An Empirical Assessment of Exposure Measurement Error and Effect Attenuation in Bipollutant Epidemiologic Models

Kathie L. Dionisio,[1] Lisa K. Baxter,[1] and Howard H. Chang[2]

[1]National Exposure Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, USA; [2]Department of Biostatistics and Bioinformatics, Emory University, Atlanta, Georgia, USA

Address correspondence to Kathie L. Dionisio, U.S. EPA, 109 T.W. Alexander Drive, Mail Code: E205-02, Research Triangle Park, NC 27709 USA. Telephone: 919-541-1321. Fax: 919-541-0239. E-mail: dionisio.kathie@epa.gov

Running head: Attenuation in bipollutant models

Acknowledgments: The authors acknowledge Janet Burke, Vlad Isakov, Jim Mulholland, Jeremy Sarnat, Stefanie Sarnat, and Haluk Ozkaynak for their contributions to the development of exposure metrics used in this analysis, and Casson Stallings and Luther Smith for assistance with SHEDS modeling runs. The U.S. EPA through its Office of Research and Development, National Exposure Research Laboratory funded and collaborated in the research described here under cooperative agreement number CR-83407301-1 to Emory University, and a U.S. EPA Clean Air Research Center grant to Emory University and the Georgia Institute of Technology (R834799). This work was also funded under a NIH grant (1R21ES022795-01A1).

Competing financial interests: The authors declare no conflicts of interest or competing financial interests.

Abstract

Background: Using multipollutant models to understand combined health effects of exposure to multiple pollutants is becoming more common. However, complex relationships between pollutants and differing degrees of exposure error across pollutants can make health effect estimates from multipollutant models difficult to interpret.

Objectives: To quantify relationships between multiple pollutants and their associated exposure errors across metrics of exposure, and use empirical values to evaluate potential attenuation of coefficients in epidemiologic models.

Methods: We used three daily exposure metrics (central-site measurements, air quality model estimates, population exposure model estimates) for 193 ZIP codes in the Atlanta, Georgia metropolitan area, from 1999-2002, for $PM_{2.5}$ and its components (EC, SO_4), O_3, CO, and NO_x, to construct three types of exposure error: $\delta_{spatial}$ (comparing air quality model estimates to central-site measurements), $\delta_{population}$ (comparing population exposure model estimates to air quality model estimates), and δ_{total} (comparing population exposure model estimates to central-site measurements). We compared exposure metrics and exposure errors within and across pollutants, and present derived attenuation factors (ratio of observed to true coefficient for pollutant of interest) for single and bipollutant model coefficients.

Results: Pollutant concentrations and their exposure errors were moderately to highly correlated (typically > 0.5), especially for CO, NO_x, and EC (i.e., "local" pollutants); correlations differed across exposure metrics and types of exposure error. Spatial variability was evident, with variance of exposure error for local pollutants ranging from 0.25–0.83 for $\delta_{spatial}$ and $\delta_{total.}$ The attenuation of model coefficients in single and bipollutant epidemiologic models relative to the true value differed across types of exposure error, pollutants, and space.

Conclusions: Under a classical exposure error framework, attenuation may be substantial for local pollutants due to $\delta_{spatial}$ and δ_{total}, with true coefficients reduced by a factor typically < 0.6 (results vary for $\delta_{population}$ and regional pollutants).

Introduction

Most epidemiologic studies of the health effects of ambient air pollution have focused on adverse effects associated with single pollutants. In reality, humans are simultaneously exposed to a complex mixture of pollutants, which can vary both spatially and temporally (Dominici et al. 2010). Epidemiological analyses that have examined multipollutant health effects have typically relied on ambient monitoring data to estimate exposures (Hoffmann et al. 2012; Tolbert et al. 2007). Measurements from federal or state ambient monitoring networks often lack spatial and temporal coverage (Goldman et al. 2010; Sarnat et al. 2010) and do not account for exposures in different microenvironments (e.g. in-vehicle and inside the home) where infiltration (Sarnat et al. 2006; Weisel et al. 2005) and indoor sources (Baxter et al. 2007; Meng et al. 2009) can contribute substantially. There is therefore a potential for exposure measurement error that can lead to effect attenuation and reduced statistical power when measurements from ambient monitors are used as the exposure estimate in an epidemiological study.

Complex relationships may exist between exposures to various pollutants, and between the exposure error associated with each pollutant. The magnitude of the exposure error may differ across pollutants (Tolbert et al. 2007). For example, pollutants with primarily local sources [e.g., carbon monoxide (CO), nitrogen oxides (NO_x), and elemental carbon (EC)] exhibit significant spatial heterogeneity (Goldman et al. 2010; Sarnat et al. 2010; Strickland et al. 2013) that may not be captured by central-site (CS) ambient monitors. Exposures estimated from ambient monitors for these pollutants may be associated with more error than monitor-based estimates for pollutants that are more spatially homogeneous [e.g., fine particulate matter ($PM_{2.5}$), sulfate (SO_4), and ozone (O_3)]. When exposure estimates do not take into account exposure factors such as time-location-activity patterns (including time spent indoors) (Monn 2001; Setton et al. 2011),

significant indoor sources (e.g., gas stoves contributing to NO_2 exposures) (Williams et al. 2012), or housing characteristics (e.g., air exchange rate (AER), or pollutant infiltration) (JA Sarnat et al. 2013), exposure error may be greater.

Previous studies have predominantly focused on quantifying and accounting for exposure error in single-pollutant models (Sarnat et al. 2010; Setton et al. 2011; Strickland et al. 2013). One study has focused on a method for analysis of health effects in multipollutant studies that is resistant to measurement error (Zeka and Schwartz 2004). Amongst other findings, the study found an association between CO and daily mortality where traditional analysis did not, suggesting that a high degree of measurement error due to spatial heterogeneity of CO concentrations may be contributing to the difference in findings. A second provides alternative methods for estimating the effect of two exposures on an outcome, which reduce bias at the cost of a small to moderate reduction in power (Schwartz and Coull 2003).

The objective of this analysis is to examine exposure errors for multiple pollutants and provide insights on the potential for bias and attenuation of effect estimates in single and bipollutant epidemiological models. We utilize this approach to examine the robustness of the association for a pollutant of interest when a second pollutant is controlled for, i.e., to examine the attenuation due to measurement errors present in both pollutants. In a previous analysis, alternative exposure estimates for ambient-generated $PM_{2.5}$, EC, SO_4, CO, NO_x, and O_3 were developed, and spatiotemporal patterns for each estimate were characterized in comparison to CS monitor measurements (Dionisio et al. 2013). The exposure estimates were used in an epidemiological study in the Atlanta metropolitan area, using a time-series design to examine the association between daily exposure to ambient air pollution and daily emergency department (ED) visits for asthma/wheeze during a four year study period (1999-2002) (SE Sarnat et al.

2013). Using a modified set of the previously generated exposure estimates, we examined the exposure error and between-pollutant relationships, and quantified potential attenuation of model coefficients in single and bi-pollutant models at the ZIP code-level for ambient-generated $PM_{2.5}$, EC, SO_4, CO, NO_x, and O_3 in Atlanta, Georgia.

Methods

Estimates of exposure

Three estimates of daily exposure to ambient $PM_{2.5}$, EC, SO_4, CO, NO_x, and O_3 were derived for 193 ZIP codes in the 20 county Atlanta, Georgia metropolitan area, for use in an epidemiologic analysis of cardiovascular and respiratory outcomes based on data from ED visits. Each metric builds on previous metrics, incorporating the coarser measurements and model estimates, and becoming increasingly more finely resolved. The three estimation approaches, or "metrics," for exposure to ambient pollution include:

1) CS: central-site measurements
2) AQ: hybrid of statistical model for regional background and dispersion model for the local contribution to ambient air quality
3) PE: stochastic population exposure model

The AERMOD (American Meteorological Society/Environmental Protection Agency Regulatory Model) dispersion model (version 09292), is used for the local contribution to the AQ metric, and the U.S. EPA's Stochastic Human Exposure and Dose Simulation (SHEDS) model (Burke et al. 2001) is used for the PE metric. The contribution from indoor sources is not included in any of the approaches, due to the desire to associate exposure to ambient pollution with the health outcome. All three approaches estimate exposures to ambient pollution at each ZIP code centroid

in the study area. Daily estimates (8-hr maximum for O_3, 24-hr average for other pollutants) from 1999-2002 are generated for the three exposure estimation approaches.

Central-site measurements

CS measurements for each pollutant are from the Southeastern Aerosol Research Characterization network, the Assessment of Spatial Aerosol Composition in Atlanta network, and the U.S. Environmental Protection Agency's (EPA's) Air Quality System monitoring network (see Supplemental Material, Figure S1). Details regarding measurement methods, imputations for filling in missing data, and previous work using these monitors to characterize background air pollution levels are provided in previous publications (Dionisio et al. 2013; Metzger et al. 2004; Tolbert et al. 2000). Daily 24-hr average concentrations of $PM_{2.5}$, EC, and SO_4 were taken directly from monitor measurements. Hourly concentrations for CO and NO_x were aggregated to 24-hr averages, and hourly concentrations for O_3 were aggregated to daily 8-hr maximum concentrations.

Air quality model estimates

Air quality model estimates were obtained by combining local- and regional-scale model results (based on CS measurements) to account for all major atmospheric processes, including local contributions (driven by local-scale variation in pollutant emissions and meteorology) and regional contributions (background levels associated with large-scale synoptic patterns). The sum of the modeled regional background contribution and the local contribution is computed hourly to obtain total modeled ambient air concentrations at each ZIP code centroid for each pollutant being studied. To obtain estimates of the regional background contribution, we modified an approach developed to provide population-weighted daily averages of ambient pollution concentrations (Ivy et al. 2008) to provide spatially resolved hourly estimates of

regional background pollution by removing local source impacts modeled by hour-of-day and day-of-week. Local-scale pollutant contributions for $PM_{2.5}$, EC, SO_4, CO, and NO_x at each ZIP code centroid were modeled using the AERMOD dispersion model version 09292 (Cimorelli et al. 2005), which simulates concentrations of pollutants directly emitted into the atmosphere. Because O_3 is formed by photochemical processes and has no direct emissions, O_3 concentrations were not modeled with AERMOD. Similarly, the SO_4 concentrations estimated from AERMOD are from direct vehicle exhaust emissions, and do not include the secondary SO_4 contribution due to photochemical transformations in the atmosphere. Further details on methodology and modeling of the regional contribution, local-scale contribution, and computation of the AQ metric estimates can be found in (Dionisio et al. 2013).

Population exposure model estimates

We used the SHEDS model (Burke et al. 2001) to derive PE model estimates of daily population exposures to ambient pollution at each ZIP code centroid. SHEDS is a stochastic population exposure model that uses a probabilistic approach to estimate personal exposures for simulated individuals of a defined population based on ambient concentrations, distributions of residential AERs and particle infiltration parameters (i.e., penetration factors and deposition rates), and time spent in various microenvironments (e.g., home, office, school, vehicle) from a large database of human activity diaries. Key inputs to the model are the AQ metric estimates described above, time-location-activity data from the U.S. EPA's Consolidated Human Activity Database (McCurdy et al. 2000), spatially varying local air exchange rates (AERs) (JA Sarnat et al. 2013), and census tract-level home-to-work commuting data (U.S. Department of Transportation Bureau of Transportation Statistics 2000; U.S. EPA 2012). Penetration and decay parameters used in the model are specific to each pollutant, but do not vary spatially or temporally (see

Supplemental Material, Tables S1 and S2). To derive model estimates for exposures to ambient pollution, consistent with the CS and AQ metrics, we excluded contributions from indoor source emissions for this analysis. For additional details, see Supplemental Material, Population exposure metric.

Statistical analyses

ZIP code-level summary statistics for each exposure metric, for each pollutant, were computed. Statistics include the annual mean normalized pollutant concentrations, and the variance across days of the normalized pollutant concentrations, for each exposure metric. To allow for comparisons across pollutants, ZIP code-specific pollutant concentrations for each exposure metric were normalized by dividing the daily pollutant concentration by the annual average CS measurement for that pollutant. The magnitude and spatial variability of normalized pollutant concentrations were then compared across pollutants and exposure metrics.

One standard approach for examining the health effects of multiple pollutants is to include each pollutant as an independent risk factor simultaneously, in a single epidemiologic model (Bell et al. 2007; Tolbert et al. 2007). To assess the impacts of exposure error on health risk estimates in a multipollutant model, the correlation between the exposure estimates, the degree of exposure error for each pollutant, and the correlation of exposure errors between pollutants must all be considered (Zeger et al. 2000; Zidek et al. 1996).

In this analysis, exposure error, δ, is calculated as the difference between two exposure metrics. Three types of exposure error ($\delta_{spatial}$, $\delta_{population}$, and δ_{total}) are presented. The exposure error due to a lack of spatial refinement in the exposure estimate is represented by $\delta_{spatial} = AQ - CS$, as our air quality models add spatial variability to the AQ metric compared to CS measurements (CS

measurements lack spatial variability as the same CS measurement was used to represent exposure in each ZIP code). Exposure error introduced when human exposure factors are not included in an exposure estimate is represented by $\delta_{population}$ = PE – AQ. Our PE metric includes variability due to human exposure factors such as time-location-activity patterns of individuals, commuting patterns, and infiltration of ambient pollutants indoors. A third type of exposure error, δ_{total} = PE – CS, represents the combined exposure error when both spatial variability and human exposure factors are not accounted for. Note that δ_{total} does not represent all potential sources of exposure error that may be present in a study, but instead represents the total exposure error that we are able to assess in this analysis. As with the pollutant concentrations, daily, ZIP code-specific estimates of exposure error are normalized by dividing by the annual average CS measurement for that pollutant, to allow for comparison across pollutants and types of exposure error. The variance calculated across days of the normalized exposure error is also presented, to aid in estimating the degree of bias and attenuation of model coefficients.

The between-pollutant Pearson correlations over time for each exposure estimation approach, and for each type of exposure error, are calculated, to provide information on the collinearity of exposure estimates and exposure error which must be accounted for in a multipollutant model. Correlations are calculated for each ZIP code individually, so that the range of correlations can be compared across the study domain.

Estimates of the level of attenuation of model coefficients for single- and bipollutant models are presented, to aid in the interpretation of future epidemiologic models including two or more pollutants. The attenuation factor (λ) for a classical error, single pollutant framework is calculated as:

$$\lambda = 1/\{1 + [var(\delta) / var(x_{fine})]\} \quad [1]$$

$$\beta_{observed} = \lambda \, x \, \beta_{true} \quad [2]$$

where δ = exposure error, x_{fine} = the exposure metric with the greater degree of refinement (i.e., increased spatial resolution, or inclusion of weighting by population factors), var(x_{fine}) = the variance across days of x_{fine}, and β = model coefficients. We assume that the related epidemiologic analysis fits a time-series model separately for each ZIP code, thus β represents the association between the health outcome and the daily pollutant exposure. For simplicity, we present the attenuation factor λ_{x_1} for pollutant x_1 in a bipollutant model, assuming pollutant x_2 has no effect ($\beta_{x_2} = 0$), given by the diagonal elements of:

$$\lambda_{x_1} = \mathbf{S} \, (\mathbf{S} + \mathbf{V})^{-1} \quad [3]$$

$$\beta_{observed,x_1} = \lambda_{x_1} x \, \beta_{true,x_1} \quad [4]$$

where **S** = covariance of the exposure metrics with the greater degree of refinement for x_1 and x_2, and **V** = covariance of the exposure errors for x_1 and x_2. For the single and bipollutant models, an attenuation factor of $\lambda = 1$ indicates no attenuation (i.e., $\beta_{observed} = \beta_{true}$) and $\lambda = 0$ (i.e., $\beta_{observed} = 0$) indicates null results. An attenuation factor of $\lambda > 1$ indicates bias away from the null, and $\lambda < 0$ indicates that the estimated coefficient will be in the opposite direction of the true effect. As an example, the λ associated with $\delta_{spatial}$ in a single pollutant model reflects the attenuation of model coefficients due to error from incomplete characterization of the spatial variation in the concentration of the pollutant in question.

All statistical analyses were completed in R version 2.15.1 (R Foundation for Statistical Computing, Vienna, Austria). All mapping was done in ArcGIS 10 (Esri, Redlands, CA).

Results

This study builds on previous work, where single-pollutant epidemiologic models were used to estimate the association between daily counts of ZIP code-level ED visits and ZIP code-specific exposures using the three metrics (JA Sarnat et al. 2013; SE Sarnat et al. 2013). Related analyses also showed that the temporal variation in the AQ measure was not always more variable than temporal variation in the CS metric (Dionisio et al. 2013). The goal of this analysis was to examine exposure error and between-pollutant relationships, and how these differ by pollutant pair and exposure metric. Using the empirical covariance structures allowed us to assess potential attenuation of model coefficients in bipollutant epidemiologic models.

