

ORIGINAL ARTICLE

Spatial variation in inversion-focused vs 24-h integrated samples of PM_{2.5} and black carbon across Pittsburgh, PA

Brett J. Tunno, Drew R. Michanowicz, Jessie LC Shmool, Ellen Kinnee, Leah Cambal, Sheila Tripathy, Sara Gillooly, Courtney Roper, Lauren Chubb and Jane E. Clougherty

A growing literature explores intra-urban variation in pollution concentrations. Few studies, however, have examined spatial variation during “peak” hours of the day (e.g., rush hours, inversion conditions), which may have strong bearing for source identification and epidemiological analyses. We aimed to capture “peak” spatial variation across a region of complex terrain, legacy industry, and frequent atmospheric inversions. We hypothesized stronger spatial contrast in concentrations during hours prone to atmospheric inversions and heavy traffic, and designed a 2-year monitoring campaign to capture spatial variation in fine particles (PM_{2.5}) and black carbon (BC). Inversion-focused integrated monitoring (0600–1100 hours) was performed during year 1 (2011–2012) and compared with 1-week 24-h integrated results from year 2 (2012–2013). To allocate sampling sites, we explored spatial distributions in key sources (i.e., traffic, industry) and potential modifiers (i.e., elevation) in geographic information systems (GIS), and allocated 37 sites for spatial and source variability across the metropolitan domain (~388 km²). Land use regression (LUR) models were developed and compared by pollutant, season, and sampling method. As expected, we found stronger spatial contrasts in PM_{2.5} and BC using inversion-focused sampling, suggesting greater differences in peak exposures across urban areas than is captured by most integrated saturation campaigns. Temporal variability, commercial and industrial land use, PM_{2.5} emissions, and elevation were significant predictors, but did not more strongly predict concentrations during peak hours.

Journal of Exposure Science and Environmental Epidemiology (2016) **26**, 365–376; doi:10.1038/jes.2015.14; published online 29 April 2015

Keywords: 24-h integrated sampling; black carbon (BC); fine particulate matter (PM_{2.5}); inversion-focused sampling; land use regression (LUR)

INTRODUCTION

In recent years, there have been a number of studies on intra-urban variation in pollution.^{1–7} Relatively few of these studies, however, have examined multiple pollutants, and none, to our knowledge, have captured spatial variation during selected hours of the day — such as during rush hours or under temperature inversion conditions. This spatial variation in “peak” exposures may have strong bearing for both refined source identification and epidemiology, as daily maximum exposures may differ substantially across an urban area, and may be particularly important for epidemiological studies of acute cardiovascular and respiratory events.

This gap in the research has been due, in large part, to limitations in sampling technology, as few land use regression (LUR) campaigns have been able to employ a fleet of monitors with the ability for temporally-controlled sampling. Because these LUR models of fine-scale variability are commonly used to derive exposure estimates for epidemiology, this sampling methodology should be well developed, including clear attention to the selection of sampling intervals (e.g., morning vs afternoon; 24-h integrated; rush hour vs full day).

Intra-urban air pollution concentrations can vary due to proximity to industrial sources, traffic density, and other site characteristics, such as population and land use, and may be modified by elevation and meteorological factors.³ During temperature inversions,

atmospheric convection and pollutant dispersion may be limited, intensifying concentrations near sources. Thus, peak (or maximum) spatial contrasts within urban areas may be expected during hours of limited atmospheric mixing and significant source activity. Better understanding these peak exposure contrasts can lead to better characterization of intra-urban air pollution gradients, source identification, and assessment of meteorological impacts on urban concentrations. More recent LURs have used active sampling and multi-pollutant approaches, often reducing the number of samples which can be taken concurrently.

Pittsburgh, PA is characterized by complex terrain, periods of heavy traffic, large industrial sources, and frequent inversion events.⁸ This combination of topography, meteorology, and emission sources results in significant spatial variability in many pollutant concentrations, including fine particulate matter (PM_{2.5}) and black carbon (BC). Although regional industrial air pollution has decreased over recent decades, local emissions inventories remain dominated by a few large legacy steel mills and coke works southeast of Pittsburgh (Edgar Thomson Steel Works and Clairton Coke Works).^{9–11} Accordingly, Pittsburgh and Allegheny County remain in federal PM_{2.5} non-attainment,^{12,13} as local regulatory monitors exceed both the average annual (> 12 µg/m³) and daily (> 35 µg/m³) National Ambient Air Quality Standards for PM_{2.5}.¹² The combination of hills and river valleys, episodes of traffic congestion, and large legacy industrial sources means that

Department of Environmental and Occupational Health, University of Pittsburgh Graduate School of Public Health, Pittsburgh, Pennsylvania, USA. Correspondence: Dr. Brett J. Tunno, Department of Environmental and Occupational Health, University of Pittsburgh Graduate School of Public Health, Bridgeside Point, 100 Technology Drive, Room 529, Pittsburgh, PA 15219-3130, USA.

Tel.: +1 724 288 3778. Fax: +1 412 624 3040.

E-mail: Bjt25@pitt.edu

Received 27 October 2014; revised 18 January 2015; accepted 23 January 2015; published online 29 April 2015

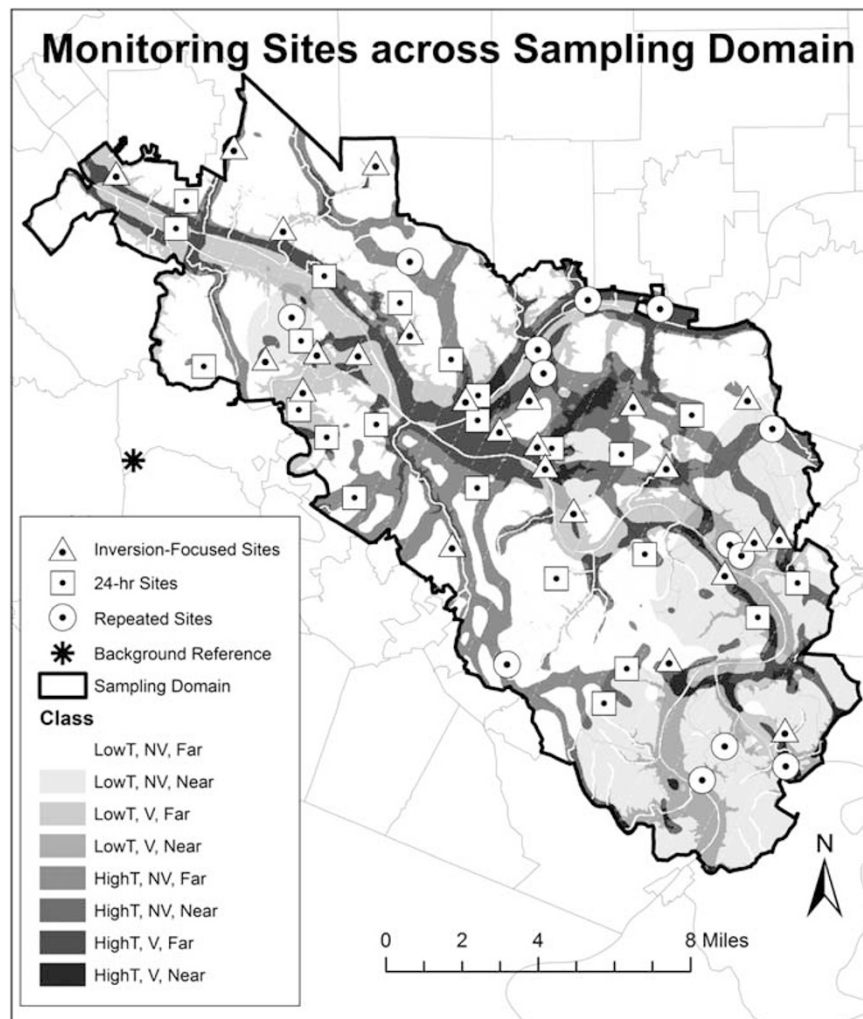


Figure 1. Domain within Allegheny County stratified according to classification system, with monitoring sites. Sites were classified by traffic density (high vs low), elevation (valley vs non-valley), and proximity to industry (near vs far).

temperature inversion events can trap pollutants along the river valleys, where local industry is concentrated and traffic emissions can accumulate,^{8,14} intensifying spatial contrasts in pollutant concentrations.^{15,16}

An improved understanding of pollution variability across Pittsburgh, with attention to this complex interplay among topography, industrial and vehicular sources, and frequent morning inversions could provide insights and reproducible models useful in other urban areas.¹⁷ In a previous study, we identified frequent morning temperature inversion events in the Pittsburgh industrial suburb of Braddock from 0700 to 1000 hours.⁸ US Environmental Protection Agency (EPA) regulatory data from multiple stations in Allegheny County suggested increased $PM_{2.5}$ concentrations between 0600 and 0700 hours, coinciding with the start of morning rush-hour traffic.

