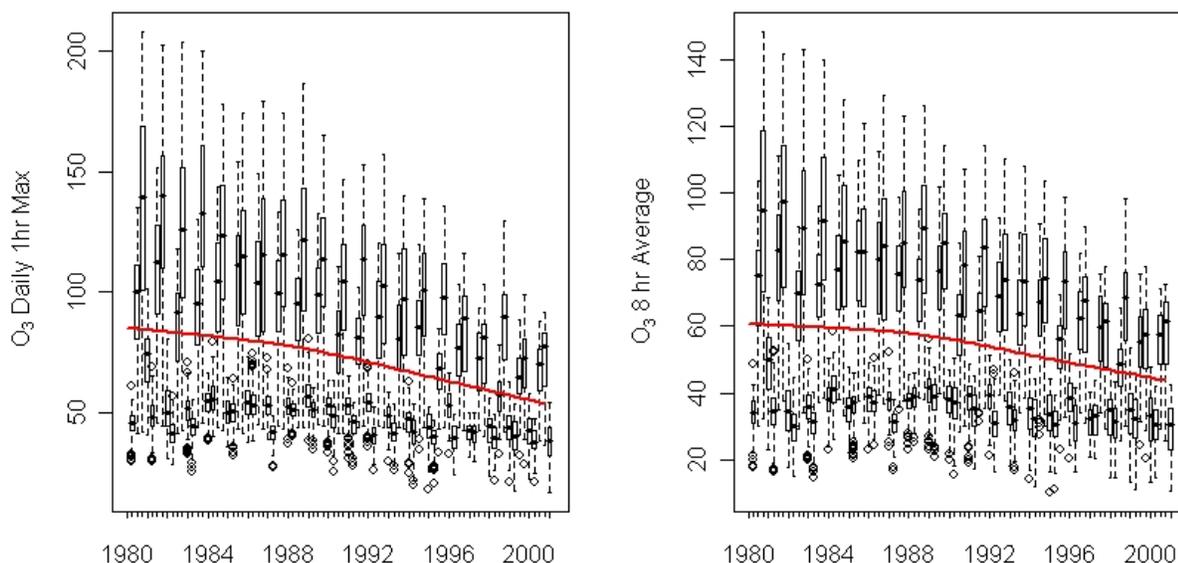


Figure S8_a: Quarterly 1-hours maximum and 8 hour average O₃ concentrations (ppb) by quarter 1980 – 2000. Box-whisker plots give medians (horizontal lines) and interquartile range (rectangle) for concentrations for quarters 2 and 3 (above the median levels) and 1 and 4 quarter. Line connects is LOESS smooth.

