Environ Health Perspect

DOI: 10.1289/EHP5311

Note to readers with disabilities: *EHP* strives to ensure that all journal content is accessible to all readers. However, some figures and Supplemental Material published in *EHP* articles may not conform to 508 standards due to the complexity of the information being presented. If you need assistance accessing journal content, please contact ehp508@niehs.nih.gov. Our staff will work with you to assess and meet your accessibility needs within 3 working days.

Supplemental Material

Moving beyond Fine Particle Mass: High-Spatial Resolution Exposure to Source-Resolved Atmospheric Particle Number and Chemical Mixing State

Qing Ye, Hugh Z. Li, Peishi Gu, Ellis S. Robinson, Joshua S. Apte, Ryan C. Sullivan, Allen L. Robinson, Neil M. Donahue, and Albert A. Presto

Table of Contents

Single particle aerosol mass spectrometer

Table S1. Summary of covariates used in building the LUR models. We also include covariates that are products of the two listed covariables in the Traffic category listed in the table. Data are from the database of Allegheny County Health Department, Pennsylvania Department of Transportation, National Emission Inventory, Pennsylvania Spatial Data Access and census.gov.

Table S2. Summary of the full LUR models.

Table S3. Summary of source-specific LUR models for traffic and cooking particle number concentration and mixing state in Pittsburgh PA built using only the sub-divided grid cells (200 m by 200 m).

Figure S1. Average mass spectra for three types of particles categorized using k-means clustering: background particles including inorganic-rich and oxygenated organic particles, traffic particles and cooking particles. Different colors refer to different chemical species.

Figure S2. Spectral comparison between mass spectra of source-resolved single particle clusters (traffic and cooking clusters) and spectra of bulk particle measurements of known sources from the AMS database. Each data point corresponds to fractional signals in the spectra. The red lines are the fitted lines.

Figure S3. Average size-dependent corrections of single particle measurement using data from a scanning mobility particle sizer co-located with the singe particle mass spectrometer at the central reference site. Error bars are the standard deviation from multiple days of co-located measurements.

Figure S4. The base map showing subdivision of the measurement areas for LUR building. Half of the ~1 km^2 areas were kept whole, while the other half were subdivided into 200 m x 200 m squares. The boundary of Pittsburgh city is shown in black. Base map sources: Light Gray Canvas Map, Esri, HERE, Garmin.

Figure S5. Vehicle density (vehicle/day/m) and restaurant density (number of restaurant/km^2) in the 20 areas. Circles are the average values of all the 5 m by 5 m rasters in the areas. Error bars are the standard deviation. Area numbers in blue are areas that we further sub-divided them into 200 m by 200 m small areas. Areas numbers in black are areas we did not sub-divide.

Figure S6. (a) An example of the difference (calculated using root mean square errors (RMSE)) between the normalized size distribution of all particles collected in the supersite and the normalized size distribution derived from particles randomly selected from the particle pool collected in supersite (more than 300,000 particles). RMSE are calculated for all size bins at different numbers of particle selected. (b) RMSE as a function of number of particles selected. When the number of particles selected are greater than 200, the RMSE curve decreases slowly. We determined that collecting 250 particles is sufficient to reproduce the size distribution of the original total particles.

Figure S7. Predicted concentration (in #/cm^3) of traffic and cooking particles from the full LUR models.

Figure S8. Comparison between concentration derived from the full LUR models and concentration derived from the source-specific LUR models. The red lines are the 1:1 lines.

Figure S9. Exposure to traffic and cooking particle number concentration in communities with different races and household income quintiles. Box and whiskers are the 5th, 25th, 50th, 75th and 95th percentile concentrations. Blue open circles are the population-weighted mean concentrations.