Here, we employed a series of saturation monitoring campaigns to characterize intra-urban variability under two different temporal regimes (inversion-focused and 24-h integrated), using a suite of programmable monitors to capture spatial variation during hypothesized peak concentration hours.¹⁴ A citywide sampling design allowed for exploration of mean variation in $PM_{2.5}$ and BC concentrations based on spatial, seasonal, and temporal differences. Seasonal pollutant-specific LUR modeling was applied to characterize intra-urban variability in air pollution concentrations, and smooth surface pollutant-specific concentration maps were

created.^{2,3,18} We hypothesized that inversion-focused sampling (0600–1100 hours, Monday through Friday) would reveal stronger spatial contrasts, and stronger impacts of local emissions sources, than would the 24-h week-long sampling scheme. We anticipated that $PM_{2.5}$ and BC concentrations would vary by sampling hours, elevation, proximity to industry, traffic density, and frequency of inversion events.

METHODS

Study Design

Sampling site allocation and classification is detailed in Shmool et al.¹⁴ Briefly, GIS-based methods were used to quantify spatial distributions of local pollution sources (i.e., proximity to industry, traffic density) and potential modifiers (i.e., elevation). These three factors hypothesized to impact upon local concentrations were dichotomized and cross-stratified to represent eight different combinations (Figure 1). Regular 100 m² lattice cells were characterized according to these cross-strata, and sampling locations ($n=37$ /year) were systematically allocated by stratified random sample to capture spatial and source variability across the Pittsburgh metropolitan area (~388 km²). The reference site, used throughout the sampling season to assess regional background pollution, was located at Settlers Cabin Park in Carnegie, PA, approximately 9 miles upwind of the study area. For comparison, an urban reference site in the community of Braddock, PA was also sampled every session.

Table 1. GIS-based source density indicators used for LUR modeling.

Source category	Covariates examined (50–1000 m concentric radial buffers)	Data source
Traffic density indicators	Mean density traffic (primary roads) Mean density traffic (primary and secondary roads) Number of signaled intersections	Pennsylvania Department of Transportation (PADOT)
Road-specific measures	Average daily traffic on nearest primary road Distance to nearest major road Summed length of primary roadways Summed length of primary and secondary roadways	PADOT
Truck, bus, and diesel	Mean density of bus traffic Distance to nearest bus route Outbound and inbound trip frequency per week summed by route Mean density of heavy truck traffic on nearest primary roadway	Google Transit (11/11 – 3/12) PADOT
Population	Census population density (blockgroup)	US Census Bureau (2010)
Land use/built environment	Total area of industrial parcels Total area of commercial parcels Total area of industrial and commercial parcels Percent developed imperviousness Land use/land cover (LULC) urban built-up total area from orthophotography	Allegheny County Assessment Data, by parcel (2011) National Land Cover Dataset (NLCD, 2006) Southwestern Pennsylvania Commission (SPC, 2011)
Industrial emissions	Mean density of total PM _{2.5} emitted per meter Mean density of total SO ₂ emitted per meter Mean density of total NO _x emitted per meter Mean density of total VOCs emitted per meter	National Emissions Inventory (NEI, 2011)
Transportation facilities	Distance to nearest active railroad Summed line length of active railroads Distance to nearest bus depot	SPC, 2011
Potential modifying factors		
Topography	Average elevation	National Elevation Dataset (NED, 2011)
Meteorology	Temperature/relative humidity Frequency of inversions Wind direction and wind speed	Obtained from sampler University of Wyoming, Department of Atmospheric Science (2011–2012) National Oceanic and Atmospheric Association (NOAA, 2011–2012)

For each year 1 (inversion-focused monitoring), 5-day (Monday through Friday) sampling session, samplers operated during 0600–1100 hours at six randomly-selected sites (25 h of sampling) to capture hours of frequent atmospheric inversions and heavy traffic, limit of detection for multiple pollutants. Six sampling sessions were performed during summer 2011 (July 25 to September 9) with 1 week skipped for logistical reasons. Sessions were repeated for winter 2012 (January 16 to February 24).

Year 2 (24-h integrated monitoring) sampling methods were similar to year 1, although samplers operated for an integrated 24-h 7-day sample of 15 min per hour (42 h of sampling). One-third of the sites ($n = 13$) from inversion-focused sampling were repeated during 24-h sampling, for direct comparison.¹⁴ Year 2 sampling was performed in summer 2012 (June 5 to July 26) and repeated in winter 2013 (January 8 to March 10), with a total of eight sites sampled per session. Due to limited equipment availability, each sampling scheme was performed during a separate year.

Monitoring Instrumentation and Quality Control

Using a temperature (20°C) and relative humidity (35%) controlled glove box (PlasLabs Model 890 THC), 37 mm Teflon filters (Pall Life Sciences) were equilibrated for 48 h and then pre-weighed and post-weighed using an ultramicrobalance (Mettler Toledo Model XP2U). Sampling units were custom-designed to capture integrated street-level samples of PM_{2.5}; Harvard Impactors (Air Diagnostics and Engineering) with 37-mm Teflon filters and a HOBO data logger (Onset Computer Corporation) were contained in waterproof Pelican cases.^{19,20} Instruments were programmed for specific hours of sampling using a chronotroller (ChronTrol Corporation). A tetraCal volumetric air flow calibrator (BGI Instruments) was used to calibrate the flow to 4.0 liters/min. A HOBO data logger recorded temperature and relative humidity at 15-min intervals. All samplers were deployed on utility poles at a height of 3 m. PM_{2.5} concentrations were calculated using presampling and postsampling Teflon net filter weights and sampling volume. BC was measured (in absorbance units) using an

EEL43M Smokestain Reflectometer (Diffusion Systems) and standard protocols.²¹

GIS-based Source Density Indicators

GIS-based covariates were calculated across a range of source indicator categories (Table 1). All analyses were conducted using ESRI ArcInfo Version 10 (Redlands, CA, USA), and all covariates summarized within concentric radial buffers surrounding each monitoring location (50–1000 m). Roadway shapefiles for Allegheny County were obtained from Pennsylvania Department of Transportation's (PennDOT) publicly available annualized average daily vehicle-count data for primary roadways. Traffic covariates included: mean kernel vehicle density, sum of traffic signaled intersections, mean density of bus and truck traffic, summed length of roadway (in feet), and total traffic kernel density (vehicle count). Point elevation at each monitoring location, and the average elevation within multiple radial buffers, was assessed using the National Oceanographic and Atmospheric Digital Elevation Model.^{22,23} For industrial emissions, PM_{2.5} (filterable plus condensable), nitrogen oxides (NO_x), sulfur dioxide (SO₂), and volatile organic chemicals (VOCs) were aggregated from the USEPA's 2011 National Emissions Inventory (NEI) over a six-county region, and estimated emissions (tons) were weighted using inverse distance weighting.²⁴ Total areas of industrial and commercial parcels within varying buffers were calculated using 2012 assessment data from the Allegheny County Office of Property Assessments. The locations and areas of commercial and industrial areas were validated using land use/land cover data derived from 2000 to 2001 orthophotography. The distance from, and summed line length of, active railroad was assessed. Census data from 2011 were obtained at the block group level, to calculate population density.²⁵

Meteorology

Atmospheric sounding data (i.e., Skew-T diagrams) were used to identify the presence of temperature inversions during the 0600–1100 sampling

Table 2. Descriptive statistics for citywide air sampling temporally-adjusted pollutant concentrations and meteorology.