Summary statistics for exposure metrics

Figure 1a presents the boxplots of ZIP code-specific normalized exposure metrics averaged across the entire study period (see Supplemental Material, Figure S2a, for an expanded version that shows the full distributions for each metric). Distributions of pollutant concentrations differ by exposure metric, with the PE estimates being consistently equal to (CO) or lower than (NO_x, EC, $PM_{2.5}$, SO_4, O_3) AQ estimates, due to the penetration and decay parameters used in the SHEDS model (Supplemental Material, Tables S1 and S2). There was no spatial variability for the CS metric, since the same CS measurement was used for all ZIP codes. However, there was considerable spatial variability [i.e., variation among the 193 ZIP code-specific estimates, as indicated in the boxplot figures by a larger interquartile range (IQR, represented by the upper and lower bounds of each box), and a larger range from the 5^{th} to 95^{th} percentiles (the lower and upper whiskers of each boxplot, respectively)] when AQ or PE modeling was used. For all pollutants except CO, PE estimates exhibited a lower degree of spatial variability than AQ estimates. Local pollutants (CO, NO_x, EC) had relatively more spatial variability in their AQ and

PE metrics than regional pollutants ($PM_{2.5}$, SO_4, O_3), which was expected due to the variation of local source emissions such as traffic at the ZIP code level.

Between-pollutant correlations of exposure metrics

Boxplots of pairwise Pearson correlation coefficients of daily, ZIP code-specific exposure metrics for local-local and regional-regional pollutant pairs are presented in Figure 1b. All local-local and regional-regional pollutant pairs showed moderate to strong positive correlations for each metric, however correlations for regional-regional pollutant pairs tended to be lower. For the regional-regional pollutant pairs, the median correlation for each pair was consistent across the three exposure metrics. In contrast, for each local-local pollutant pair, the correlation coefficient for CS measurements was lower than the median correlation for the AQ and PE metrics. Correlations of local-regional pollutant pairs were more varied and typically weaker than local-local and regional-regional pollutant pair correlations, with the exception of correlations of CO, NO_x, and EC with $PM_{2.5}$ (Supplemental Material, Figure S2b).

Spatial variability (described by the width of the boxplot) was present to varying degrees for correlations within the AQ and PE metrics, with more spatial variability present for local-local pollutant correlations than regional-regional pollutant correlations, especially for the CO-EC and NO_x-EC pairs (Figure 1b). The degree of spatial variability for regional-regional pollutant pairs was similar for both the AQ and PE metrics. There was no spatial variability present for the between-pollutant correlations of exposure for the CS metric, since the same CS measurement was used for each ZIP code.

Summary statistics for exposure error

Figure 2a compares the magnitude and spatial variability of the three types of normalized exposure error ($\delta_{spatial}$, $\delta_{population}$, and δ_{total}) across pollutants (see Supplemental Material, Figure S3a for the full distribution). The distribution of exposure error across ZIP codes was mostly negative (indicating the exposure metric with a greater degree of refinement had a lower magnitude), though exposure errors were positive for a small number of ZIP codes. The magnitude of the exposure error varied by type of error, with the absolute value of exposure error greater for $\delta_{population}$ and δ_{total} than for $\delta_{spatial}$ for regional pollutants, and mixed results for local pollutants. With $\delta_{spatial}$ near zero for the regional pollutants (median absolute value of $\delta_{spatial}$ across ZIP codes less than 0.12, indicating similar magnitude for CS measurements and AQ estimates), their total exposure error (δ_{total}) consisted mostly of exposure error due to human exposure factors ($\delta_{population}$, indicating greater differences in magnitude for AQ estimates relative to PE estimates.)

To assess the potential for spatially differential exposure error, we compared the spatial variability of exposure errors across ZIP codes. With the exception of $\delta_{population}$ for CO, the spatial variability of exposure error was greater for local pollutants than regional pollutants (Figures 2a, 2c). For local pollutants, spatial variability was present to varying degrees across all types of error [smallest range of 5^{th} to 95^{th} percentiles of normalized exposure error -0.64 to -0.13 (EC, $\delta_{population}$), largest range -0.85 to 1.73 (NO_x, $\delta_{spatial}$)], with the exception of $\delta_{population}$ for CO, which was near zero due to use of a penetration factor of 1 (i.e., assuming free flow of outdoor and indoor air) in the SHEDS model for CO (Supplemental Material, Table S2). In contrast, regional pollutants exhibited little spatial variability across types of exposure error and the degree of spatial variability was consistent within a pollutant, across types of error.

Between-pollutant correlations of exposure error

The collinearity of exposure error was examined based on Pearson correlations between daily exposure error for local-local and regional-regional pollutant pairs (Figure 2b, see Supplemental Material, Figure S3b, for local-regional pairs). The correlation of exposure error was highly dependent on both pollutant pair and type of exposure error. Between-pollutant correlations of exposure error were mostly positive, though there were some ZIP codes with negative correlations, especially for CO. Correlation of exposure error due to a lack of spatial refinement ($\delta_{spatial}$) was moderate to strong for local-local pollutant pairs (median correlation over all ZIP codes ranging from 0.65 to 0.76), and relatively weak for regional-regional pollutant pairs (median correlation ranging from 0.03 to 0.21). Correlation for $\delta_{population}$ showed a near opposite trend, with weak, negative correlations of $\delta_{population}$ for CO-NO_x and CO-EC (-0.13 and -0.19 respectively), and moderate to strong positive correlations of $\delta_{population}$ for NO_x-EC (0.85) and the regional-regional pollutant pairs (ranging from 0.52 to 0.77). The magnitude of the correlation of total exposure error (δ_{total}) between local-local and regional-regional pollutant pairs varied, with median correlations of δ_{total} across ZIP codes ranging from 0.35 to 0.72 (Table 1, Figure 2b).

Local-local and regional-regional pollutant pairs showed a moderate degree of spatial variability in the correlation of $\delta_{spatial}$ (Figure 2b). The patterns of spatial variability of the correlation of $\delta_{population}$ are more varied, with local-local pollutant pairs showing a larger degree of spatial variability compared to regional-regional pollutant pairs (5^{th} to 95^{th} percentile for correlation coefficients of 0.56 to 0.93 for NO_x-EC, -0.42 to 0.63 for CO-NO_x, and -0.46 to 0.59 for CO-EC). Though there was a large range of correlations across ZIP codes for $\delta_{population}$ for CO-NO_x and CO-EC in particular, the bulk of the correlations across the study area were relatively weak (25^{th} to 75^{th} percentile for correlation coefficients of -0.27 to 0.21 for CO-NO_x and -0.37 to 0.17 for

CO-EC. As reflected in comparisons of $\delta_{spatial}$ and $\delta_{population}$, we see greater spatial variability in the correlation of δ_{total} for the local-local pollutant pairs, and very little spatial variability in the correlation of δ_{total} for the regional-regional pairs.

Variance of exposure error

For regional pollutants ($PM_{2.5,}$ SO_4, and O_3), variance across days of the normalized exposure error had very little spatial variability (i.e., boxplots of the variance of normalized exposure error are narrow), and was less than 0.20 for any type of error in any ZIP code (Figure 3a). In comparison, with the exception of $\delta_{population}$ for CO, variance of the exposure error, and spatial variability of the variance, was present for local pollutants (Figure 3, see Supplemental Material, Figure S4, for the full distribution). For the local pollutants, the magnitude and spatial variability of the variance of normalized error differed depending on pollutant and type of error, with the variance of $\delta_{spatial}$ and $\delta_{population}$ for NO_x having the largest range of spatial variability, while the variance of exposure error for EC exhibited more modest spatial variability.

Attenuation of model coefficients

By compiling empirically determined parameters related to the between-pollutant relationships and their associated exposure error (Table 1), and utilizing Equation 3, we were able to quantify the potential attenuation of model coefficients in a bipollutant model. Table 1 presents the median values across all ZIP codes of the correlations over time and variances across days for pollutant concentrations and their associated exposure errors. The individual ZIP code-specific values of these parameters (Figures 1b, 2b, and 3a, for full range of parameter values across all ZIP codes see Supplemental Material, Figures S2a, S3b, and S4) were used to calculate attenuation factors presented in Figure 4.

Figure 4 presents the potential attenuation factors for single and bipollutant epidemiologic models, based on empirical estimates of the relationships between exposure metrics and their exposure error. Attenuation factors presented for bipollutant models assume that one pollutant has a true effect on the health outcome, while the other pollutant has no effect. For $\delta_{spatial}$, we see a clear distinction between local and regional pollutants, with more attenuation (typically $\lambda < 0.6$) for both single and bipollutant models of local pollutants, and less attenuation for regional pollutants (typically $\lambda > 0.6$) (Figure 4a; noting that $\lambda = 1$ indicates no attenuation (i.e., $\beta_{observed} = \beta_{true}$), $\lambda = 0$ indicates null results, $\lambda > 1$ indicates bias away from the null, and $\lambda < 0$ indicates that the estimated coefficient will be in the opposite direction of the true effect). Addition of a co-pollutant appears to increase attenuation. Results for $\delta_{population}$ and δ_{total} are more varied, with attenuation factors depending on the pollutant and co-pollutant (Figures 4b, 4c). For $\delta_{spatial}$ and δ_{total}, we see notable spatial variability in the attenuation factors (evidenced by wider boxplots) for local pollutants (except for δ_{total} for NO_x). For $\delta_{population}$, and regional pollutants for $\delta_{spatial}$ and δ_{total}, the degree of spatial variability depends on the type of exposure error, pollutant, and co-pollutants.

For comparison, we present the attenuation factors for a bipollutant model with one local (NO_x), and one regional ($PM_{2.5}$) pollutant (Figure 4d), which shows significant differences in the attenuation factor across types of exposure error, but smaller differences between single and bipollutant models. Attenuation factors for bipollutant models for all local-regional pollutant pairs are presented in Supplemental Material, Figure S5. Results occasionally showed bias away from the null ($\lambda > 1$) for some bipollutant combinations because of the strong correlations in both pollutant concentrations and exposure errors.

Discussion

Improved understanding of the degree of exposure error among pollutants and their dependent structure is needed to properly interpret results from epidemiologic models including multiple pollutants. Through examination of three different exposure metrics, and three types of associated exposure measurement errors, we were able to empirically estimate bipollutant relationships and the potential for attenuation of model coefficients in related bipollutant epidemiologic models. For bipollutant models with local-local pollutant pairs, $\delta_{spatial}$ and δ_{total} were likely to introduce attenuation of model coefficients due to high correlations between local pollutant concentrations ($corr(x_1, x_2) > 0.80$ for all local-local pollutant pairs), unequal and non-zero variance of the exposure error for each pollutant ($0.25 < var(\delta) < 0.83$), and moderate to high correlation of the exposure error for each pollutant pair ($corr(\delta_1, \delta_2) > 0.52$ excepting CO-NO_x for δ_{total}). For regional-regional pollutant pairs, attenuation of model coefficients was likely to be minimal, due to relatively low variance of the exposure error ($var(\delta) < 0.16$ for all regional pollutants and types of exposure error). The empirical quantification of the above parameters resulted in a predicted attenuation factor due to $\delta_{spatial}$ that was typically < 0.6 for single and bipollutant models of local pollutants, with less attenuation for regional pollutant models (typically $\lambda > 0.6$), and more varied results for $\delta_{population}$ and δ_{total}.

The mean over all ZIP codes of AQ metric estimates incorporating both regional background and local pollution contributions are similar in magnitude to CS measurements, though particularly for local pollutants AQ metric estimates can exhibit spatial variability depending on local traffic patterns within the ZIP code. With the exception of CO, PE metric estimates for each pollutant are lower than their corresponding CS measurement, due to infiltration and decay parameters incorporated into the SHEDS human exposure model, and the inclusion of time-activity data

based on diaries which indicated that individuals spend the majority of their time indoors. PE metric estimates for CO are similar to AQ metric estimates, as the penetration parameter for CO was set to 1 (i.e., assuming full penetration of CO from the outdoor to the indoor environment). Pollutant contributions from indoor sources were not included in this study, thus the PE metric represents indoor and outdoor exposures to ambient pollution originating outdoors only.

Air quality models introduce spatial variability into AQ exposure estimates that is not captured when a single CS measurement is used for all ZIP codes in a study area. Spatial variability was much greater for pollutants with predominantly local sources (CO, NO_x, EC) compared to pollutants dominated by regional source contributions ($PM_{2.5}$, SO_4, O_3). This increase in spatial variability for local pollutants was mainly due to differences in traffic volume and patterns among different ZIP codes. Between-pollutant correlations were strong for local-local pollutant pairs, and moderate to strong for regional-regional pairs, reflecting the common emissions sources contributing to pollutant concentrations within each pair.

As expected, total exposure error (δ_{total}) for regional pollutants was made up mostly of exposure error due to human exposure factors (e.g., time-activity patterns, AER in the home), with a small contribution from unmeasured spatial variability. In contrast, for the local pollutants NO_x and EC, there were substantial exposure error contributions from both human exposure factors and spatial heterogeneity in ambient concentrations. For CO, we see a near zero contribution from $\delta_{population}$ (due to full penetration of CO indoors).

Potential impact on epidemiologic model coefficients

Here we discuss the impact of attenuation on epidemiologic model coefficients in a multipollutant model. In a multipollutant model, the absolute magnitude of this bias will depend

on the variance of the exposure error, the correlation between exposure estimates, and the correlation between exposure errors.

This analysis builds upon the hypothetical simulation presented in Zeger et al. (2000) of predicted bias in regression coefficients in a bipollutant epidemiologic model. In a bipollutant model, we may not be concerned with bias if two regional pollutants are included, due to the near-zero ($\delta_{spatial}$) and very low ($\delta_{population}$ and δ_{total}) variance of exposure error for regional pollutants (Figure 3a, Table 1). However in a bipollutant model including two local pollutants, there is the potential for bias and attenuation of model coefficients due to a higher degree of variance of exposure error. The effect in bipollutant models including one local and one regional pollutant will vary, depending on the pollutant pair. In addition, empirically determined attenuation factors for single- and bipollutant models show that the potential for attenuation in the estimated effects can be quite substantial for many pollutants and exposure error types, in particular for local pollutants (with the exception of $\delta_{population}$ for CO) (Figure 4).

In addition to the potential for bias, results presented here show that spatial variability is present in the exposure error for local pollutants, and in the between-pollutant correlations of exposure error for local-local pollutant pairs. Figure 3b visually displays how the variance of spatial exposure error for NO_x changes across the study domain, with variance of spatial exposure error highest in the urban core (within and immediately surrounding the blue circular line indicating a major road), lowest in the central ring of our study domain (ring surrounding the urban core), and increasing slightly again as you extend to the western boundary of the study domain. These results highlight the importance of characterizing intra-urban variations in exposure to avoid spatially varying differential exposure error. This is a particular concern when examining effect modification of air pollution health risks obtained without spatially resolved exposure estimates.

For example, observed effect modification by ZIP code-level socioeconomic measures (SE Sarnat et al. 2013), which exhibit strong spatial patterns, may be due at least in part to varying degrees of attenuation bias from spatially differential exposure error.

Finally, when multiple pollutants are included simultaneously in a model of associations with health outcomes, bias away from the null may also occur. "Effect transfer" (Zidek et al. 1996) occurs when two correlated pollutants are measured with differential exposure error, and the effect of the pollutant measured with more error is transferred to the pollutant measured with less error. In this case a pollutant without an effect on an outcome may become associated with it.

Limitations

Limitations of this study include uncertainties in the more refined exposure metric estimates (including that small area variations in pollutant concentrations may not be resolved due to sparsely distributed measurements used as inputs), and the exclusion of the influence from indoor sources. While it is commonplace to use exposure to ambient sources as a proxy for an individual's total exposure in an epidemiologic study, the inclusion of indoor sources would further enhance study findings.

Results presented here may be generally applicable to study areas with similar source contributions (e.g. local sources predominantly traffic related) and housing characteristics (e.g., low AERs). For any study area, methods and models presented here may be applied if appropriate input data sets are used. With the exception of the locally derived AERs, all Atlanta-specific input data sets (e.g., central site pollutant measurements, traffic patterns, local emissions) were extracted from larger, publically available databases maintained by federal and state agencies, thus similar input data sets for any study area could be compiled. If local AERs were

not available, estimates could be made based on published distributions of AERs from various parts of the country.

Though the magnitudes of effects may differ, we expect that general conclusions from our analysis will be applicable to other geographic areas. For example, most study regions will have some pollutant concentrations dominated by regional sources which are likely to remain spatially homogeneous, and some pollutant concentrations dominated by local sources which are likely to be spatially heterogeneous within the study area. Thus we believe that our conclusions about the spatial variability of exposure error being present, and the general likelihood of bias due to measurement error for certain pollutants are likely to apply across studies.