1. Single particle aerosol mass spectrometer

Particles were sampled by an aerosol mass spectrometer (AMS) capable of detecting single particle was used in this study. The aerodynamic lens of the AMS inlet with a flow rate of 1.3-1.4 cc/s. In the particle time-of-flight region, a mechanical chopper with two 1% opening slits operates at 145 Hz. Under typical conditions, at most one particle is in the particle time-of-flight region during any given chopper cycle. T At the end of the particle time-of-flight region, particles will strike a heated surface of 600 ^{o}C that vaporizes the molecules. All vapors are ionized by 70 eV electron ionization and the resulting ions are analyzed by a time-of-flight mass spectrometer (Tofwerk AG). The time particles spend between the chopper and the heater establishes the aerodynamic particle size (D_{va}).

The single particles are acquired in the event trigger (ET) mode. In this mode, mass spectra from individual particles are saved if the signals they generate surpass any of the user pre-defined signal thresholds, which are called regions of interest (ROIs). The ROIs used here are m/z 41 to m/z 43 $\geq$ 4 ions; m/z 45 to m /z 150 $\geq$ 5 ions and m /z 36 $\geq$ 3 ions. These ROIs thresholds are chosen because they do not trigger too many background events, nor do they miss too many actual particle events. Further analysis was conducted to remove any “false positive” for particle events. To categorize the detected particles, we first perform k-means clustering with k = 30 for all the normalized spectra of the detected particles then combine the clusters with similar spectra. To identify the source of each cluster, we compare the average spectra with the spectra from particles of known sources in the AMS database (http://cires1.colorado.edu/jimenez-group/AMSsd/).

Table S1: Summary of covariates used in building the LUR models. We also include covariates that are products of the two listed covariables in the Traffic category listed in the table. Data are from the database of Allegheny County Health Department, Pennsylvania Department of Transportation, National Emission Inventory, Pennsylvania Spatial Data Access and census.gov.

Major Category	GIS Category	Units	Circular Buffer Radius	Description
Traffic	Road length	m	25, 50, 100, 300, 500, 1000	Length of all roads, Length of major roads
	Inverse distance to the nearest road	1/m	N/A	Inverse distance to nearest road, Inverse square distance to nearest road
	Annual average daily traffic (AADT) on nearest road	veh/day	N/A	AADT on nearest road, AADT on nearest major road, Diesel truck AADT on nearest road, Diesel truck AADT on nearest major road
	Vehicle density	veh m/day	25, 50, 100, 300, 500, 1000	Vehicle density on all roads, Vehicle density on major roads, Diesel truck density on all roads, Diesel truck density on major roads
	Bus fuel consumption	kg fuel/day	25, 50, 100, 300, 500, 1000	Bus fuel consumption
	Rail length	m	25, 50, 100, 300, 500, 1000	Rail length
	Traffic land use zoning	m^2	100, 300, 500, 1000, 5000	Utility/transport land use area
Restaurant	Point density of restaurants	km^{-2}	100, 300, 500, 1000, 5000	Number of restaurants per unit area
	Inverse distance to the restaurants	1/m	N/A	Euclidean inverse distance to nearest restaurant, Euclidean inverse square distance to nearest restaurant
Industry	Point density of industry sources (NEI)	km^{-2}, lb km^{-2}	1000, 1500, 3000, 5000, 7500, 10000, 15000, 20000, 30000	Number of facilities per unit area, Annual pollutant emissions per unit
	Inverse distance to nearest industrial sources	1/m	N/A	Euclidean inverse distance to nearest facility, Euclidean inverse square distance to nearest facility
	Pollution emission	lb	N/A	Pollution emission at nearest corresponding facility

	Inverse distance weighted annual emissions	lb	N/A	Inverse distance weighted annual emissions
	Industry land use zoning	m^2	100, 300, 500, 1000, 5000	Industrial land use area
Elevation	Elevation	m	N/A	Elevation
Others	Land use zoning variables	m^2	100, 300, 500, 1000, 5000	Residential land use area, Commercial land use area, Agricultural land use area, Vacant/Forest land use area
	Population		100, 300, 500, 1000, 5000	Number of inhabitants
	Housing		100, 300, 500, 1000, 5000	Number of households