	Summer 2011	Winter 2012	Summer 2012	Winter 2013
PM _{2.5} (mean, SD, min–max $\mu\text{g}/\text{m}^3$)	14.35 (SD 3.97) (1.33–22.71)	12.76 (SD 2.97) (8.02–20.10)	13.94 (SD 2.01) (11.26–22.59)	11.26 (SD 2.01) (8.01–18.92)
BC (mean, SD, min–max abs)	1.64 (SD 0.91) (0.02–4.64)	1.34 (SD 0.53) (0.70–2.72)	1.06 (SD 0.36) (0.61–2.47)	0.93 (SD 0.35) (0.50–2.15)
Temperature (mean, min–max $^{\circ}\text{F}$)	69.26 (61.9–78.1)	34.07 (19.8–40.7)	76.99 (70.0–82.0)	34.07 (25.1–44.5)
Relative humidity (mean, min–max %)	85.85 (75.5–98.7)	78.55 (69.5–87.7)	61.59 (46.2–79.2)	72.19 (55.3–85.2)
Wind speed (mean, min–max m/s)	1.79 (1.3–2.2)	3.20 (2.3–3.9)	2.66 (2.2–3.5)	4.21 (3.2–5.0)
Wind direction (percentage of sessions)	33% W 33% SW 17% SE 17% NE	50% W 33% W 17% N	33% SW 33% SW 17% W 17% S	33% NW 33% W 33% SW
Inversion presence (percentage of sessions)	50% 2 inversions 33% 3 inversions 17% 1 inversion	50% 3 inversions 33% 4 inversions 17% 2 inversions		

hours (year 1) and the 24-h integrated hours (year 2).²⁶ A binary inversion metric was created in which inversion presence on each sampling day was determined. Hourly meteorological data (e.g., wind speed, wind direction, temperature, precipitation, ceiling height) were downloaded from the National Climate Data Center in TD-3505 (ISHD — full archival) format. Radiosonde upper air data was collected at the Pittsburgh National Weather Service station located in Moon Township, PA, approximately 15 miles west (upwind by predominant wind direction) of Pittsburgh and was obtained from the National Oceanic and Atmospheric Administration.²⁷

Temporal Adjustment

PM_{2.5} and BC concentrations at distributed monitoring sites were temporally adjusted using the background reference concentration, which ran concurrently with all sampling sessions. Because all monitoring sites could not be sampled in the same week, temporal adjustment was performed using a reference site, to allow for comparisons between sites. This adjustment also allows us to assess the temporal contribution in the LUR models. Temporal adjustment methods and sensitivity tests are detailed elsewhere¹⁴ PM_{2.5} concentration at a particular monitoring location was divided by the specific weekly PM_{2.5} concentration from the background site then multiplied by the seasonal average PM_{2.5} concentration from the background site.

Quality Assurance/Quality Control

A seventh monitoring session was performed each season to co-locate monitors at four randomly-selected sites. Field blanks were deployed during each session, and all concentrations were blank-corrected. Hourly PM_{2.5} concentrations were obtained from Allegheny County Health Department (ACHD) for the Liberty, Lawrenceville and Avalon EPA Air Quality System (AQ5) monitoring locations. ACHD data were aggregated to match sampling hours each session, and used as additional validation sites, to corroborate temporal trends observed at our background and urban reference sites. BC data was not available from ACHD monitors.

Statistical Analysis

Descriptive statistics, scatterplots, and histograms were used to characterize distributions of PM_{2.5} concentrations and BC absorbance, as well as spatial covariates (e.g., traffic density, elevation, industrial emissions) and temporal covariates (temperature, relative humidity, wind speed, wind direction, frequency of inversions) (Table 2). Before modeling, bivariate source–pollutant correlation analyses were performed. Data analysis and model building were performed separately for PM_{2.5} and BC and for summer and winter seasons. Statistical analyses were conducted using the Proc GLM command in SAS version 9.3 (SAS Institute, Cary, NC, USA).

LUR models were implemented using manual forward step-wise linear regression to assess raw PM_{2.5} and BC concentrations for the summer and winter seasons, using an adapted version of the modeling approach described in Clougherty et al.¹⁹ The full set of source indicator covariates (Table 1) were tested individually for each pollutant by year and by season.

First, bivariate correlation coefficients (Spearman rho) were examined, and the two covariates with the highest correlations from each source category were individually incorporated, ordered by strength of the bivariate correlation. Temporal trends in PM_{2.5} and BC were first incorporated into LUR models using the sampling session-specific background concentration. Source terms with the strongest univariate correlation with the temporally-adjusted pollutant were then incorporated individually, in descending order by strength of the bivariate correlation. Regression models were sequentially fit to assess overall model improvement at each stage, using the coefficient of determination (R^2) and removing non-significant covariates in the order of descending P -value, until all terms were significant ($P < 0.05$). Covariates were removed, at any stage, if variance inflation factor became > 2.0 . Finally, we tested modification of each significant source term effect by inversion frequency, elevation, and wind speed (median dichotomized).

LUR model residuals were mapped to identify systematic spatial variation and locations poorly predicted by LUR, suggesting incorporation of additional covariates (i.e., inverse distance to NEI sites, elevation). Semivariograms of residuals were created in GIS, and residuals were mapped against latitude and longitude coordinates (decimal degrees) of monitoring locations to explore residual patterns. Spatial autocorrelation in residuals was tested using Moran's I statistic. For each final model, a spatial R^2 (roughly, the proportion of spatial variance explained by LUR terms) was derived, using final LUR covariates to predict temporally-adjusted concentrations.

Predicted PM_{2.5} concentrations and BC absorbance were mapped across a regular $100 \times 100 \text{ m}^2$ grid, smoothed using inverse distance weighting (IDW), allowing spatial influence from nearest 100 grid cell centroids. Isolines were calculated to connect points of equal concentration across the IDW surfaces, for visualization. Contour intervals of $2 \mu\text{g}/\text{m}^3$ for PM_{2.5} and 0.5 abs units for BC were selected for display.

Sensitivity Analyses

Covariate selection was sensitivity-tested using scatterplots to assess fit between each significant predictor and raw pollutant concentrations, to ensure that candidate covariates selected captured variability across the range of the data, not reliant on outliers or influential points. Tree structures and Random Forest automated methods were performed to corroborate covariate selection. A scatterplot of each retained term was tested against the residual of the prior model in the sequential model-building process to check for outliers. Model residuals were examined to ensure normality. Spatial autocorrelation across the residuals of the distributed sites was determined using Moran's I , and spatial correlations were determined using generalized additive models (GAM) to determine whether residual smoothing was required.

Sensitivity of the model-building procedure to temporal adjustment was assessed by comparing two methods (background only vs background and urban reference site adjustment) and by modeling covariates against temporally-adjusted pollutant concentrations to determine the percentage of spatial variation across the eight LUR models. Backwards elimination from multivariate linear models including all covariates with significant bivariate correlations with pollutant concentrations further corroborated model structure and covariate selection. For validation of LUR predictions, a random 20% of sites ($n = 7$) were removed from the analysis, and the LUR

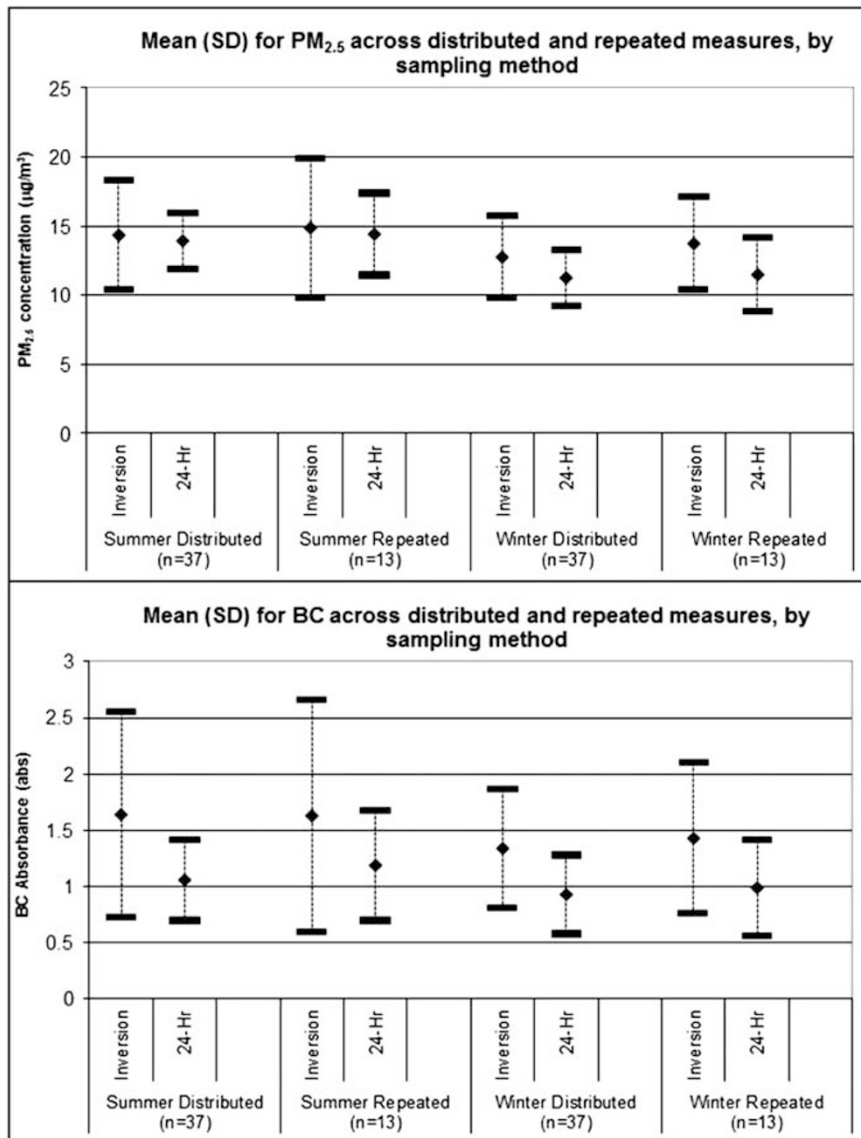


Figure 2. Heightened spatial contrasts in temporally adjusted PM_{2.5} concentrations and BC absorbance were found under inversion-focused monitoring compared with 24-h integrated. The same pattern was found across repeated sites that were monitored under both sampling regimes.

model was re-fit and used to predict pollutant concentrations at withheld sites. Analyses were performed in SAS, v 9.3, GIS ArcInfo, v 10.1 (ESRI, Redlands, CA, USA), R statistical software v 2.12.1, and Microsoft Excel 2010.