In calculating the attenuation factor, we assumed a classical exposure measurement error framework. We recognize this is a strong assumption, but feel it is more appropriate than assuming a Berkson error framework because the CS does not necessarily represent "average" exposure for any ZIP code on any given day. Since exposure measurement error is likely to contain both classical and Berkson type errors, depending on the pollutants and study design, the assumption of a solely classical error framework implies limited applicability. Moreover, our assessment of attenuation assumes the effect estimate is not subject to residual confounding, the association between pollutant concentration and the health outcome is linear, and there is no effect modification between the pollutant association in a bipollutant model. Further, we have implicitly assumed that the only bias present is additive (current analysis does not consider multiplicative bias); this should not impact the regression slope. Lastly, though empirical covariance structures and exposure errors have been used to quantify potential attenuation in bipollutant models (assuming only one pollutant has an effect on the health outcome), this analysis does not address the potential for effect transfer in a bipollutant model when both

pollutants have an effect on the health outcome. A simulation study including the covariance structures of data presented here is warranted to quantify the effect on model coefficients in a multipollutant model.

In addition to the role of exposure error, additional factors must be considered as researchers further investigate epidemiologic analyses including multiple pollutants. These include the possibility of non-linear relationships of the various pollutants with the health outcome, interaction or synergism amongst pollutants included in a single epidemiologic model, and the possibility of the high correlation we have seen amongst pollutants leading to one pollutant appearing to be associated with the health outcome in an epidemiologic model when a correlated pollutant is the true causal association. A future simulation study that examines the applicability of the classical exposure measurement error framework and the degree of effect attenuation and transfer is warranted.

Conclusions

This study is one of the first to quantify the effects of correlated exposure measurement error in bipollutant models (Chang et al. 2011). To the authors' knowledge, this is the first to look in detail at the effects of spatial variation using dispersion models and stochastic personal exposure simulators in a multipollutant context. We have used empirical relationships to show the potential for bias, particularly effect attenuation, in epidemiologic model coefficients for bipollutant models, particularly for local pollutants (CO, NO_x, EC), due to the presence of variance in the exposure error, and correlation between pollutants and their errors. Further, we have seen evidence of the potential for spatially varying attenuation and bias due to the spatial variability present in these parameters on the ZIP code-level. As researchers move towards

multipollutant approaches, we must recognize the potential effects on model coefficients depending on the relationships that exist between pollutants and their error.

References

Baxter LK, Clougherty JE, laden F, Levy JI. 2007. Predictors of concentrations of nitrogen dioxide, fine particulate matter, and particle constituents inside of lower socioeconomic status urban homes. J Expo Sci Environ Epidemiol 17:433-444.

Bell ML, Kim JY, Dominici F. 2007. Potential confounding of particulate matter on the short-term association between ozone and mortality in multisite time-series studies. Environ Health Perspect 115:1591-1595.

Burke JM, Zufall MJ, Özkaynak H. 2001. A population exposure model for particulate matter: Case study results for $PM_{2.5}$ in Philadelphia, PA. J Expo Anal Environ Epidemiol 11:470-489.

Chang HH, Peng RD, Dominici F. 2011. Estimating the acute health effects of coarse particulate matter accounting for exposure measurement error. Biostatistics 12:637-652.

Cimorelli AJ, Perry SG, Venkatram A, Weil JC, Paine RJ, Wilson RB, et al. 2005. Aermod: A dispersion model for industrial source applications. Part i: General model formulation and boundary layer characterization. J Appl Meteorol 44:682-693.

Dionisio KL, Isakov V, Baxter L, Sarnat JA, Sarnat SE, Burke J, et al. 2013. Development and evaluation of alternative approaches for exposure assessment of multiple air pollutants in Atlanta, Georgia. J Expo Sci Environ Epidemiol 23:581-592.

Dominici F, Peng RD, Barr CD, Bell ML. 2010. Protecting human health from air pollution: Shifting from a single-pollutant to a multipollutant approach. Epidemiology 21:187-194.

Goldman GT, Mulholland JA, Russell AG, Srivastava A, Strickland MJ, Klein M, et al. 2010. Ambient air pollutant measurement error: Characterization and impacts in a time-series epidemiologic study in Atlanta. Environ Sci Technol 44:7692-7698.

Hoffmann B, Luttmann-Gibson H, Cohen A, Zanobetti A, Souza Cd, Foley C, et al. 2012. Opposing effects of particle pollution, ozone, and ambient temperature on arterial blood pressure. Environ Health Perspect 120:241-246.

Ivy D, Mulholland JA, Russell AG. 2008. Development of ambient air quality population-weighted metrics for use in time-series health studies. J Air Waste Manag Assoc 58:711-720.

McCurdy T, Glen G, Smith L, Lakkadi Y. 2000. The national exposure research laboratory's consolidated human activity database. J Expo Anal Environ Epidemiol 10:566-578.

Meng QY, Spector D, Colome S, Turpin B. 2009. Determinants of indoor and personal exposure to $PM_{2.5}$ of indoor and outdoor origin during the RIOPA study. Atmos Environ 43:5750-5758.

Metzger KB, Tolbert PE, Klein M, Peel JL, Flanders WD, Todd K, et al. 2004. Ambient air pollution and cardiovascular emergency department visits. Epidemiology 15:46-56.

Monn C. 2001. Exposure assessment of air pollutants: A review on spatial heterogeneity and indoor/outdoor/personal exposure to suspended particulate matter, nitrogen dioxide and ozone. Atmos Environ 35:1-32.

Sarnat JA, Sarnat SE, Flanders WD, Chang HH, Mulholland J, Baxter L, et al. 2013. Spatiotemporally-resolved air exchange rate as a modifier of acute air pollution related morbidity in Atlanta. J Expo Sci Environ Epidemiol 23:606-615.

Sarnat SE, Coull BA, Ruiz PA, Koutrakis P, Suh HH. 2006. The influences of ambient particle composition and size on particle infiltration in Los Angeles, CA, residences. J Air Waste Manag Assoc 56:186-196.

Sarnat SE, Klein M, Sarnat JA, Flanders WD, Waller LA, Mulholland JA, et al. 2010. An examination of exposure measurement error from air pollutant spatial variability in time-series studies. J Expo Sci Environ Epidemiol 20:135-146.

Sarnat SE, Sarnat JA, Mulholland J, Isakov V, Özkaynak H, Chang H, et al. 2013. Application of alternative spatiotemporal metrics of ambient air pollution exposure in a time-series epidemiological study in Atlanta. J Expo Sci Environ Epidemiol 23:593-605.

Schwartz J, Coull BA. 2003. Control for confounding in the presence of measurement error in hierarchical models. Biostatistics 4:539-553.

Setton E, Marshall JD, Brauer M, Lundquist KR, Hystad P, Keller P, et al. 2011. The impact of daily mobility on exposure to traffic-related air pollution and health effect estimates. J Expo Sci Environ Epidemiol 21:42-48.

Strickland MJ, Gass KM, Goldman GT, Mulholland JA. 2013. Effects of ambient air pollution measurement error on health effect estimates in time-series studies: A simulation-based analysis. J Expo Sci Environ Epidemiol e-pub:doi: 10.1038/jes.2013.1016.

Tolbert PE, Klein M, Metzger KB, Peel J, Flanders WD, Todd K, et al. 2000. Interim results of the study of particulates and health in Atlanta (sophia). J Expo Sci Environ Epidemiol 10:446-460.

Tolbert PE, Klein M, Peel JL, Sarnat SE, Sarnat JA. 2007. Multipollutant modeling issues in a study of ambient air quality and emergency department visits in Atlanta. J Expo Sci Environ Epidemiol 17:S29-S35.

U.S. Department of Transportation Bureau of Transportation Statistics. 2000. Census transportation planning package (ctpp) 2000. Part 3 - journey to work. Available: http://transtats.bts.gov.

U.S. EPA. 2012. Total risk integrated methodology (trim) air pollutants exposure model documentation (trim.Expo/apex, version 4.5), volume ii: Technical support document. EPA-452/B-12-001b. Research Triangle Park, NC:US EPA Office of Air Quality Planning and Standards.

Weisel CP, Zhang J, Turpin BJ, Morandi MT, Colome S, Stock TH, et al. 2005. Relationships of indoor, outdoor, and personal air (RIOPA): Part 1, collection methods and descriptive analyses.Res Rep Health Eff Inst.

Williams R, Jones P, Croghan C, Thornburg J, Rodes C. 2012. The influence of human and environmental exposure factors on personal NO_2 exposures. J Expo Sci Environ Epidemiol 22:109-115.

Zeger SL, Thomas D, Dominici F, Samet JM, Schwartz J, Dockery D, et al. 2000. Exposure measurement error in time-series studies of air pollution: Concepts and consequences. Environ Health Perspect 108:419-426.

Zeka A, Schwartz J. 2004. Estimating the independent effects of multiple pollutants in the presence of measurement error: An application of a measurement-error-resistant technique. Environ Health Perspect 112:1686-1690.

Zidek JV, Wong H, Le ND, Burnett R. 1996. Causality, measurement error and multicollinearity in epidemiology. Environmetrics 7:441-451.

Table 1. Parameters impacting attenuation and bias in bivariate pollutant models of pairs of local pollutants (CO, NO_x, EC) or pairs of regional pollutants ($PM_{2.5}$, SO_4, O_3)[a]

Parameter	CO-NO_x	CO-EC	NO_x-EC	$PM_{2.5}$-SO_4	$PM_{2.5}$-O_3	SO_4-O_3
AQ Corr(x_1,x_2)	0.96	0.86	0.88	0.76	0.52	0.62
PE Corr(x_1,x_2)	0.86	0.84	0.80	0.76	0.49	0.60
$\delta_{spatial}$						
Var(δ_1)[b]	0.25	0.25	0.83	0.04	0.04	0.05
Var(δ_2)[b]	0.83	0.30	0.30	0.05	0.02	0.02
Corr(δ_1, δ_2)	0.73	0.65	0.76	0.21	0.03	0.11
$\delta_{population}$						
Var(δ_1)[b]	0.00	0.00	0.32	0.09	0.09	0.10
Var(δ_2)[b]	0.32	0.05	0.05	0.10	0.11	0.11
Corr(δ_1, δ_2)	-0.13	0.85	0.85	0.77	0.52	0.62
δ_{total}						
Var(δ_1)[b]	0.25	0.80	0.80	0.12	0.12	0.16
Var(δ_2)[b]	0.80	0.33	0.33	0.16	0.16	0.16
Corr(δ_1, δ_2)	0.35	0.72	0.72	0.70	0.41	0.57

[a]All values presented are median across all ZIP codes. The first pollutant indicated in each pair corresponds to x_1, the second to x_2.

[b]Var(δ) represents variance of normalized exposure error.

Figure Legends

Figure 1. Relationships between three exposure metrics. For each box, n=193, the bottom and top of the box represent 25th and 75th percentiles, the band near the middle of the box is the median, and the ends of the whiskers are the 5th and 95th percentiles.

Normalized exposure estimates from three methods. Note figure has been zoomed in for clarity. Supplemental Material, Figure S2a includes full extent of data and outliers.

Between-pollutant correlations of exposure for local-local and regional-regional pollutant pairs. Supplemental Material, Figure S2b, presents local-regional pollutant pairs.

Figure 2, Relationships between three types of exposure error. For each box, n=193, the bottom and top of the box represent 25th and 75th percentiles, the band near the middle of the box is the median, and the ends of the whiskers are the 5th and 95th percentiles.

Normalized exposure error. Note figure has been zoomed in for clarity. Supplemental Material, Figure S3a, includes full extent of data and outliers. ($\delta_{spatial}$ = AQ – CS; $\delta_{population}$ = PE – AQ; δ_{total} = PE – CS)

Between-pollutant correlations of exposure error for local-local and regional-regional pollutant pairs. Supplemental Material, Figure S3b, presents local-regional pollutant pairs.

$\delta_{spatial}$ for NO_x. Colored regions represent ZIP codes in the study area, blue and brown lines indicate major roads. Legend is grouped by percentile, where 5% = -0.85; 25% = -0.66; 50% = -0.18; 75% = 0.63; and 95% = 1.73.

Figure 3. Variance of exposure error. For each box, n=193, the bottom and top of the box represent 25th and 75th percentiles, the band near the middle of the box is the median, and the ends of the whiskers are the 5th and 95th percentiles.

Variance of normalized exposure error. Note figure has been zoomed in for clarity. Supplemental Material, Figure S4 includes full extent of data and outliers. ($\delta_{spatial}$ = AQ – CS; $\delta_{population}$ = PE – AQ; δ_{total} = PE – CS)

Variance of $\delta_{spatial}$ for NO_x. Colored regions represent ZIP codes in the study area, blue and brown lines indicate major roads. Legend is grouped by percentile, where 5% = 0.58; 25% = 0.72; 50% = 0.83; 75% = 1.14; and 95% = 4.14.

Figure 4. Attenuation of model coefficients in a classical error, single pollutant framework, and in bipollutant models, assuming one pollutant has an effect, and one pollutant has no effect. For each box, n=193, the bottom and top of the box represent the 25th and 75th percentiles, the band near the middle of the box is the median, and the ends of the whiskers are the 5th and 95th percentiles.

a) $\delta_{spatial}$ for local-local and regional-regional pollutant pairs.
b) $\delta_{population}$ for local-local and regional-regional pollutant pairs.
c) δ_{total} for local-local and regional-regional pollutant pairs.
d) $\delta_{spatial}$, $\delta_{population}$, and δ_{total} for a local-regional pollutant example.

Supplemental Material, Figure S5, presents all local-regional pollutant pairs. Solid boxplots are attenuation factors for single pollutant models, dashed boxplots are attenuation factors for bipollutant models. The first row of x-axis labels indicate the pollutant effect being considered

(e.g., “CO”). The second row of x-axis labels indicates the relevant model and the presence or absence of co-pollutants (e.g., “CO, CO+NO_x, CO+EC”). For example, for the set of 3 boxplots representing attenuation factors for CO, the 1st boxplot is the set of attenuation factors for CO in a single pollutant model, the 2nd boxplot is the attenuation factors for CO in a bipollutant model with NO_x, assuming NO_x has no effect, and the 3rd boxplot is the attenuation factors for CO in a bipollutant model with EC, assuming EC has no effect. $\lambda = 1$ indicates no attenuation, $\lambda = 0$ indicates null results.

Figure 1a

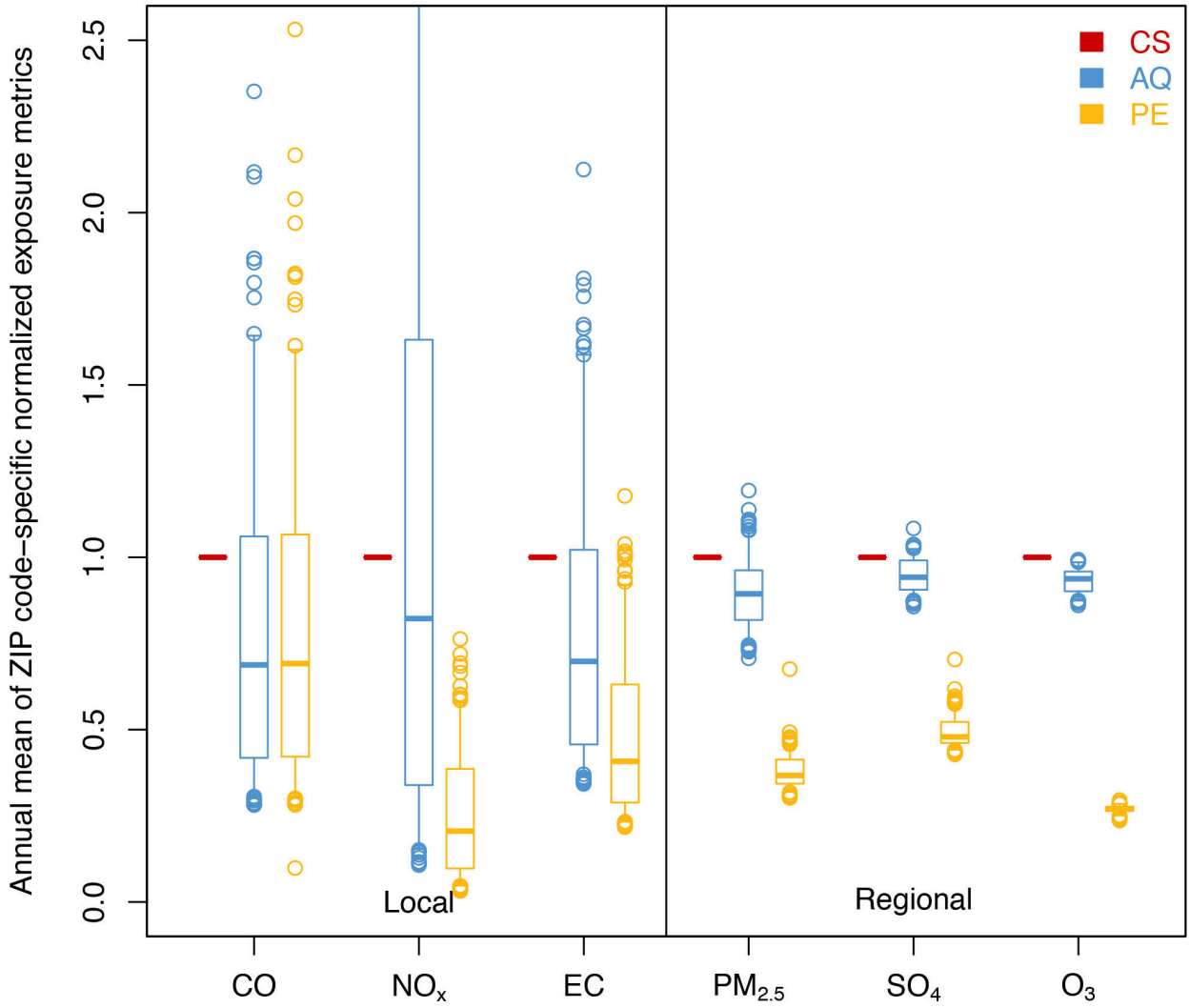

Annual mean of ZIP code−specific normalized exposure metrics
0.0
0.5
1.0
1.5
2.0
2.5
CS
AQ
PE
Local
Regional
CO
NOx
EC
PM2.5
SO4
O3