Table S2: Summary of the full LUR models

	Covariates selected[a]	Coefficients	Model/partial r^2	RMSE[b]	AME[c]
Traffic	Vehicle density in all roads (100 m)	25	0.49	1905	1507
	Utility and traffic land use area (5000 m)	1.49×10^{-3}	0.11		
	Intercept	-543	Model: 0.60		
Cooking	Restaurant counts (100 m)	11.3	0.30	1418	1098
	Major road length (1000 m)	0.0361	0.18		
	Inverse distance to the nearest National Emission Inventory PM source	2.03×10^{5}	0.17		
	Population density (1000 m)	0.148	0.05		
	Intercept	-792	Model: 0.70		

[a] Data sources of covariates are from the databases of the Allegheny County Health Department, the Pennsylvania Department of Transportation, the National Emission Inventory, the Pennsylvania Spatial Data Access, and census.gov..
[b] RMSE: root mean square error
[c] AME: absolute mean error

Table S3: Summary of source-specific LUR models for traffic and cooking particle number concentration and mixing state in Pittsburgh PA built using only the sub-divided grid cells (200 m by 200 m)

	Covariates selected[a]	**Coefficient**	**Model/partial r^2**	**10-fold validation r^2**	**RMSE**[b] **(#/cm^3)**	**AME**[c] **(#/cm^3)**
Traffic	Vehicle density in all roads (100 m)	20.97	0.35	0.43	2050	1641
	Diesel annual average daily travel * squared inverse distance to the nearest road (100 m)	316.9	0.18			
	Intercept	1731	Model: 0.53			
Cooking	Restaurant counts (100 m)	12.00	0.36	0.55	1534	1191
	Major road length (1000 m)	0.09	0.26			
	Residential land use area (100 m)	0.14	0.05			
	Intercept	-1252	Model: 0.66			
Mixing State ($1 - \chi$)	Major road length (1000 m)	6.16×10^{-6}	0.55	0.66	0.06	0.05
	House density (300 m)	8.87×10^{-5}	0.05			
	NEI emission weighted point source density (15 km)	4.65×10^{-5}	0.09			
	Intercept	0.186	Model: 0.69			

[a] Data sources of covariates are from the databases of the Allegheny County Health Department, the Pennsylvania Department of Transportation, the National Emission Inventory, the Pennsylvania Spatial Data Access, and census.gov..
[b] RMSE: root mean square error
[c] AME: absolute mean error