RESULTS

Inversion-Focused Monitoring (Year 1, Summer 2011/Winter 2012)
 Temporally-adjusted PM_{2.5} concentrations and BC absorbance were, on average, higher in the summer compared with winter, using inversion-focused sampling. Regional inversions were observed on one to three days during each summer session and on two to four days during each 5-day winter session (Table 2).

24-h Integrated Monitoring (Year 2, Summer 2012/Winter 2013)
 Temporally-adjusted PM_{2.5} concentrations and BC absorbance were, on average, higher in the summer compared with winter in 24-h sampling. PM_{2.5} concentrations and BC absorbance were lower, on average, using 24-h sampling, compared with inversion-focused measures, for both summer and winter (Table 2). Differences

between PM_{2.5} concentrations by sampling scheme were statistically significant for winter ($P=0.04$), but not for summer ($P=0.70$) (Supplementary Figure S1) We found statistically significant elevated BC levels for inversion-focused sampling in both the summer ($P=0.002$) and winter ($P=0.0002$) (Supplementary Figure S2) Greater variability in pollutant concentrations was detected in the inversion-focused campaigns compared with 24-h integrated (Figure 2).

Comparison of Repeated Sites

A subset ($n=13$) of monitoring sites were sampled using both sampling regimes, for direct comparability. Temporally-adjusted PM_{2.5} concentrations in both seasons were higher under inversion-focused sampling (summer mean = 14.88 (SD = 5.04) µg/m³; winter mean = 13.74 (SD = 3.37) µg/m³) than for 24-h sampling (summer mean = 14.42 (SD = 2.96) µg/m³; winter mean = 11.49 (SD = 2.66) µg/m³). Greater range of measured concentrations was found under inversion-focused sampling (Table 2, Figure 2).

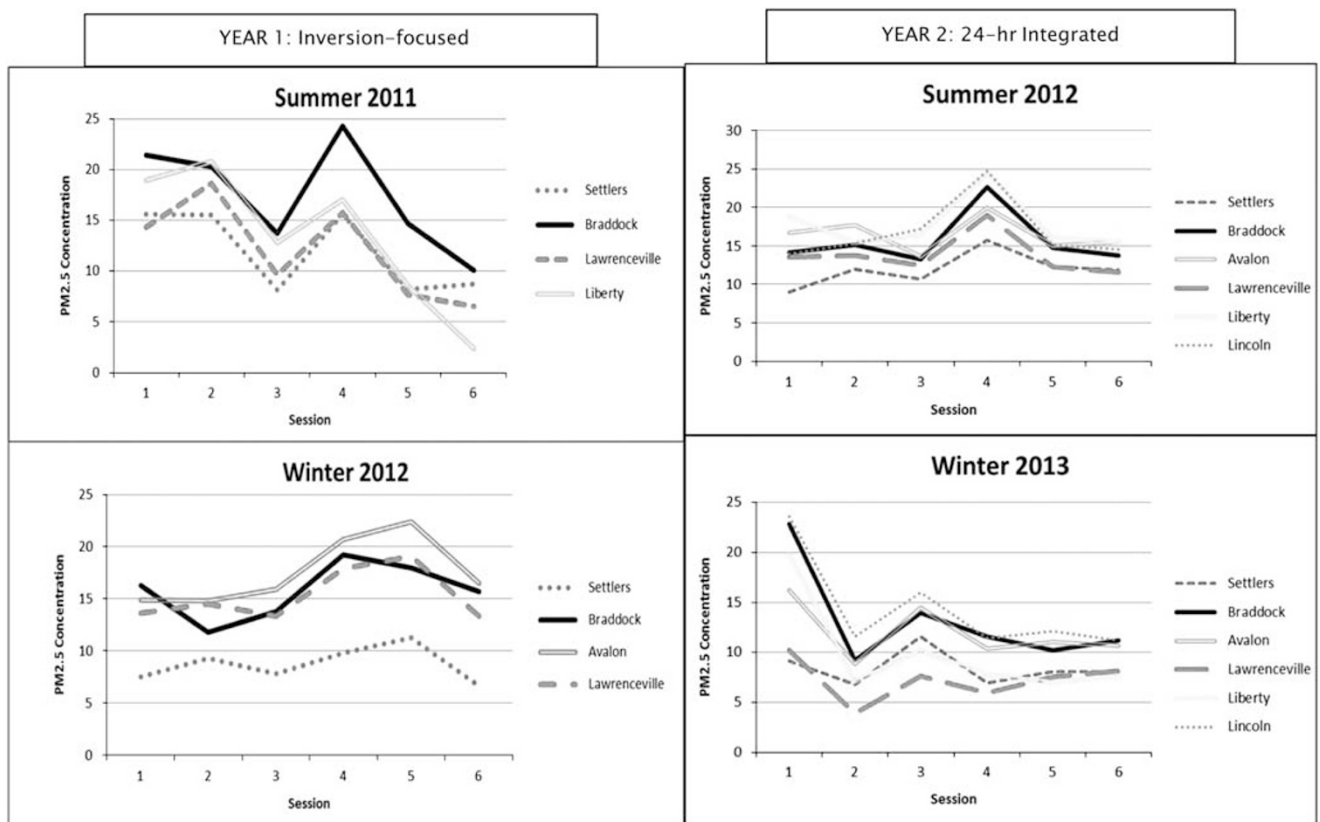


Figure 3. Temporal trends across reference sites for all sampling seasons.

Concentrations were similar using each sampling method at the background reference site during both seasons (summer 24-h = 11.92 (SD = 3.99) $\mu\text{g}/\text{m}^3$ vs summer inversion-focused = 11.91 (SD = 2.22) $\mu\text{g}/\text{m}^3$); (winter 24-h = 8.43 (SD = 1.76) $\mu\text{g}/\text{m}^3$ vs winter inversion-focused = 8.64 (SD = 1.73) $\mu\text{g}/\text{m}^3$). BC results were similar to those found for $\text{PM}_{2.5}$ concentrations.

$\text{PM}_{2.5}$ Variability at Reference Sites

Season-specific $\text{PM}_{2.5}$ temporal trends at the background reference site (Settlers), used for LUR modeling, were corroborated by trends observed at an urban reference site (Braddock) and by EPA sites ($n=2$ for inversion-sampling, $n=4$ for 24-h sampling) (Figure 3). Temporal trends from our reference sites and EPA monitoring sites were similar in both years.

Meteorology

For the specific sampling hours, under both sampling methods, wind direction was predominantly from the west, and wind speeds were generally higher during winter than summer (Figure 4). For summer inversion-focused sampling, session 6 winds were from the south. Higher variability in wind direction occurred during 24-h integrated summer sampling, where three sessions had non-westerly winds. The frequency of inversions within a given sampling session was not a significant predictor for either pollutant in any season ($P > 0.05$), and likewise an interaction term between inversion frequency and source terms was not significant.

$\text{PM}_{2.5}$ LUR Modeling

In all season-specific $\text{PM}_{2.5}$ LUR models, a substantial portion of variability in concentrations was explained by the background reference site (Table 3). For inversion-focused summer sampling, the

spatial pattern in $\text{PM}_{2.5}$ was predicted by commercial and industrial land use within a 200-m buffer (Table 3). For 24-h summer sampling, spatial variance in $\text{PM}_{2.5}$ was predicted by the IDW of $\text{PM}_{2.5}$ emissions and wind direction (interquartile range (IQR) $\text{PM}_{2.5}$ increase of 1.62 $\mu\text{g}/\text{m}^3$ with west/northwesterly winds; Table 3, Figure 5). For inversion-focused winter sampling, the LUR included land use and average wind speeds (m/s) ($-1.89 \mu\text{g}/\text{m}^3 \text{PM}_{2.5}$ per unit (m/s) increase in wind speeds; Table 3). For 24-h winter sampling, the number of signaled intersections within a 750-m buffer, industrial $\text{PM}_{2.5}$ emissions, and industrial land use contributed toward explaining spatial variation in $\text{PM}_{2.5}$ (Table 3; Figure 5).