Figure 1b

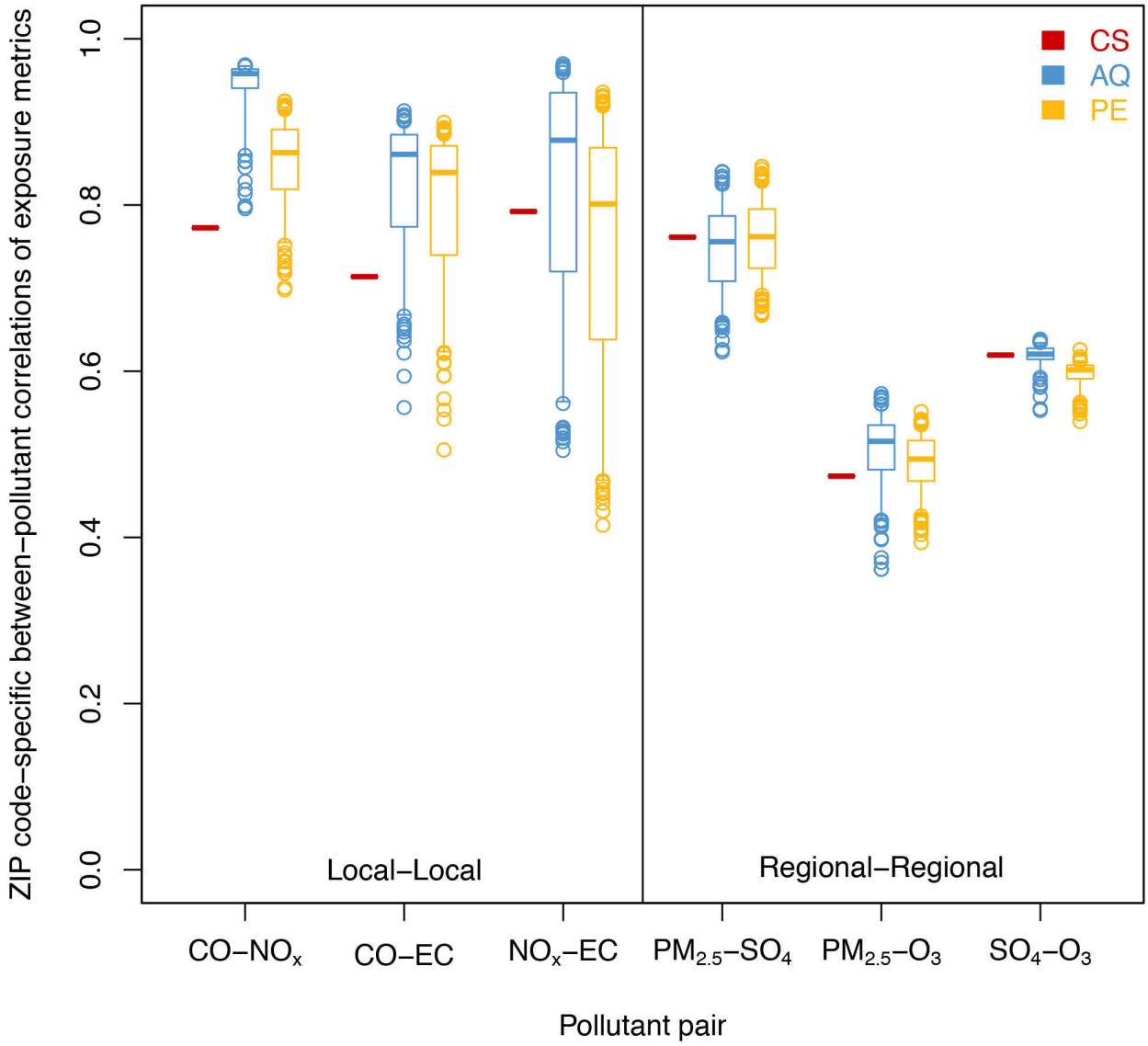

ZIP code–specific between–pollutant correlations of exposure metrics
0.0
0.2
0.4
0.6
0.8
1.0
CS
AQ
PE
Local–Local
Regional–Regional
CO–NOx
CO–EC
NOx–EC
PM2.5–SO4
PM2.5–O3
SO4–O3
Pollutant pair

Figure 2a

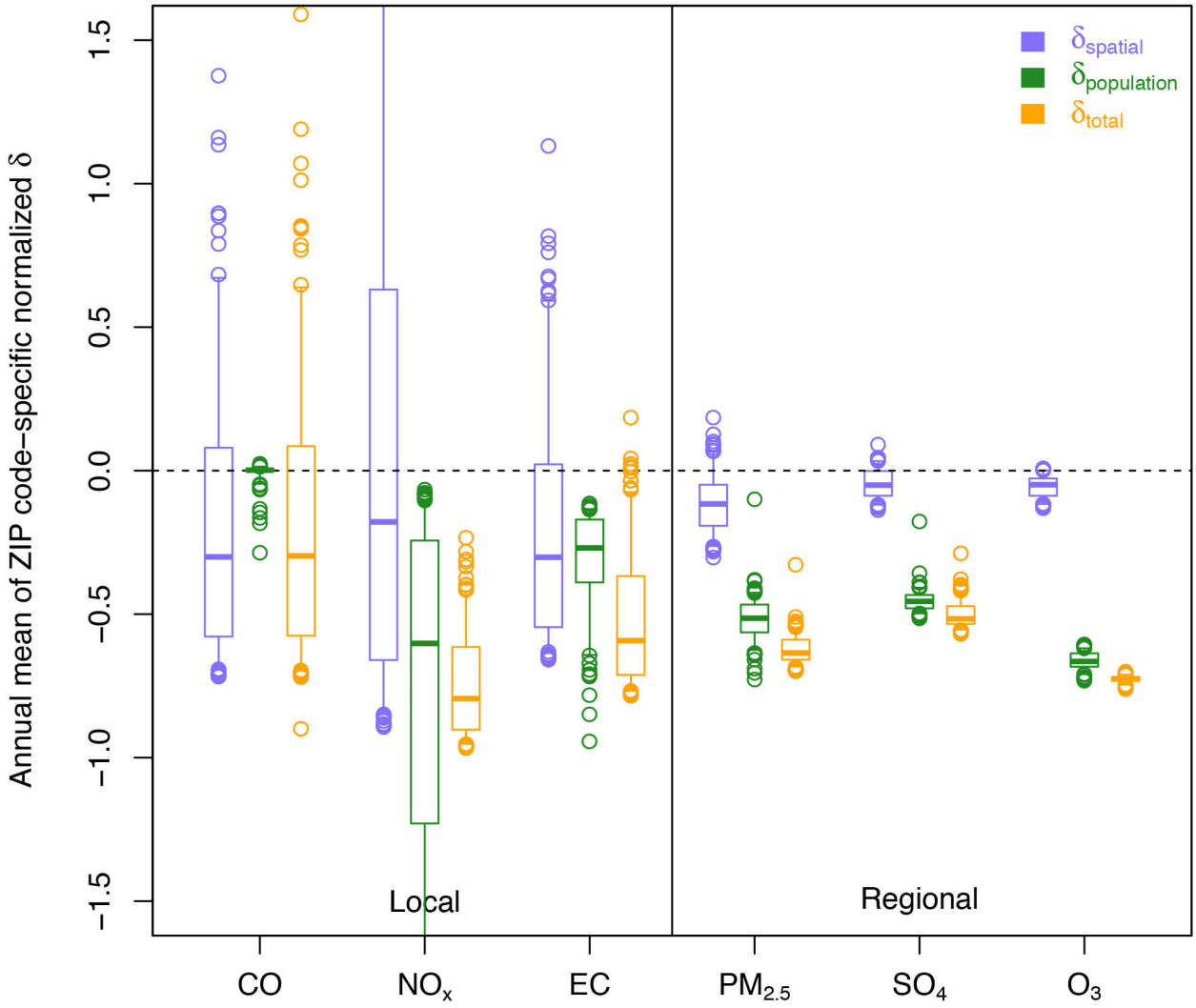

Annual mean of ZIP code−specific normalized δ
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
δspatial
δpopulation
δtotal
Local
Regional
CO
NOx
EC
PM2.5
SO4
O3

Figure 2b

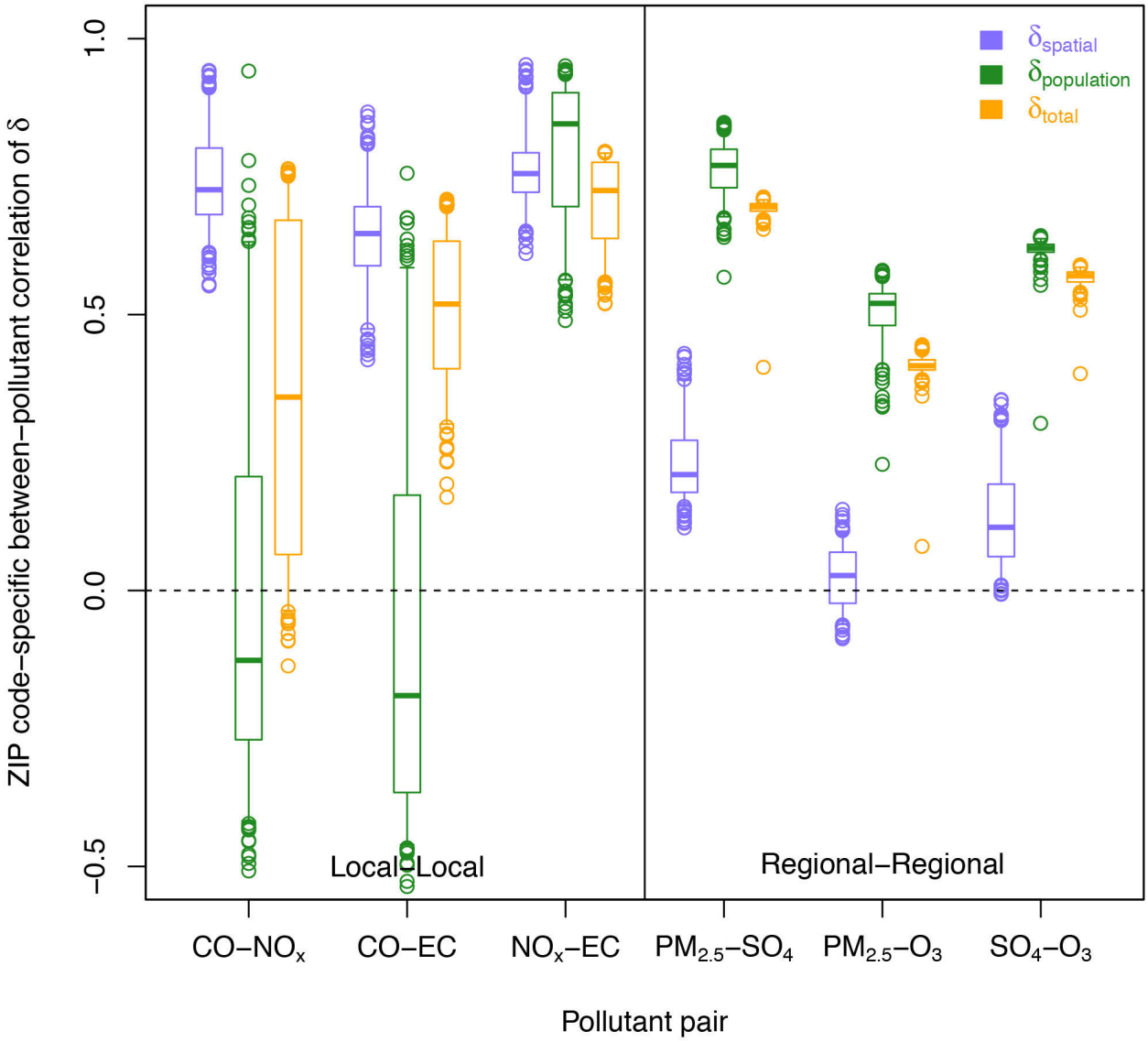

ZIP code−specific between−pollutant correlation of δ
1.0
0.5
0.0
−0.5
δspatial
δpopulation
δtotal
Local−Local
Regional−Regional
CO−NOx
CO−EC
NOx−EC
PM2.5−SO4
PM2.5−O3
SO4−O3
Pollutant pair