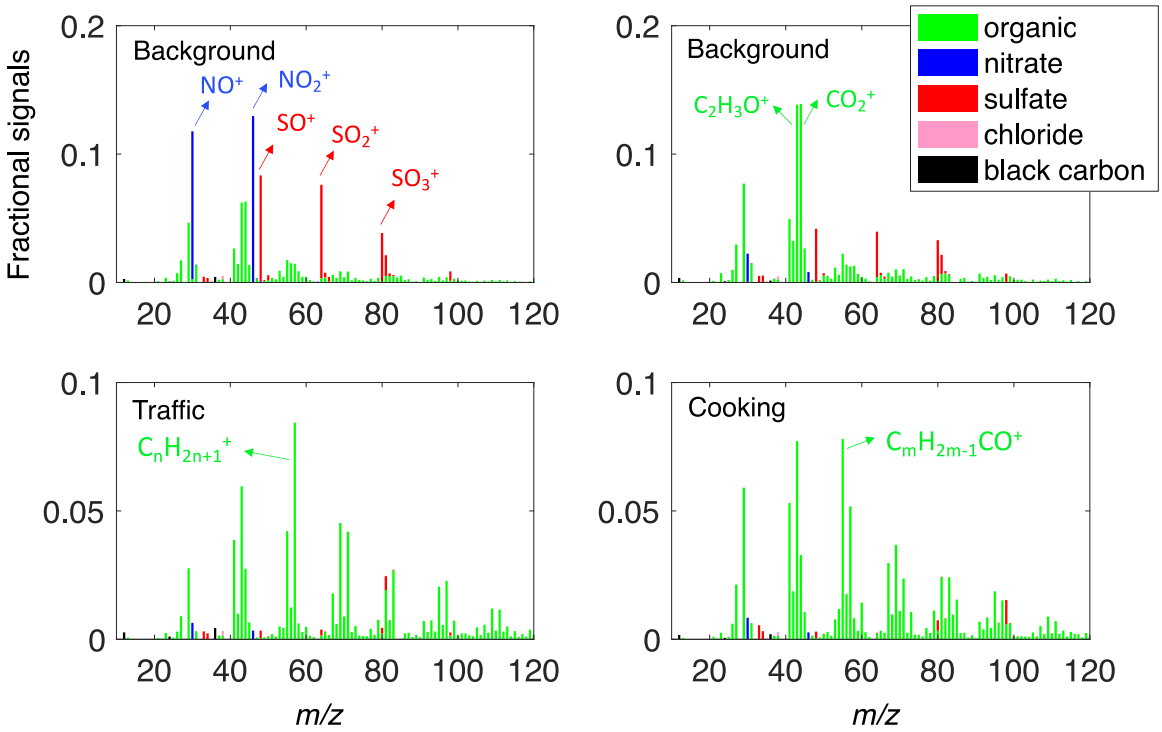

Figure S1: Average mass spectra for three types of particles categorized using k-means clustering: background particles including inorganic-rich and oxygenated organic particles, traffic particles and cooking particles. Different colors refer to different chemical species.

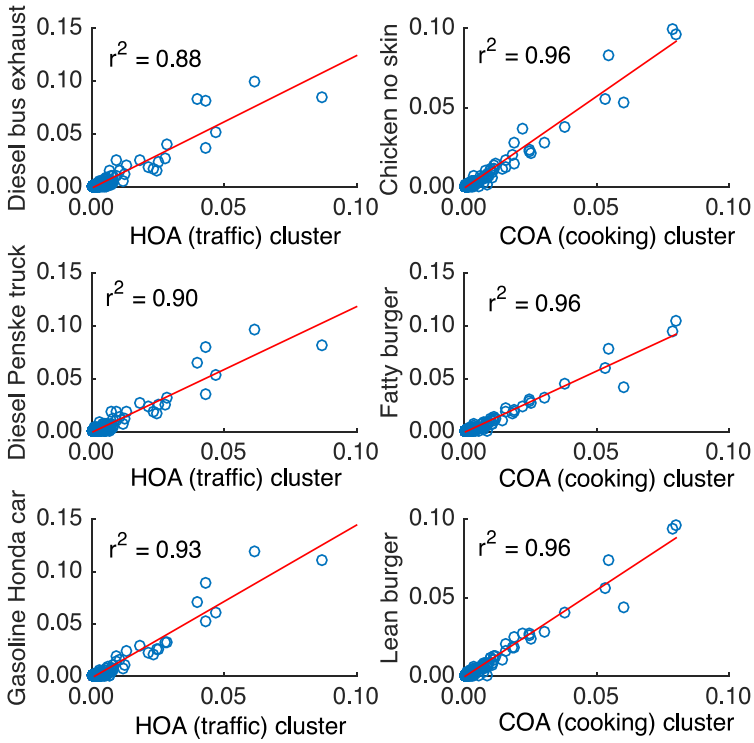

Figure S2: Spectral comparison between mass spectra of source-resolved single particle clusters (traffic and cooking clusters) and spectra of bulk particle measurements of known sources from the AMS database. Each data point corresponds to fractional signals in the spectra. The red lines are the fitted lines.