Although heightened spatial contrasts were seen during inversion-focused sampling, a greater proportion of variability in $\text{PM}_{2.5}$ was explained by source terms and meteorology for 24-h sampling in both seasons (Figure 5). For the summer, the spatial R^2 was 0.34 for inversion-focused sampling, notably lower than the R^2 of 0.60 for 24-h sampling. During winter, the spatial R^2 was also lower for inversion-focused sampling (0.40), compared with the R^2 of 0.56 for 24-h sampling. Specifically, industrial emissions and wind direction were significant predictors under 24-h sampling but not under inversion-focused monitoring.

BC LUR Modeling

The weekly temporal term from the background reference site explained lesser variability in each BC model, relative to $\text{PM}_{2.5}$ models (Table 4). For inversion-focused summer sampling, the spatial pattern in BC was predicted by industrial land use and elevation within a 1000-m buffer (an IQR decrease of -0.58 abs per unit increase in elevation) (Table 4). For 24-h summer sampling, BC was predicted by the IDW of $\text{PM}_{2.5}$ emissions, commercial and industrial land use within a 200-m buffer, and

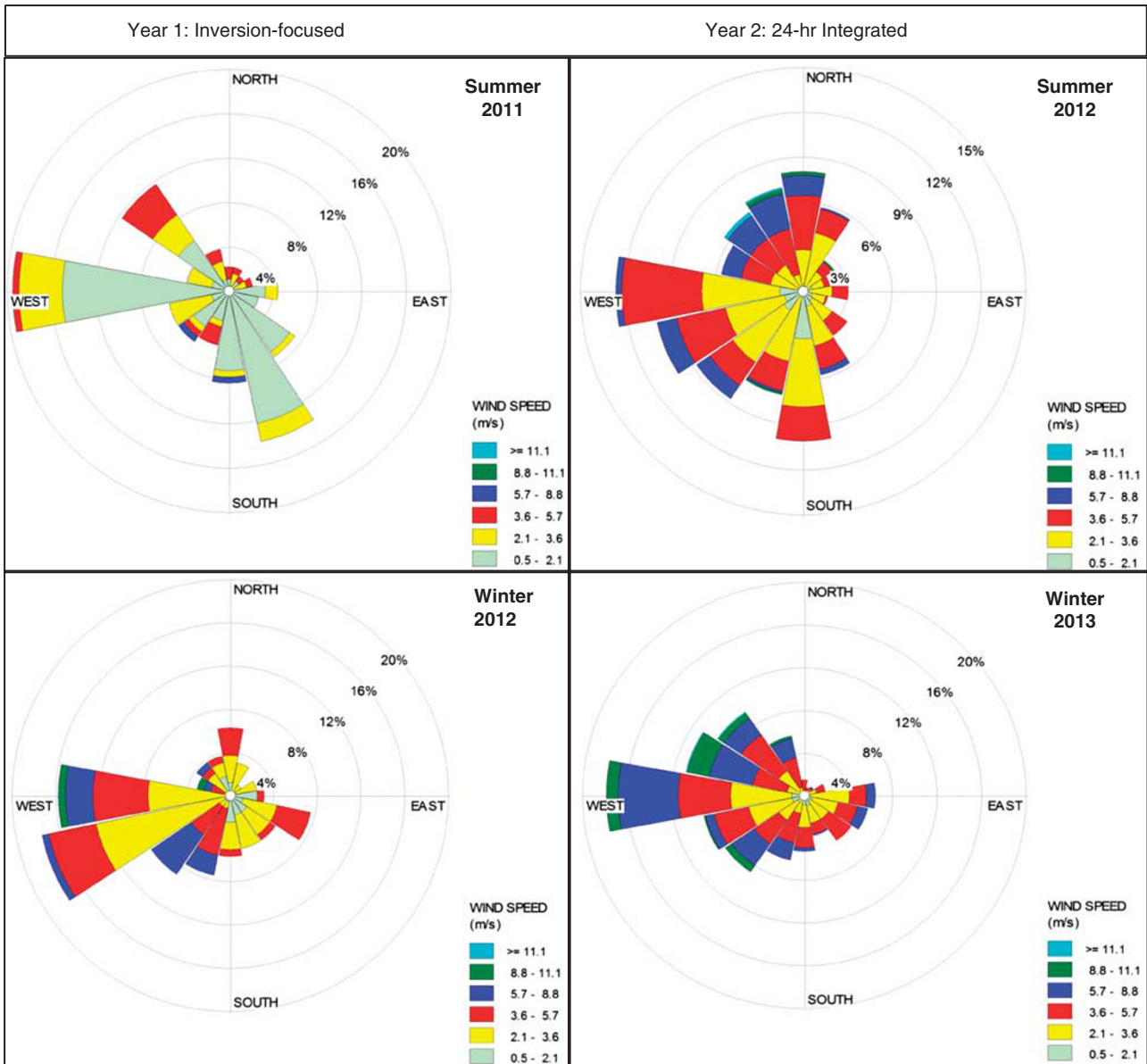


Figure 4. Wind rose diagrams for summer (top) and winter (bottom) across all hours of sampling.

wind direction (0.25 abs increase with west/northwesterly winds) (Table 4; Figure 6).

For inversion-focused winter sampling, industrial land use, signaled intersections within a 500-m buffer, average wind speed (m/s) (0.33 abs decrease per unit increase in wind speeds), and elevation explained spatial variation (Table 4). For 24-h winter sampling, the final model included the IDW of PM_{2.5} emissions and industrial land use (Table 4) (Figure 6). Additional LUR output for PM_{2.5} and BC models can be found in Supplementary Tables S1 and S2.

During summer months, a greater percentage of variability was explained for 24-h sampling. For winter sampling, more variability was explained under inversion sampling. A greater range of variability in BC absorbance was observed under the inversion-focused, vs the 24-h integrated sampling scheme, during both summer and winter (Figure 6). For summer, the inversion-focused spatial R^2 was 0.51, notably lower than the R^2 of 0.70 for 24-h sampling. During winter, the inversion-focused spatial R^2 was higher (0.75) than for 24-h sampling (0.54).

Sensitivity Analyses

Scatterplots ensured final models were not driven by outliers or influential points; one influential site was removed from summer inversion-focused modeling. Tree structures and Random Forest automated methods were used to confirm covariates that were incorporated and retained in final models. Moran's I and GAM indicated no spatial autocorrelation in the residuals across distributed monitoring sites. LUR models with temporal adjustment using the average concentration at both the urban and background reference sites provided comparable results to models using only the regional background site concentration. Removal of a random subset (20%) of monitoring sites did not significantly change any of the eight models, and predicted concentrations of the withheld sites were within 10–15% of the measured concentrations.

DISCUSSION

We found significant spatial variability in PM_{2.5} and BC ambient concentrations across 37 metropolitan sites under both inversion-

Table 3. Land use regression (LUR) covariates and model fits for inversion-focused and 24-h PM_{2.5}.

Covariates	LUR model		
	β (P-value)	IQR conc. increase ^a	Seq R ² ^b
<i>Inversion-focused summer PM_{2.5} (μg/m³)^c</i>			
Intercept	-1.11 (0.48)	—	—
Weekly reference PM	1.17 (< 0.0001)	—	0.66
Land use (Com+Ind) at 200 m	8.1×10^{-5} (0.0004)	2.86	0.77
<i>Inversion-focused winter PM_{2.5} (μg/m³)</i>			
Intercept	9.68 (0.01)	—	—
Weekly reference PM	0.91 (0.0006)	—	0.47
Land use (Com+Ind) at 200 m	5.3×10^{-5} (0.0002)	1.90	0.59
Wind speed (m/s)	-1.89 (0.007)	-0.79	0.68
<i>24-H summer PM_{2.5} (μg/m³)</i>			
Intercept	-2.44 (0.05)	—	—
Weekly reference PM	1.20 (< 0.0001)	—	0.65
IDW of PM _{2.5} emissions	2.27 (< 0.0001)	0.95	0.80
Wind direction			
Blowing from NW/W	1.62 (0.0005)	1.62	—
Blowing from SW/S	—	—	0.86
<i>24-H winter PM_{2.5} (μg/m³)</i>			
Intercept	-1.61 (0.21)	—	—
Weekly reference PM	1.26 (< 0.0001)	—	0.52
Signaled intersections within 750 m	0.14 (< 0.0001)	0.84	0.63
Land use (industry) at 750 m	5.9×10^{-6} (0.01)	0.82	0.79
IDW of PM _{2.5} emissions	1.70 (0.003)	0.71	0.84

^aIQR concentration increase = $\beta \times$ IQR of source indicator. ^bSeq R² is the sequential model fit for each additional term incorporated into the model. ^cOne influential point removed for LUR modeling. Bold values are the percentages of explained pollutant variability according to final LUR models.