Figure 2c

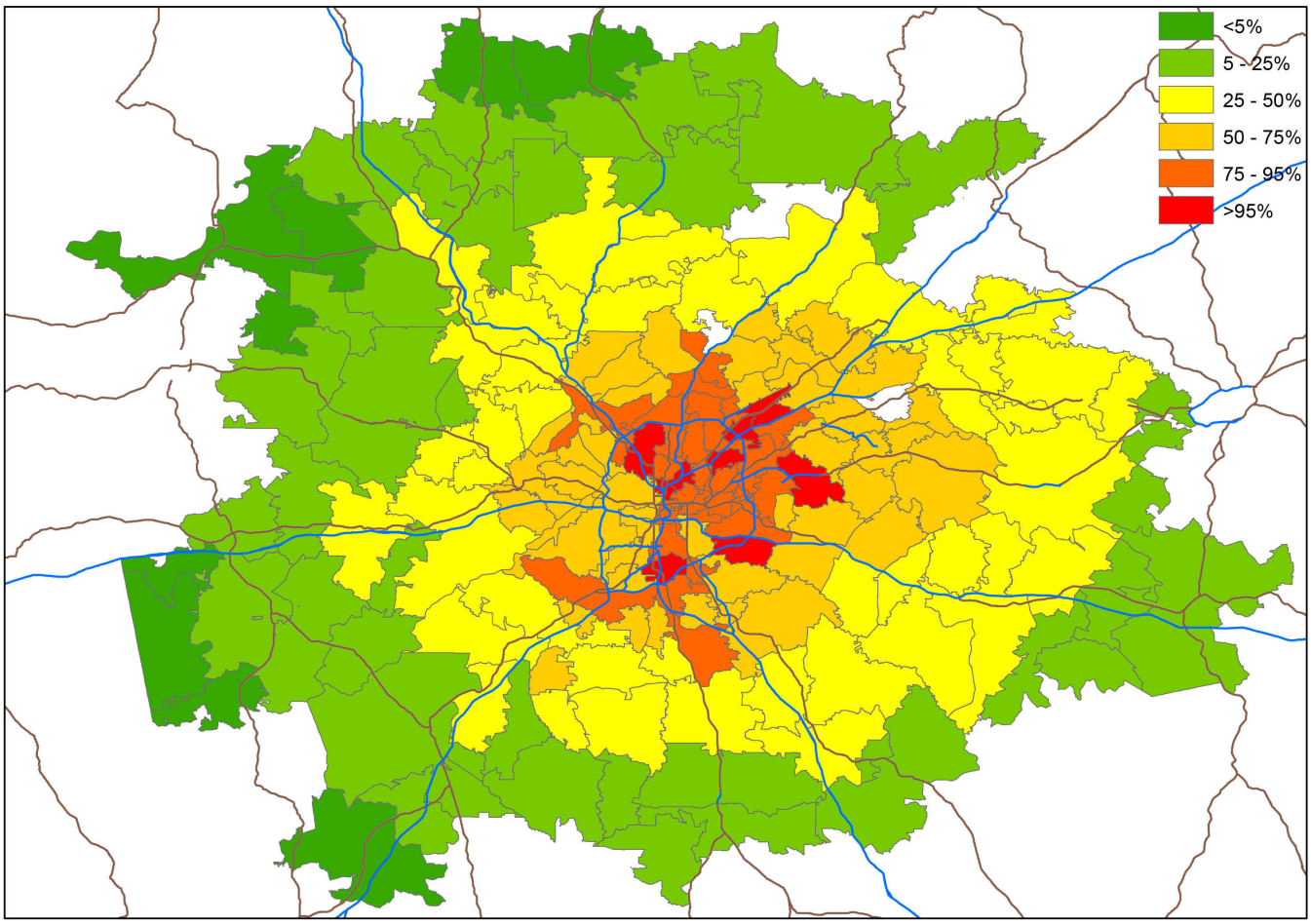


Figure 3a

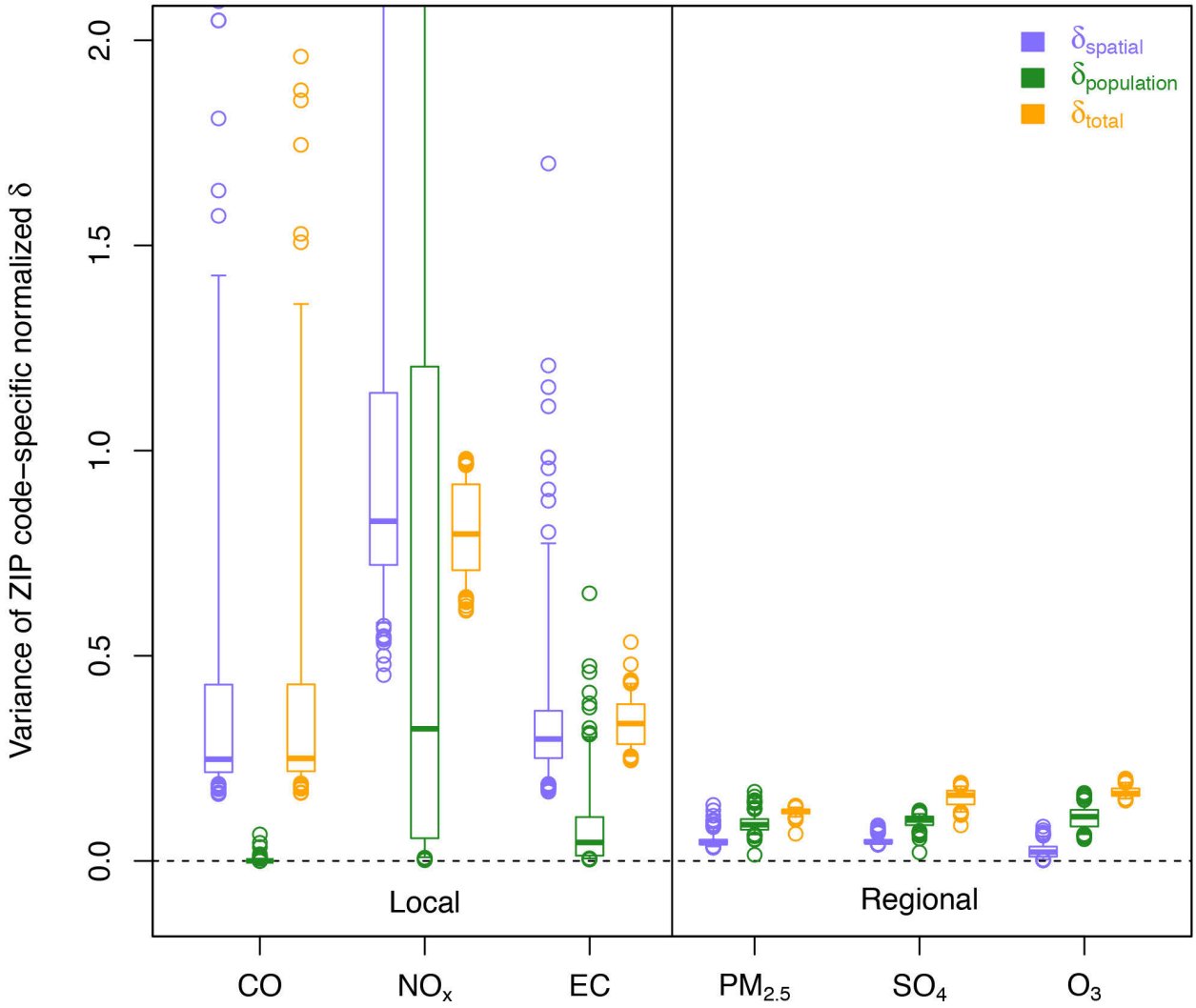

Variance of ZIP code−specific normalized δ
0.0
0.5
1.0
1.5
2.0
δspatial
δpopulation
δtotal
Local
Regional
CO
NOx
EC
PM2.5
SO4
O3

Figure 3b

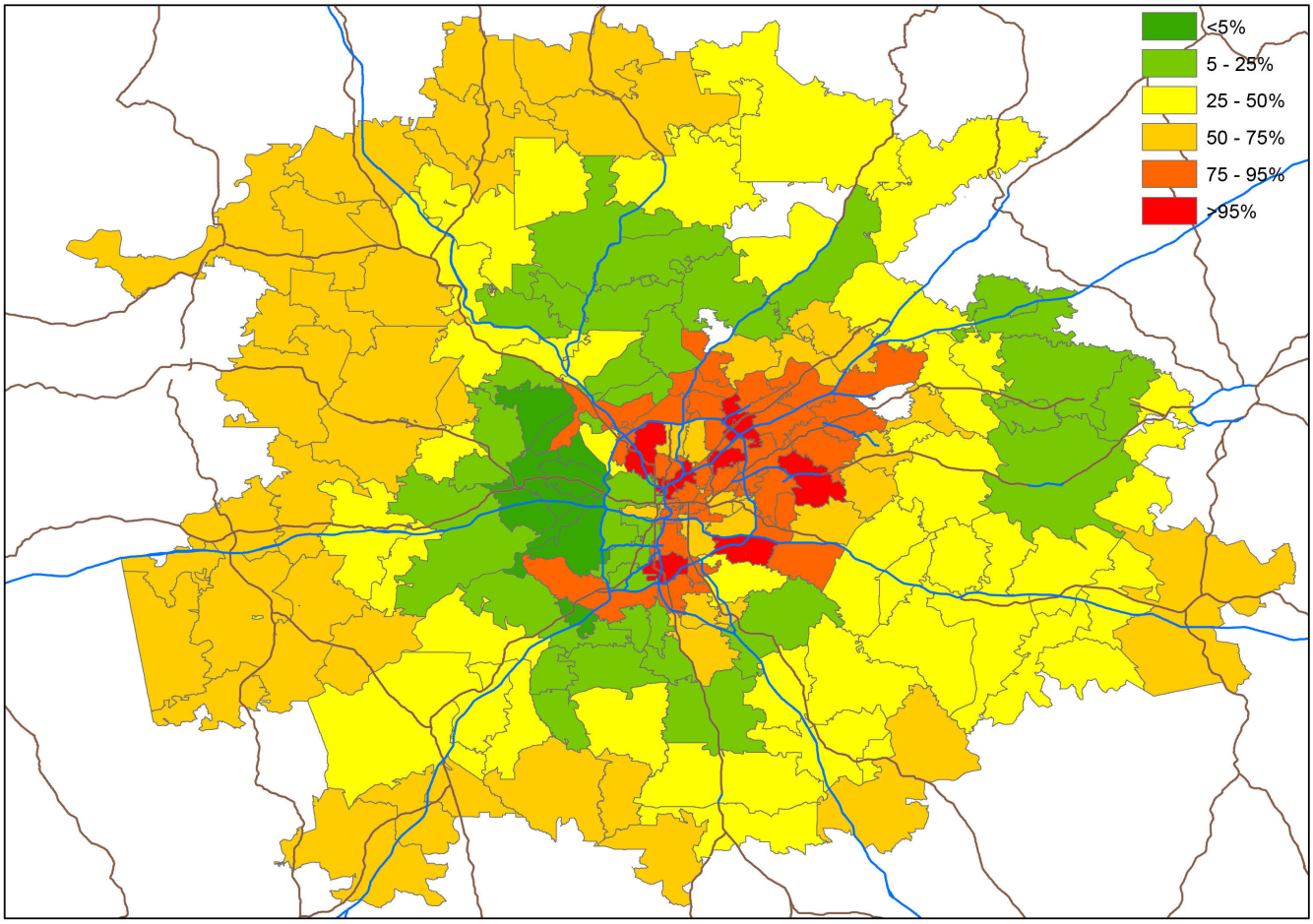


Figure 4a

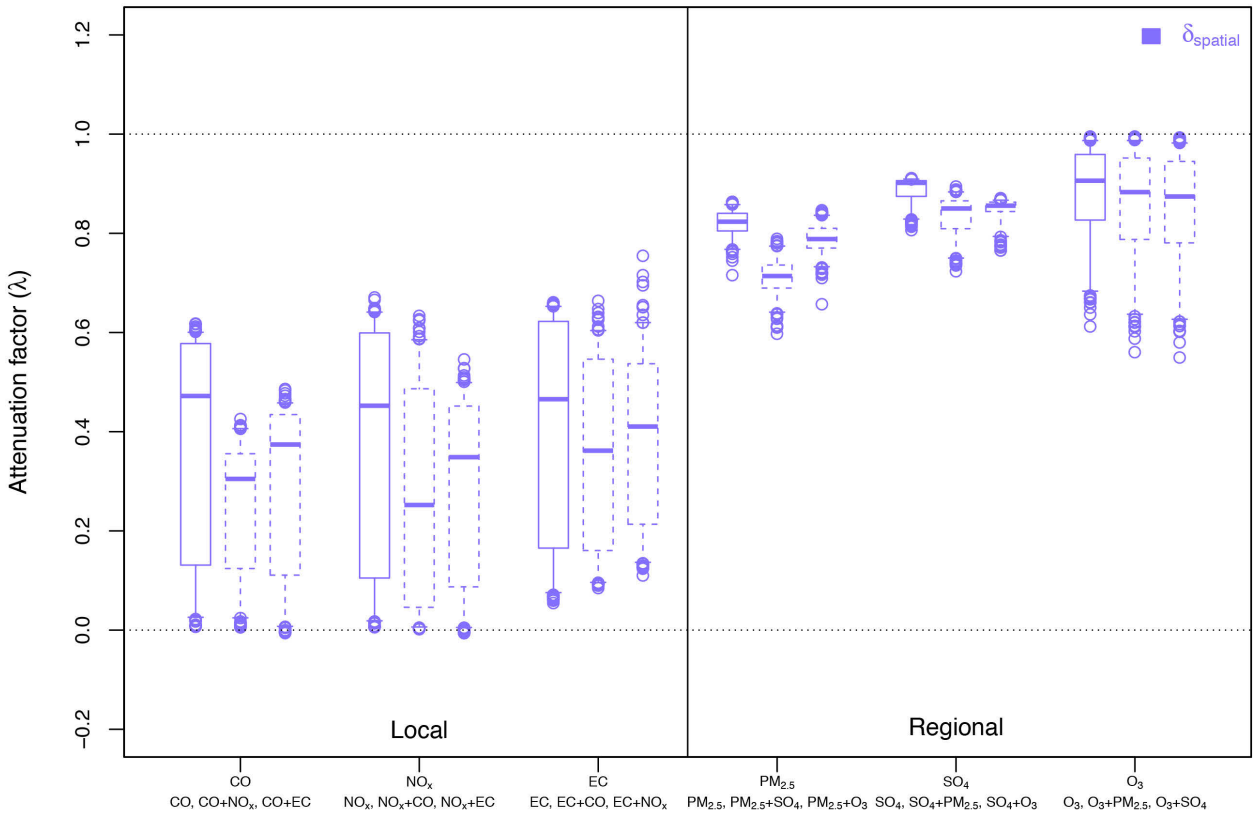

Attenuation factor (λ)
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2
δspatial
Local
Regional
CO
CO, CO+NOx, CO+EC
NOx
NOx, NOx+CO, NOx+EC
EC
EC, EC+CO, EC+NOx
PM2.5
PM2.5, PM2.5+SO4, PM2.5+O3
SO4
SO4, SO4+PM2.5, SO4+O3
O3
O3, O3+PM2.5, O3+SO4

Figure 4b

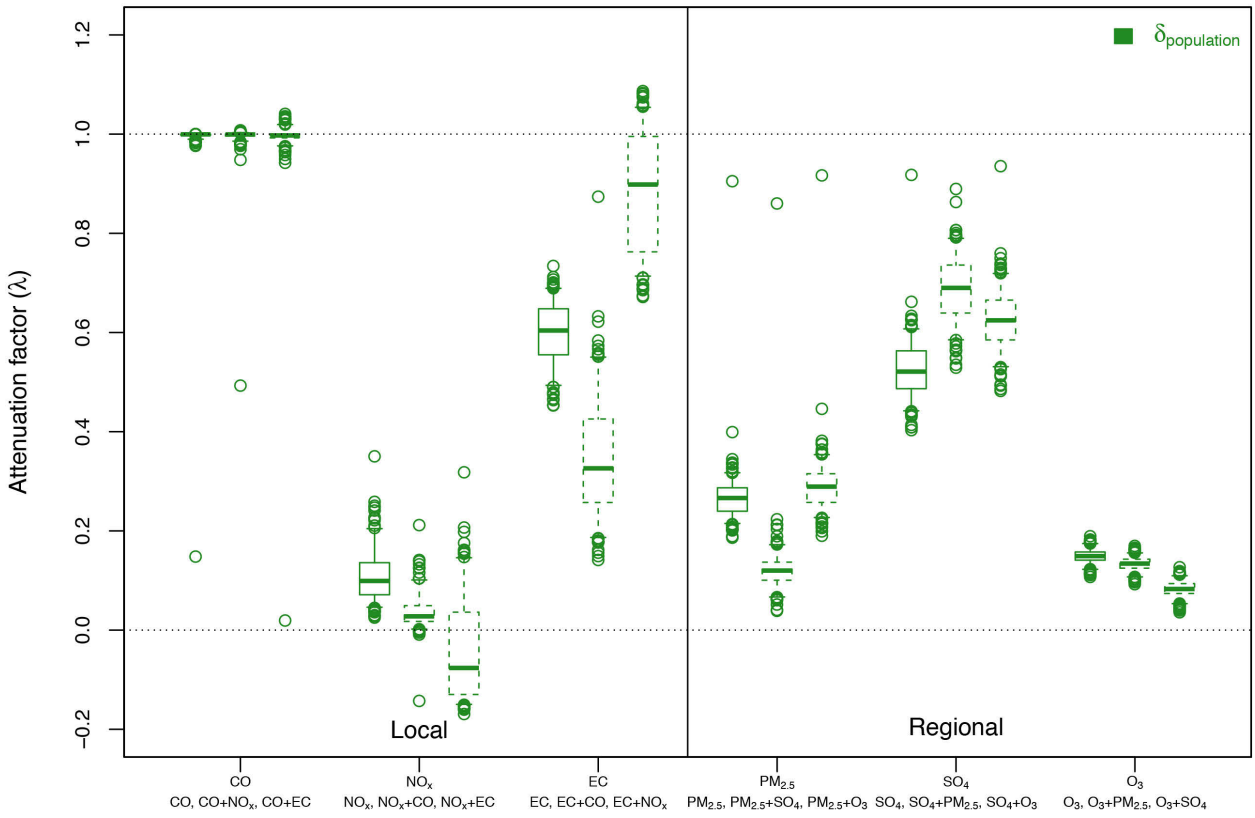

Attenuation factor (λ)
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
δpopulation
Local
Regional
CO
CO, CO+NOx, CO+EC
NOx
NOx, NOx+CO, NOx+EC
EC
EC, EC+CO, EC+NOx
PM2.5
PM2.5, PM2.5+SO4, PM2.5+O3
SO4
SO4, SO4+PM2.5, SO4+O3
O3
O3, O3+PM2.5, O3+SO4