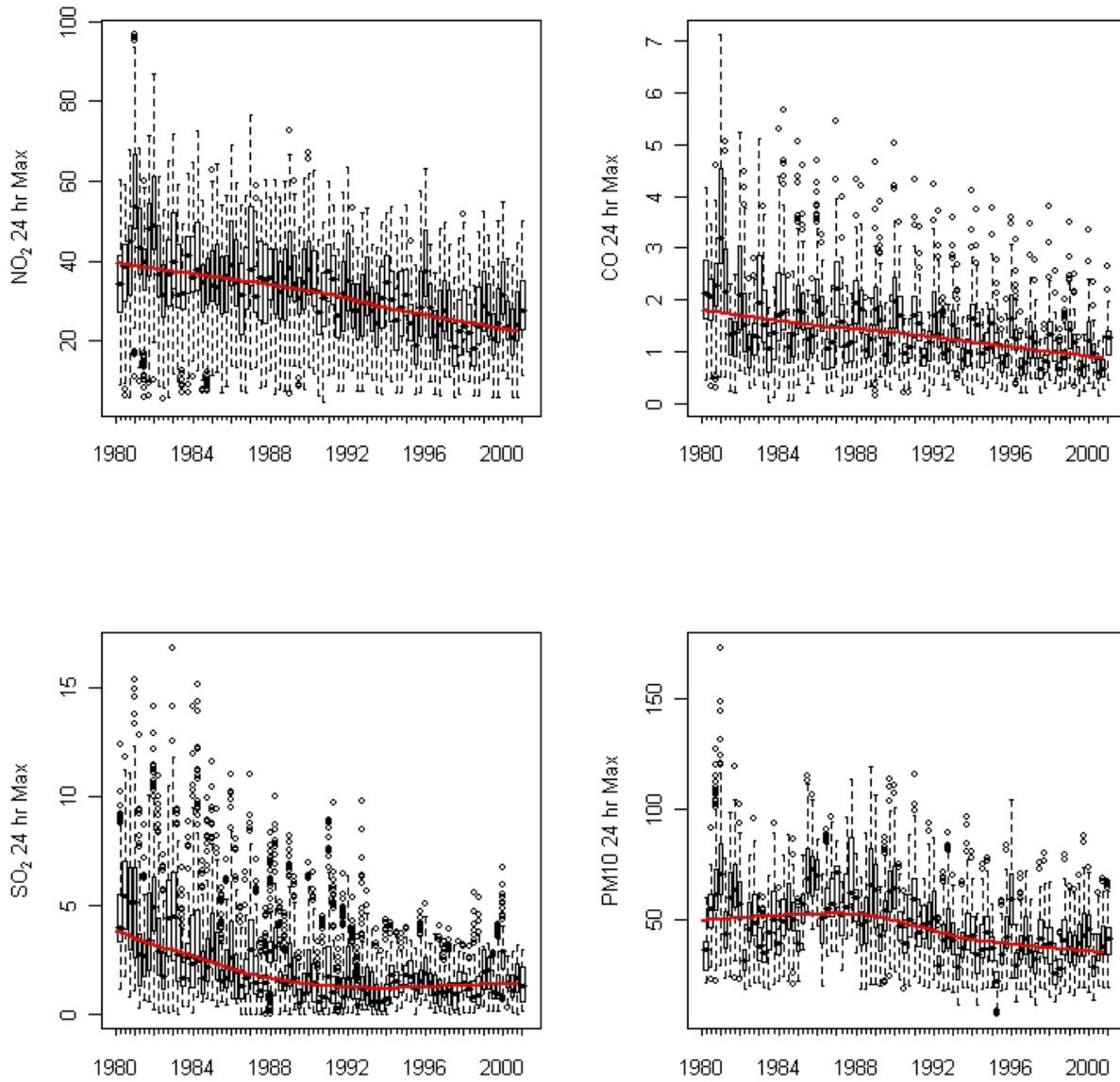


Figure S8_b: Quarterly concentrations of CO (ppm), NO₂ (ppb), PM₁₀ (μg/m³) and SO₂ (ppb). Note the lack of clear seasonal variation for NO₂ and the quarter 1,4 increases in PM₁₀ and CO.

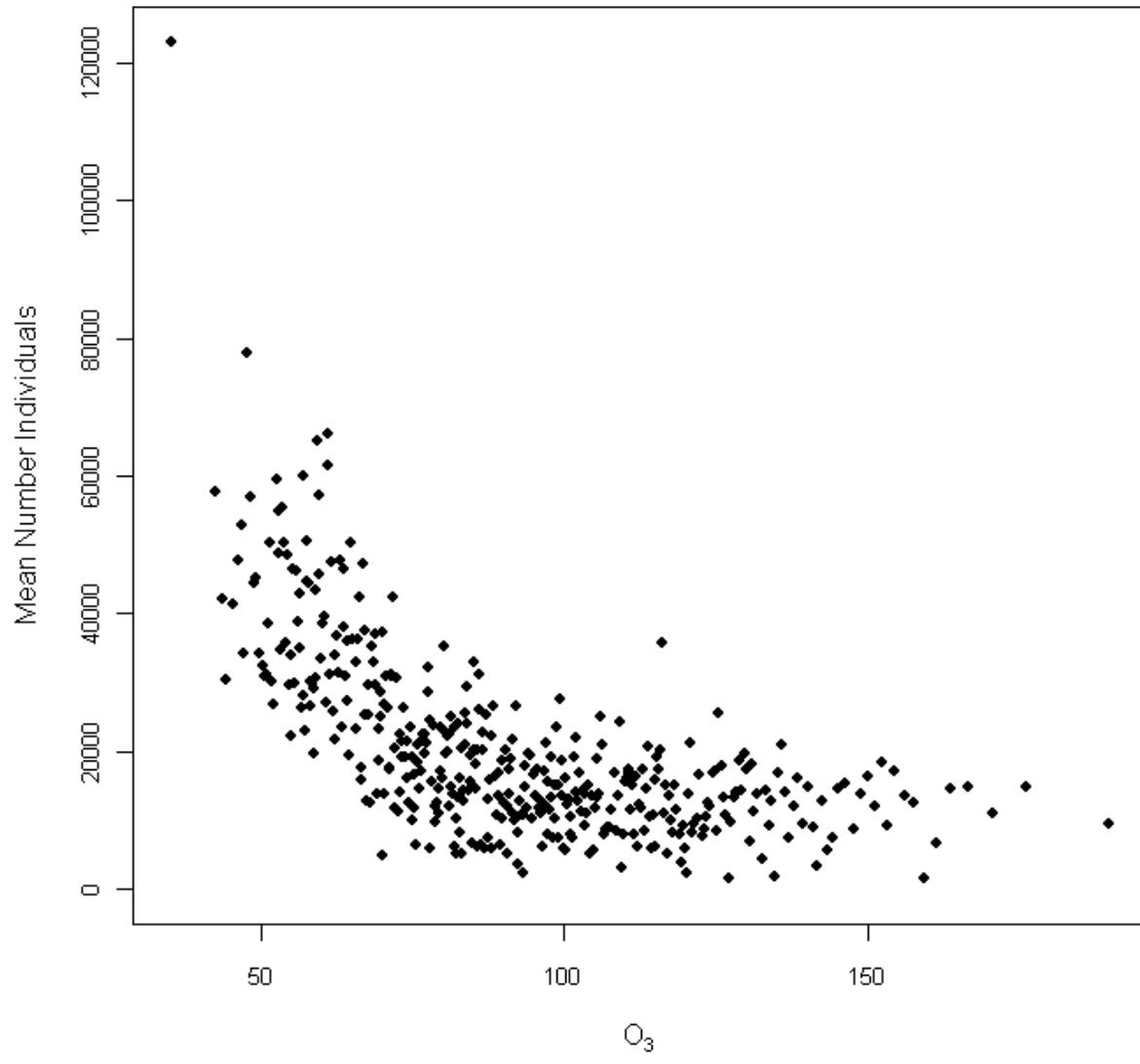


Figure S9: Distribution of population number by 400 quantiles of quarterly 1-hour maximum O₃ over quarters 2 and 3 years 1983-2000.