Table 4. Land use regression (LUR) covariates and model fits for inversion-focused and 24-h black carbon (BC).

Covariates	LUR Model		
	β (P-value)	IQR conc. increase ^a	Seq R ² ^b
<i>Inversion focused summer BC (abs)^c</i>			
Intercept	2.18 (0.11)	—	—
Weekly reference BC	2.53 (0.0007)	—	0.26
Land use (industry) at 750 m	3.0×10^{-6} (0.01)	0.41	0.57
Elevation at 1000 m	-0.009 (0.01)	-0.58	0.64
<i>Inversion focused winter BC (abs)</i>			
Intercept	3.10 (0.12)	—	—
Weekly reference BC	0.60 (0.82)	—	0.04
Land use (industry) at 750 m	1.9×10^{-6} (0.0006)	0.26	0.42
Signaled intersections within 500 m	0.05 (0.01)	0.14	0.60
Elevation at 1000 m	-0.005 (0.02)	-0.32	0.67
Wind speed (m/s)	-0.30 (0.001)	-0.33	0.76
<i>24-H summer BC (abs)</i>			
Intercept	-0.31 (0.36)	—	—
Weekly reference BC	1.55 (0.01)	—	0.12
IDW of PM _{2.5} emissions	0.36 (< 0.0001)	0.15	0.52
Land use (Com+Ind) at 200 m	4.5×10^{-6} (< 0.0001)	0.13	0.64
Wind direction			
Blowing from NW/W	0.25 (0.001)	0.25	—
Blowing from SW/S	—	—	0.74
<i>24-H winter BC (abs)</i>			
Intercept	-0.09 (0.56)	—	—
Weekly reference BC	1.31 (< 0.0001)	—	0.28
IDW of PM _{2.5} emissions	0.38 (0.001)	0.16	0.61
Land use (industry) at 750 m	1.0×10^{-6} (0.02)	0.14	0.67

^aIQR concentration increase = $\beta \times$ IQR of source indicator. ^bSeq R² is the sequential model fit for each additional term incorporated into the model. ^cOne influential point removed for LUR modeling. Bold values are the percentages of explained pollutant variability according to final LUR models.

focused and 24-h integrated sampling schemes for two summer and two winter seasons (June 2011 to March 2013). As hypothesized, the inversion-focused sampling approach revealed greater spatial contrasts in concentrations across a region of complex terrain, in both PM_{2.5} and BC, compared with 24-h sampling. These differences were greater by sampling scheme

than by season for both pollutants. However, GIS-based source terms did not explain more concentration variability under the inversion-focused approach compared with 24-h sampling.

The hours we hypothesized as “peak” exposure hours were selected based on: (1) morning rush hour congestion, and (2) the likelihood of inversion presence, based on our previous mobile

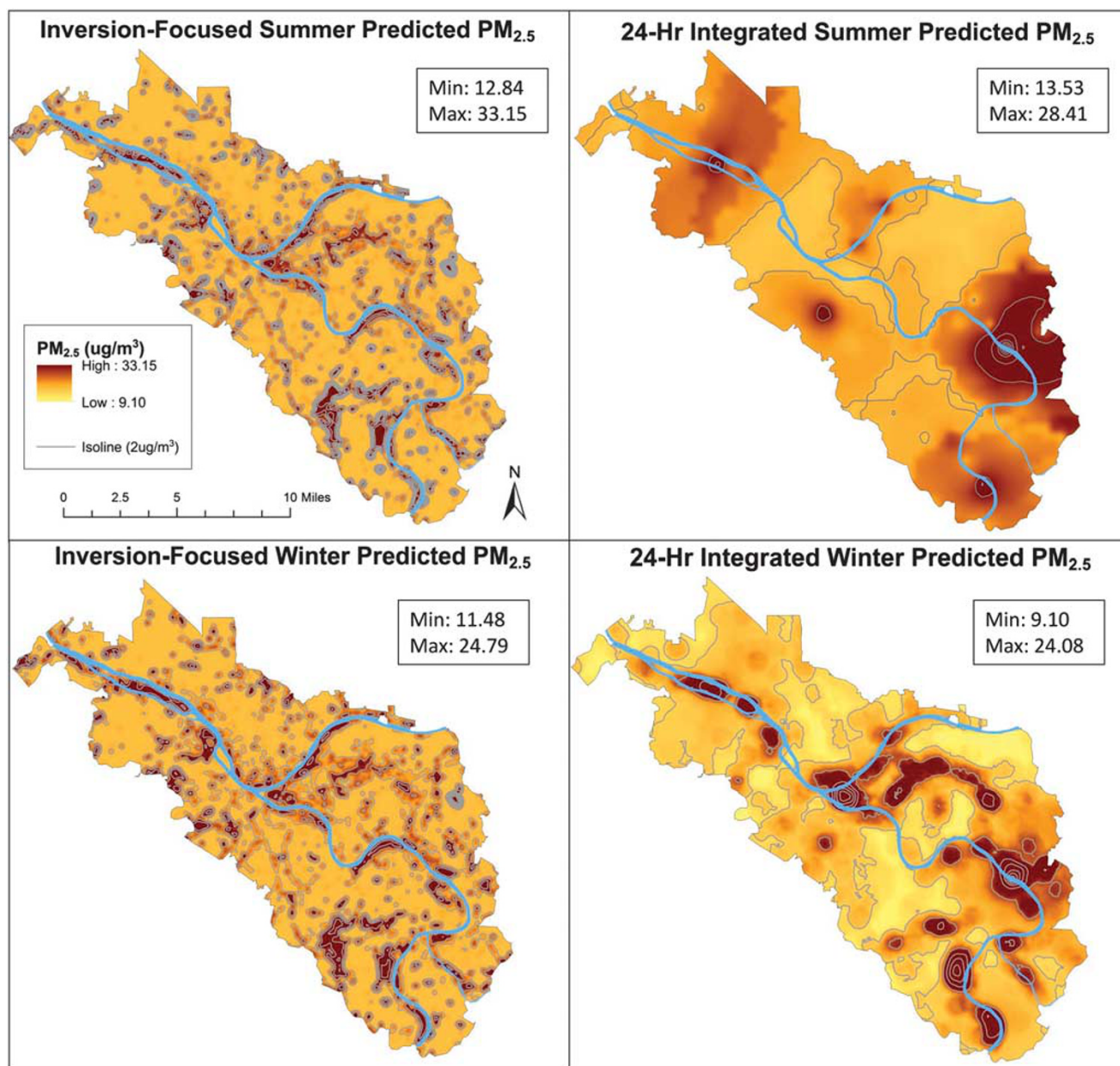


Figure 5. Seasonally-averaged predicted $PM_{2.5}$ exposure surface maps for inversion-focused summer and winter (left) and 24-h integrated summer and winter (right) sampling. For the 24-h integrated summer $PM_{2.5}$ map, wind directions were assumed to be W/NW (predominant wind direction), as a covariate in the LUR model. For the inversion-focused winter $PM_{2.5}$, wind speeds were averaged across the season for these specific sampling hours and applied to all sampling locations, as a covariate in the LUR model.

monitoring campaign.⁸ Our results indicate that there is greater spatial variation in exposures during these hours, which can be captured using this systematic saturation approach.

$PM_{2.5}$ temporal trends at our background reference monitor were consistent with regulatory monitors for both seasons and years regardless of sampling scheme. Under both sampling schemes, higher $PM_{2.5}$ concentrations and BC absorbance were generally found at sites nearer to industry or at lower elevation (i.e., in valleys).