Figure 4c

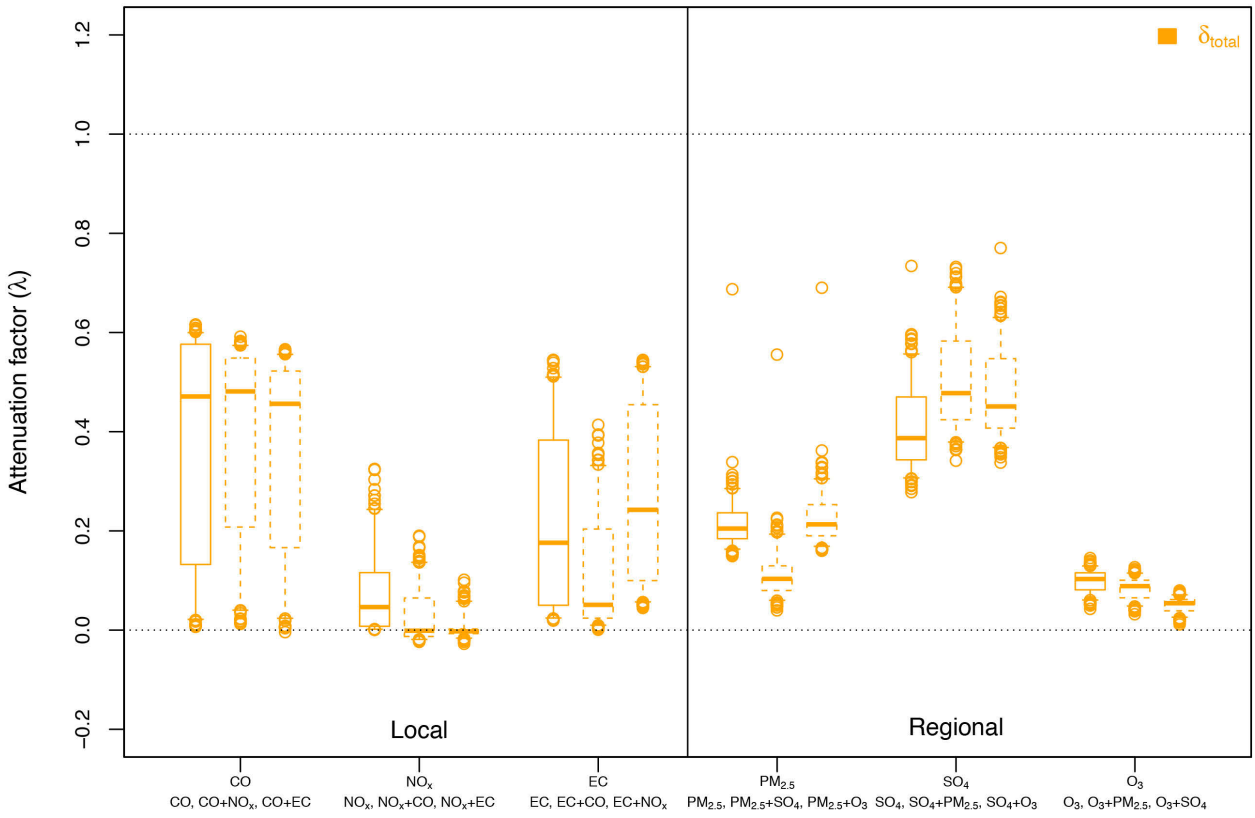
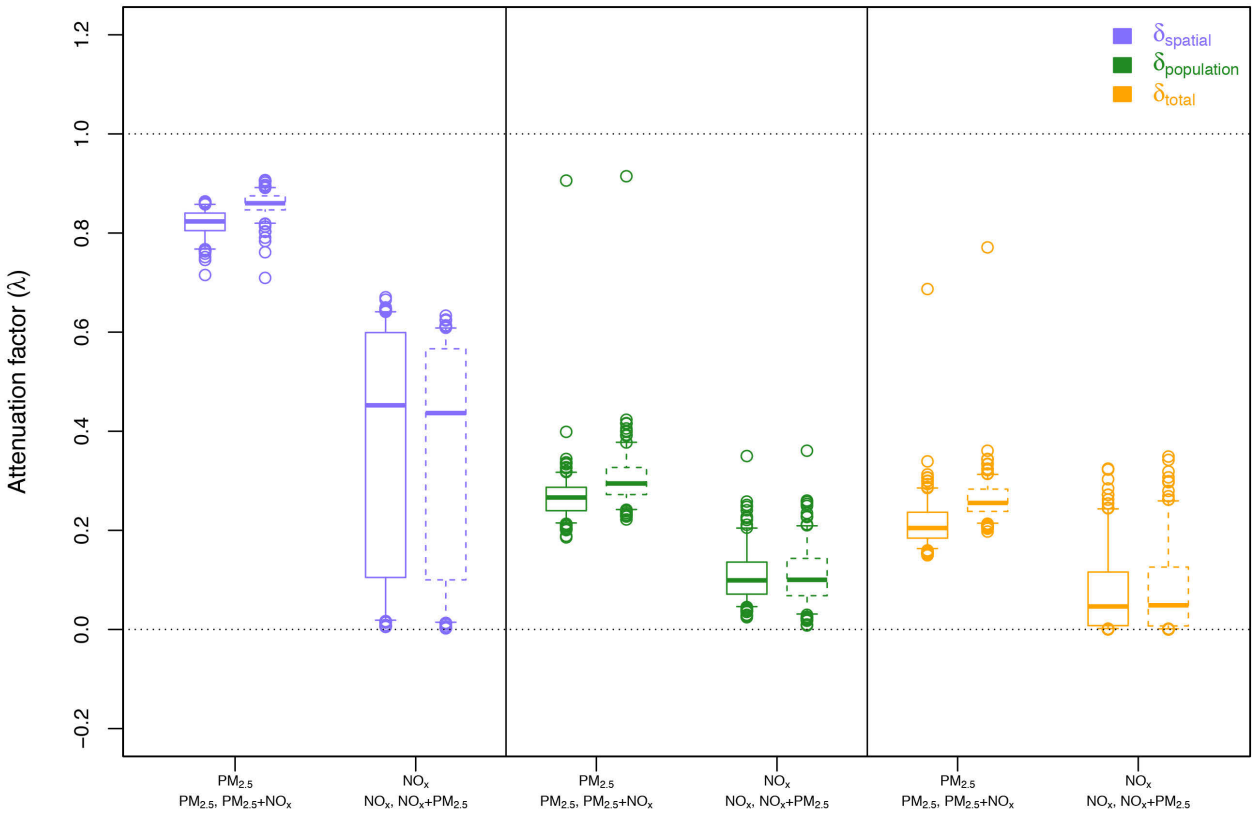

Attenuation factor (λ)
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
δtotal
Local
Regional
CO
CO, CO+NOx, CO+EC
NOx
NOx, NOx+CO, NOx+EC
EC
EC, EC+CO, EC+NOx
PM2.5
PM2.5, PM2.5+SO4, PM2.5+O3
SO4
SO4, SO4+PM2.5, SO4+O3
O3
O3, O3+PM2.5, O3+SO4

Figure 4d


Attenuation factor (λ)
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
δspatial
δpopulation
δtotal
PM2.5
PM2.5, PM2.5+NOx
NOx
NOx, NOx+PM2.5
PM2.5
PM2.5, PM2.5+NOx
NOx
NOx, NOx+PM2.5
PM2.5
PM2.5, PM2.5+NOx
NOx
NOx, NOx+PM2.5

Supplemental Material

An Empirical Assessment of Exposure Measurement Error and Effect Attenuation in Bipollutant Epidemiologic Models

Kathie L. Dionisio, Lisa K. Baxter, and Howard H. Chang

Table of Contents **Page**

Population exposure metric

Ambient pollutant concentrations supplied at ZIP code centroids (as described in the CS metric section above), were used to represent ambient outdoor concentrations for calculation of exposure at the home, outside the home, and the home garage for groups of simulated individuals. Individuals were sampled by census tract as is normally done in SHEDS Air. The target was for 100 simulated individuals to be assigned to each of the 193 ZIP codes of interest. To obtain ZIP code level exposure estimates, individuals were also assigned to a ZIP code polygon. Because it is common for census tracts to overlay two or more ZIP codes, an area weighted scheme was used to define the likelihood of an individual in a specific census tract being assigned to a specific ZIP code.

To account for exposure during commuting and travel, each simulated individual's work location was selected stochastically based on probabilities derived from the 2000 U.S. Census Tract-toTract commuting data. Individuals can commute to any census tract overlapping the ZIP codes of interest. Workers commuting out of the sampling area (~3% of workers) are removed from the population and replaced with another worker that commutes within the sampling area. Travel other than work is limited to the closest 20 census tracts within the sampling area, and within 10 miles of an individual's home census tract. Each individual is assigned to 3 nearby census tracts to which they may travel for a variety of reasons (e.g. shopping, church, other facilities related to microenvironments in the CHAD database). Some facilities are always in the same census tract for an individual (e.g. church), and others are assigned randomly each trip (e.g. retail stores). Air quality during travel is based on the ZIP code associated with the origin of the travel.

Penetration and decay parameters can be found in Tables S1 and S2.

Table S1. Input parameters for the SHEDS model: $PM_{2.5}$, SO_4, and EC.

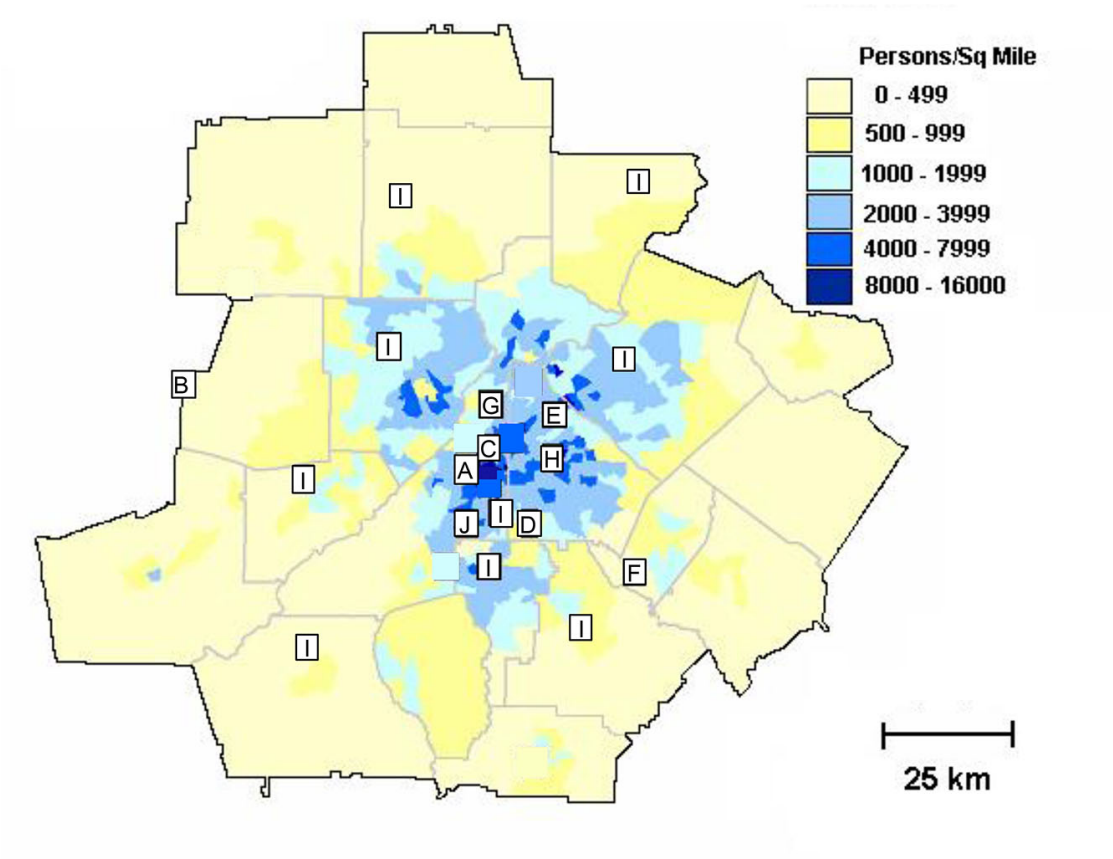
Micro-environment *Equation parameter or sub-environment*	**$PM_{2.5}$**	**SO_4**	**EC**
Home Indoor[a]			
Penetration, P (unitless)[b]	P_i = N[0.89, 0.06] P_d = N[P_i, 0.03*P_i] Limits = [0.76, 1]	P_i = N[0.95, 0.07] P_d = N[P_i, 0.03*P_i] Limits = [0.80, 1]	P_i = N[0.96, 0.04] P_d = N[P_i, 0.03*P_i] Limits = [0.87, 1]
Deposition rate, k (h^{-1})[c]	k_i = N[0.39, 0.09] k_d = N[k_i, 0.05*k_i] Limits = [0.19, 0.58]	k_i = N[0.19, 0.04] k_d = N[k_i, 0.05*k_i] Limits = [0.10, 0.28]	k_i = N[0.10, 0.02] k_d = N[k_i, 0.05*k_i] Limits = [0.06, 0.14]
Indoors[d,e,f]			
Attached garage	$C_{ambient}$	$C_{ambient}$	$C_{ambient}$
Office	Slope = 0.18 Intercept = 3.6 Residual = N[0, 2.9]	Scaling Factor = N[0.450, 0.382] Limits = [0, -]	Scaling Factor = N[0.390, 0.339] Limits = [0, -]
Store	Slope = 0.75 Intercept = 9 Residual = N[0, 2.1]	Scaling Factor = N[0.479, 0.223] Limits = [0, -]	Scaling Factor = N[0.606, 0.351] Limits = [0, -]
School	Slope = 0.6 Intercept = 6.8 Residual = N[0, 5.4]	Scaling Factor = N[0.511, 0.241] Limits = [0, -]	Scaling Factor = N[0.550, 0.152] Limits = [0, -]
Restaurant	Slope = 1.0 Intercept = 9.8 Residual = N[0, 10]	Scaling Factor = N[1.0, 0.457] Limits = [0, -]	Scaling Factor = N[0.367, 0.153] Limits = [0, -]
Parking garage	Same as Other	Same as Other	Same as Other
Service station	$C_{ambient}$	$C_{ambient}$	$C_{ambient}$
Other	Slope = 0.85 Intercept = 8.4 Residual = N[0, 4]	Scaling Factor = N[0.340, 0.391] Limits = [0, -]	Scaling Factor = N[0.590, 0.423] Limits = [0, -]
Outdoors[a,f]			
Home	$C_{ambient}$	$C_{ambient}$	$C_{ambient}$
Near street[e]	Scaling Factor = N[0.974, 0.185] Limits = [0.7, 1.3]	$C_{ambient}$	$C_{ambient}$
Other	$C_{ambient}$	$C_{ambient}$	$C_{ambient}$
In-vehicle[a,d,e]	Slope = 0.7125 Intercept = 0 Residual = N[0, 6.64] Limits = [-12, 20]	Scaling Factor = N[0.961, 0.129] Limits = [0, -]	Scaling Factor = N[1.633, 1.088] Limits = [0, -]

[a]N[x,y] = Normal[mean, standard deviation]. [b]P_i = individual penetration, P_d = daily penetration, with exposure concentrations calculated using mass-balance. [c]k_i = individual decay, k_d = daily decay, with exposure concentrations calculated using mass-balance. [d]When "slope" and "intercept" are provided, implemented as Slope * $C_{ambient}$ + Intercept + Residual using linear regression. [e]When "scaling factor" is provided, implemented as Scaling Factor * $C_{ambient}$, with Limits = [minimum, maximum], implemented using linear regression; see footnote 8 for definition of $C_{ambient}$. [f]$C_{ambient}$ = concentration of the pollutant in question, in the ambient (outdoor) air of the microenvironment under consideration.

Table S2. Input parameters for the SHEDS model: O_3, CO, and NO_x.

Micro-environment *Equation parameter or sub-environment*	**O_3**	**CO**	**NO_x**
Home Indoor			
Penetration, P (unitless)[a]	P_d = 1	P_d = 1	P_d = 1
Deposition Rate, k (h^{-1})[b,c,d]	k_d = LN[2.51, 1.53] Limits = [0.95, 8.05]	k_d = 0	k_d = U[1.02, 1.45]
Indoors[c,e,f,g,h]			
Attached garage	Same as home indoors	$C_{ambient}$	$C_{ambient}$
Office	P_d = 1 k_d = LN[2.51, 1.53] Limits = [0.95, 8.05] AER = LN[1.109, 3.015] Limits = [0.07, 13.8]	$C_{ambient}$	$C_{ambient}$
Store	P_d = 1 k_d = LN[2.51, 1.53] Limits = [0.95, 8.05] AER = LN[1.109, 3.015] Limits = [0.07, 13.8]	$C_{ambient}$	$C_{ambient}$
School	P_d = 1 k_d = LN[2.51, 1.53] Limits = [0.95, 8.05] AER = LN[1.109, 3.015] Limits = [0.07, 13.8]	$C_{ambient}$	$C_{ambient}$
Restaurant	P_d = 1 k_d = LN[2.51, 1.53] Limits = [0.95, 8.05] AER = LN[1.109, 3.015] Limits = [0.07, 13.8]	$C_{ambient}$	$C_{ambient}$
Parking garage	Scaling Factor = N[0.755, 0.203] Limits = [0.422, 1.0]	$C_{ambient}$	$C_{ambient}$
Service station	$C_{ambient}$	$C_{ambient}$	$C_{ambient}$
Other	Same as office, store, school, restaurant	$C_{ambient}$	$C_{ambient}$
Outdoors[f,g,h]			
Home	$C_{ambient}$	$C_{ambient}$	$C_{ambient}$
Near street[e]	Scaling Factor = N[0.755, 0.203] Limits = [0.422, 1.0]	$C_{ambient}$	$C_{ambient}$
Other	$C_{ambient}$	$C_{ambient}$	$C_{ambient}$
In-vehicle[d,f,g,h]	Scaling Factor 1 = N[0.755, 0.203] Limits = [0.422, 1] Scaling Factor 2 = N[0.3, 0.232] Limits = [0.1, 1]	$C_{ambient}$	Scaling Factor = U[0.6, 1]

[a]P_i = individual penetration, P_d = daily penetration, with exposure concentrations calculated using mass-balance. [b]k_i = individual decay, k_d = daily decay, with exposure concentrations calculated using mass-balance. [c]LN[x,y] = Lognormal[geometric mean, geometric standard deviation]. [d]U[x,y] = Uniform[minimum, maximum]. [e]AER = air exchange rate. [f]N[x,y] = Normal[mean, standard deviation]. [g]When "scaling factor" is provided, implemented as Scaling Factor * $C_{ambient}$, with Limits = [minimum, maximum], implemented using linear regression; see footnote 8 for definition of $C_{ambient}$. [h]$C_{ambient}$ = concentration of the pollutant in question, in the ambient (outdoor) air of the microenvironment under consideration.