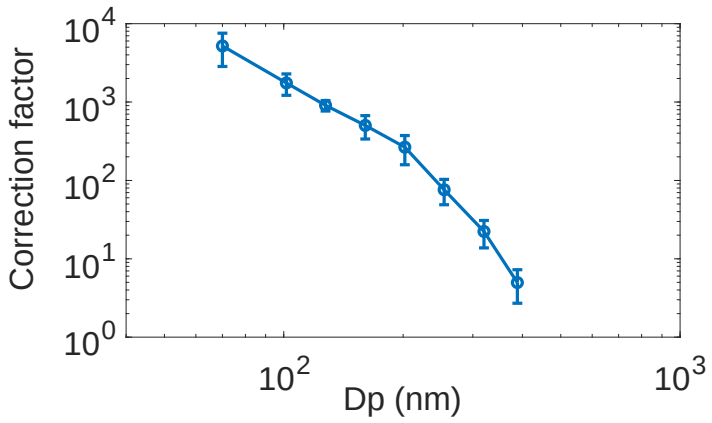

Figure S3: Average size-dependent corrections of single particle measurement using data from a scanning mobility particle sizer co-located with the singe particle mass spectrometer at the central reference site. Error bars are the standard deviation from multiple days of co-located measurements.

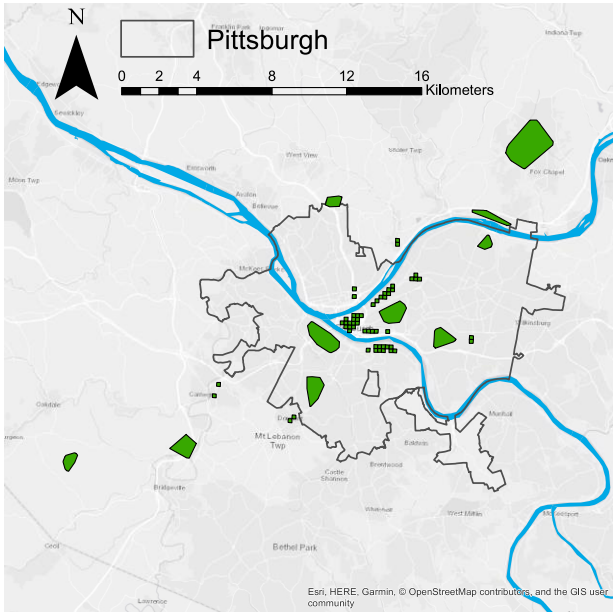

Figure S4: The base map showing subdivision of the measurement areas for LUR building. Half of the ~1 km^2 areas were kept whole, while the other half were subdivided into 200 m x 200 m squares. The boundary of Pittsburgh city is shown in black. Base map sources: Light Gray Canvas Map, Esri, HERE, Garmin.

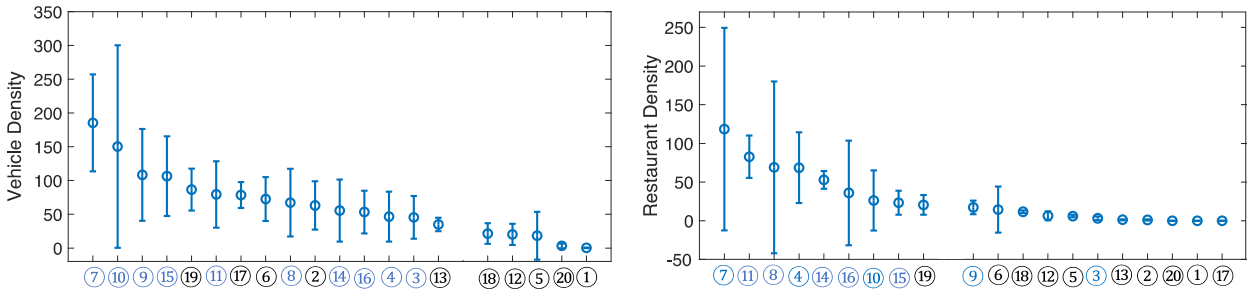

Figure S5: Vehicle density (vehicle/day/m) and restaurant density (number of restaurant/km^2) in the 20 areas. Circles are the average values of all the 5 m by 5 m rasters in the areas. Error bars are the standard deviation. Area numbers in blue are areas that we further sub-divided them into 200 m by 200 m small areas. Areas numbers in black are areas we did not sub-divide.