$PM_{2.5}$ models were comparable across seasons, and in the LUR models, 47–66% of the spatial variability in concentrations was explained by the background reference site, regardless of sampling method, although GIS-based source terms generally explained a greater percentage of variability in the 24-h integrated

models. The combination of commercial and industrial parcels were stronger than commercial or industrial parcels alone, suggesting higher activity in an area with more traffic and trucks may lead to heightened pollutant levels. One explanation for identifying weaker source effects in the “peak” hour models is that the spatial covariates generally available for LUR represent long-term source averaging (e.g., annual average emissions or average daily traffic), and thus may not accurately capture peak concentrations such as during morning rush hours. Because static GIS-based covariates may not be sufficient, more sophisticated covariates such as rush hour traffic information, vehicle fleet breakdown, or site-specific source dispersion information may better complement our temporally purposeful sampling scheme. Relatedly, 7-day samples may be better predicted by annual

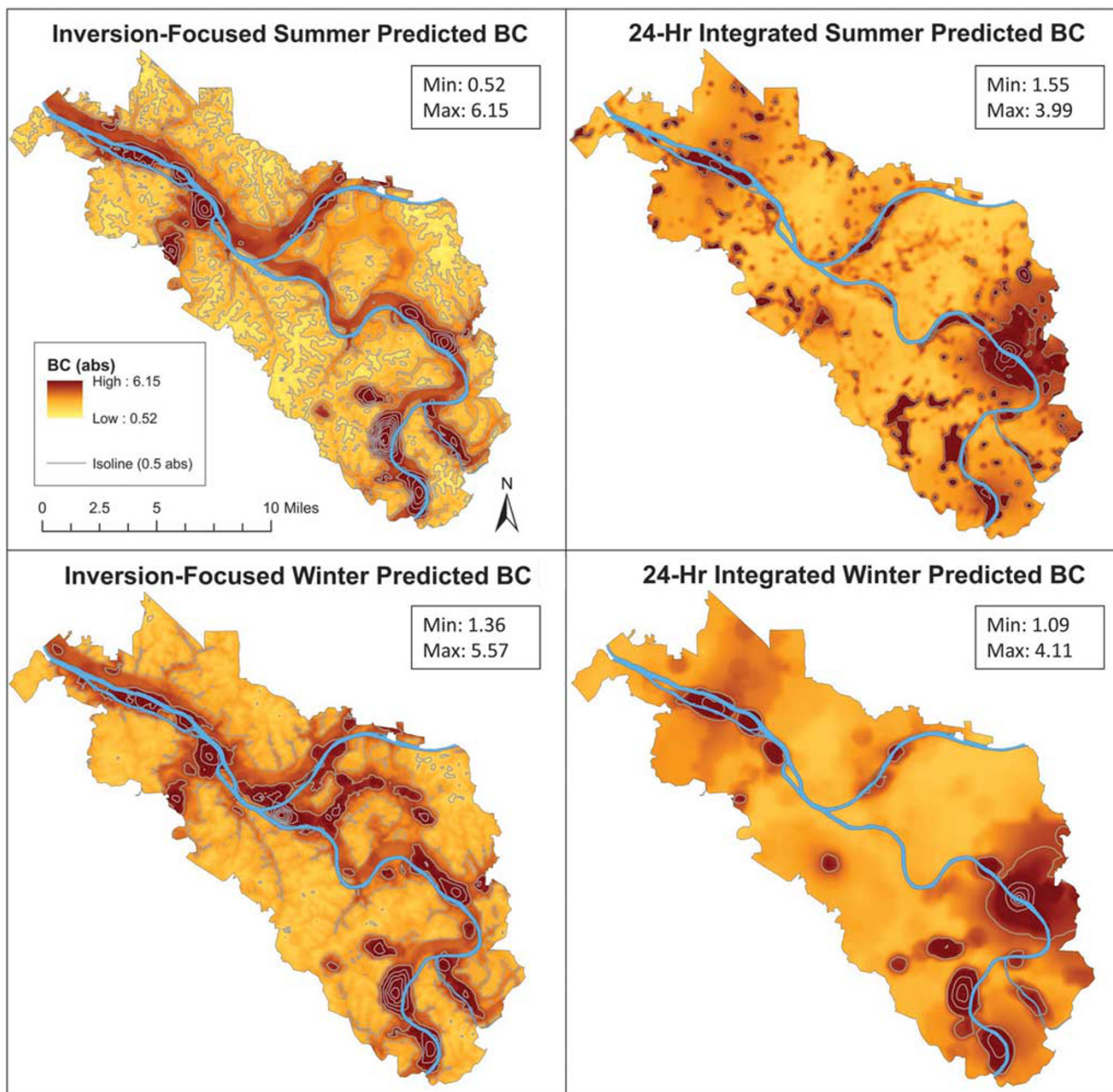


Figure 6. Seasonally-averaged predicted BC exposure surface maps for inversion-focused summer and winter (left) and 24-h integrated summer and winter (right) sampling. For the 24-h integrated summer BC, wind directions were assumed to be W/NW (predominant wind direction), as a covariate in the LUR model. For the inversion-focused winter BC, wind speeds were averaged across the season for these specific sampling hours and applied to all sampling locations, as a covariate in the LUR model.

average covariates (e.g., traffic density) than are 5-day samples—an effect which is not accounted for by using period-specific reference site data.

BC models were also generally comparable across seasons, with less variability explained by the background reference site ($R^2 = 0.04 - 0.28$) and greater observed local influence of land use, industrial emissions, and elevation. Number of signaled intersections was the only traffic indicator found significant in any model — for winter BC under inversion-focused sampling and for winter $PM_{2.5}$ under 24-h sampling. Other studies have found significant traffic contributions, as total and truck traffic densities were found for wintertime $PM_{2.5}$ and BC in the New York City Community Air Survey (NYCCAS) and traffic density within a 300-m buffer was found in Los Angeles.^{14,19}

Inversions and Meteorology

Although we found greater differences in concentrations across sites under inversion-focused sampling, frequency of inversions was not a significant predictor or source modifier. We previously found that inversion frequency modified the effect of elevation but did not remain in final inversion-focused LUR models here. This may be due to the low inversion variation between sampling sessions, and high frequency of inversions in our data set — producing minimal variation across sessions, and making it difficult to observe effects. In addition, variation in intensity of inversion events was not captured in our meteorological data — which may be more important than inversion presence. Finally, our inversion characterization (based on upwind, regional data

from Pittsburgh International Airport) may be incompatible with fine-scale spatial effects within our domain.

Meteorological factors such as higher temperature and relative humidity have been found to have a role in higher PM_{2.5} concentrations in the eastern United States, and although temperature and relative humidity were collected at each sampling unit, contemporaneous with sampling events, these factors were not significant predictors in LUR modeling. Changes in temperature and relative humidity may be accounted for in the background temporal term used in the models.²⁸ Given less spatial variance in PM_{2.5} across our sites, relative to the temporal contribution from the upwind reference monitor, this may indicate a greater relative influence of regional contribution and secondary formation.

Elevation and Complex Terrain

In our domain, elevation is spatially confounded with pollution sources, as many industries and highways in our region are located within the river valleys. Our stratified random sampling approach was designed to help distinguish these two potentially important drivers of local concentrations and examine the role of elevation in trapping pollutants during inversion events (0600–1100 hours). However, source effects were not significantly modified by elevation in LUR models. The most significant elevation covariate was at the largest buffer (1000 m) — possibly a proxy for low source intensity in locations of complex terrain (e.g., sharp elevation gradients), rather than hypothesized topography–meteorology interactions. For BC, the significance of the elevation term may reflect the influence of local sources relative to long-range transport, given the lesser amount of variability explained by the background concentration.

Sampling Density

Our study in Pittsburgh included 37 sampling locations across 388 km² (a sampling density of 0.10 sites/km²). Although our domain was large — and included urban, suburban, and semi-rural sites — our sampling density is within the range of other LURs. Though some (e.g., NYCCAS) captured 150 sites in a 777 km² domain (0.19 sites/km²),²⁰ many LURs have been developed from much sparser data (e.g., in Los Angeles, 23 sampling sites across 98,500 km² (0.0002 sites/km²), and in Vancouver, 80 sampling sites across 2199 km² (0.04 sites/km²)).^{6,7}

Limitations

One key limitation of our study was the logistical inability, due to equipment availability, to capture inversion-focused and 24-h integrated samples simultaneously at each site. Nevertheless, concentrations from the 13 repeated sites were correlated across years and seasons, reference site data using each method was consistent with regulatory monitors, and temporal adjustments explained comparable variability in both years. Data were obtained from the meteorological station at Pittsburgh International Airport — rather than within our domain — to assess inversion frequency, wind speeds, and wind direction across sampling sites. Inversion presence was assessed as the percentage of days on which an inversion was detected, although we may have missed some events, as environmental soundings are collected only twice per day. The overall high frequency of inversion events resulted in low variation across sessions and sites. Finally, both “low” and “high” elevation sites were assigned the same inversion frequency metric, likely misclassifying inversion presence at sites of very different elevation. We examined inversion effects separately for low and high elevation sites, although this analysis may have had insufficient variation.¹⁴

Strengths and Implications

This inversion-focused study allowed us to determine spatial variation during selected hours of the day (hypothesized frequent inversion/high source-intensity hours of 0600–1100 hours) — hours during which we previously detected elevated PM_{2.5} in a mobile monitoring campaign in a nearby industrial community.⁸ Using a unique programmable sampling system and a structured modeling approach, our results show that local industrial sources of PM_{2.5} and elevation were key pollutant predictors during these inversion-focused hours. The range of meteorological data examined is a strength of our approach, although hourly meteorological data, and better measures of inversion intensity and local mixing height (especially, across complex terrain), would have improved specificity.