A	Jefferson Street, SEARCH (*1,2,3,4,5*)
B	Yorkville, SEARCH (*1,2,3,4,5*)
C	Georgia Tech, AQS (*1*)
D	South DeKalb, STN & ASACA (*1,3,4*)
E	Tucker, AQS & ASACA (*1,3,4*)
F	Conyers Monastery, AQS (*1,3*)
G	Roswell Road, AQS (*2*)
H	Dekalb Tech, AQS (*2*)
I	various ozone monitors, AQS (*3*)
J	Fort McPherson, ASACA (*4*)

Figure S1. Map of metropolitan Atlanta with monitoring site locations and population density. Letters reference monitor locations. The table identifies station name, network, and air pollutants monitored, with air pollutants indicated by numbers [1 = NO_2/NO_x, 2 = CO, 3 = O_3, 4 = $PM_{2.5}$ mass, 5 = $PM_{2.5}$ composition (SO_4, EC)]. Population density is from 2000 Census data.

Figure S2a

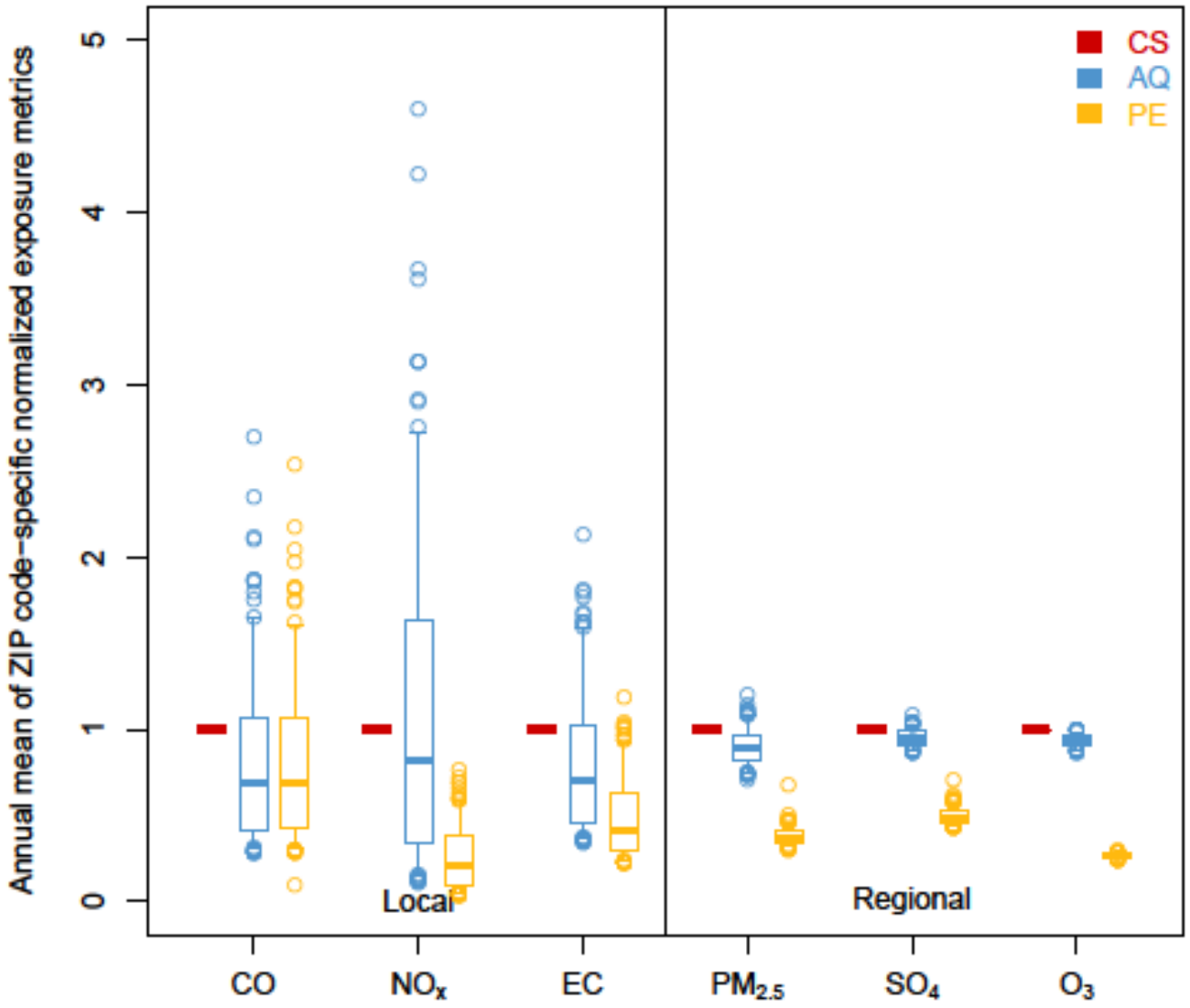


Figure S2a. Normalized exposure estimates from three methods: full extent of data and outliers (zoomed out version of Figure 1a; n=193 for each box). The bottom and top of the box represent 25th and 75th percentiles, the band near the middle of the box is the median, and the ends of the whiskers are the 5th and 95th percentiles.

Figure S2b

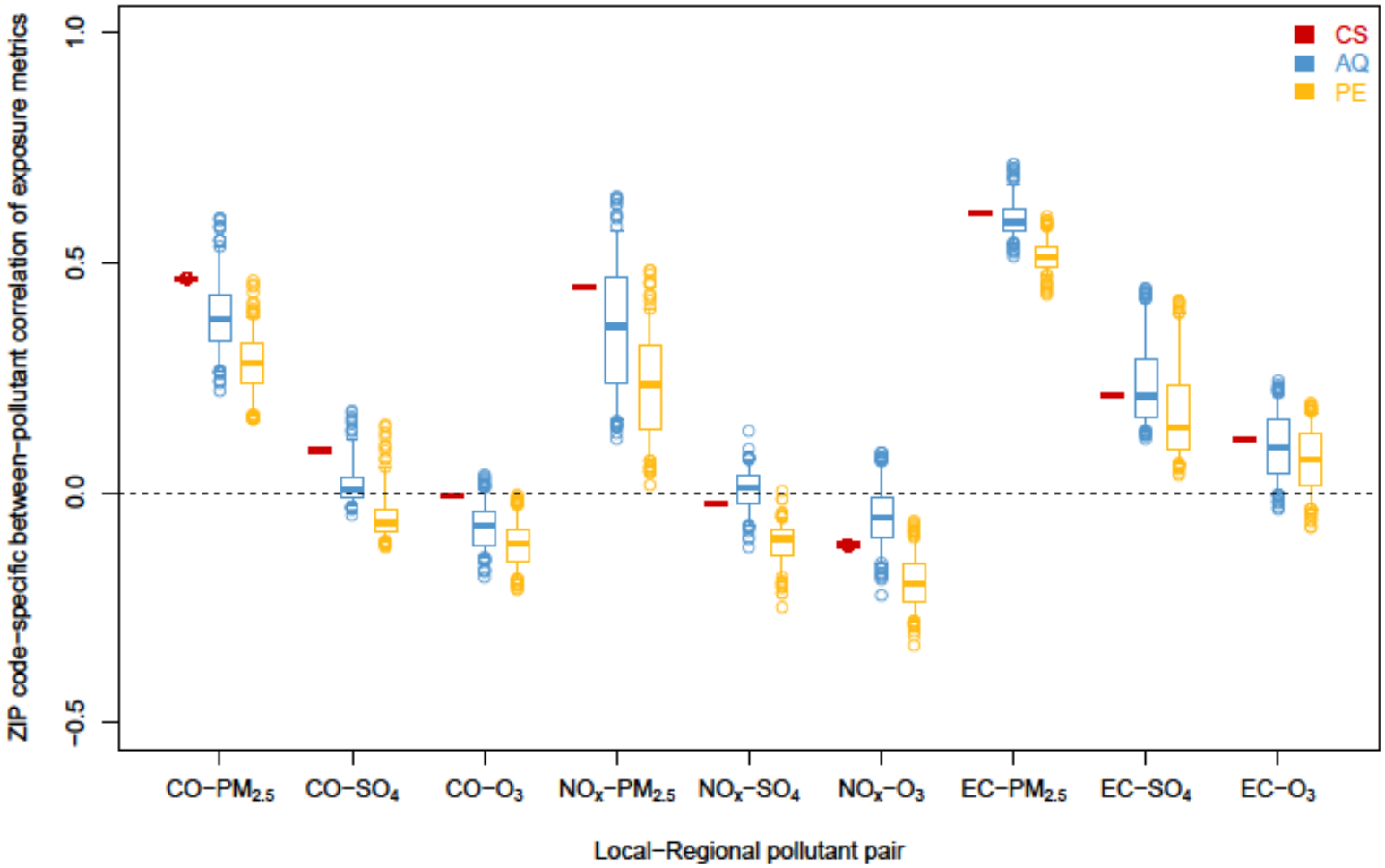


Figure S2b. Between-pollutant correlations of exposure for local-regional pollutant pairs; n=193 for each box. The bottom and top of the box represent 25th and 75th percentiles, the band near the middle of the box is the median, and the ends of the whiskers are the 5th and 95th percentiles.

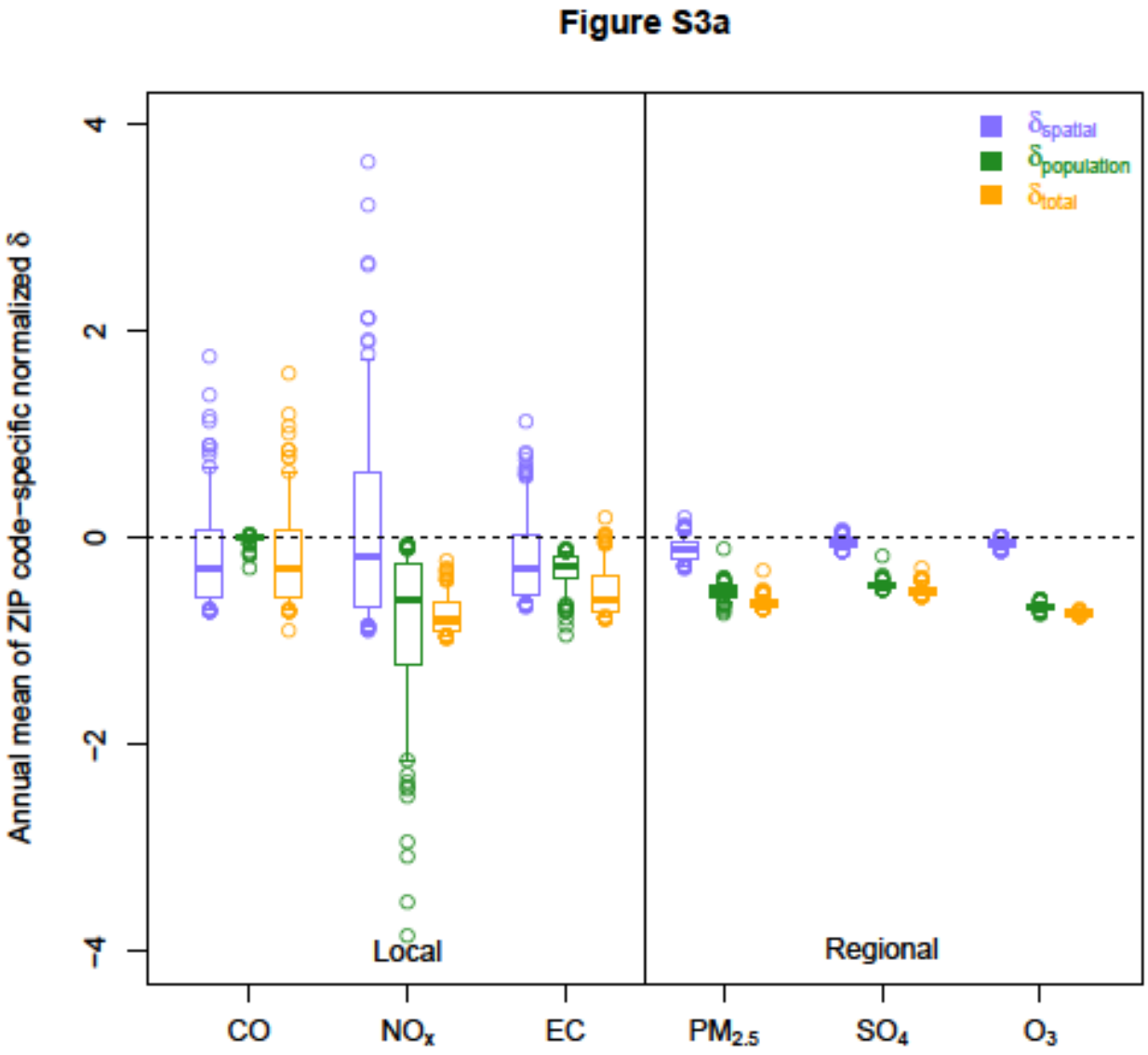


Figure S3a. Normalized exposure error: full extent of data and outliers (zoomed out version of Figure 2a; n=193 for each box). The bottom and top of the box represent 25th and 75th percentiles, the band near the middle of the box is the median, and the ends of the whiskers are the 5th and 95th percentiles.

Figure S3b

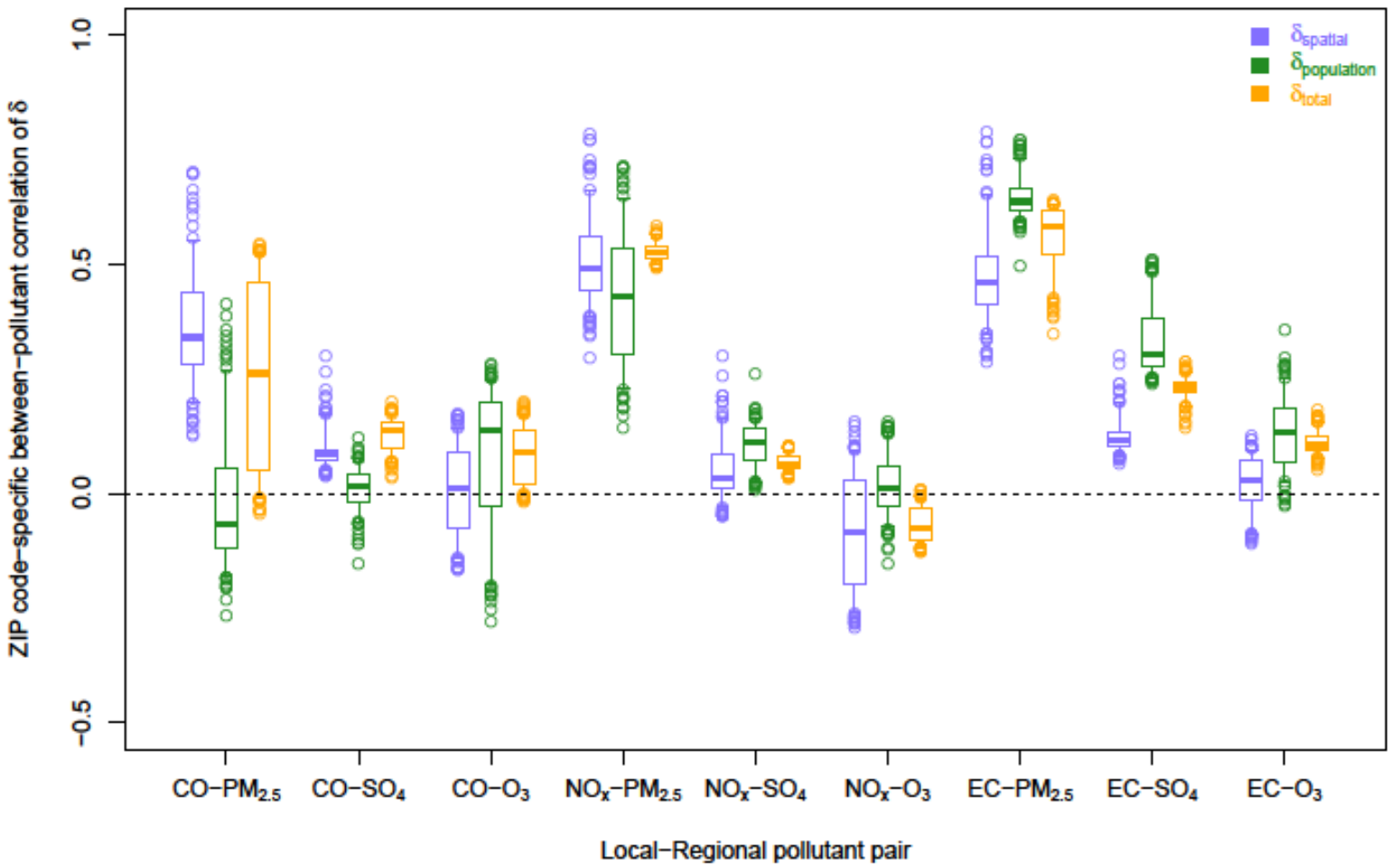


Figure S3b. Between-pollutant correlations of exposure error for local-regional pollutant pairs; n=193 for each box. The bottom and top of the box represent 25th and 75th percentiles, the band near the middle of the box is the median, and the ends of the whiskers are the 5th and 95th percentiles.