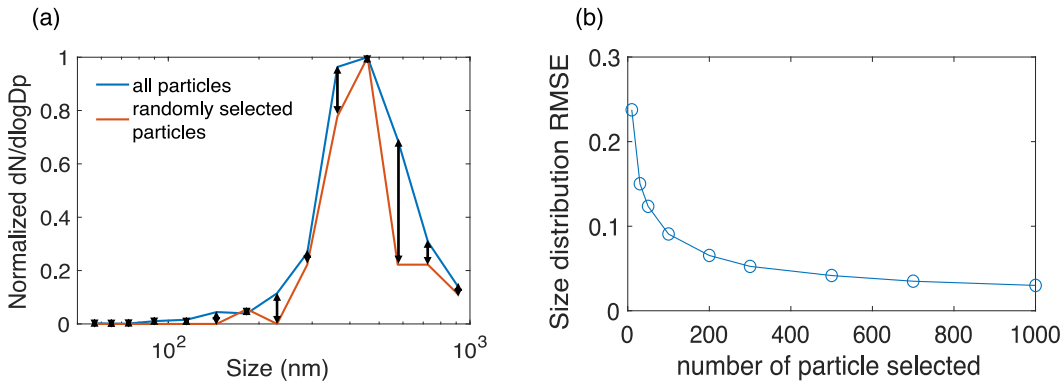

Figure S6: (a) An example of the difference (calculated using root mean square errors (RMSE)) between the normalized size distribution of all particles collected in the supersite and the normalized size distribution derived from particles randomly selected from the particle pool collected in supersite (more than 300,000 particles). RMSE are calculated for all size bins at different numbers of particle selected. (b) RMSE as a function of number of particles selected. When the number of particles selected are greater than 200, the RMSE curve decreases slowly. We determined that collecting 250 particles is sufficient to reproduce the size distribution of the original total particles.

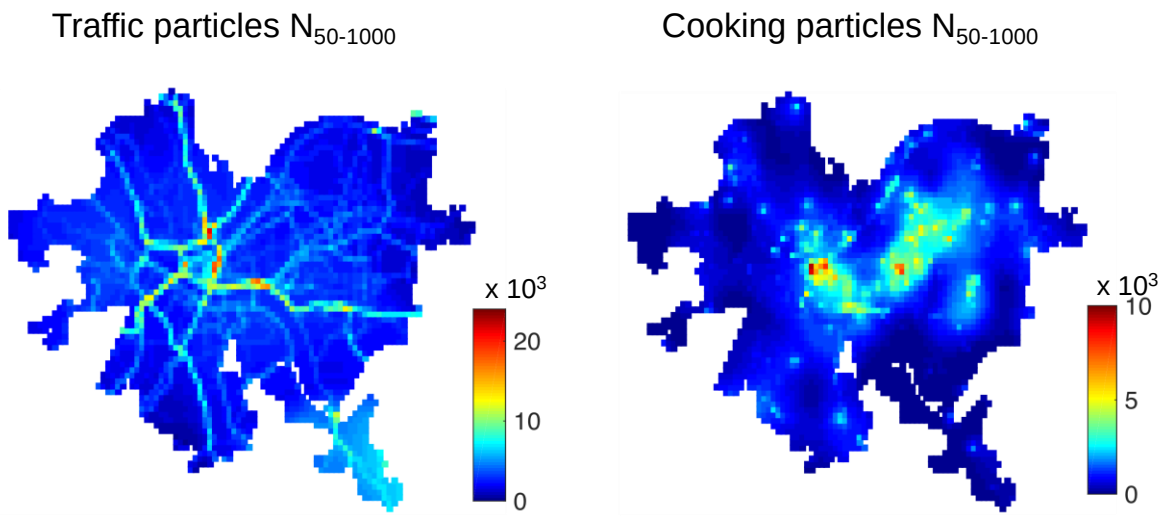

Figure S7: Predicted concentration (in #/cm^3) of traffic and cooking particles from the full LUR models.