Across the metropolitan sampling domain, summer PM_{2.5} concentrations and BC absorbance were significantly higher than winter concentrations under both sampling methods. Inversion-focused sampling produced greater spatial contrast across sites, although the source terms in these LUR models did not explain more variability compared with the 24-h sampling method. These LUR models will be used to derive exposure estimates for future local epidemiological studies, to compare the predictive power of 24-h vs “peak” exposure contrasts in explaining chronic disease outcomes. Our sampling approach could be replicated in other cities, with an emphasis on attempting to observe greater pollutant spatial contrast during hypothesized peak hours, which may be more predictive of some health outcomes.

CONFLICT OF INTEREST

The authors declare no conflict of interest

ACKNOWLEDGEMENTS

This work was supported in part by internal University of Pittsburgh Department of Environmental and Occupational Health funds and the Heinz Foundation. We thank Jeff Howell and Rebecca Dalton for field work assistance, Kyra Naumoff Shields for assistance with sampling permits and other logistics, and John Gorczynski and Holger Eisl for training on sampling equipment. We also thank the Pittsburgh Department of Public Works (Alan Asbury and Mike Salem), Parks Department of Allegheny County, and Duquesne Light Company for monitoring permissions and the Allegheny County Health Department for regulatory monitoring data and support.

REFERENCES

- 1 Ryan P.H, LeMasters G.K. A review of land-use regression models for characterizing intraurban air pollution exposure. *Inhal Toxicol* 2007; **19**(Suppl 1): 127–133.
- 2 Jerrett M, Burnett RT, Ma R, Pope CA 3rd, Krewski D, Newbold KB *et al*. Spatial analysis of air pollution and mortality in Los Angeles. *Epidemiology* 2005; **16**: 727–736.
- 3 Clougherty JE, Wright RJ, Baxter LK, Levy JI. Land use regression modeling of intra-urban residential variability in multiple traffic-related air pollutants. *Environ Health* 2008; **7**: 17.
- 4 Ross Z, English PB, Scalf R, Gunier R, Smorodinsky S, Wall S, Jerrett M. Nitrogen dioxide prediction in Southern California using land use regression modeling: potential for environmental health analyses. *J Expo Sci Environ Epidemiol* 2006; **16**: 106–114.
- 5 Jerrett M, Arain MA, Kanaroglou P, Beckerman B, Crouse D, Gilbert NL. Modeling the intraurban variability of ambient traffic pollution in Toronto, Canada. *J Toxicol Environ Health A* 2007; **70**: 200–212.
- 6 Abernethy RC, Allen RW, McKendry IG, Brauer M. A land use regression model for ultrafine particles in Vancouver, Canada. *Environ Sci Technol* 2013; **47**: 5217–5225.
- 7 Moore DK, Jerrett M, Mack WJ, Künzli N. A land use regression model for predicting ambient fine particulate matter across Los Angeles, CA. *J Environ Monit* 2007; **9**: 246–252.
- 8 Tunno BJ, Shields KN, Liou P, Chu N, Kadane JB, Parmanto B *et al*. Understanding intra-neighborhood patterns in PM_{2.5} and PM₁₀ using mobile monitoring in Braddock, PA. *Environ Health* 2012; **11**: 76.
- 9 Allegheny County Health Department (ACHD). *Air Quality Annual Report for 2011 with 1991–2011 ACHD, Pittsburgh, PA. Trends*, 2011.

- 10 Liou P.J, Georgopoulos P.G. New jersey: a case study of the reduction in urban and suburban air pollution from the 1950s to 2010. *Environ Health Perspect* 2011; **119**: 1351–1355.
- 11 Dominici F, Peng RD, Zeger SL, White RH, Samet JM. Particulate air pollution and mortality in the United States: did the risks change from 1987 to 2000? *Am J Epidemiol* 2007; **166**: 880–888.
- 12 Environmental Protection Agency (EPA). Area Designations for 2006 Fine Particle (PM_{2.5}) Standards, 2009. 19 July 2012 [cited 6 October 2012]; Available from <http://www.epa.gov/pmdesignations/2006standards/final/region3.htm>. EPA, Research Triangle Park, NC.
- 13 Kelly M, Besselman M *Point Source Emission Inventory Report*. Allegheny County Health Department: Pittsburgh, PA, USA, 2007.
- 14 Shmool J.L et al. Saturation sampling for spatial variation in multiple air pollutants across an inversion-prone metropolitan area of complex terrain. *Environ Health* 2014; **13**: 28.
- 15 Wallace J, Kanaroglou P. The effect of temperature inversions on ground-level nitrogen dioxide (NO₂) and fine particulate matter (PM_{2.5}) using temperature profiles from the Atmospheric Infrared Sounder (AIRS). *Sci Total Environ* 2009; **407**: 5085–5095.
- 16 Wallace J, Corr D, Kanaroglou P. Topographic and spatial impacts of temperature inversions on air quality using mobile air pollution surveys. *Sci Total Environ* 2010; **408**: 5086–5096.
- 17 Basagana X, Aguilera I, Rivera M, Agis D, Foraster M, Marrugat J et al. Measurement error in epidemiologic studies of air pollution based on land-use regression models. *Am J Epidemiol* 2013; **178**: 1342–1346.
- 18 Levy JI, Clougherty JE, Baxter LK, Houseman EA, Paciorek CJHEI Health Review Committee. *Evaluating heterogeneity in indoor and outdoor air pollution using land-use regression and constrained factor analysis*. *Res Rep Health Eff Inst* 2010; **152**: 5–80 discussion 81–91.
- 19 Clougherty JE, Kheirbek I, Eisl HM, Ross Z, Pezeshki G, Gorczynski JE et al. Intra-urban spatial variability in wintertime street-level concentrations of multiple combustion-related air pollutants: the New York City Community Air Survey (NYCCAS). *J Expo Sci Environ Epidemiol* 2013; **23**: 232–240.
- 20 Matte TD, Ross Z, Kheirbek I, Eisl H, Johnson S, Gorczynski JE, Kass D et al. Monitoring intraurban spatial patterns of multiple combustion air pollutants in New York City: design and implementation. *J Expo Sci Environ Epidemiol* 2013; **23**: 223–231.
- 21 ISO TC 146/SC3., ISO 9835:1993, Ambient air - Determination of a black smoke index. International Organization for Standardization; Geneva, Switzerland. 1993.
- 22 Gesch DB. The National Elevation Dataset. (ed). Maune D 2nd edn. *Digital Elevation Model Technologies and Applications: The DEM Users Manual*. American Society for Photogrammetry and Remote Sensing: Bethesda, Maryland, USA, 2007.
- 23 Gesch D, Oimoen M, Greenlee S, Nelson C, Steuck M, Tyler D, The National Elevation Dataset: Photogrammetric Engineering and Remote Sensing. Vol. 68. American Society for Photogrammetry and Remote Sensing: Bethesda, Maryland, USA, 2002.
- 24 EPA. 2011 National Emissions Inventory Data. 28 August 2013 [cited 10 August 2012]; Available from <http://www.epa.gov/ttnchie1/net/2008inventory.html>.
- 25 United States Census. 2010 Census Data. 2011; Available from http://www.census.gov/geo/www/2010census/centerpop2010/blkgp/CenPop2010_Mean_BG42.txt.
- 26 University of Wyoming. Environmental Soundings. 2011 [cited 20 March 2012]; Available from <http://weather.uwyo.edu/upperair/naconf.html>.
- 27 National Oceanic and Atmospheric Administration (NOAA)National Climatic Data Center. 2012 [cited 20 March 2012]; Available from http://cdo.ncdc.noaa.gov/pls/plclimprod/cdomain.selectdataelem?p_nDataSetId=11&p_cSubQueryBy=STATION&p_cPoeOptions=ADVANCED&p_asubqueryitems=72520094823.
- 28 Chu N, Kadane J, Davidson C. Using statistical regressions to identify factors influencing PM_{2.5} Concentrations: The Pittsburgh supersite as a case study. *Aerosol Sci Technol* 2010; **44**: 766–774.



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/4.0/>

Supplementary Information accompanies the paper on the Journal of Exposure Science and Environmental Epidemiology website (<http://www.nature.com/jes>)