Figure S4

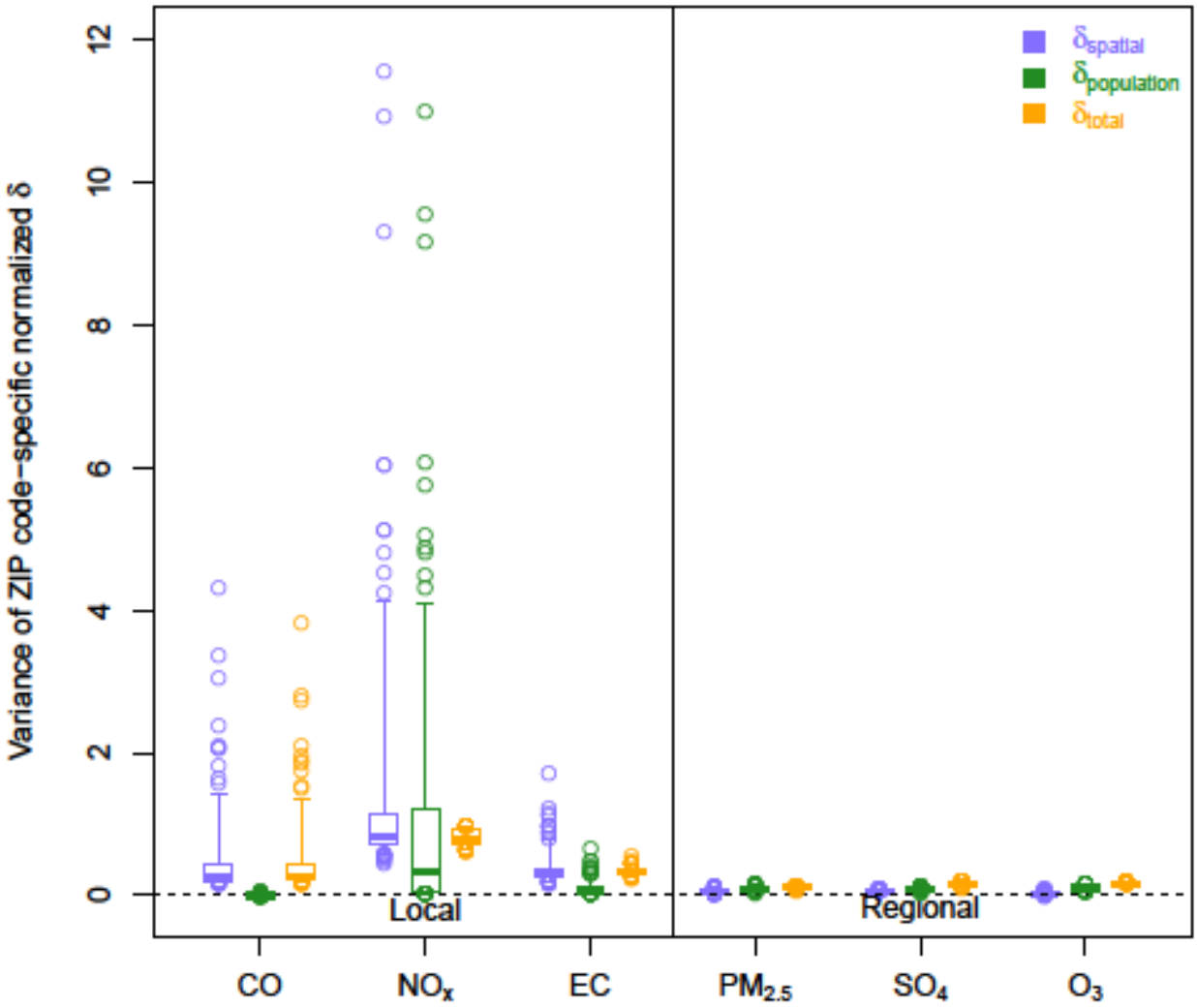


Figure S4. Variance of normalized exposure error: full extent of data and outliers (zoomed out version of Figure 3a; n=193 for each box). The bottom and top of the box represent 25th and 75th percentiles, the band near the middle of the box is the median, and the ends of the whiskers are the 5th and 95th percentiles.

Figure S5a

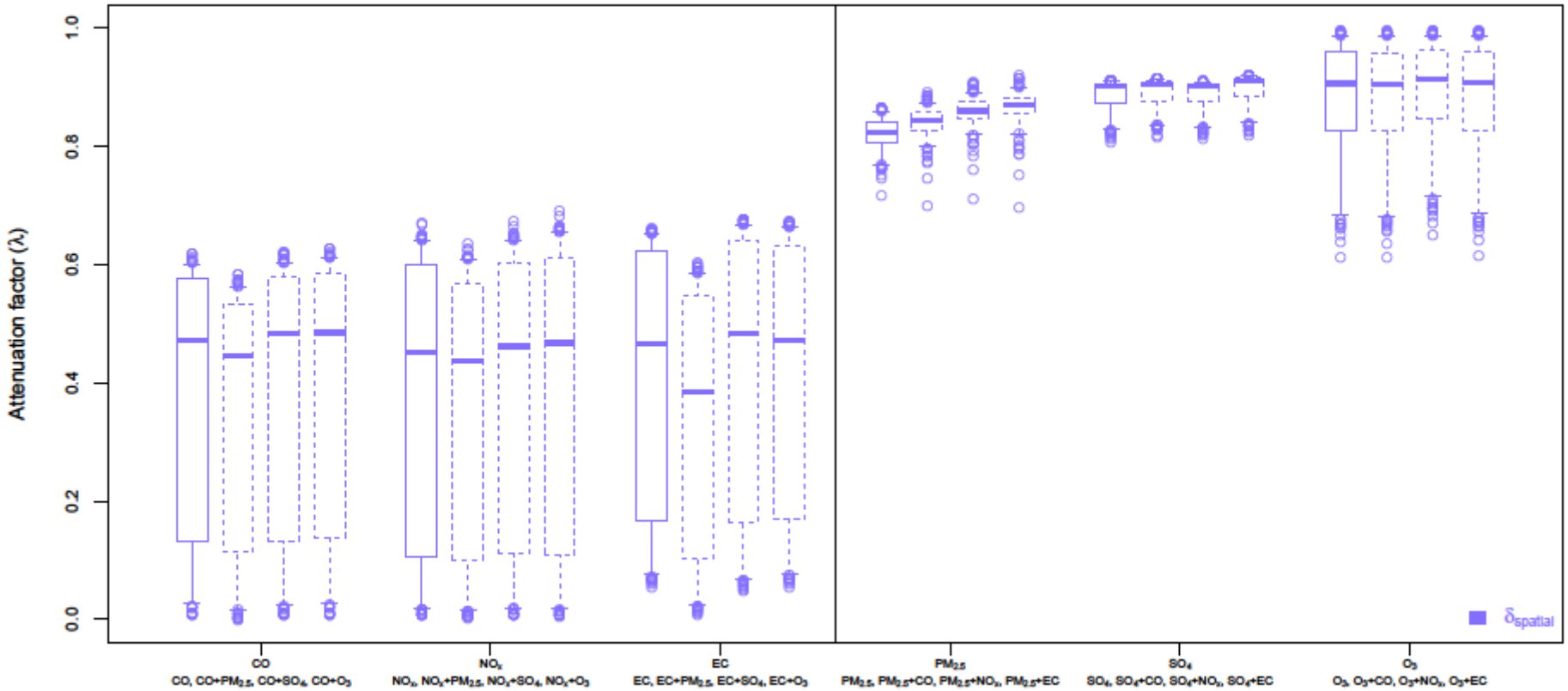


Figure S5a. Attenuation of model coefficients in a classical error, single pollutant framework, and in bi-pollutant models, assuming one pollutant has an effect, and one pollutant has no effect; $\delta_{spatial}$ for local-regional pollutant pairs; n=193 for each box. The bottom and top of the box represent 25th and 75th percentiles, the band near the middle of the box is the median, and the ends of the whiskers are the 5th and 95th percentiles.

Figure S5b

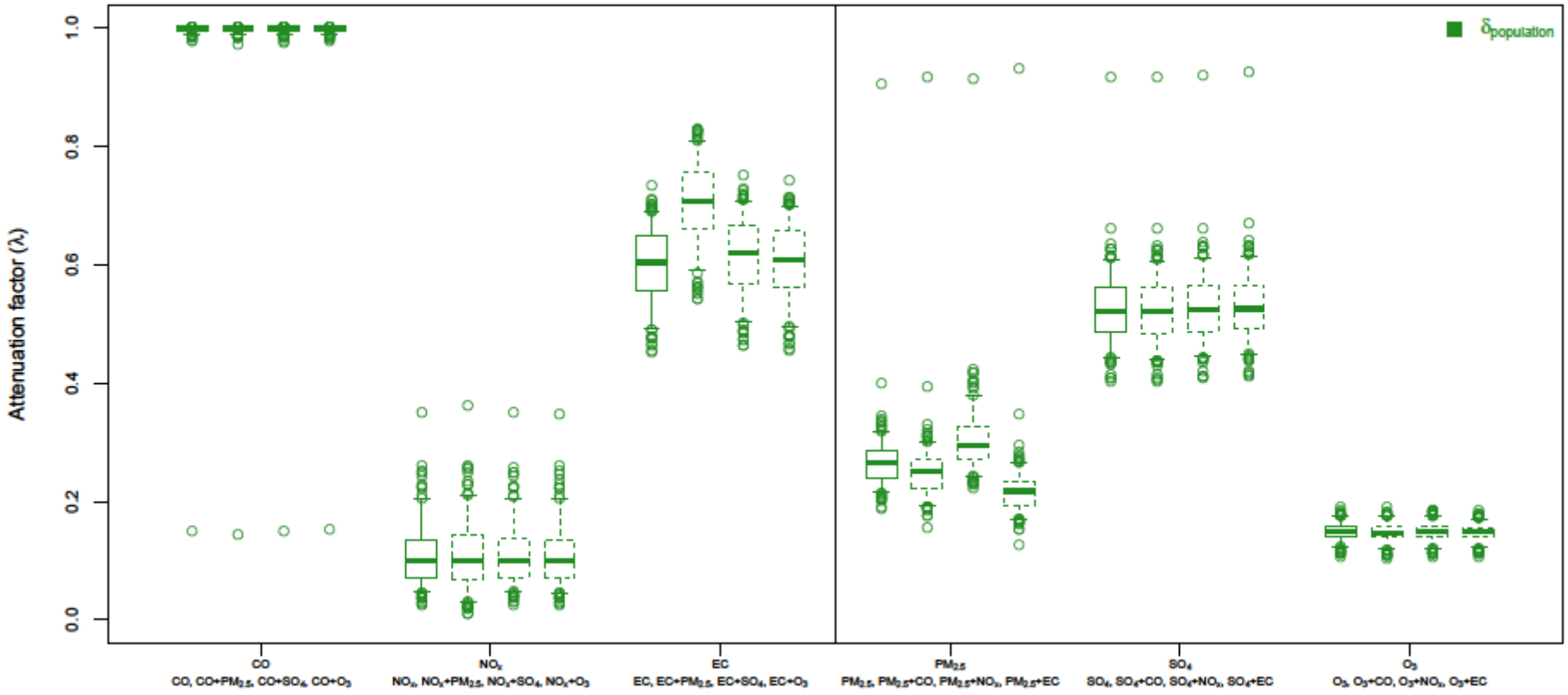


Figure S5b. Attenuation of model coefficients in a classical error, single pollutant framework, and in bi-pollutant models, assuming one pollutant has an effect, and one pollutant has no effect; $\delta_{population}$ for local-regional pollutant pairs; n=193 for each box. The bottom and top of the box represent 25th and 75th percentiles, the band near the middle of the box is the median, and the ends of the whiskers are the 5th and 95th percentiles.

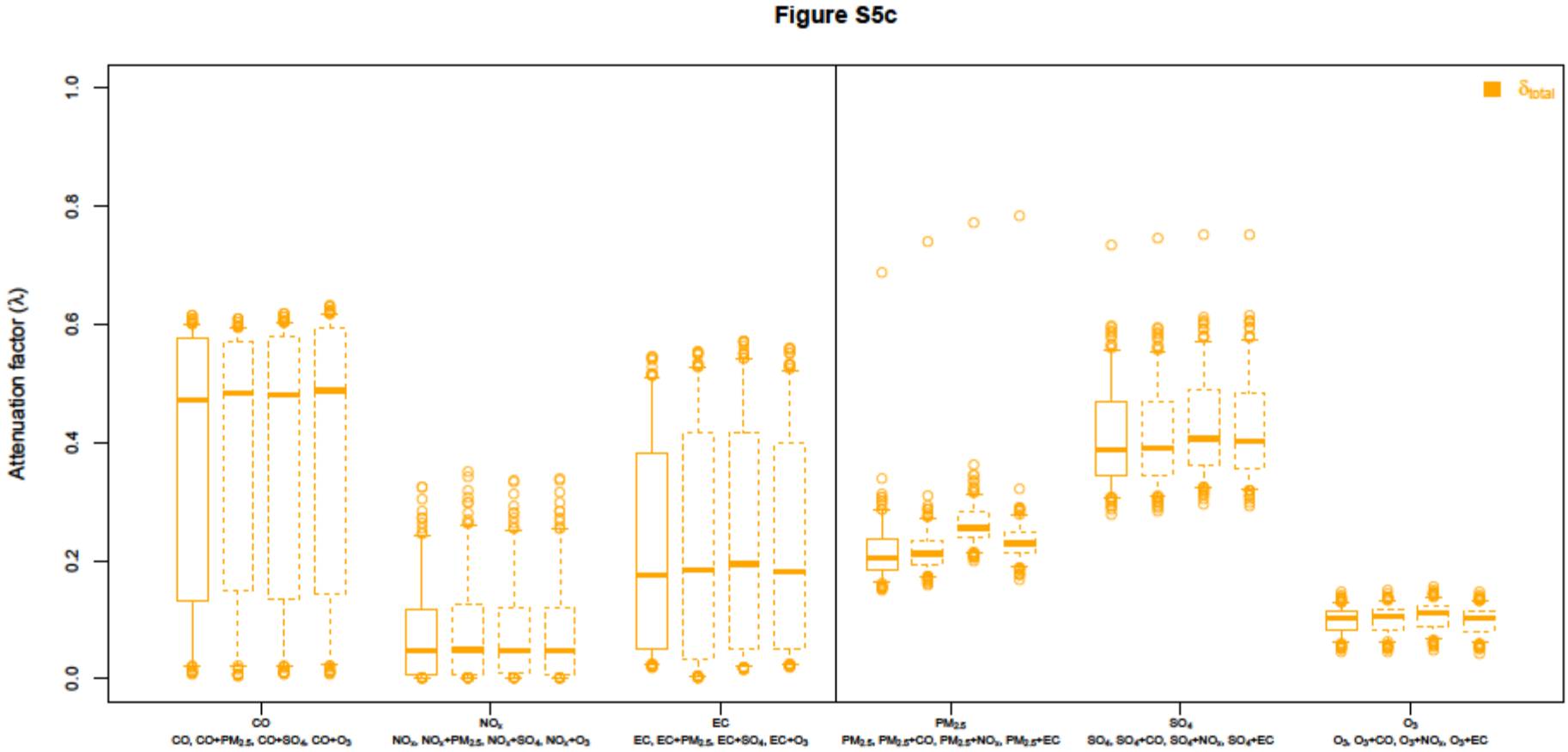


Figure S5c. Attenuation of model coefficients in a classical error, single pollutant framework, and in bi-pollutant models, assuming one pollutant has an effect, and one pollutant has no effect; δ_{total} for local-regional pollutant pairs; n=193 for each box. The bottom and top of the box represent 25th and 75th percentiles, the band near the middle of the box is the median, and the ends of the whiskers are the 5th and 95th percentiles.