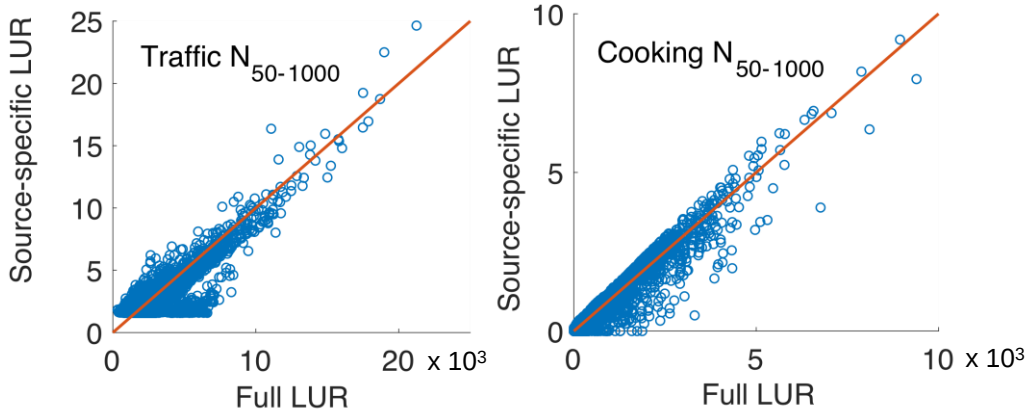

Figure S8: Comparison between concentration derived from the full LUR models and concentration derived from the source-specific LUR models. The red lines are the 1:1 lines.

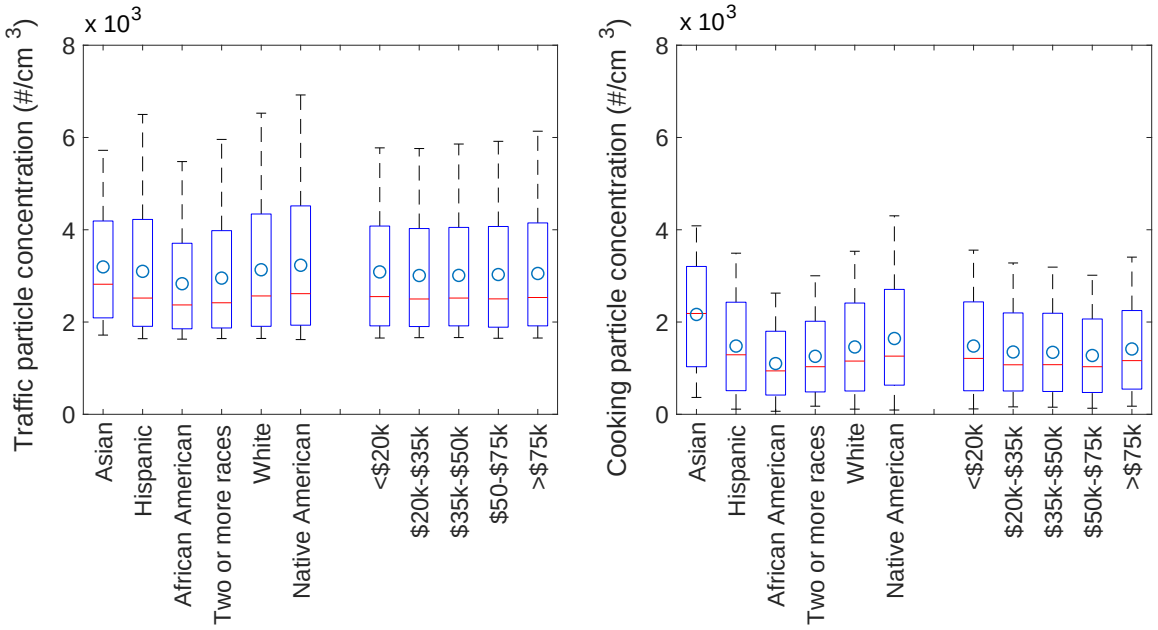

Figure S9: Exposure to traffic and cooking particle number concentration in communities with different races and household income quintiles. Box and whiskers are the 5th, 25th, 50th, 75th and 95th percentile concentrations. Blue open circles are the population-weighted mean concentrations.